Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8765335, 28 pages
https://doi.org/10.1155/2022/8765335

*Review Article*

# A Survey of Data Aggregation Protocols for Energy Conservation in WSN and IoT

**Beneyaz Ara Begum and Satyanarayana V. Nandury**

*CSIR-Indian Institute of Chemical Technology, Hyderabad, India and Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India*

Correspondence should be addressed to Beneyaz Ara Begum; beneyaz.8553@csiriict.in

Received 6 May 2022; Revised 22 July 2022; Accepted 30 September 2022; Published 29 October 2022

Academic Editor: Junaid Shuja

Conservation of energy has been a major concern for Wireless Sensor Networks (WSNs) and IoT applications. Several strategies were devised, aimed at optimizing energy consumption in these applications, based on: (a) use of low-powered hardware devices, (b) deploying mobile/relay agents for data collection, (c) clustering, and (d) data aggregation. Amongst these, data aggregation is widely acknowledged as an important tool to conserve energy in WSN and IoTs. The paper provides a comprehensive survey of various data aggregation strategies, discusses the efficacy of these strategies in handling issues that are typical to WSN and IoT applications. These issues severely impact the performance metrics such as: energy efficiency, latency, fault-tolerance, network throughput, and network lifetime. Therefore, to optimize the data aggregation approach, an application developer needs to arrive at optimal tradeoffs between these parameters. A major contribution of the paper is to present a holistic review of data aggregation approaches emphasizing the effect of topology, security, mobility, interference, and fault-tolerance in WSN and IoTs. Based on gap areas in literature, we throw open few challenges and present them as "posers", and put-forth suggestions for further research.

## 1. Introduction

The emergence of low-powered wireless embedded sensors has brought to fore their utility for remote data capture and sensing applications. This has greatly contributed to the proliferation of Internet of Things (IoT) in applications such as: smart cities, defense, surveillance, healthcare, agriculture, power grids, etc. Central to this development is a battery-powered embedded wireless sensor that typically comprises a transceiver, antenna, microcontroller, and the sensing mechanism. The wireless sensors are endowed with the ability aggregate, process, compute, communicate, and network with external agents like other wireless sensors, actuators, and IoT devices. Considering the limitless number of IoT devices that can potentially be networked through WSNs, the mechanisms to (a) regulate the bidirectional flow of data between the wireless sensors and IoT devices, (b) maintain the integrity, correctness, freshness, and temporal relations of the data flow, and (c) handle large chunks of data from various devices; become imperative. Data aggregation has emerged as an effective mechanism to address the above issues. In data aggregation, a group of nodes designated as data aggregators, perform aggregation on the data received from their subsidiary nodes, and relay only an aggregate to a high-end computational platform, either through a local Base Station (BS) or through an internet cloud, or a combination of both.

The physically dispersed sensor nodes in a WSN are usually independent, but collaboratively cooperate to drive an application. The stand-alone nature of the sensor nodes and IoT devices in a resource constrained environment, throws open a plethora of challenges, primarily aimed at developing energy-efficient, reliable, and robust data aggregation mechanisms. Due to the heterogeneous and mobile nature of various sensors, actuators, and IoT devices; network topology plays a significant role in dictating the data aggregation strategies in WSN and IoTs. Further, due to the dense deployment of these devices, the signals transmitted by them on wireless medium are susceptible to interference. Another challenge encountered by the sensor nodes and IoT devices, stems from the fact that these devices carry limited battery power and hence are prone to energy drain-outs, which results in faults due to node failures. Amidst the presence of different networks, protocols, topologies, and large user base in WSN and IoT applications, security and privacy are major challenges that demand special focus. With the advent of new technologies like edge computing, AI and ML, deep learning, bioinspired learning, and advanced network services and infrastructure like cloud and 5G networks; new approaches for data aggregation are receiving the attention of researchers. The paper discusses these technologies and issues related to their use in data aggregation.

While there have been quite a few contributions that surveyed data aggregation, most of these surveys broadly cover approaches that are specific either to WSN or IoT applications [1–6]. Abdulzahra et al. [7] provided a comprehensive survey of data aggregation methods and protocols in WSNs for IoT applications. They discussed the application of performance metrics such as resource (data and energy) efficiency, network topology (cluster, tree, chain and grid), and network lifetime in evaluating various data aggregation approaches in WSNs. Ali et al. [8], have carried out a detailed analysis of data aggregation techniques aimed at reducing power consumption and network traffic. In IoT applications, issues related to network heterogeneity and node mobility need to be factored into the data aggregation schemes. Saeedi et al. [9] analyzed the performance of these schemes.

The contributions of these surveys are summarized in Table 1. In comparison to these surveys, the focus of our paper is on providing an in-depth review of various data aggregation schemes, that address (a) topology, (b) mobility, (c) interference, (d) fault-tolerance, and (e) security issues; in the context of their applicability to both WSN and IoT applications. In addition, the purpose of our work is to highlight the trade-off issues that are necessary to bring about optimization in terms of energy efficiency, latency, interference, etc. in both WSN and IoT applications.

The main contribution of the paper is to provide:

(i) A holistic review of data aggregation approaches for WSNs and IoTs, with particular emphasis on topology, mobility, interference, fault-tolerance, and security issues.

(ii) Articulate different data aggregation schemes and highlight the gap areas, to throw open few research challenges as "posers" for further research.

## 2. Brief Overview of WSN, IoT, Data Aggregation

WSNs and IoTs have drawn considerable research interest due to their far reaching impact in applications related to monitoring of environment, habitat, agriculture, healthcare, hazardous, and disaster-prone regions. Due to the ease of installation at unmanned, harsh terrains and also due to their versatility in monitoring remote locations, WSNs are increasingly being used for ubiquitous sensing, communication, computation, and control. WSNs find ready applications in aerospace, target tracking, military reconnaissance and surveillance, infiltration detection and assessment, ubiquitous computing and smart cities, health monitoring etc. [1, 10].

As the name suggests, WSNs comprise a network of wireless sensors often referred to as nodes, deployed at strategic locations to remotely sense and monitor a phenomenon (or phenomena) of interest in the deployment region. A wireless sensor node comprises small low-powered and cost-effective sensing devices equipped with radio transceivers for wireless communication. Due to their miniaturized size, the wireless nodes do not require energy intensive infrastructure for collecting data. Unlike wireless adhoc networks, where two or more nodes can communicate without any central command, the nodes in WSNs owe their allegiance to the BS, to which they route all their sensed data. The BS analyses the data received from the nodes and draws inference on the phenomena being monitored by the WSN. To uphold this characteristic, the nodes exploit the inherent features of WSN, like flexible topology and self-organizing capability, to successfully route their sensed data to BS, even in the presence of node/link failures, packet drop, radio interference etc. Some of the key features of WSN illustrated in Figure 1.

As WSN and internet became popular, the concept of multiple smart devices connected to the internet has started gaining ground. By late 1990s, it became imminent that *anything and everything* would eventually be connected to the internet. The idea of internet of things (IoT), which originated from this concept was developed in parallel to WSN during its formative stage. However, it is the advances in WSN technology, that fueled the growth of IoT, as the advances in WSN were readily assimilated into IoT technologies [11]. IoTs have evolved as a network of physical and virtual *things* supported by a strong internet cloud backbone, to facilitate sensing and actuation, communication and computation, and storage and retrieval of *anything and everything*. A typical IoT application comprises a set WSNs and IoT devices networked together through an internet cloud as shown in Figure 2. The basic differences between the characteristics of WSN and IOT are presented in Table 2.

*2.1. Overview of Data Aggregation.* A key strategy to leverage the low-cost advantage of miniaturized wireless sensors and devices is to optimize their limited battery energy by minimizing the number of wireless transmissions in the network. While the number of data packets generated by each node/device may be relatively small, the quantum of packets generated collectively by all nodes is significantly large. This leads to multifold increase in network traffic. Due to limited storage capacity, a sensor node/device may not have large enough buffer to accommodate the incoming data, which leads to packet drop once the buffer is full. In addition to loss of data, it involves substantial load on network traffic, as nodes are forced to retransmit the dropped packets. This

Table 1: Areas covered in recent surveys on data aggregation.

| Survey | Areas covered |
|---|---|
| Jesus et al. [1] | (i) Reviews various distributed data aggregation algorithms.<br>(ii) Presents a computational taxonomy of existing aggregation techniques based on different types of aggregation functions. |
| Rahman et al. [2] | (i) Provides comparative analysis of data aggregation techniques with reference to energy dissipation, network lifetime, throughput, latency, etc.<br>(ii) Analyses the performance of LEACH and LEACH-C protocols in IoT in terms of the above performance metrics. |
| Salman and Jain, [3] | (i) Surveys communication standards for routing, network and session layer protocols, and the applicability for IoT.<br>(ii) Discusses management and security protocols in addition to the current challenges in IoT.<br>(iii) Presents insight into IoT data link protocols for carrier aggregation. |
| Lin et al. [4] | (i) Highlights the relationship between cyber-physical systems and IoT in existing architectures (viz., IoTSDN, SOA, middleware architectures).<br>(ii) Discusses security and privacy issues in fog/edge computing-based IoT for real-world applications (smart grid, smart transportation, and smart healthcare).<br>(iii) Provides an insight to challenges in resource allocation in fog/edge computing-based IoT. |
| Ray [5] | (i) Presents a survey of IoT architectures to facilitate developer's requirements and security.<br>(ii) Compares the existing IoT supported architectural platforms.<br>(iii) Presents a case study of IoT cloud platform for agricultural, health, and smart society domains. |
| Dehkordi et al. [6] | (i) Reviews advanced data integration and clustering techniques.<br>(ii) Highlights the advantages and challenges of structure-based and structure-less data aggregation protocols in terrestrial, underwater, and underground WSNs, in terms of energy, bandwidth, performance, and delivery ratio. |
| Abdulzahra et al. [7] | (i) Presents an overview of data aggregation techniques.<br>(ii) Reviews the existing data aggregation mechanisms in terms of their topology, resource efficiency, network life time, the approach followed and their objectives. |
| Ali, et al. [8] | (i) Presents a classification of some of the recent works based on the nature of data and data sets (multimedia, data packets, encrypted data, etc.).<br>(ii) Surveys data aggregation techniques in the context of network life-time, network capacity, eliminating data redundancy, security etc.<br>(iii) Presents a review of data aggregation techniques of recent works by evaluating the energy consumed by the nodes. |
| Saeedi et al. [9] | (i) Highlights the advantages, limitations, and challenges of data aggregation.<br>(ii) Surveys the data aggregation approaches to bring out the differences between flat and hierarchical networks.<br>(iii) Presents a survey of recent publications on data aggregation, and a comparative study in terms of network topology (cluster, tree, and flat), node type (homogeneous/heterogeneous), aggregator type (static/mobile), centralized/distributed algorithm, application domain, etc. |

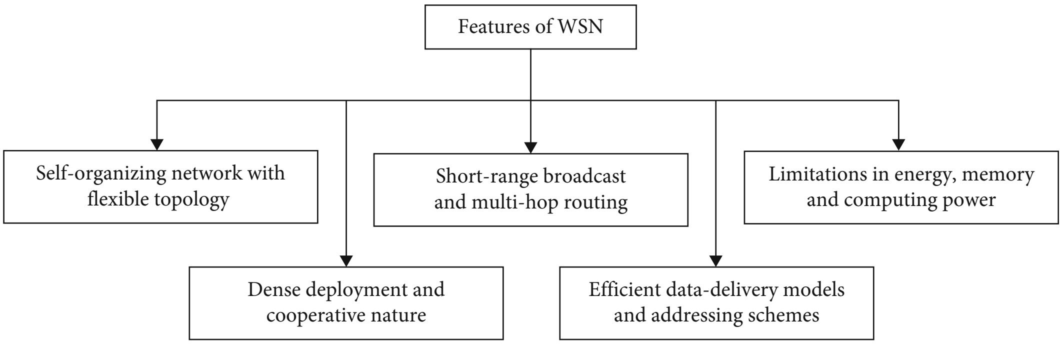

Features of WSN
Self-organizing network with flexible topology
Short-range broadcast and multi-hop routing
Limitations in energy, memory and computing power
Dense deployment and cooperative nature
Efficient data-delivery models and addressing schemes

Figure 1: Features of wireless sensor networks.

results in faster energy drain-out of nodes/devices besides leading to undesirable consequences like increased latency. To get over this problem, data aggregation is usually employed. Data aggregation involves integration of correlated data at intermediary nodes designated as aggregators. Depending on application requirement, the aggregator

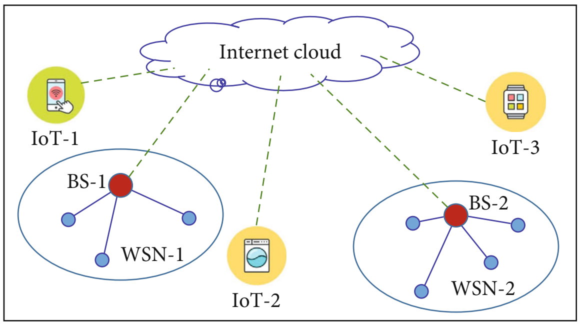


Figure 2: Typical IoT application network.

employs an appropriate aggregation function *viz.*, numeric or statistical operators such as min, max, sum, average etc., on the data it receives from its subsidiary nodes.

Data Aggregation is an essential tool in IoT applications. To illustrate its use, consider an application where functions like power generation and distribution, waste management, surveillance, smart metering etc., need to be monitored in a smart-city, as shown in Figure 3. Also consider the smart-city to have a smart WSN-based ubiquitous backbone infrastructure to handle big data generated by thousands of sensors, actuators, smart meters, etc., deployed across the smart-city [10]. To monitor the humongous amount of data generated by the IoT devices, the IoT application performs stage-wise data aggregation at a local-level or at application-level initially, (e.g. power generation and distribution, waste management, surveillance, and smart meters) and later at global level. The localized aggregates are integrated thorough an IoT cloud infrastructure.

Besides reducing the number of data packets and easing the network traffic, data aggregation also reduces the latency and the power consumed by the nodes in the network. To illustrate this, consider the WSN shown in Figure 4(a), where the leaf nodes $S_7$, $S_1$, $S_6$, $S_5$, $S_4$, $S_3$, and $S_2$ capture their sensed information in one data packet each in a given time period. It is assumed that other nodes in the network act as relay or aggregator nodes. To address channel contention, a timeslot allocation scheme is charted out, where a node is permitted to either receive or transmit only one packet per timeslot allocated to it. As per this scheme, the leaf nodes $S_7$, $S_1$, $S_6$, $S_4$, and $S_2$ are allocated timeslot $T_1$, and $S_5$, $S_3$, are allocated timeslot $T_2$ to transmit their sensed data to their parent node. The time slots for intermediary nodes, $S_8$, $S_9$, $S_{10}$, and $S_{11}$ are accordingly adjusted so as to avoid channel contention for performing both reception and relay operations.

Thus, the relay node $S_{10}$ receives the sensed data packet from its child nodes $S_6$ and $S_{10}$ in time slots $T_1$, and $T_2$, respectively, and relays these packets to $S_{11}$ in time slots $T_3$ , and $T_4$. Similarly, the nodes $S_8$, $S_9$, and $S_{11}$ adjust their transmission slots and the BS receives all data packets only after 12 timeslots as shown in Figure 4(a).

The node $S_{11}$ being closest to BS, acts as its gateway node. In the absence of data aggregation mechanism, node $S_{11}$ is forced to transmit large number of data packets compared to other leaf or intermediary nodes, and takes a minimum of 6 time-slots ($T_2$, $T_8$, $T_9$, $T_{10}$, $T_{11}$, and $T_{12}$) to relay all packets that it receives. Due to this delay, there is good chance that some of the incoming packets are dropped if the $S_{11}$ buffer is full. This forces the source nodes to retransmit the dropped packets. Further, as $S_{11}$ is involved in relaying data to BS continuously, it drains out much faster than other nodes in the network. Being one of the main gateways to the BS, there is a strong likelihood that the network collapses as soon as the energy of node $S_{11}$ drains out. This can amicably be addressed through data aggregation scheme as illustrated in Figure 4(b), where the leaf nodes $S_7$, $S_1$, $S_6$, $S_4$, and $S_2$ transmit in timeslot $T_1$, while the nodes $S_5$, $S_3$ transmit in timeslot $T_2$. The aggregator nodes $S_8$, $S_9$, $S_{10}$, and $S_{11}$ relay just one data (aggregated) packet to their parent nodes. While aggregator nodes $S_8$ and $S_{10}$ transmit their aggregates in timeslot $T_3$, $S_9$ transmits its aggregate to $S_{11}$ in timeslot $T_4$ after performing aggregation on the packet received from $S_4$ (timeslot $T_1$) and the aggregate from $S_8$ (timeslot $T_3$).

The aggregator node $S_{11}$ transmits its aggregate in timeslot $T_5$. Thus, the traffic across the network is relatively reduced, and the BS receives an aggregate of the data sensed by the leaf nodes in relatively less number of timeslots i.e. 5 timeslots. To further illustrate the advantage of data aggregation, assume that each node $S_n$ in the example, can transmit only one packet in a given time slot $T_i$, which is of 5 time units duration. If data aggregation is not performed the BS receives the data packets from all leaf nodes in 60 time units (12 timeslots), and the network traffic witnesses a total of 20 transmissions. On the other hand, if aggregation is performed BS receives the information from source nodes in 25 time units (5 timeslots) and the overall network traffic is reduced to 11 transmissions.

*2.2. Taxonomy of Data Aggregation Protocols.* Data aggregation protocols define the standard operational procedures to: (a) aggregate the sensed data based on an aggregation function, (b) handle the communication of data and control messages, and (c) route the aggregates to the BS/internet cloud. The primary objectives of data aggregation protocols are to eliminate redundant data transmission from source nodes to BS/internet cloud, maintain the accuracy of the data while performing aggregation, and improve the lifetime of WSN/IoT. Considering the diversity of application and widely varying nature of operation, there can be several classifications to data aggregation protocols based on (a) WSN/IoT topology, (b) interference models (c) security, and (d) network dynamics as shown in Figure 5. A taxonomy of the protocols based on above classification for WSN and IoT is shown in Figure 6. The protocols based on above classification are surveyed in subsequent sections.

## 3. Data Aggregation Protocols Based on Topology, Security, and Mobility

*3.1. Data Aggregation Protocols Based on Topology.* Data aggregation protocols based on network topology is categorized into flat-based and hierarchical-based protocols.

Table 2: Characteristics of WSN and IoT.

| Characteristics | WSN | IoT |
|---|---|---|
| Number of devices | Typically restricted to few wireless sensors to thousands of sensors | Theoretically no limit |
| Topology | Adhoc, hierarchical (tree, cluster, ring, chain, and grid) flat (flooding and forwarding) | Adhoc, largely heterogeneous |
| Radio channel access | CSMA-CD, CSMA-CA TDMA, TDMA/CDMA, and LORA | CSMA-CD, CSMA-CA, TDMA, TDMA/CDMA, and LORA |
| Communication | Wireless | Wireless, internet, edge computing, and fog/cloud computing |
| Security | The wireless transmissions are not always encrypted, hence the communication security is moderate | Since all data transmissions are through internet, the application has the same security as the one provided by the ISP, which is usually high. If the transmissions are through fog/cloud, the security is very high. |
| Keys | Symmetric (ZigBee, WirelessHART), asymmetric (ISA 100.11a) | Symmetric (ZigBee, WirelessHART), asymmetric (ISA 100.11a) |
| Interface to external world | Through BS, centralized, and distributed | Through internet, cloud, and distributed |
| Scalability | Moderate scalability | Highly scalable |
| Protocols | ZigBee, WirelessHART, ISA 100.11a, WiFi, and mmWave (2.4GHz, 5GHz, 6GHz upper, 6GHz lower, 24GHz, and 60GHz), LoRaWAN RF (868 MHz) LoRaWAN RF (900 MHz)<br>3G/4G/5G Mobile data, Bluetooth low energy (2.4GHz) | **Datalink protocol:** IEEE 802.15.4e, EEE 802.11ah, WirelessHART (TDMA), Z-wave Bluetooth low energy (2.4GHz), ZigBee smart energy, DASH7, and HomePlug. G.9959, IPv6, LTE-A, LoRaWAN RF (868 MHz, 900 MHz), NB-IoT, DECT/ULE, EnOcean, and 3G/4G/5G Mobile data<br>**Network layer routing protocols:** RPL, CORPL, CARP, and E-CARP<br>**Network layer encapsulation protocol:** 6LoWPAN, 6TiSCH, 6Lo, IPv6 over G.9959, and IPv6 over Bluetooth low energy<br>**Session layer protocols:** MQTT, SMQTT, AMQP, CoAP, XMPP, DDS<br>WiFi, and mmWave (2.4, 5, 6, GHz Upper and Lower, 24GHz, and 60GHz) [9] |
| Autonomy of devices | Moderate | Highly autonomous |
| Deployment and coverage | The sensors are usually deployed through a predefined strategy. As WSNs are highly application oriented, the deployment strategy ensures maximum coverage. | As the communication is through internet, control on deployment and coverage does not exist. |
| Signal and data processing | While signal processing is performed by the sensors, data processing is performed by the BS. Therefore, the computational capacity of the BS can be a limiting factor. | Signal and data processing activities are performed through internet and cloud, and hence the IoT devices need not carry high-end processors for computation. IoTs are good candidates for handling Bigdata. |
| Mobility | Mobility is restricted, as sensor nodes rely on BS for all communication and high-end computation needs. | IoT devices rely on internet/cloud for communication and high-end computation. Therefore, IoT can accommodate high mobility of devices. |

*3.1.1. Data Aggregation Protocols Based on Flat Topology.* Flat networks are topology-free and the nodes are not bound by hierarchy. The nodes are generally assumed to possess same functionality and capability in terms of battery and computational powers except in IoTs, where the devices differ widely from each other. All devices in Flat networks maintain same network state information and keep track of their one-hop neighbor. This helps the nodes to collaborate and relay aggregated data to the BS. Based on communication methodology, the routing schemes are classified into (a) Flooding, (b) Forwarding, and (c) Data-centric based routing.

Sensor Protocols for Information via Negotiation (SPIN) is the most prominent data aggregation protocol amongst Flat networks. SPIN is a three-stage protocol where the basic operations are performed using control messages *viz.* ADV (advertise), REQ (request), and DATA (message to be transmitted). The nodes incorporate resource adaptation technique to determine when to participate in negotiation process. During negotiation, each node polls to know the

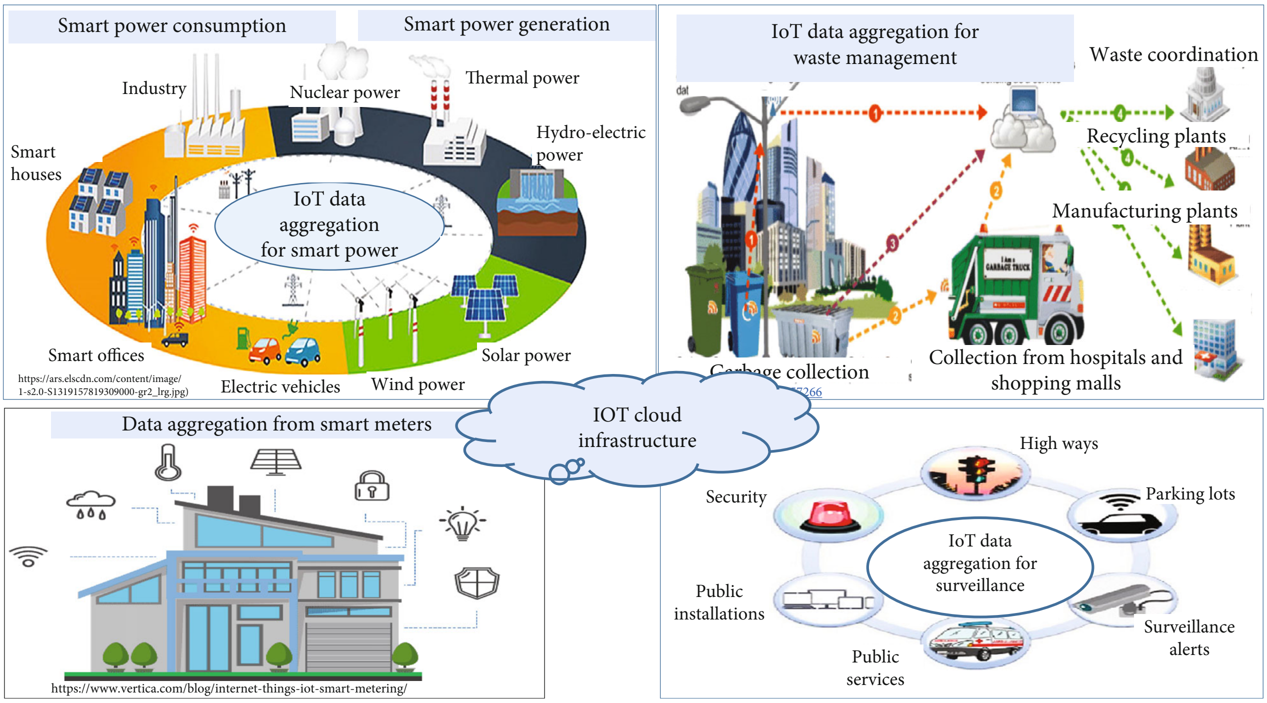
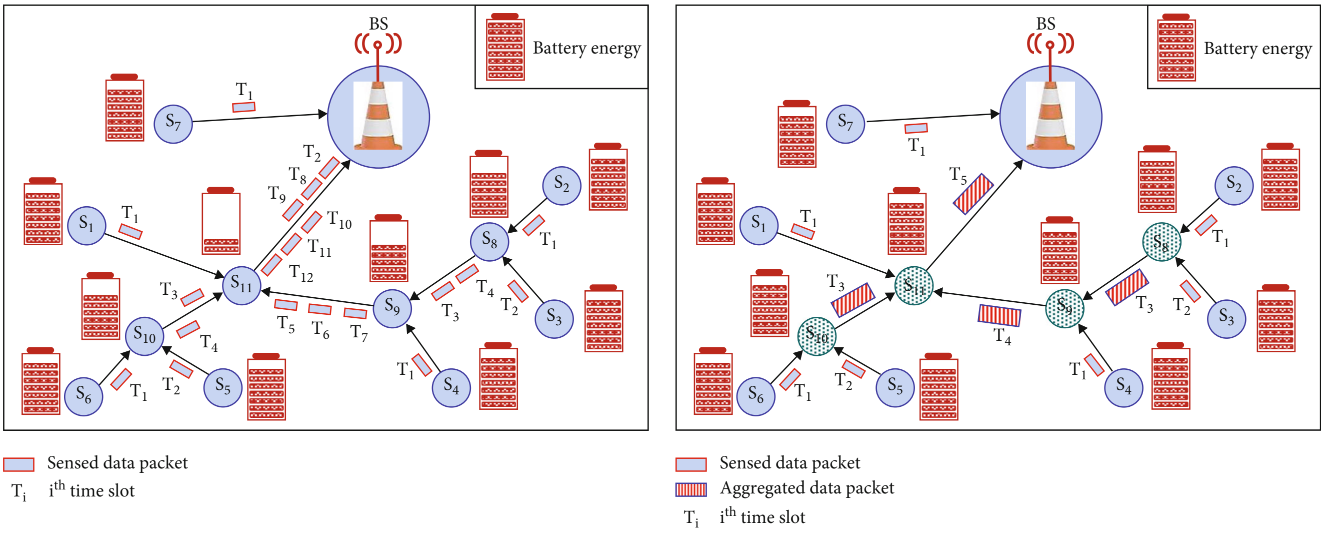


Figure 3: Illustration: IoT data aggregation.



(a) Without data aggregation

(b) In-network data aggregation

Figure 4: Illustration: reduction of latency and energy and data consumption due to data aggregation.

current energy level, so that nodes with lower energy can cut down low priority tasks. SPIN manages topological changes locally by making forwarding decisions based on the knowledge of its one-hop neighbours. Due to this feature, SPIN can be extended to IoT applications where nodes are mobile. While direct implementation of SPIN protocol for IoT application is not widely reported, certain variants of SPIN like M-SPIN, SPIN-BC, and SPIN-RL can be applied to IoT. The family of SPIN and their features are summarized in Table 3.

*3.1.2. Data Aggregation Protocols Based on Hierarchy.* Flat networks may sometimes lead to excessive computation and communication overheads, resulting in energy depletion. To reduce these overheads several hierarchical data aggregation protocols were proposed, which can be categorized into: (i) Cluster-based, (ii) Chain-based, (iii) Tree-based, (iv) Grid-based, and (v) Ring-sector based networks. The most popular data aggregation approach in hierarchical networks is the cluster-based approach, particularly for IoT applications. In this approach, the region of interest is

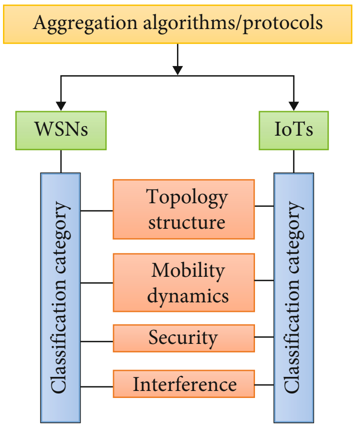


Figure 5: Classification of data aggregation protocols for WSN and IoT applications.

fragmented into clusters with one of the nodes designated as Cluster Head (CH). The CH collects data from all cluster nodes, performs aggregation, and routes its aggregate to the BS, either through direct transmission or with the help of neighboring CH nodes. Based on this, several protocols were proposed, the most popular being LEACH and HEED, which are briefly described below.

*(1) Low Energy Adaptive Clustering Hierarchy (LEACH).* The LEACH protocol facilitates the formation of adaptive, self-configuring clusters with localized control, and application-specific data aggregation or compression techniques. The nodes that are in close proximity with one another form a cluster, where one of the nodes is chosen as the Cluster Head (CH) through an election process. Periodic election ensures that the elected CH makes way for a new CH, once its own energy falls below a predefined threshold. Depending on the methodology adopted for cluster formation, CH selection and communication mechanism; several variants of LEACH were developed, some of which are described in Table 4.

*(2) Hybrid Energy Efficient Distributed Clustering (HEED).* In HEED all nodes are assumed to have similar functionality and possess discrete transmission power levels. Unlike LEACH, the node with maximum residual energy and minimum communication overhead is selected as the CH. The neighboring nodes are mapped based on their proximity to the CH. However, if a node falls in the range of two or more CHs, the CH with minimum intracluster communication overhead is chosen as its CH. Both LEACH and HEED have emerged as popular protocols for hierarchy based networks and have attracted the attention of researchers to devise several variants as shown in Table 4.

*3.1.3. Bioinspired Selection of Cluster Head.* CH selection has been an important topic for research in all cluster based data aggregation approaches. The main strategy is to select a CH, based on residual energy, cluster density, proximity to BS, etc. However, the challenge is to determine the metrics for CH selection and arrive at an optimum mix to select the most suitable CH for a given application. In addition, issues such as load balancing, coverage, network life-time, hop-count distance to BS, delay latency, etc., need to be considered. This leads to a multiobject optimization problem. In IoT applications where the nodes are mobile, the transmission energies required by a node varies widely. Therefore, estimation of residual energies of nodes and their hop-count to the BS is nondeterministic. Due to the nondeterministic nature of these variables, the traditional deterministic approaches involved in cluster formation and CH selection encounter severe limitations. To counter this, several Bioinspired techniques such as: Multiobjective Optimization Algorithm (SMS-EMOA), Nondominated Sorting Genetic Algorithm (NSGA-II), S-Metric Selection Evolutionary and MultiObjective Evolutionary Algorithm by Decomposition (MOEA/D) have emerged as popular approaches for selection of CH. These bioinspired approaches define a fitness function based on optimization of multiple objectives. The fitness function is determined iteratively for all potential candidates that are in contention to be selected as the CH. When the iterations reach a saturation point, the node with the best fitness function is selected as the CH [28–31].

Ahmad et.al [32], propose a Honey Bee algorithm to form clusters and select an appropriate CH in mobile WSNs, which can be extended to IoT applications. The bees have the combined responsibility of forming nonoverlapped clusters and identifying most suitable set of CHs for performing data aggregation. In this approach the population is divided into two groups. The *onlooker bees* (control packets) are responsible for identifying the food source (nodes) based on node energy, direction and speed of the mobile node and node degree. The data packets are represented by the *employed bees* that are responsible for nectar collection (data aggregation). The node with higher degree, energy, and uniform distance from other nodes becomes a better candidate to be selected as the CH.

In Flying Adhoc Networks (FANETS), due to the continuous movement of the Unmanned Aerial Vehicles (UAV), which are considered as nodes in the network, the topology keeps changing at a rapid pace. Khan et.al [33], make use of Glow-worm Swarm Optimization (GOS) and Krill Heard (KH) algorithm to form clusters. The UAV with the best fitness function is selected as the CH, which performs data aggregation.

*3.2. Security and Mobility in Data Aggregation.* WSNs and IoT networks are vulnerable to security attacks due to the involvement of multiple entities like sensors, actuators, communication and computational devices, etc.; and also due to physical interactions with external environment and user-base. Therefore, while security issues pose unique challenges to data aggregation [34], especially in defense and mission-critical applications; issues related to privacy are of prime importance in applications where data that is privy to an individual need to be preserved. Hardware and software solutions do help ward off these threats to a limited extent, but counter measures in the form of encryption and decryption, secure key management, secure routing, etc., are more effective and relatively easier to implement.

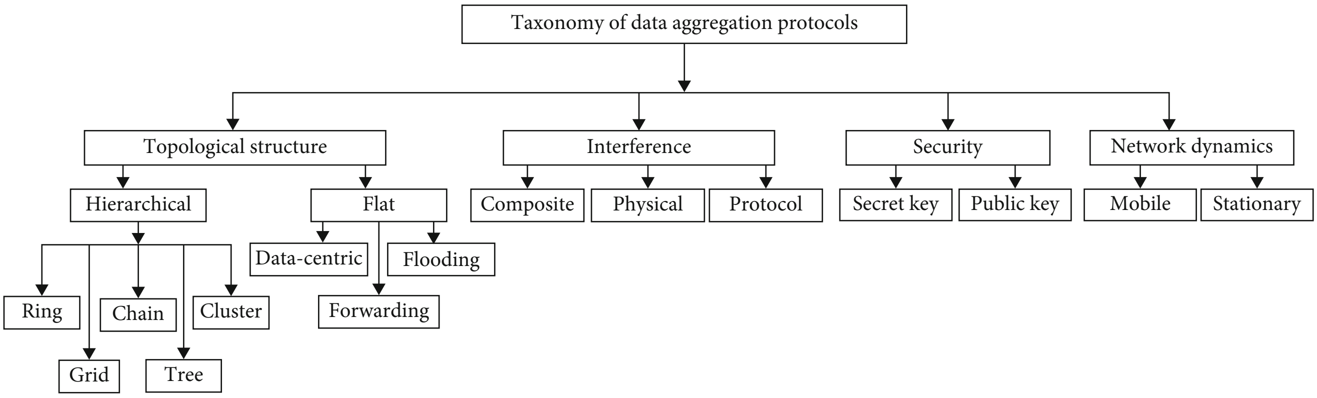


Figure 6: Taxonomy of data aggregation protocols.

In general, network security prevents injection of malicious attacks and guarantees integrity and confidentiality in data aggregation. Data encryption, authentication, attack detection, etc., are some of the popular methods traditionally used to provide network security. However, direct use of these methods for data aggregation in WSN and IoT applications may suffer from severe limitations due to multitude devices, heterogeneity, and network topologies. To improve network performance, data aggregation is combined with security goals to ensure availability, confidentiality, freshness, lifetime, and integrity of data transmitted during the delivery process. A comprehensive review of secure data aggregation (SDA) in WSNs is provided in [35]. The authors present a comparative analysis of SDA protocols by categorizing them into five different security mechanisms *viz.* slicing SDA, confidence SDA, encryption SDA, privacy SDA, and anomaly detection SDA, security goals, and network topologies. To save communication costs and protect private data, data decomposition technique is implemented instead of slicing in privacy-preserving data aggregation protocols [36].

Rezaeibagha et.al [37] present an efficient and provably secure scheme as a first step toward secure data aggregation for handling data collection and analysis of IoT wireless body sensors. A novel cryptographic accumulator based on authenticated additive homomorphic encryption was developed, which can collect and aggregate data from IoT wireless wearable devices. The encrypted data can be used for analysis in an encrypted form so that the information is not revealed. Some of the recent security protocols developed for WSN and IoT applications are presented in Table 5.

In applications that demand large geographical spread, the nodes are located quite far apart from each other. In such applications, energy consumed by the nodes to communicate with their immediate neighbors is quite large, resulting in faster battery drain-out. To counter this, WSNs employ mobile agents that go around the network to collect data from each node in the network and perform aggregation [21, 38]. However, when the demand is to establish interconnection between mobile wireless devices, as in the case of IoTs, issues related to mobility pose severe challenges to the existing data aggregation schemes. A review of data aggregation schemes based on mobility in presented in Table 6.

## 4. Data Aggregation in IoT

Faster access to WiFi and internet connectivity has made it possible to establish connection between two or more devices (*things*) at any point in time. This has led to the proliferation of IoTs into several application areas like healthcare, environment, surveillance [39], network traffic monitoring, etc. The recent spurt in the number of devices that can be connected to the cloud, has paved way for the development of protocols with low latency transactions. Due to close similarity with WSNs, IoTs have thrived on borrowing the concepts of WSN for developing strategies for data aggregation protocols. Unlike WSNs, where the data generated is singularly in-tune with its intended application, data generated by IoT devices encompass a wider range of application areas. To handle such heterogeneous data, IoTs rely on a collection of aggregator nodes that report to multiple sinks/BSs. More so, in Machine to Machine (M2M) communication scenarios where both sensor and actuator devices work in tandem [40].

Due to the deployment of diverse range of IoT devices and their geographic spread, IoTs rely on cloud server environment for computation and interpretation of data, and for actuation. Such platforms are particularly useful in fog computations [41] where low-latency fog devices (nodes) deployed close to the IoT network edge, play a large role in operations related to control, computation, and storage. The all-encompassing nature of IoTs demand pervasive security and privacy at both aggregator and cloud server ends. A privacy preserving data aggregation scheme that makes use of data slicing approach is presented in [42]. In this scheme, each IoT device in a group (comprising $n$ nodes/devices) slices its data randomly into $n$ segments and forwards the slices to other $n-1$ nodes/devices in the group using symmetric encryption, while retaining one of such slice for itself. The aggregator node aggregates the slices received along with its own slice and forwards the same to its group aggregator by employing homomorphic and AES encryption. Haseeb-Ur-Rehman et al. [43], analyze the sensor cloud frameworks by considering critical issues such as heterogeneity, scalability, service availability, data accessibility, and security. Based on this study, a comparison of

Table 3: Variants of SPIN Protocol.

| Protocols | Features | Advantages, limitations |
|---|---|---|
| SPIN-1, SPIN-2 [12] | (i) Ensures that only the correct data queried is transmitted.<br>(ii) SPIN-2 is an extension of SPIN-1. It incorporates threshold-based resource awareness mechanism. | (i) SPIN-1 is simple to implement.<br>(ii) Allows a node to negotiate only when it can complete all operations without reaching a low energy-threshold level. |
| SPIN-BC (broadcast) [13]<br>SPIN-PP (point-to-point),<br>SPIN-EC (energy consumption awareness) [13]<br>SPIN-RL [13] | SPIN-BC is specifically designed to work in broadcast medium.<br>SPIN-PP is designed for point-to-point communication in lossless medium. Nodes that reach low-energy threshold, do not participate in negotiation. While the nodes cannot forbid themselves from receiving ADV and REQ messages, they are capable of stopping the transmission of data message.<br>SPIN-RL is a reliable version of SPIN-BC for lossy medium. | SPIN-BC works in environments where nodes have sufficient energy but are susceptible to transmission errors in lossy medium.<br>(i) SPIN-PP is resilient to frequent changes in topology.<br>(ii) It is adaptable to lossy, mobile, and unconfigured networks.<br>(iii) SPIN-PP is ideal for applications where nodes have adequate energy. SPIN-EC incorporates energy conservation heuristics and prevents receipt of DATA messages for nodes with energy below a threshold.<br>(iv) SPIN-RL selectively limits the frequency of retransmission of same data, thus ensuring transmission of reliable data. |
| SPMS (shortest path minded SPIN) [14] | (i) Uses shortest path to reach the destination.<br>(ii) Determines maximum transmission radius or zones for each node. | (i) Optimizes energy consumption by adjusting the transmitting power level of nodes on the basis of distance.<br>(ii) Node failure is detected and recovered through a backup path. |
| Modified SPIN (M-SPIN) [15] | (i) SPIN is designed to support quick reliable response and selective transmission. It adopts a distance discovery phase to find distance in terms of hop distance to BS.<br>(ii) Data dissemination is moderated by the hop-count distance between nodes. It facilitates only one-way transmission of data from nodes to the BS. | (i) The energy consumption is reduced due to simplex communication between nodes and BS. However, the computational complexity is more due to frequent energy level status updates after every transaction.<br>(ii) Compared to SPIN-BC, M-SPIN takes relatively longer time to calculate the hop distance.<br>(iii) Suitable for alarm monitoring applications. |
| Secure SPIN [16] | (i) Uses PSAC (personal sensor authentication code) to generate sensor node's privacy key.<br>(ii) BS maintains privacy keys of all nodes and CHs.<br>(iii) Hash function is used to generate the message authentication code (MAC) session key for maintaining data freshness. Session key ensures that no adversary can replay the old messages. | (i) Provides multilevel security amongst: (i) node and CH (ii) CH and BS, and (iii) BS to all nodes.<br>(ii) CDMA codes help maintain secure communication.<br>(iii) Data integrity is ensured through MAC.<br>(iv) PSAC ensures data confidentiality and authentication.<br>(v) Requires relatively smaller processing power and memory for data authentication and integration. |
| S-SPIN [17] | MAC scheme is used to guarantee correctness and integrity of the messages. Each sensor has a resource manager to keep track of resource consumption. | Secure against existential forgery attack. |

different cloud frameworks in the context of reliability, energy efficiency, and latency, is presented in the paper.

As IoTs proliferate into social networking sites, geo-tagging of data generated in Social Internet of Things (SIoT) and performing data aggregation, has emerged as a major area of research. Shuja et al. [44] present a hierarchical clustering framework for geo-tagged data, and determine optimal parameters to handle various types of data sets, cluster sizes, tokenization techniques, etc. A systematic and comprehensive approach to study and analyze the importance of network lifetime in data aggregation for IoT applications was carried out by Pourghebleh et al. [45].

Compressed sensing which is a signal processing tool for efficient data acquisition and signal reconstruction has

Table 4: Variants of LEACH and HEED Protocols.

| Protocol | Description, features | Advantages/limitations |
|---|---|---|
| LEACH-C LEACH-Centralized [18] | (i) LEACH-C considers centralized single-hop homogeneous network. CHs are elected by the BS based on their average energy level and distance from BS. The BS determines the clusters using simulated annealing algorithm.<br>(ii) Ensures uniform distribution of clusters. | BS carries out load balancing to ensure uniform energy consumption across the network. Network lifetime is more, compared to LEACH. |
| TL-LEACH Two-Level Leach [19] | TL-LEACH considers a two-level structure comprising primary and secondary CHs in a homogenous multihop network, with the primary CH communicating the final aggregate to BS. | Due to smaller transmit distances the energy consumption is relatively reduced. |
| E-LEACH Energy LEACH [20] | (i) E-LEACH considers homogenous single hop network.<br>(ii) Equal probability for all nodes to become CH in first round. After first round, residual energy of nodes is considered to select CHs in next round. | (i) Improved CH selection procedure.<br>(ii) Prolonged network lifetime compared to LEACH. |
| M-LEACH Multi-hop LEACH [20] | (i) M-LEACH adopts multihop routing between CH and BS via other CHs in a homogenous multihop network.<br>(ii) Other CHs in the route act as relay nodes to convey the information. | Longer network life time compared to both LEACH and E-LEACH protocols. |
| LEACH-ME LEACH-Mobile-Enhanced [21] | (i) LEACH-ME protocol considers homogenous location-aware network that supports mobility. CH rotation/election is based on a membership function (MF). A node with least MF, minimal transition count and energy level above a defined threshold is selected as CH.<br>(ii) Nodes detached from a cluster due to mobility get connected to other clusters.<br>(iii) Remoteness is determined using reference point group mobility model. | Improved communication over LEACH-Mobile. LEACH-ME is useful particularly when the node speed and angular deviation from the current state are unpredictable. |
| iHEED (integrated HEED) [22] | (i) Suitable for both source and data driven applications.<br>(ii) Follows integrated data aggregation where each node aggregates its own data and then routes the aggregate to the CH, either directly or through a parent.<br>(iii) Parent selection module estimates the link cost for each neighbor node, based on its proximity to the BS.<br>(iv) Communication quality is determined by data losses and link symmetry. | (i) Reduces contention on communication channels.<br>(ii) Network lifetime is prolonged due to smaller clustering interval.<br>(iii) Minimal clustering effect.<br>(iv) Periodic reclustering delays first node death by pushing each node in and out of the routing overlay. |
| hetHEED-1, 2, 3 [23] | (i) CH is determined based on node residual energy and (ii) Cluster node density.<br>(ii) In hetHEED-1, all sensor nodes have uniform energy levels. In hetHEED-2, the nodes have two energy levels while nodes in hetHEED-3 have three energy levels. | The network lifetime of hetHEED-3 is > than hetHEED-2 > than hetHEED-1 with minimal energy dissipation. |
| hetHEED-FL-1, FL-2, FL-3 [23] | Fuzzy logic approach is used to determine a CH by considering: (i) residual energy of each node, (ii) hop distance from a node to BS, and (iii) node density. | Precise information of the residual energy status is not necessary to determine the CH. hetHEED-FL-3 provides the longest lifetime amongst the three variants. |
| Modified Leach [24] | The modified LEACH discovers its neighbors and then forms clusters. The maximum number of cluster members is set to a desired value which is equal to the node degree of CH. | The modified LEACH can adapt to changes in network and hence, the nodes can adaptively change their transmission range to ensure the network connectivity. |

Table 4: Continued.

| Protocol | Description, features | Advantages/limitations |
|---|---|---|
| Energy Efficient Modified LEACH [25] | Unlike traditional LEACH, a CH can again assume the role of CH if its energy is more than a predefined threshold energy limit. A threshold energy limit is introduced for CH selection. The nodes can bid for CH selection by generating a priority value. | The energy efficient modified LEACH was found to be more effective than LEACH for IoT applications. |
| FOI-LEACH (field observation instruments LEACH) [26] | FOI-LEACH is primarily developed for improving the routing amongst field instruments. The protocol is applicable to heterogeneous networks, with rechargeable field instruments. CH election is based on residual energy, rechargeable energy of nodes and the proximity with BS. The FOI-LEACH alleviates the "hotspot" problem, to extend the lifetime of network nodes. | This protocol can be made directly applicable to IoT devices to enhance their routing capability. |
| S-LEACH (sectored LEACH) [27] | Divides the network area into sectors. Sectored network shrink's transmission distance, limits sensing and transmitting area, and provides equal energy consumption distribution for CH.<br>(i) Self-organized CH selection based on the node's residual energy without BS contribution reduces network overload.<br>(ii) Minimized transmission distance reduces energy consumed and prolongs network lifetime. | AI (particle swarm optimization (PSO) algorithm and genetic algorithm (GA)) techniques can be used for nodes distribution to improve IoT network life-time and attain high packet delivery ratio. |

shown great promise for data aggregation in IoT. However, energy efficiency and recovery fidelity are often compromised. Therefore, optimal data aggregation aimed at maximizing IoT network lifetime by minimizing constrained on-board resource utilization continues to be a challenging task. To address this, Amarlingam et al. [62] have developed a Light Weight Compressed Data Aggregation (LWCDA) algorithm, which fragments the entire network into non-overlapping clusters. Clustering results in localized data compression which reduces the number of data transmissions by the cluster. As the number of IoT devices connected through internet tend to become large, the task arrival rates are random and intermittent. To study this, Metzger et al. [63] presented a survey of various approaches to map common IoT network properties to different types of network traffic. Further, they derive guidelines for assuming Poisson process for aggregated periodic IoT traffic.

As IoT sensors pervade all walks of life, the network density increases, and humongous amount of data is generated, which also includes large chunks of redundant data. Under this scenario, one of the main impediments to the development of efficient data aggregation algorithms is to aggregate nonredundant data. Idrees et.al [64] developed a two-level data aggregation mechanism IDiCoEK using divide-conquer algorithm to eliminate the redundant data. In this approach, the CH applies enhanced K-means approach that weeds out redundant data and determines the best representative data to be aggregated.

Constrained by limited data storage space, IoT devices tend to rely on cloud network for data storage, which results in higher transmission overheads leading to faster depletion of energy [4, 65]. To overcome these constraints, Edge computing offers localized access to data to offset transmission delays between the IoT device and the cloud. Edge computing in IoT facilitates: (a) decentralized data aggregation by the IoT/sensor devices, and (b) localized computation at the IoT device. Such arrangement, on one hand helps better management of resources by the end-users, and on the other hand, it ensures uniform availability of resources across all IoT devices and users. Further, when such resources are brought to the edge, the IoT/sensor devices expend less energy for activities such as data aggregation and communication [66].

Ghosh et.al [67] combined edge and cloud computing platforms to reduce data transfer overheads and delay latency through edge computing, while taking advantage of cloud for ML and DL for Human Activity Recognition (HAR) application. Instead of waiting for the arrival of all data and then performing aggregation, sliding window concept is used, which significantly reduces latency and energy consumed.

## 5. Data Aggregation Protocols Based on Interference

When two or more transmitter nodes within the radio interference range of a receiver node, make simultaneous transmissions, the receiver node receives distorted/corrupt or interfered signal [68–70]. In this Section we outline two most popular approaches (a) Protocol and (b) Physical, that model the interference phenomenon in WSN and present a review of other data aggregation approaches that attempt to mitigate the impact of interference. As IoTs tend to be a conglomeration of several WSNs, the data aggregation schemes developed for WSN to address interference phenomenon, are also of interest to IoTs. In this context, some

Table 5: Security in data aggregation protocols.

| Protocol | Description, features | Applicability to IoT |
|---|---|---|
| RPIDA [46] (Recoverable rivacy-preserving Integrity-assured Data Aggregation) | (i) Combines pH and aggregate HMAC techniques along with data aggregation to provide both end-to-end data privacy and data integrity for data aggregation.<br>(ii) Handles false data injection in cluster WSN at sensor nodes and CHs aggregators. | Prevents eavesdropping in IoT. It can also be used to detect malicious sensor devices. |
| Sen-SDA [47] (Secure Data Aggregation): Supports cluster based three-tier heterogeneous topology and follows asymmetric cryptosystem. | (i) The CH collects time-stamped CT-signature pair from its member nodes, verifies the timestamp and signature, and transmits time-stamped aggregates to BS.<br>(ii) HE and IBS ensure end-to-end confidentiality and hop-by-hop authentication, respectively.<br>(iii) CHs and BS execute batch verification using binary quick search (BQS) to detect invalid signatures. | High communication and computation overheads.<br>The rigid three-tier topology may not be suitable for IoTs. |
| SESDA [48] (Secure Energy-saving Data Aggregation): Supports heterogeneity with. High-end CHs and low-end member nodes. | (i) CH shares separate symmetric keys with its member nodes and the BS.<br>(ii) BS decrypts CT and extracts sensed data from aggregates after verifying its data integrity.<br>(iii) CHs employ MAC to filter out bogus/false data packets.<br>(iv) Implements Okamoto-Uchiyama HE to provide end-to-end data confidentiality against adversary attack, viz., eavesdropping, replay, injection, unauthorized aggregation, and CH attacks. | SESDA is applicable for large-scale WSNs, and hence, can be employed for IoT applications.<br>Reduces decryption delay. |
| MODA (multifunctional secure data aggregation) [49] supports tree topology. Uses differential encoding and applies asymmetric key EC-EG based additive HE for secure aggregation. | (i) Suitable for data mining WSNs that require secure data aggregation.<br>(ii) Encoding of raw data into vectors preserves value, order, and context.<br>(iii) Homomorphic encryption enables CT aggregation and end-to-end security.<br>(iv) RODA (enhanced RandOm selected encryption based data aggregation), a variant of MODA, has reduced communication cost and relatively low security.<br>(v) Another complementary of MODA, the compression-based data aggregation (CODA) also reduces communication cost but the aggregation accuracy is low. | The data mining feature of MODA makes it attractive for IoTs.<br>However, the trade-off between security, communication cost, and accuracy need to addressed. |
| LBOA (Location based Secure Outsourced Aggregation for IoT) [50] | (i) Used in location-critical applications where data aggregation at each location is outsourced and the application demands privacy and confidentiality of location as well as the location strategy.<br>(ii) Public-key encryption is used to protect the privacy of location and order-preserving encryption the confidentiality of user's location strategy. | LBOA can readily be adopted for IoT applications.<br>The security features satisfy the privacy and confidentiality requirements. |
| CBDA [51] (Chain-Based Data Aggregation) | (i) Sensor nodes are organized as a tree topology. Leaf nodes reconnect sequentially with each other to form chain topologies.<br>(ii) Tail nodes of the chain provide data privacy to collected data by slicing them into fragments. Mitigate risks of sliced data by injecting fake fragments to interfere with adversaries.<br>(iii) Its semihonest model can withstand eavesdropping and tolerate collusion attacks. | Applicable to IoT and IoT to achieve privacy-preserving data aggregation |

Table 5: Continued.

| Protocol | Description, features | Applicability to IoT |
|---|---|---|
| FESDA (Fog-Enabled Secure Data Aggregation in Smart Grid IoT Network) [52] | (i) Fog nodes (FN) are employed to aggregate HE data from SMs.<br>(ii) FN use HMAC secret key for each SM, to ensure data integrity and source authentication.<br>(iii) Paillier cryptosystem is used to prevent FNs from extracting user consumption data.<br>(iv) FESDA is resilient to false data injection attacks. | FESDA can readily be adopted for IoT applications. |
| ESDTA (Efficient and Secure Data Transmission and Aggregation) [53] | (i) Delimiter based message aggregation and extraction provides enhanced aggregation efficiency and data security.<br>(ii) Employs the SMA algorithm at the Mobile node and SMD algorithms at FN, to ensure secure data aggregation and data forwarding of healthcare parameter values.<br>(iii) Preserves data integrity and protects against threats viz., data fabrication and replay attack.<br>(iv) Computation at FN minimizes storage and computational cost at the cloud server. | Used in internet of medical things (IoMT) for remote health monitoring. |
| EEDAM (Energy-Efficient Data Aggregation Mechanism) [54] Decentralized cluster topology and blockchain secure edge services with minimum delay | (i) Member nodes for clustering at the CH are analyzed using a fuzzy similarity matrix.<br>(ii) Member nodes employ sleep scheduling to reduce data redundancy, network traffic jamming, and transmission costs.<br>(iii) Fuzzy based data aggregation in the IoT layer.<br>(iv) Uses edge computing to provide on-demand trusted services to IoT.<br>(v) Cloud server has integrated blockchain and the edge is validated by the blockchain. | It can be used in 6G to ensure reliable transmission of sensor elements for IoT system |

∗ HE- Homomorphic Encryption, IBS- Identity-Based Signature, pH- privacy homomorphism, MAC- message authentication code, HMAC- Hash-MAC, CT- ciphertexts, EC-EG - elliptic curve ElGamal, SM - Smart Meter, SMA-Secure Message Aggregation, SMD- Secure Message Decryption.

of the data aggregation schemes developed for WSN to address the interference phenomenon, which are also of interest to IoTs are reviewed in this section.

### 5.1. Review of Data Aggregation Techniques to Address Interference in WSN.

The primary objective of most approaches that model interference is to establish basic conditions for successful communication between two nodes in terms of: (a) the Euclidean distance between the receiver and an unintended transmitter, and (b) the SINR threshold at the receiver end. The two most popular approaches to model interference are (a) Protocol Interference model, based on Euclidean distance between two communicating nodes and (b) Physical Interference model, based on SINR.

The Protocol Interference (PRI) model defines a protocol for identifying the source nodes, whose transmissions can potentially interfere with the radio transmissions of an unintended receiver node. PRI is a graph-based model where a link between any two nodes can be established if the Euclidean distance between them is $\leq R$, where $R$ is communication range. In a Unit Disk Graph (UDG) of radius $R$, the communication between nodes $n_1$ and $n_2$ is successful if there is no other transmitting node within a certain interference range $R_I$ (where $R_I \geq R$) from $n_2$.

The Physical Interference model assumes that all nodes in the network possess same transmission power $P$. If a node $S_1$ transmits with power $P$ to node $S_2$, the power of the transmission received by $S_2$ is given by: $P_{S_2}(S_1) = P * \min\left(1, \|S_1, S_2\|^{-\alpha}\right)$, where $\|S_1, S_2\|$ is the Euclidean distance between $S_1$ and $S_2$, $\alpha \geq 2$ is the path-gain exponent. The path gain from node $S_1$to $S_2$ is generally set as $\min\left(1, \|S_1, S_2\|^{-\alpha}\right) \leq 1$. The transmission from $S_1$ to $S_2$ is successful if and only if the SINR (signal-to-noise plus interference ratio) at $S_2$ is greater than a threshold $\beta$.

$$\mathrm{SINR}_{S_2}(S_1) = \frac{P_{S_2}(S_1)}{N_0 + \sum_{S_i \epsilon N'} P_{S_2}(S_i)} \geq \beta, \tag{1}$$

where $\beta$ (threshold SINR) >0, $N_0$ $\geq 0$ is the varying background noise and $S_i$ is the set of transmitting nodes $N'$ other than $S_2$ i.e. $(N' \epsilon (N - S_2))$, that transmit in the same time slot as that of $S_1$.

The Protocol and Physical Interference models, triggered the development of several models on how to deal with interfering transmissions in WSN. However, not much effort has been devoted toward developing models that effectively address: (a) Exposed Station (ES) problem, (b) Hidden Terminal (HT) problem, (c) identifying the links that can potentially interfere with communication between neighboring nodes, and (d) devising schemes for Interference-Fault Free Transmission (IFFT) schedule. To tide over these

Table 6: Mobility related data aggregation protocols.

| Protocol | Features | Limitations/advantages, applicability for IoT |
|---|---|---|
| EEMSRA [55] (Energy-Efficient Mobile Sink Routing Algorithm) Hierarchical, cluster. Multihop. Controlled sink mobility. | (i) Cross-layer protocol that operates in coordination with the MAC layer and is based on LEACH. The Mobile sink broadcasts it's next projected cluster visit in order to enable the network to update routes prior to the sink's actual arrival at the cluster.<br>(ii) CHs perform aggregation before transmitting data to mobile BS. | (i) Avoids the bottleneck problem of funnel model.<br>(ii) Requires the BS to have knowledge about its short-term trajectory.<br>(iii) Has better energy efficiency than LEACH.<br>(iv) Need for controlled BS mobility is a limiting factor, and hence may not be applicable for IoT. |
| HexDD [56] (Hexagonal Cell-Based Data Dissemination) Hierarchical, Homogeneous (neighborhood). Random sink mobility. Hexagonal grid.<br>Multiple sink support. | (i) Constructs a virtual grid infrastructure with highways in a honeycomb tessellation.<br>(ii) Query matching invokes data transmission in reverse path to BS. The forwarding paths along the diagonals are shared by all source-BS pairs. It allows similar data to meet at the common border nodes.<br>(iii) Data from multiple sources can be aggregated and replaced by a single data packet and forwarded towards BS. | (i) Suitable for intelligent sensor network applications.<br>(ii) Honeycomb architecture keeps the traffic flow in all regions of the network nearly balanced.<br>(iii) Replicating data on the border cells decreases the cost of data look-up and the data delivery latency.<br>(iv) Nodes lying on borderlines and on the center cell suffer from hotspot problem. This can be avoided by adjusting the size of borderlines and shape of central region based on network size and traffic.<br>(v) Applicable for IoT. |
| BRH-MDG [57] (Bounded Relay Hop Mobile Data Gathering) Mobile node/collector.<br>Controlled movement of mobile node. | (i) Data is gathered by a mobile node/collector by resorting to polling mechanism, where a mobile collector starts its tour from a static BS located inside or outside in the sensing field.<br>(ii) PP temporarily buffers the aggregated data and uploads the aggregates upon the arrival of mobile collector before forwarding to the BS. | (i) A tradeoff need to be arrived between energy saving and data gathering latency by striking a balance between relay hop count of data aggregate and the tour length of the mobile collector.<br>(ii) Buffer overflow problem occurs due to delay in the arrival of the mobile collector to polling points.<br>(iii) The need to have a mobile collector makes it unsuitable for IoT. |
| TCBDGA [58] (Tree-Cluster-based Data-Gathering Algorithm)<br>Multiple trees rooted at RP. Mobile BS starts periodically and has unlimited energy. | (i) Weighted trees are constructed by taking into account residual energy of 1-hop neighbor, the number of its 2-hop neighbors, and the distance to the BS.<br>(ii) The mobile BS stop at at some locations named Rendevous points (RP) and sub-RPs to collect and aggregate the data from the neighboring nodes. The RPs are reselected after a certain data collection rounds/period. | (i) Balances evenly the load of entire network, reduces the energy consumption, alleviates the hotspot problem, and prolongs the network lifetime.<br>(ii) Sub-tree decomposition into sub-RPs and normal nodes balances network energy consumption.<br>(iii) The need for mobile BS makes it unsuitable for IoT. |
| TTDD-QL [59] Grid structure, multicast routing, multi-hop, mobile sink, hierarchical, Sensor nodes Proactive. | (i) TTDD-QL is based on Q-learning to find most energy efficient path from dissemination node to the BS. | (i) Two-level aggregation reduces communication overhead between dissemination nodes (higher tier) and BSs (lower tier). It supports mobility of BSs.<br>(ii) Soft-states timer installed at dissemination nodes balances the overhead of periodic upstream update messages generation.<br>(iii) Ability to disseminate information across multiple BSs, makes TTDD a good candidate for use IoT. |

Table 6: Continued.

| Protocol | Features | Limitations/advantages, applicability for IoT |
|---|---|---|
| MSRP [60] (Mobile Sink based Routing Protocol) Cluster, stationary sensor nodes, between sink and CH. | (i) Mobile sink in first movement cycle determines CHs that are closer than a specified distance threshold. For subsequent cycles the residual energy of CHs maintained in cluster head residual energy table (CHRET) by the sink, is considers.<br>(ii) Mobile BS visits the CHs with higher energy to gather their aggregates. | (i) Energy efficient and extends network lifetime.<br>(ii) Suitable of delay-tolerant applications.<br>(iii) Do not guarantee that mobile BS visits all the CHs within a time bound.<br>(iv) Not applicable for IoT. |
| TSVA-CP-ABE scheme (time-sensitive and verifiable data aggregation) [61]<br>IoT data aggregation system | (i) Attribute-based encryption to achieve efficient access control in edge-assisted mobile crowd sensing.<br>(ii) IoT devices perform outsourced computing and edge nodes perform verification and filtration of aggregated data.<br>(iii) A mobile crowd sensing platform (MCP) encrypts the time-sensitive task data as CTs.<br>(iv) Edge nodes perform aggregation on collected CTs. | (i) Suitable for edge-assisted mobile crowd sensing where mobile devices are equipped with smart sensing computing and communication capabilities, and context-aware applications. |

issues, Beneyaz et al. [70] develop a new holistic framework - the Composite Interference Mapping (CIM) model that maps the nodes that potentially interfere with each and every node in the network. Armed with the potential interference map, all active links can coordinate to schedule their transmissions so as to get over the ES and HT problems, and also determine an IFFT schedule.

Due to the presence of intermediary nodes in a tree topology rooted at the BS, the transmission of aggregated data need to be carefully coordinated to avoid interference from other concurrent transmissions. The Minimum Latency Aggregation Scheduling (MLAS) algorithm aims to minimize the effect of wireless interferences and the number of time slots required to aggregate the data [71]. Several data aggregation scheduling algorithms have been developed using either Protocol or Physical Interference models or by using a combination of both, to address interference in WSN and IoTs [72]. Algorithms developed on the concepts of link/node coloring, dominator-dominatee, nearest-neighbor, link-length diversity, competitor nodes in the interference range, etc., have emerged as the most popular approaches. We briefly discuss of some of these algorithms in this section.

For physical interference model, Li et al. [73] have proposed a time-efficient data aggregation distributed MLAS algorithm, where the latency depends on network radius or depth of the aggregation tree, and node degree of the communication graph. The MLAS takes into consideration primary interference problem by allowing only disjoint set of links. This work assumes that all nodes transmit with a constant power $P$. The power of the signal received by a node $x$ from a transmitting node $y$ located at a distance $r$ is given by $\mathscr{P}_x(y) = P * \min\left(1, r^{-\alpha}\right)$ where $\alpha$ is the path gain exponent. This expression indicates that the power of received signal decreases as the distance increases. If $N_0$ is the variance in background noise, the transmitted signal is successfully received if $SINR \geq \beta$ (a threshold value) and the distance $r = (P/\beta N_0)^{-1/\alpha}$. A network of $n$ nodes is modelled as a graph $G(n, \delta r)$ with communication links of Euclidean distance $\leq \delta r$, where $\delta \, \epsilon \, (0, 1)$, and only nodes that are within a distance $\leq \delta r$ are allowed to communicate with each other. If the transmission link $> \delta r$, the probability of interference with other transmissions taking place in the same time slot is high. The scheduling algorithm developed in this paper is based on distributed synchronous message passing model with unicast communication and it ensures that every datum is aggregated only once without violating the SINR. The MLAS employs a link coloring scheme to determine sets of disjoint transmitting nodes, and identifies noninterfering nodes in a given timeslot from a CDS-based aggregation tree discussed in [74]. To further improve the performance of the algorithm and reduce/compress the scheduling latency of MLAS, a Compressed Scheduling algorithm is also developed. In this algorithm, the scheduled links from previous time-slots are merged into a single slot, such that there is no interference amongst the scheduled links. The authors employ an approach to generate a subset of feasible set of links using path scheduling algorithm as discussed in [75]. The subset is then expanded by adding noninterfering nodes that obey the SINR-threshold. Merging multiple links into a single time slot not only reduces overall number of timeslots required for aggregation but also lowers the latency compared to the distributed algorithm.

Two interference-free, TDMA scheduling approaches: Centralized Improved Aggregation Scheduling (CIAS) and DIAS (Distributed-IAS), to minimize delay during data aggregation under protocol interference are discussed in [76]. As per the approach, the network topology is organized as a CDS or Cluster-based data aggregation tree rooted at the topology/network center. The CDS acts as the backbone of the network. In the CIAS approach, data from dominatees are aggregated by their dominators. Aggregation process progresses level-by-level in a bottom–up manner between dominators within two hops level, which are connected by connectors. The connectors collect aggregated data from the dominators in the lower level, and transmit the aggregates to the dominator nodes in level above. As CIAS approach depends on the CDS tree structure, it may not be amenable to changes in networks that need to accommodate dynamic topologies. To address this, a distributed approach that allows aggregation of data in an interleaved manner was presented. In the first step the data from dominatees is aggregated by their dominators. Later, the aggregates of dominators in lower level are aggregated by their higher level dominators at two-hop distance. The aggregator node at higher level, then schedules its transmission greedily after it has aggregated data from all its child nodes, without waiting for its same-level neighbors to complete their aggregation. This interleaved mechanism increases the number of simultaneous transmissions and minimizes aggregation delay. Every node in the CDS tree maintains (i) set of competitor nodes that are in its interference range, and (ii) ready competitor set, a subset of competitor nodes, which are active. Based on this information, a node is allowed to transmit its data only if its transmission does not interfere with the set of ready competitor nodes. It is critical to understand the limit of many-to-one information flows, and devise efficient data collection algorithms to improve the performance of WSN. Chen et al. [77] present an approach to establish theoretical upper and lower-bounds for evaluating *data collection* capacity. Data collection capacity quantified in bits/sec, is a measure of how fast the BS can collect non-interfered data from its sensor nodes. To establish the bounds, the approach considers different interference models such as Protocol and Physical Interference models, disk graph and Gaussian Channel models. The approach assumes that: (i) there is no spatial correlation amongst the sensed data, and (ii) the transmission and interference ranges of nodes are identical. Further, the models assume that packet size ($b$ bits) and transmission rate ($W$) of packets in a given transmission path are fixed. The nodes can resort to concurrent transmission only if they are spatially separated, and their transmissions do not cause any interference. For every given path $P_i$, the number of slots $\tau_i$ required to collect one data packet at the BS is equal to the product of the path length and its maximum interference number $\Delta_i$ (the number of nodes in the interference range), *i.e.*, $\Delta_i|P_i|$. The number of time slots $\tau$ required to collect data from all nodes along all paths is given by$\Delta n$, where $\mathit{\Delta}$ *is* max $\{\Delta_1, \Delta_2, \cdots, \Delta_c\}$ and $c$ is the number of leaf nodes in the tree. As the length of each slot $t = b/W$, the overall delay encountered in data collection along all branches is,$D = \tau t$ and the data collection capacity $C$ is given by: $C = nb/D = W/\Delta$. Therefore, the lower bound for capacity is $\Theta\,(W/n)$ and upper-bound is $\Theta\,(W)$.

The MLAS problem under the physical interference model is also investigated in [78]. Two MLAS algorithms a centralized Nearest-Neighbor Aggregation Scheduling (NN-AS) and a distributed Cell-Aggregation Scheduling (Cell-AS) scheduling were developed. The centralized NN-AS algorithm is proposed as a benchmark for distributed MLAS algorithm. NN-AS constructs the aggregation tree in a phase-by-phase manner by determining the transmission set at each phase. At each round of aggregation, a node in the transmission set identifies its nearest neighbor that is not connected to other nodes; and establishes a link with it. The links thus formed, are scheduled using non-linear power assignment. At the end of each round, the nodes that have transmitted are removed from the transmission set. This process is repeated with the reduced node set using the nearest neighbor criterion until a single sensor node is left, which then transmits the aggregate data to the BS in a single hop. For networks where centralized approach is not practical, a distributed Cell-AS algorithm, which divides the network into hexagonal cells is employed. For any given cell, the node closest to the BS is selected as the Head Node (HN). Aggregation is performed by the HN by extracting the data from its neighbors with the help of *pulling* mechanism. After aggregation the HNs pool up into larger hexagonals until the entire area is covered. Unlike NN-AS algorithm, the Cell-AS algorithm strategically divides the network based on link-length diversity. Therefore, it obviates the need to possess global interference information for scheduling. Various popular works on handling interference in data aggregation are summarized in Table 7.

*5.2. Review of Data Aggregation Techniques to Address Interference in IoT.* As IoT embraces different types protocols, technologies and topologies; the impact of interference is quite significant in data aggregation. Unlike WSN where all sensor nodes operate on the same network, in IoT, the devices may operate on technologies that use different frequency bands on the RF spectrum. At the same time, some wireless technologies such as Bluetooth, WiFi, and ZigBee may operate in the same 100 MHz bands in 2.4 GHz to 2.5 GHz range. In such cases, a WiFi network can interfere with another neighboring WiFi network, even if the two networks are not in direct communication with each other. In addition, any neighboring non-IoT device (e.g. a microwave) utilizing the same frequency as that of an IoT device can interfere with the WiFi network. Similarly, there may be several proprietary products with unique, non-standard frequencies that can be the potential sources of RF interference. A major challenge is to detect and diagnose interference remotely. Additionally, when IoT devices cannot communicate with one another due to interference issues they try to establish communication repeatedly, and multiple such attempts lead to battery drain-out. One solution to manage interference is such cases is to use IoT systems only in cellular bands where the RF environment is well planned and coordinated. However, such systems are more expensive than those deployed in the unlicensed bands. A second solution is to use IoT systems that are well-separated in frequency.

Issues due to limited radio frequency spectrum and bandwidth availability, etc., force the IoT devices to share narrow spectrum with overlapped frequencies. The issue is further compounded by the fact that in some applications, the IoT devices might be located in close proximity to one other. In such scenario, the communication between the IoT devices is highly prone to interference. To counter this,

Table 7: Data aggregation protocols based on interference models for WSN and IoT.

| Interference model | Algorithm complexity and description |
|---|---|
| Wan et al. [82]<br>(i) Follows protocol interference model MLAS in synchronous multihop networks.<br>(ii) Constructs three centralized data aggregation schedules, SAS, PAS and E-PAS with same transmission and interference range $\rho = 1$. | (i) Upper bound latencies of the three algorithms are:<br>(a) SAS (sequential aggregation scheduling): $15R + \Delta - 4$,<br>(b) PAS (pipelined aggregation scheduling): $2R + O(\log\ R) + \Delta$,<br>(c) E-PAS (enhanced-pipelined aggregation scheduling): $1 + O(\log\ R/3\sqrt{R}))R + \Delta$.<br>Where $R$ is radius and $\Delta$ is maximum node degree.<br>(ii) The algorithms implement CDS for routing by constructing maximal independent set (MIS) induced by a BFS ordering of vertices. |
| Li et al. [73]<br>(i) Follows physical interference model.<br>(ii) MLAS algorithm constructs a BFS data aggregation tree and prepares a collision-free schedule. | (i) The latency of MLAS schedule is bounded by $O(R + \Delta)$ time-slots.<br>(ii) CDS tree used for routing by constructing MIS induced by a BFS ordering of vertices.<br>(iii) Deployment pane is partitioned into grids and a link coloring approach is followed to resolve interference. Latency of scheduling is further improved by compressive scheduling which merges links scheduled in different slots to a single slot without violating SINR. |
| Li et al. [83]<br>(i) Follows physical interference model.<br>(ii) Presents two algorithms for MLAS<br>(iii) Cell-AS, and NN-AS use distributed and centralized data approaches, respectively. | (i) Latency for cell aggregation scheduling cell-AS for arbitrary WSN topology is $O(K)$ time slots, where $K$ is the logarithm of the ratio between the lengths of the longest and shortest links in the network.<br>(ii) Cell-AS exploits link diversity combined with coloring of cells to avoid interference and to minimize the aggregation latency.<br>(iii) Latency for nearest-neighbor aggregation scheduling (NN-AS) is $O(\log^3 n)$ time slots (where n is the total number of nodes). The deployment pane is divided into hexagonal cells. |
| Orsson et al. [84]<br>(i) Follows physical interference model.<br>(ii) Focuses on connecting arbitrary point set into a strongly connected diagraph. | (i) The algorithm "schedule" connects the arbitrarily oriented MST in O (log n) slots.<br>(ii) The algorithm "schedule" considers both unidirectional and bidirectional (half-duplex) communication between the nodes of a link in the same slot. |
| Yousefi et al. [85]<br>(i) Follows protocol interference model.<br>(ii) Considers fixed tree-based topology and connected network.<br>(iii) Minimizes time latency by generating a collision-free schedule with least number of time slots. | (i) FAST (collision-free minimum latency aggregation scheduling algorithm for tree-based WSNs) generates aggregation schedules under the consideration that collisions generally occur at the receivers.<br>(ii) Latency of distributed TDMA-based FAST is upper-bound by $12R + \Delta - 2$<br>(iii) To avoid collision, the parents schedule transmissions based on negotiation with each child node and decide time slots for transmissions according to their priorities.<br>(iv) FAST adapts waiting policy, where each node in the backbone waits for the arrival of the target schedules from all its higher-ranked neighbors in order to successfully determine the applicable transmission slots for its children. FAST outperforms Clu-DDAS as the network density increases. |
| Bushnaq et al. [86]<br>Follows physical interference model and slotted ALOHA | (i) A data aggregation algorithm is developed for data collection by an unmanned aerial vehicle that hovers over a finite region of interest.<br>(ii) With hovering and travelling times as tradeoffs, the algorithm heuristically arrives at an optimum number of hovering locations, data collection time and number samples to be collected, without compromising the data accuracy. |

Table 7: Continued.

| Interference model | Algorithm complexity and description |
|---|---|
| Nabi et al. [87]<br>(i) Follows physical interference model.<br>(ii) Spatial & temporal models to characterize SINR distribution in large-scale grid-based IoT networks with synchronous periodic traffic. | (i) IoT devices are modelled as spatially interacting phase-type arrival/departure (pH/pH/1) queues for packet generation, transmission scheduling, and rate-sensitive SINR-based packet departure.<br>(ii) The model considers the impact of inter-device spacing, directional antenna & radiation pattern, packet sizes, power control and data transmission rate on data aggregation by arriving at optimal transmission reliability by considering data granularity, transmission delay as some of the trade-off parameters. |
| Fitzgerald et al. [40]<br>(i) Follows physical interference model.<br>(ii) Formulates 1 K model data collection and nK model for joint optimization of data aggregation and dissemination. | (i) Uses mixed-integer programming (MIP) to arrive at energy-optimal data aggregation and routing of sensor measurement data in IoT edge networks.<br>(ii) Optimization of network accounts for energy cost in terms of both minimal total energy usage, and min-max per-node energy usage (computation of aggregation functions).<br>(iii) The algorithm however, does not account for reliability of IoT systems in terms of redundancy of data transmitted over multiple destination nodes and multiple redundant paths. |
| An et al. [71]<br>Collision (interference) model similar to protocol model | (i) A constant-factor approximation algorithm (ASIoT-aggregation scheduling in IoTs) is developed to address the MLAS problem in IoTs.<br>(ii) The algorithm builds a data aggregation tree from connected dominating set. It then schedules dominatees first, and repeatedly schedules dominators and connectors until all nodes in the aggregation tree are scheduled.<br>(iii) Latency is approximated to be $\leq \Delta - 1 + 15D$, where $D$ is network diameter $\leq 2R$. |
| Cai et al. [80]<br>Follows protocol interference model | (i) The algorithms developed achieve reduction in latency through (a) even distribution of aggregator nodes so as to maximize the coverage and (b) determination of collision–free communication schedule, in battery-free wireless sensor networks (BF-WSNs) and IoTs.<br>(ii) The algorithms are extended for the study of BF-WSNs with multiple channels. |

it is important to devise Interference Management Mechanisms (IMM) that eliminate the narrowband IoT (NB-IoT) and move toward long-term evolution (LTE) systems. Furthermore, as IoTs comprise several heterogeneous devices and networks that use different protocols, the data aggregation approaches developed for WSN are not directly applicable to IoT. The IMMs can be broadly classified into two categories: (a) IMMs that use resource partitioning to isolate mutually interfering transmissions, and (b) IMMS that use signal processing strategies to facilitate concurrent transmission of multiple interfering signals. While the first category may lead to poor spectrum efficiency, the second category demands accurate Channel State Information (CSI) to implement the IMM. To counter this, a novel Interference Steering (IS) method is introduced [79] where a steering signal is generated to steer the interference imposed orthogonal to the original intended signal and thus neutralizing the effect of interference on the intended receiver.

Fitzgerald et al. [40] present a Mixed-Integer Programming (MIP) formulation for providing energy optimal routing and data aggregation at multiple BSs in IoT edge networks. Further, the MIP formulations assess the minimum total energy and min-max energy required for data aggregation and dissemination. The approach also presents schemes for scheduling of transmissions under Physical Interference model to attain optimized throughput. The MLAS problem in IoTs is addressed in [71]. A constant-factor approximation Aggregation Scheduling in IoTs (ASIoT) is developed to manage scheduling of heterogeneous devices. ASIoT employ Collision (Interference) model to identify conflicting nodes. The aggregation tree is in the form of CDS. The dominatees in the aggregation tree are scheduled using modified first-fit scheduling algorithm. Both the dominators and connectors connecting the dominators schedule their aggregated data using level-based scheduling. Theoretical analysis of ASIoT algorithm generates a schedule with latency $\leq \Delta - 1 + 15 * D$, i.e., $\leq \Delta - 1 + 15 * 2R$, where $D \leq 2R$ is network diameter and $R$ is network radius.

Battery-Free Wireless Sensor Networks (BF-WSNs) are increasingly becoming popular for IoT applications. A popular MLAS approach that addresses coverage requirement and latency reduction in BF-WSNs; is to randomly choose *q* percent of nodes for communication and aggregation purposes, and select Aggregator Nodes (AN) based on their residual energy. However, this approach does not guarantee equitable distribution of ANs across the network, especially in applications where IoT device-density is non-uniform. Such non-equitable distribution could leave some nodes orphaned, while many others may end up being squeezed to a limited number of aggregators. To counter this, two scheduling algorithms based on bottom-up and top-down approaches that guarantee equitable distribution of aggregators was proposed in [80]. The theoretical analysis and simulations results suggest that the algorithms perform relatively better in terms of latency. Data Aggregation for IoT networks are often mired with issues that demand adaptability to dynamic requirements of the network, particularly synchronization. This is addressed by deploying self-configurable sensors that perform time synchronization on the data sensed. However, if the volume of data generated is quite large, time synchronization turns out to be extremely complicated. To counter this, mechanisms that determine the optimal sample size and sample probability for calculating approximate values need to be designed [81].

## 6. Fault Diagnosis and Fault-Tolerant Mechanisms

*6.1. Fault Detection and Diagnosis.* A probabilistic distributed localized fault detection algorithm to identify the status of faulty sensors is described in [88]. Locally, each sensor node in the network identifies its own status and categorizes as, *good* or *faulty*, and based on this assessment, the neighbors of the node either support or oppose the claim. The approach focuses on hardware faults such as calibration systematic error, random noise error, etc., with an assumption that the system software and application software are not prone to faults. A similar approach is discussed in [89] using *judgment* principle [88] to judge the status of other sensors through parent-child relationship.

In WSN applications involving large geographic spread, a centralized approach to detect faults may lead to degradation of service. In such applications faults are diagnosed by following a distributed approach where the sensed data and the data received are compared for deviations. To validate sensor readings, a BIT (Built-In Test) diagnosis method with spatial correlated weighted adaptation is usually adopted. Whilst BIT methods are used to detect hard faults (readings beyond operating rage), methods that additionally involve spatial correlation with weighted adaptation are used to detect soft faults (inrange or slow drift faults). Depending on how close the sensor readings are to the minimum or maximum operating range, the BIT method determines the performance degradation of sensors.

*6.2. Fault-Tolerance.* Larrea et al. [90] present three fault-tolerant hierarchical data aggregation algorithms to address intra-region and inter-region process faults. The algorithms consider a predefined QoS metric and a battery depletion threshold, for selecting a reliable fault-tolerant aggregator based on distributed Omega Failure Detector Model [91]. In this approach a node that encounters minimum number incarnations (*crashes*) is chosen as a super-aggregator (SA). After every round of aggregation, the energy level of the SA is checked against a *battery depletion threshold* to assess the necessity of electing a new SA. To detect process failures, a *suspect* list of processes that might fail is maintained. The processes in the suspect list periodically broadcast I-AM-ALIVE message in order to proclaim that they have not crashed. A popular approach to address issues such as energy efficiency, integrity and fault-tolerance in a multi-sensor hierarchical clustered WSN; is to assign a predefined weight-factor to each node. In the event of sensor failure or its likelihood, the weight of the node is adaptively decreased. The CH validates the integrity of acquired data and performs a weighted-average data aggregation. This approach however, does not zero-out faulty sensors, but decreases the contribution of faulty sensors in data aggregation. Younis et al. [92] analyze network topology management techniques for tolerating node faults, and suggest the use of techniques to develop robust mechanisms for failure detection, restoration of connectivity and offsetting the effect of node mobility in recovery schemes.

A fault-tolerant data aggregation protocol comprising (i) aggregation scheduling and (ii) amendment strategy is presented in [93]. The data is aggregated according to CDS-based aggregation scheduling discussed in [74]. The data aggregation scheduling consists of two phases. The first phase involves a single-hop aggregation schedule, which aggregates data from leaf nodes (dominatees) to the dominators. The single-hop aggregation schedule is based on Iterative Minimum Covering (IMC). In the second phase, aggregation is carried out layer-by-layer using SAS algorithm [82]. In the event of an intermediate node failure, the amendment strategy is implemented that aims to minimize the number of nodes affected indirectly due to presence of faulty nodes in the network.

Zhang et al. [94] have proposed a TDMA-based fault-tolerant scheduling (FTS), where every node maintains information about its Backup Parent Set (BPS). The data collection process starts by identifying a *start-point* node which is closest to the BS in terms of time slots/hops. A Maximum Non-Interference set of nodes that can transmit concurrently without interference is determined. To identify the status of nodes, a coloring scheme is adopted, where all next hop nodes in the BFS tree are initially are colored white. A node that does not have any packets to transmit is colored black. After every transmission, the color of the node is updated iteratively. For better reliability every successful receipt of data is duly acknowledged by the parent node. To restore communication, in the event of failure of the parent, the child node randomly selects one node from its BPS as its new parent node. However, in the FTS algorithm, the BPS contains only the neighboring nodes that lead to the parent of the faulty node to find alternate path. Therefore, this strategy precludes other alternate paths that might lead to BS.

In the absence of a robust mechanism to detect interference-faults, quite often the receiver node aggregates the distorted or interfered faulty signals that it receives, and relays it to the BS. This leads to propagation of faulty, *interference ridden* signals across the network. The present approaches primarily focus on addressing node/link failures and not on faults induced by signal interference. To this end, Begum and Nandury [95–97] have developed algorithms to determine Interference-Fault Free Transmission (IFFT) schedules. The algorithms attempt to accomplish IFFT by: (a) identifying nodes whose transmissions can potentially interfere with other ongoing transmissions of neighboring nodes, (b) mapping interference effect of each transmitting node on other on-going transmissions in the network, and (c) identifying the constituent transmitting nodes from the received inference-ridden signal. A component-graph based self-healing algorithms were developed in [98] that attempt to restore the overall structure of a hierarchical data aggregation tree, without the use of redundant resources. Besides tolerating interference-faults and node failures, the algorithms maximize the number of transmissions and minimize the energy consumed per round compared to other algorithms.

Large IoTs generally depend on WSNs to gather data in their field of interest. Due to heterogeneity of the sensor nodes and IoT devices, cluster-based routing is preferred compared to individual device-centric routing of information. However, in the event of failure of CH, the data aggregated by this CH is lost. To counter this, Lin et al. [99] utilize virtual CH, which serves as a back-up to all CHs in case of their failure. Flow-graph modeling is then used to retrieve information from the virtual CH.

In several applications, communication between nodes is constrained due to poor or intermittent bandwidth availability. Grining et.al [100] develop a privacy preserving algorithm that utilizes limited communication between the nodes and preserves their privacy. IoT applications that rely on fog/cloud to access data storage and computational resources often encounter faults due to non-reporting of data by an IoT device. In such eventuality, the fault-tolerance feature makes an estimate of the data from past record. This strategy is also used to make data estimates if a node malfunctions or reports false data [101]. As data is shared and aggregated across the IoT devices, issues related to data privacy are of prime importance. Although the data is encrypted to facilitate secure transmission of the data, for better reliability, the data aggregation algorithms need to be fault-tolerant. C. Xu et.al [102] present fault-tolerant privacy preserving algorithm that aggregates time-series data. The algorithm accommodates failure of periodic data uploads from IoT devices and tolerates arbitrary aggregation functions without much loss in accuracy. While utilizing the fog/cloud resources, data aggregation is prone to unsolicited injection of false data. H.M. Khan et.al [103] develop a privacy preserving data aggregation algorithm to safeguard against such FDI (false data injection) attacks in fog-enabled smart grids.

## 7. Requirements and Tradeoffs in Data Aggregation Approaches

Data aggregation approaches in WSN and IoT applications encounter several challenges & requirements related to: (a) maintaining the accuracy of the data aggregates, (b) data corruption due to transmission losses and radio interference, (c) communication delays, (d) network lifetime, (e) energy constraints, (f) temporal data, etc. These challenges primarily arise due to limited computational & communication capabilities, memory storage, and battery-energy of sensor nodes & IoT devices. With little or no access to frequent battery recharge option, the sensor nodes/devices may fail or function erroneously, which might lead to incorrect aggregation of sensed data. Considering the underlying complexities involved in handling these challenges, it may not always be possible to find a unilateral data aggregation scheme that satisfies all these requirements, which at times, may be contradictory to each other. For example, in pursuit of accuracy, if the aggregator has to wait for too long a period till all nodes/devices have transmitted their sensed data to the aggregator, it might lead to poor latency. Similarly, if the data aggregation scheme, while performing aggregation drops data packets corrupted due to radio interference, it might affect both data accuracy and latency parameters. As a consequence, the design of data aggregation algorithms that aim to address these challenges is inherently demanding; as one needs to consider trade-offs amongst various optimization parameters such as latency, data accuracy, reliability, energy efficiency, etc. A thorough understanding of these parameters is necessary to define appropriate optimization functions. Based on these functions, data aggregation strategies can be drawn to bring about the desired tradeoff. Issues related to optimization parameters and tradeoffs are briefly enumerated in this Section.

QoS is an essential requirement to assure guaranteed performance of the identified quality parameters [104]. Due to wide ranging nature of applications, the primary QoS parameters can be classified as: (i) coverage – the number of sensors required and their deployment, (ii) sensing mechanism, (iii) data accuracy, (iv) network life-time, (v) time criticality, and (vi) reliability. The QoS parameters that are of prime importance for data aggregation are latency, timeliness and accuracy. To guarantee data accuracy, the QoS mechanism must ensure data freshness while performing aggregation.

Stankovitch [105] proposed the major research challenges for IoTs, namely: massive scaling, architecture & its dependencies, creating knowledge & big data, robustness & openness, security & privacy, and human-in the-loop. Each of these challenges primarily focus on new problems that arise for future IoT systems. Out of the challenges identified, scaling, architecture/topology and handling of big data are critical for data aggregation in WSN and IoTs.

*7.1. Optimization Parameters.* Most tradeoffs in WSN and IoTs focus on optimization parameters that aim to: (a) accommodate heterogeneity, (b) prolong the network lifetime, (b) enhance data accuracy and energy efficiency, (c) shorter latency and temporal correctness, (d) QoS, etc. Some of these critical parameters are discussed below:

*7.1.1. Accommodating Heterogeneity.* To accommodate the inherent heterogeneity in IoT, while at the same time leverage the homogeneity in WSNs; parameters such as: (a) network topology, coverage and capacity, (b) computational capability, (c) data storage, etc., play an important role in arriving at optimization strategies for data aggregation. While coverage refers the extent up to which a sensor node or an IoT device can reach out to other neighbouring devices, the capacity of a network refers to the amount of traffic that the network can handle. For IoT application, which essentially is a conglomerate of several WSNs with distinct protocols and heterogeneous IoT devices; too large a coverage space of sensors might be counter-productive. Larger coverage might result in reception of unsolicited data from IoT devices and sensors, leading to interference, high network traffic load, transgression into the domain of other neighbouring IoTs, and poor energy conservation. On the other hand, a smaller coverage value results in poor connectivity and higher data transmission range, leading to faster depletion of node energy. Further, transgression into other domains might lead to interference amongst the devices in the overlapping domains. This issue gains prominence when several low-power devices like electric meters, smart watches, etc., need to be networked through datalink protocols. While most datalink protocols like Sigfox and LoRaWAN, perform admirably well in the absence of interference in high coverage networks; there is a drastic fall in their performance when these protocols are subjected to interference [106]. Therefore, coverage and capacity are key optimizing parameters to accommodate heterogeneity, especially in the presence of interference [107].

*7.1.2. Prolong Network Lifetime.* Network Life Time (NLT) is measure of the overall health of a WSN or an IoT application. Based on the criticality and nature of application, there are several definitions for NLT. In dense WSNs, where node failures do not significantly impact its functioning, NLT is defined in terms of the number of aggregation rounds, or the time-interval till the energy drain-out of the last healthy node. On the other hand, in applications where failure of a single node can cripple the whole network, NLT is defined as the time interval until the first sensor has drained off its energy. Studies to estimate NLT assume prime importance for giving QoS guarantees. In general, NLT depends on the communication load of the network. The communication load in turn, depends on optimal usage of computational resources and the quantum of information processed by the node/device. Therefore, the key optimization parameters to minimize communication load are ready availability of (a) computational resources, and (b) information.

Edge computing is a popular approach to ensure ready availability of computational resources to every aggregating node in the WSN/IoT. Edge computing works on the paradigm, where the computational resources and storage are made readily available or brought to the edge/doorstep of the node/device that requires these resources [65]. Another approach to prolong NLT is to employ caching schemes, where the nodes cache the frequently used information. In cooperative caching scheme, the nodes/IoT devices mutually share their cached data with their neighbors. This eliminates the need for a node to route its query all the way to BS each time it needs an information. Identification of the information that needs to be cached is a critical factor for optimal use of caching schemes [108, 109]. To make best use of the resources, the IoT network needs to identify the type of resources to be optimized before assessing the type of edge computing–cloud, fog, or mobile-edge to be deployed.

*7.1.3. Data Accuracy and Energy Efficiency.* Data accuracy to a large extent is application-specific and is a measure of how close the aggregate is to the actuals, as recorded by the sensor nodes/IoT devices. For example, in a routine environment monitoring application, data accuracy depends on the accuracy of the data sensed, transmitted and aggregated. Therefore, the optimization parameters for such application are dependent only on the sensing mechanism of the sensor; its interference-fault free transmission and data aggregation. However, in defense and space applications, more than one type of sensor may be required to accurately determine a phenomenon. In such applications, the data accuracy depends on optimizing the number of sensors and their locations, besides the sensing mechanisms of each sensor.

Most data aggregation schemes assume that the energy consumption across the network is uniform and devise unilateral schemes that try to optimize the energy consumption. However, in IoTs there is a wide variation in the energy overheads due to the presence of heterogeneous devices, networks and protocols. Hence, the energy load imposed by the nodes across the IoT is nonuniform in nature. As a consequence, it is extremely complicated to model the energy consumption pattern while developing data aggregation algorithms. Therefore, for IoTs, it is necessary to have an application-specific definition of energy-efficiency in order to evaluate the efficacy of a data aggregation algorithm. Accordingly, a data aggregation algorithm is said to be energy-efficient, if the aggregation strategy arrives at an optimum trade-off with respect to other optimization parameters such as latency, time-criticality, data delivery ratio, NLT, etc.

*7.1.4. Latency and Temporal Correctness.* WSNs often encounter delays in transmission and reception of data packets due to factors such as: (i) packet-drop that force retransmissions, (ii) network congestion, (iii) node/link failures, etc. In applications where data sensed by the nodes is time varying, the prime requirement is to guarantee timely delivery of the sensed information to the BS before a predefined time deadline. While accuracy is enhanced if the aggregation is performed only after the aggregator receives data from all its sensor nodes, it is of little or limited significance if the BS receives this aggregate beyond a threshold time limit. Thus, there exists a trade-off between how long an aggregator node needs to wait to perform data aggregation, versus how quickly the BS needs to receive 'fresh data' from the aggregator node. Further, the energy load imposed on the network to accumulate data from all sensor nodes might result in faster depletion of node energy. Therefore, in the quest to achieve better accuracy, the WSN application may lose out on its effectiveness in terms of its NLT. So, there is a trade-off that exists between data accuracy versus data freshness and its effectiveness [110].

*7.1.5. Interference.* Industrial IoT (IIoT) are driven by high precision and any deviation in the data emanating from the sensors can severely impact the process conditions. The critical role of real-time interference detection and classification mechanisms that rely on IIoT devices is analyzed in [111]. The trade-offs between performance and feasibility were analyzed in connection with the implementation on low-complexity IIoT devices.

Malicious interference due to jammers and/or unintentional interference due to neighboring IoT devices working on same frequency is difficult to handle, as it may not be possible to identify the source of interference. To this end, Sparber et al. [112] present a new approach - DynCCA, where the impact of unintentional interference is estimated and compared with a clear channel assessment threshold that is computed at run-time. The run-time threshold computed is used to dynamically mitigate the malicious and unintentional interference beyond the threshold limit.

*7.2. Resource Optimization and Tradeoffs.* Different strategies have been proposed for resource optimization in WSNs *viz.*, (i) construction of broadcast and multicast trees for data dissemination, (ii) identification of a head node to represent a cluster or group of nodes, (iii) variable transmission power, (iv) manipulation of sleep/awake states (v) switching off transceivers of idle nodes, (vii) strategic placement of source nodes in data-centric aggregation, (viii) assigning fair bandwidth to ease traffic congestion, etc. The protocols so developed, can be broadly classified under the categories (a) strategic structuring of nodes and resource optimization, (b) medium access control, and (c) broadcast and multicast trees [113]. Some of the strategies and protocols are described below.

The number of source nodes and their position in the network play an important role in optimizing the energy consumption during data aggregation. Data aggregation protocols based on topological structure like LEACH and HEED along with their variants; PEGASIS and PEDAP etc., strategically position the data aggregators so as to minimize the communication overhead and maximize the energy efficiency. Most strategies rely on choosing a node closest to the BS as the aggregator.

Event-driven applications tend to produce unexpected load in the network when intermediary nodes witness a sudden spurt in incoming data packets. This leads to traffic congestion, packet drop and reduction in network throughput. To address these issues, a strategy to dynamically assign additional (fair) bandwidth to congestion hotspots which is commensurate with the traffic inflow at each node is presented in [114, 115]. For this strategy to be effective, we need to determine the busyness of the channel by considering the total length of busy periods. Smaller the *busyness ratio* of an intermediary node, lower is the traffic load. A higher busyness ratio indicates node congestion, which needs to be eased-out by fair allocation of bandwidth. Another strategy to accomplish fair bandwidth allocation is to estimate the traffic-flow in a channel and determine a priority index to indicate the importance of a node's transmission. Based on the traffic-flow and the priority-index, fair bandwidth is allocated to reduce congestion and improve energy-efficiency.

## 8. Few Posers and Conclusion

While reviewing the literature and the work carried out in data aggregation, we have come across few gap areas that need to be explored further. We list out some of these areas which can well become topics to be explored in future by researchers working in WSN.

*8.1. Energy Efficiency, Energy Optimization, Network Life Time.* Most energy-aware routing protocols and data aggregation protocols make an assumption that the residual energy of each node in the WSN is known *a priori*. It is this assumption, that has driven researchers to devise energy-aware protocols (e.g. LEACH, HEED, PEGASIS, PAMAS) that center around judicious selection of aggregator nodes/CHs, construction of minimum energy MST, etc., without paying much heed to understand how the information of residual energy of a sensor node is communicated to other nodes in the network. Presently, schemes like e-Scan that piggyback the residual information on control/data packets are in vogue. The efficacy of these schemes to get accurate information of the residual energies, geographic location, connectivity with other nodes, etc., in order to select the most suitable node as lead aggregator or CH, need to be studied. This issue is more pertinent to IoT applications where due to mobility and diverse nature of IoT devices, it may not be possible to deterministically estimate the residual energy and the geographical location of the nodes in the network. In such applications, the selection of CH is a major challenge and needs to depend on heuristic algorithms. The heuristics depend on arriving at a tradeoff between various parameters for determining the node with optimal residual energy, with better connectivity to other nodes in the network. In such applications we propose the use of bioinspired heuristics [28–33, 116] and fuzzy-based methods to dynamically estimate the residual-energies of nodes and choose the best amongst them as lead node. To estimate the residual energy of nodes and their connectivity, the fitness function (in the case of bio-inspired algorithms) and fuzzy classifiers need to factor the number transmissions and receptions, idle time, and computational overhead to grade the nodes according to their residual energies.

*8.2. Effect of Routing Protocols, Network Topology and Network Resources on Data Aggregation.* In most MAC protocols, the sleep schedules are extraneously imposed through SYNC message. This forces a receiving node $n$ to synchronize its sleep schedule with neighboring nodes. Instead, can one bring in modifications to the protocols, where the node $n$ dynamically estimates its sleep schedule based on transmission loads of its neighboring nodes? This way, the node $n$ gets up from sleep, just-in-time to receive transmission from a neighboring node. As a safeguard against endless sleep schedule, the node $n$ can fix an upper bound for its sleep time, after which, the node wakes up to broadcast "I AM ALIVE" message to communicate its health status.

While Data Centric (DC) routing is an effective tool to aggregate data of similar nature from closely placed nodes; the energy efficiency tends to be suboptimal if the nodes are located far apart. On the other hand, Address-Centric (AC) routing loses its distinct advantage while aggregating data of closely located nodes. Can we have a hybrid protocol that combines the advantages of both AC and DC routing, where DC is used for aggregating data from closely located nodes and AC for far away nodes? Such protocol shall find ready application in IoT devices, where the placement of nodes is not predefined. In some zones the nodes may be closely packed facilitating data-centric routing, while in other zones the nodes may be sparsely located.

Determining an energy efficient routing paths in IoT applications throws open a number of research challenges. The challenges are primarily due to the fuzzy nature of routing transactions, energy-unaware devices, node mobility, etc. In such scenarios we suggest that fuzzy classifiers with appropriate membership functions and fitness functions based on bioinspired techniques be developed to (a) estimate the current network traffic load to facilitate the election of lead aggregator node, (b) provide options for a node to transmit its sensed data to BS via alternate low-latency paths, and (c) arrive at optimal energy resource conservation strategies.

*8.3. Effect of Network Resources on Data Aggregation.* Allocating a fair bandwidth to nodes to control congestion in network traffic is too fair in the sense that it does not distinguish between intermediary nodes and near-sink nodes. Therefore, this scheme allocates bandwidth in proportion to the network load of each node. However, it may be worthwhile to explore if assigning higher priority to near-sink nodes compared to intermediary nodes eases the network traffic. Further, if buffer_full/buffer_available signals can be broadcast by a near sink node, the downstream nodes can look for alternate routes to relay their data, in case the nearest node is not available. The two modifications suggested may potentially increase the overall throughput besides preventing packet drop by nodes whose buffer is full.

*8.4. Effect of AI and ML, Bigdata Analytics, and 5G on WSN and IoT.* The ever increasing popularity of IoT has thrown a plethora of challenges to connect *anything to everything*. The volume of data that can be collected presents limitless opportunities to develop analytics to seek solutions, which were hitherto inconceivable due to the empirical and NP hard nature of problems being encountered. With multifold increase in global data, the present day 4G technologies like LTE, are not capable of meeting the requirements demanded by the use of IoT. Therefore, 5G technology is likely to become indispensable, and might emerge as the standard for all "connected things".

While collection of voluminous amount of data from IoT devices is unlikely to pose any unsurmountable challenges due to 5G technology, a major issue is to ensure data quality, integrity, and availability, which are of prime importance for data-centric approaches for informed decision making. In this perspective, in addition to connectivity, compute and control issues of IoT, one needs to ponder on other open research issues on how to handle data availability, data redundancy, data integrity and authenticity, data transactions, real-time data, mobile data, etc. If these issues are amicably addressed, the data can be used to develop a host of data analytics for the IoT application domain.

Intelligence at the edge is one of the latest IoT trends. Edge computing reduces latency when sending data from a large number of devices to the cloud. Instead of sending the data to the cloud for analysis and action, decisions, and data processing can happen at the edge. This reduces traffic through the network and provides additional gains in performance [117, 118]. As new generation networks and protocols such as 5G, IPV6 keep emerging, there is a need to integrate these approaches with an IoT application. Such integration is essential for seamless exchange of information across the network, security, and for QoS. Martinez et al. [118] advocates the development of a network that integrates WSN with 5G, TCP/IP (IPv6) protocols with IoT, for secure exchange of information with QoS guarantees.

*8.5. Exploiting Interference as a Tool to Enhance Fault-Tolerant Features of WSNs.* Successive Interference Cancellation (SIC) techniques attempt to retrieve information from a node that receives interfered data transmissions from multiple nodes. Nodes implementing SIC identify the strongest signal, decode it and subtract this data from the mixed signal. The process continues iteratively to retrieve next strongest data signal from the remaining signal. Li et al. [119] discuss an Efficient Minimum Approximation Successive Interference Cancellation (EMA-SIC) algorithm that can recover data from multiple simultaneous senders under Physical interference model.

A new paradigm is proposed, where interference is *not* treated as a source of noise that is unintentionally received/ overheard by a node, but to consider the interfered signal as information received from different source nodes that needs to be deciphered. Armed with the knowledge of potential interferers based on the CIM model [70] and the interfered signal retrieved through successive interference cancellation techniques [119] new approaches could be devised to optimize the number of backup copies to be scheduled as per the primary-backup approach of fault-tolerance.

*8.6. Fault-Tolerance.* Fault-recovery schemes based on Omega failure detector help identify the time of failure, based on reception of the last "I AM ALIVE" message from a process in *suspect set*. Recovery mechanisms are initiated in the event of nonreception of this message. However, these schemes do not address a scenario, where a process $p$ sends "I AM ALIVE" message to process $q$, but for some reason like buffer full, traffic congestion, etc., the process $q$ fails to receive the message. We put forth two posers for further work on Omega failure detection models. The first is on, "how to develop schemes that address the scenario discussed above". The second is to explore if the Omega detection model can be used to detect node/link faults in addition to process faults. Solutions to the two posers help the data aggregation algorithms to identify the correct data to aggregate. The concept of watch-dog timers that were primarily used to detect hardware faults can also be explored to detect the time of failure of a process.

Reliable network connectivity being one of the prime requirements for data aggregation in WSNs and IoT, a "network-reliability" index, that works as a metric for network reliability needs to be established. While an attempt to study the impact of energy depletion and node aging [120] was carried out by associating aging with battery discharge, a more formal effort is warranted. The proposed network-reliability index shall factor some common traits of WSN and IoT devices like, inherent redundancy, heterogeneity, application domain, etc. The state of this index at any given point in time, may act as an indicator for remaining life assessment studies of the sensors and IoT devices.

## 9. Conclusion

In this paper an effort is made to present an in-depth review of literature on data aggregation in WSN and IoT. A brief overview of the fundamentals of various data aggregation approaches for WSN and IoT applications is presented. The key features, advantages, and disadvantages of various data aggregation approaches and protocols based on WSN/ IoT network topology, security, mobility, interference, and fault-tolerance are reviewed. Certain gap areas, where further work or abstraction is required, and the suggestions to handle some these issues are presented as posers for further research to be carried out.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

It is stated that the authors do not have conflict of interest.

## Acknowledgments

The work is of relevance to the Grant-in-Aid Project (GAP) being funded by the Department of Science & Technology (DST), Ministry of S&T and Earth Sciences, Government of India. The project is assigned with the number GAP 0729 by my Institute CSIR-Indian Institute of Chemical Technology, Hyderabad, India. In all GAP category projects, the funding agency (DST, Govt. of India) encourages the Principal Investigator to publish his/her work in reputed journals. The funder in such GAP projects is seldom involved in manuscript preparation and submission to a Journal.

## References

[1] P. Jesus, C. Baquero, and P. S. Almeida, "A survey of distributed data aggregation algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 381–404, 2015.

[2] H. Rahman, N. Ahmed, and I. Hussain, "Comparison of data aggregation techniques in internet of things (IoT)," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1296–1300, Chennai, India, March 2016.

[3] T. Salman and R. Jain, "A survey of protocols and standards for internet of things," *Advanced Computing and Communications*, vol. 1, no. 1, 2017.

[4] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.

[5] P. P. Ray, "A survey on internet of things architectures," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 3, pp. 291–319, 2018.

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

[7] S. A. Abdulzahra and A. K. M. Al-Qurabat, "Data aggregation mechanisms in wireless sensor networks of IoT: a survey," *International Journal of Computing and Digital System*, 2021.

[8] S. S. Ali, N. Giweli, A. Dawoud, and P. W. C. Prasad, "Data aggregation techniques in wireless sensors networks: a survey," in *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, Sydney, Australia, November 2016.

[9] I. D. I. Saeedi and A. K. M. Al-Qurabat, "A systematic review of data aggregation techniques in wireless sensor networks," *Journal of Physics: Conference Series*, vol. 1818, no. 1, article 012194, 2021.

[10] S. V. Nandury and B. A. Begum, "Smart WSN-based ubiquitous architecture for smart cities," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2366–2373, Kochi, India, August 2015.

[11] M. Y. Kathjoo, F. A. Khanday, and M. T. Banday, "A comparative study of WSN and IoT," in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, pp. 1–5, Bangalore, India, February 2018.

[12] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking - MobiCom '99*, pp. 174–185, Seattle Washington, USA, August 1999.

[13] J. Kulik, R. B. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, no. 2/3, pp. 169–185, 2002.

[14] G. Khanna, S. Bagchi, and Y.-S. Wu, "Fault tolerant energy aware data dissemination protocol in sensor networks," in *International Conference on Dependable Systems and Networks, 2004*, pp. 795–804, Florence, Italy, 2004.

[15] Z. Rehena, S. Roy, and N. Mukherjee, "A modified SPIN for wireless sensor networks," in *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, pp. 1–4, Bangalore, India, January 2011.

[16] D. Xiao, M. Wei, and Y. Zhou, "Secure-SPIN: secure sensor protocol for information via negotiation for wireless sensor networks," in *2006 1ST IEEE Conference on Industrial Electronics and Applications*, pp. 1–4, Singapore, May 2006.

[17] L. Tang and Q. L. Li, "S-SPIN: a provably secure routing protocol for wireless sensor networks," in *2009 International Conference on Communication Software and Networks*, pp. 620–624, Chengdu, China, February 2009.

[18] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[19] V. Loscri, G. Morabito, and S. Marano, "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)," in *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference, 2005*, pp. 1809–1813, Dallas, TX, USA, September 2005.

[20] F. Xiangning and S. Yulin, "Improvement on LEACH protocol of wireless sensor network," in *2007 International Conference on Sensor Technologies and Applications (SENSORCOMM 2007)*, pp. 260–264, Valencia, Spain, October 2007.

[21] G. S. Kumar, P. M. V. Vinu, and K. P. Jacob, "Mobility metric based LEACH-mobile protocol," in *2008 16th International Conference on Advanced Computing and Communications*, pp. 248–253, Chennai, India, December 2008.

[22] O. Younis and S. Fahmy, "An experimental study of energy-efficient routing and data aggregation in sensor networks," in *Proc. Int. Workshop on Localized Commun. And Topology Protocols for Ad Hoc Netw. (LOCAN)*, pp. 50–57, Washington, DC, USA, November 2005.

[23] S. Chand, S. Singh, and B. Kumar, "Heterogeneous HEED protocol for wireless sensor networks," *Wireless Personal Communications*, vol. 77, no. 3, pp. 2117–2139, 2014.

[24] T. N. Nguyen, C. V. Ho, and T. T. T. Le, "A topology control algorithm in wireless sensor networks for IoT-based applications," in *2019 International Symposium on Electrical and Electronics Engineering (ISEE)*, pp. 141–145, Ho Chi Minh City, Vietnam, October 2019.

[25] T. M. Behera, U. C. Samal, and S. K. Mohapatra, "Energy-efficient modified LEACH protocol for IoT application," *IET Wireless Sensor Systems*, vol. 8, no. 5, pp. 223–228, 2018.

[26] J. Huo, X. Deng, and H. M. M. al-Neshmi, "Design and improvement of routing protocol for field observation instrument networking based on LEACH protocol," *Journal of Electrical and Computer Engineering*, vol. 2020, Article ID 8059353, 19 pages, 2020.

[27] F. A. B. Mohammed, N. Mekky, H. H. Suleiman, and N. A. Hikal, "Sectored LEACH (S-LEACH): an enhanced LEACH for wireless sensor network," *IET Wireless Sensor Systems*, vol. 12, no. 2, pp. 56–66, 2022.

[28] K. Miranda, S. Zapotecas-Martínez, A. López-Jaimes, and A. García-Nájera, "A comparison of bio-inspired approaches for the cluster-head selection problem in WSN," in *Advances in Nature-Inspired Computing and Applications. EAI/Springer Innovations in Communication and Computing*, S. Shandilya, S. Shandilya, and A. Nagar, Eds., pp. 165–187, Springer Int. Publishing, 2019.

[29] P. Nayak, K. Kavitha, and N. Khan, "Cluster head selection in wireless sensor network using bio-inspired algorithm," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pp. 1690–1696, Kochi, India, October 2019.

[30] Z. Wang, Y. S. Ong, J. Sun, A. Gupta, and Q. Zhang, "A generator for multiobjective test problems with difficult-to-approximate pareto front boundaries," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 4, pp. 556–571, 2019.

[31] A. García-Nájera, S. Zapotecas-Martínez, and K. Miranda, "Analysis of the multi-objective cluster head selection problem in WSNs," *Applied Soft Computing*, vol. 112, article 107853, 2021.

[32] M. Ahmad, A. A. Ikram, I. Wahid, M. Inam, N. Ayub, and S. Ali, "A bio-inspired clustering scheme in wireless sensor networks: BeeWSN," *Procedia Computer Science*, vol. 130, pp. 206–213, 2018.

[33] A. Khan, F. Aftab, and Z. Zhang, "BICSF: bio-inspired clustering scheme for FANETs," *IEEE Access*, vol. 7, pp. 31446–31456, 2019.

[34] M. S. Yousefpoor, E. Yousefpoor, H. Barati, A. Barati, A. Movaghar, and M. Hosseinzadeh, "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: a comprehensive review," *Journal of Network and Computer Applications*, vol. 190, article 103118, 2021.

[35] X. Liu, J. Yu, F. Li, W. Lv, Y. Wang, and X. Cheng, "Data aggregation in wireless sensor networks: from the perspective of security," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6495–6513, 2020.

[36] X. Liu, J. Yu, X. Zhang, Q. Zhang, and C. Fu, "Energy-efficient privacy-preserving data aggregation protocols based on slicing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, Article ID 19, 2020.

[37] F. Rezaeibagha, Y. Mu, K. Huang, and L. Chen, "Secure and efficient data aggregation for IoT monitoring systems," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8056–8063, 2021.

[38] N. Gupta and V. Gupta, "A review on sink mobility aware fast and efficient data gathering in wireless sensor networks," in *2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Spring)*, pp. 1–4, Dehradun, India, April 2016.

[39] F. Santamaria, P. Raimondo, M. Tropea, F. De Rango, and C. Aiello, "An IoT surveillance system based on a decentralised architecture," *Sensors*, vol. 19, no. 6, p. 1469, 2019.

[40] E. E. Fitzgerald, M. Pióro, and A. Tomaszwski, "Energy-optimal data aggregation and dissemination for the internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 955–969, 2018.

[41] A. Ullah, G. Said, M. Sher, and H. Ning, "Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN," *Peer-to-Peer Networking and Applications*, vol. 13, no. 1, pp. 163–174, 2020.

[42] Y. Pu, J. Luo, C. Hu et al., "Two secure privacy-preserving data aggregation schemes for IoT," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 3985232, 11 pages, 2019.

[43] R. M. A. Haseeb-Ur-Rehman, M. Liaqat, A. H. M. Aman et al., "Sensor cloud frameworks: state-of-the-art, taxonomy, and research issues," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 22347–22370, 2021.

[44] J. Shuja, M. A. Humayun, W. Alasmary, H. Sinky, E. Alanazi, and M. K. Khan, "Resource efficient geo-textual hierarchical clustering framework for social IoT applications," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25114–25122, 2021.

[45] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the internet of things: a systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23–34, 2017.

[46] L. Yang, C. Ding, and M. Wu, "RPIDA: recoverable privacy-preserving integrity-assured data aggregation scheme for wireless sensor networks," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 9, no. 12, 2015.

[47] K. A. Shim and C. M. Park, "A secure data aggregation scheme based on appropriate cryptographic primitives in heterogeneous wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2128–2139, 2015.

[48] J. Cui, L. Shao, H. Zhong, Y. Xu, and L. Liu, "Data aggregation with end-to-end confidentiality and integrity for large-scale wireless sensor networks," *Peer-to-Peer Networking and Applications*, vol. 11, no. 5, pp. 1022–1037, 2018.

[49] P. Zhang, J. Wang, K. Guo, F. Wu, and G. Minc, "Multi-functional secure data aggregation schemes for WSNs," *Ad Hoc Networks*, vol. 69, pp. 86–99, 2018.

[50] J. Zhang, Y. Zong, C. Yang, Y. Miao, and J. Guo, "LBOA: location-based secure outsourced aggregation in IoT," *IEEE Access*, vol. 7, pp. 43869–43883, 2019.

[51] S. Hu, L. Liu, L. Fang, F. Zhou, and R. Ye, "A novel energy-efficient and privacy-preserving data aggregation for WSNs," *IEEE Access*, vol. 8, pp. 802–813, 2020.

[52] A. Saleem, A. Khan, S. U. R. Malik et al., "FESDA: fog-enabled secure data aggregation in smart grid IoT network," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6132–6142, 2020.

[53] M. Azeem, A. Ullah, H. Ashraf et al., "FoG-oriented secure and lightweight data aggregation in IoT," *IEEE Access*, vol. 9, pp. 111072–111082, 2021.

[54] A. Ahmed, S. Abdullah, M. Bukhsh, I. Ahmad, and Z. Mushtaq, "An energy-efficient data aggregation mechanism for IoT secured by blockchain," *IEEE Access*, vol. 10, pp. 11404–11419, 2022.

[55] X. Yuan and R. Zhang, "An energy-efficient mobile sink routing algorithm for wireless sensor networks," in *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, Wuhan, China, September 2011.

[56] A. T. Erman, A. Dilo, and P. Havinga, "A virtual infrastructure based on honeycomb tessellation for data dissemination in multi-sink mobile wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, Article ID 17, 2012.

[57] M. Zhao and Y. Yang, "Bounded relay hop mobile data gathering in wireless sensor networks," *IEEE Transactions on Computers*, vol. 61, no. 2, pp. 265–277, 2012.

[58] C. Zhu, S. Wu, G. Han, L. Shu, and H. Wu, "A tree-cluster-based data-gathering algorithm for industrial WSNs with a mobile sink," *IEEE Access*, vol. 3, pp. 381–396, 2015.

[59] N. Wang and W. Hsu, "Energy efficient two-tier data dissemination based on Q-learning for wireless sensor networks," *IEEE Access*, vol. 8, pp. 74129–74136, 2020.

[60] H. Basumatary and M. K. D. Barma, "Analysis of mobile sink based routing protocols in wireless sensor networks," *International Journal of Computational Intelligence & IoT*, vol. 2, no. 3, 2019.

[61] T. Zhang, X. Song, L. Xiongfei, H. Zheng, Y. Han, and Q. Kai, "Towards time-sensitive and verifiable data aggregation for mobile crowdsensing," *Security and Communication Networks*, vol. 2021, Article ID 6679157, 14 pages, 2021.

[62] M. Amarlingam, P. K. Mishra, P. Rajalakshmi, S. S. Channappayya, and C. S. Sastry, "Novel light weight compressed data aggregation using sparse measurements for IoT networks," *Journal of Network and Computer Applications*, vol. 121, pp. 119–134, 2018.

[63] F. Metzger, T. Hobfeld, A. Bauer, S. Kounev, and P. E. Heegaard, "Modeling of aggregated IoT traffic and its application to an IoT cloud," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 679–694, 2019.

[64] A. K. Idrees, A. K. M. Al-Qurabat, C. A. Jaoude, and W. L. Al-Yaseen, "Integrated divide and conquer with enhanced k-means technique for energy-saving data aggregation in wireless sensor networks," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, June 2019.

[65] J. Ren, Y. He, G. Huang, G. Yu, Y. Cai, and Z. Zhang, "An edge-computing based architecture for mobile augmented reality," *IEEE Network*, vol. 33, no. 4, pp. 162–169, 2019.

[66] H. Xue, B. Huang, M. Qin, H. Zhou, and H. Yang, "Edge computing for internet of things: a survey," in *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, pp. 755–760, Rhodes, Greece, November 2020.

[67] A. M. Ghosh and K. Grolinger, "Edge-cloud computing for internet of things data analytics: embedding intelligence in the edge with deep learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2191–2200, 2021.

[68] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[69] P. Cardieri, "Modeling interference in wireless ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 551–572, 2010.

[70] B. A. Begum and N. V. Satyanarayana, "Composite interference mapping model for interference fault-free transmission in WSN," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2119–2125, Kochi, India, August 2015.

[71] M. K. An, H. Cho, B. Zhou, and L. Chen, "Minimum latency aggregation scheduling in internet of things," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 395–401, Honolulu, HI, USA, February 2019.

[72] Y. Shi, Y. T. Hou, J. Liu, and S. Kompella, "Bridging the gap between protocol and physical models for wireless networks," *IEEE Trans. on Mobile Computing*, vol. 12, no. 7, pp. 1404–1416, 2013.

[73] X. Y. Li, X. Xu, S. Wang et al., "Efficient data aggregation in multi-hop wireless sensor networks under physical interference model," in *2009 IEEE 6th International Conference on*

*Mobile Adhoc and Sensor Systems*, pp. 353–362, Macau, China, October 2009.

[74] P. J. Wan, K. M. Alzoubi, and O. Frieder, “Distributed construction of connected dominating set in wireless ad hoc networks,” *Mobile Networks and Applications*, vol. 9, no. 2, pp. 141–149, 2004.

[75] O. Goussevskaia, R. Wattenhofer, M. M. Halldorsson, and E. Welzl, “Capacity of arbitrary wireless networks,” in *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pp. 1872–1880, Rio de Janeiro, Brazil, April 2009.

[76] X. Xu, X. Y. Li, X. Mao, S. Tang, and S. Wang, “A delay-efficient algorithm for data aggregation in multihop wireless sensor networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 1, pp. 163–175, 2011.

[77] S. Chen, S. Tang, M. Huang, and Y. Wang, “Capacity of data collection in arbitrary wireless sensor networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 52–60, 2012.

[78] H. Li, C. Wu, Q. Hua, and F. C. M. Lau, “Latency-minimizing data aggregation in wireless sensor networks under physical interference model,” *Ad Hoc Networks*, vol. 12, pp. 52–68, 2014.

[79] Z. Li, Y. Liu, K. G. Shin, J. Liu, and Z. Yan, “Interference steering to manage interference in IoT,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10458–10471, 2019.

[80] Z. Cai and Q. Chen, “Latency-and-coverage aware data aggregation scheduling for multihop battery-free wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1770–1784, 2021.

[81] J. Li, M. Siddula, X. Cheng, W. Cheng, Z. Tian, and Y. Li, “Approximate data aggregation in sensor equipped IoT networks,” *Tsinghua Science and Technology*, vol. 25, no. 1, pp. 44–55, 2020.

[82] P. Wan, S. C. Huang, L. Wang, Z. Wan, and X. Jia, “Minimum-latency aggregation scheduling in multihop wireless networks,” in *MobiHoc '09: Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, pp. 185–194, New Orleans, LA, USA, 2009.

[83] H. Li, Q. S. Hua, C. Wu, and F. C. M. Lau, “Minimum-latency aggregation scheduling in wireless sensor networks under physical interference model,” in *MSWIM '10: Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, pp. 360–367, Bodrum, Turkey, October 2010.

[84] M. H. Orsson and P. Mitra, “Wireless connectivity and capacity,” in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 516–526, Kyoto, Japan, January 2012.

[85] H. Yousefi, M. Malekimajd, M. Ashouri, and A. Movaghar, “Fast aggregation scheduling in wireless sensor networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3402–3414, 2015.

[86] O. M. Bushnaq, A. Celik, H. ElSawy, M. Alouini, and T. Y. Al-Naffouri, “Aerial data aggregation in IoT networks: hovering & traveling time dilemma,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Abu Dhabi, United Arab Emirates, December 2018.

[87] Y. Nabil, H. ElSawy, S. al-Dharrab, H. Mostafa, and H. Attia, “Data aggregation in regular large-scale IoT networks: granularity, reliability, and delay tradeoffs,” *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17767–17784, 2022.

[88] J. Chen, S. Kher, and A. Somani, “Distributed fault detection of wireless sensor networks,” in *DIWANS '06: Proceedings of the 2006 workshop on Dependability issues in wireless ad hoc networks and sensor networks*, pp. 65–72, Los Angeles, CA USA, 2006.

[89] X. Xu, W. Chen, J. Wan, and R. Yu, “Distributed fault diagnosis of wireless sensor networks,” in *2008 11th IEEE International Conference on Communication Technology*, pp. 148–151, Hangzhou, China, November 2008.

[90] M. Larrea, C. Martin, and J. J. Astrain, “Hierarchical and fault-tolerant data aggregation in wireless sensor networks,” in *2007 2nd International Symposium on Wireless Pervasive Computing*, pp. 531–536, San Juan, PR, USA, February 2007.

[91] T. D. Chandra, V. Hadzilacos, and S. Toueg, “The weakest failure detector for solving consensus,” *Journal of the ACM (JACM)*, vol. 43, no. 4, pp. 685–722, 1996.

[92] M. Younis, I. F. Senturk, K. Akkaya, S. Lee, and F. Senel, “Topology management techniques for tolerating node failures in wireless sensor networks: a survey,” *Computer Networks*, vol. 58, pp. 254–283, 2014.

[93] Y. Feng, S. Tang, and G. Dai, “Fault tolerant data aggregation scheduling with local information in wireless sensor networks,” *Tsinghua Science and Technology*, vol. 16, no. 5, pp. 451–463, 2011.

[94] L. Zhang, Q. Ye, J. Cheng et al., “Fault-tolerant scheduling for data collection in wireless sensor networks,” in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 5345–5349, Anaheim, CA, December 2012.

[95] B. A. Begum and S. V. Nandury, “Component-based self-healing algorithm with dynamic range allocation for fault-tolerance in WSN,” in *ICCCT-2017: Proceedings of the 7th International Conference on Computer and Communication Technology*, pp. 58–65, Allahabad, India, November 2017.

[96] B. A. Begum and S. V. Nandury, “Self-reconfiguration aggregation scheduling to recover from node and interference faults in WSNs,” in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1541–1546, Chennai, India, March 2017.

[97] B. A. Begum and S. V. Nandury, “Interference-fault free data aggregation in tree-based WSNs,” in *2016 International Conference on Information Technology (ICIT)*, pp. 333–341, Bhubaneswar, India, December 2016.

[98] B. A. Begum and S. V. Nandury, “Component based self-healing approach for fault-tolerant data aggregation in WSN,” *IEEE Access*, vol. 10, pp. 73503–73520, 2022.

[99] J. Lin, P. R. Chelliah, M. Hsu, and J. Hou, “Efficient fault-tolerant routing in IoT wireless sensor networks based on bipartite-flow graph modeling,” *IEEE Access*, vol. 7, pp. 14022–14034, 2019.

[100] K. Grining, M. Klonowski, and P. Syga, “On practical privacy-preserving fault-tolerant data aggregation,” *International Journal of Information Security*, vol. 18, no. 3, pp. 285–304, 2019.

[101] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, “A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT,” *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[102] C. Xu, R. Yin, L. Zhu et al., “Privacy-preserving and fault-tolerant aggregation of time-series data with a semi-trusted authority,” *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12231–12240, 2022.

[103] H. M. Khan, A. Khan, F. Jabeen, and A. U. Rahman, "Privacy preserving data aggregation with fault tolerance in fog-enabled smart grids," *Sustainable Cities and Society*, vol. 64, article 102522, 2021.

[104] P. Kale and M. J. Nene, "Data aggregation trees with QoS in sensor networks," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1–5, Bombay, India, 2019.

[105] J. A. Stankovic, "Research directions for the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, 2014.

[106] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovacs, P. Mogensen, and M. Sorensen, "Interference impact on coverage and capacity for low power wide area IoT networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, San Francisco, CA, USA, March 2017.

[107] S. Aggarwal and A. Nasipuri, "Survey and performance study of emerging LPWAN technologies for IoT applications," in *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)*, pp. 069–073, Charlotte, NC, USA, October 2019.

[108] Z. Zhang, C. Lung, I. Lambadaris, and M. St-Hilaire, "IoT data lifetime-based cooperative caching scheme for ICN-IoT networks," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–7, Kansas City, MO, USA, May 2018.

[109] J. A. G. de Brito, J. R. d. P. Junior, F. d. R. Henriques, and L. S. de Assis, "Topology control optimization of wireless sensor networks for IoT applications," in *WebMedia '19: Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, pp. 477–480, Rio de Janeiro, Brazil, October 2019.

[110] T.-D. Nguyen, D.-T. Le, V.-V. Vo, M. Kim, and H. Choo, "Fast sensory data aggregation in IoT networks: collision-resistant dynamic approach," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 766–777, 2021.

[111] S. Grimaldi, A. Mahmood, S. A. Hassan, G. P. Hancke, and M. Gidlund, "Autonomous interference mapping for industrial internet of things networks over unlicensed bands: identifying cross-technology interference," *IEEE Industrial Electronics Magazine*, vol. 15, no. 1, pp. 67–78, 2021.

[112] T. Sparber, C. A. Boano, S. S. Kanhere, and K. Römer, "A mitigating radio interference in large IoT networks through dynamic CCA adjustment," *Open Journal of Internet of Things (OJIOT)*, vol. 1, no. 1, pp. 103–113, 2017.

[113] S. Moulik, S. Misra, and C. Chakraborty, "Performance evaluation and delay-power trade-off analysis of ZigBee protocol," *IEEE Transactions on Mobile Computing*, vol. 18, no. 2, pp. 404–416, 2019.

[114] C. J. Raman and V. James, "FCC: fast congestion control scheme for wireless sensor networks using hybrid optimal routing algorithm," *Cluster Computing*, vol. 22, Supplement 5, pp. 12701–12711, 2019.

[115] P. Yarde, S. Srivastava, and K. Garg, "A delay abridged judicious cross-layer routing protocol for wireless sensor network," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pp. 634–638, Singapore, February 2019.

[116] M. Esmaeili and S. Jamali, "A survey: optimization of energy consumption by using the genetic algorithm in WSN based internet of things," *CiiT International Journal of Wireless Communication*, vol. 8, no. 2, pp. 65–72, 2016.

[117] J. Ni, X. Lin, and X. S. Shen, "Toward edge-assisted internet of things: from security and efficiency perspectives," *IEEE Network*, vol. 33, no. 2, pp. 50–57, 2019.

[118] S. H. Martínez, P. O. J. Salcedo, and B. S. R. Daza, "IoT application of WSN on 5G infrastructure," in *2017 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1–6, Marrakech, Morocco, May 2017.

[119] H. Li, C. Wu, D. Yu, Q. Hua, and F. Lau, "Aggregation latency-energy tradeoff in wireless sensor networks with successive interference cancellation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 11, pp. 2160–2170, 2013.

[120] D. Bruneo, S. Distefano, F. Longo, A. Puliafito, and M. Scarpa, "Evaluating wireless sensor node longevity through Markovian techniques," *Computer Networks*, vol. 56, no. 2, pp. 521–532, 2012.