QUEUE AWARE CONGESTION AVOIDANCE FOR HEALTHCARE DATA SHARING IN INTERNET OF THINGS

By

Muhammad Zafarullah

NATIONAL UNIVERSITY OF MODERN LANGUAGES

ISLAMABAD

JANUARY,2026

QUEUE AWARE CONGESTION AVOIDANCE FOR HEALTHCARE DATA SHARING IN INTERNET OF THINGS

By

Muhammad Zafarullah

PhD Computer Science, National University of Modern Languages, Islamabad, 2025

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In **Computer Science**

To

FACULTY OF ENGINEERING & COMPUTING

NATIONAL UNIVERSITY OF MODERN LANGUAGES ISLAMABAD

THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computing for acceptance.

Thesis Title:

QUEUE AWARE CONGESTION AVOIDANCE FOR HEALTHCARE DATA SHARING IN INTERNET OF THINGS

Submitted By: Muhammad Zafarullah **Registration #:** 4 PHD/CS/S20

Doctor of Philosophy in Computer Science

Title of the Degree

Computer Science

Name of Discipline

Dr. Ata Ullah _____

Name of Research Supervisor Signature of Research Supervisor

Dr. Fazli Subhan _____

HOD, Computer Science Department Signature of HOD

Dr. Noman Malik _____

Name of Dean (FEC) Signature of Dean (FEC)

Maj Gen Shahid Mahmood Kayani HI(M), Retd _____

Name of Rector Signature of Rector

26TH JANUARY,2026

AUTHOR'S DECLARATION

I Muhammad Zafarullah

Son of Manzoor Ahmad

Discipline Computer Science

Candidate of Doctor of Philosophy in Computer Science at the National University of Modern Languages do hereby declare that the thesis Queue Aware Congestion Avoidance for Healthcare Data Sharing in Internet of Things submitted by me in partial fulfillment of PhD degree, is my original work and has not been submitted or published earlier. I also solemnly declare that it shall not, in the future, be submitted by me for obtaining any other degree from this or any other university or institution. I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be canceled and the degree revoked.

Signature of Candidate

Muhamamd Zafarullah

Name of Candidate

26TH JANUARY,2025

Date

Abstract

Title: Queue Aware Congestion Avoidance for Healthcare Data Sharing in Internet of Things

The Internet of Healthcare Things (IoHT) emerged in order to assist with healthcare operations and patient monitoring by collecting continuous data from health sensors attached to the body of the patients. Because of the huge volume of the data collected from the sensor, it is difficult to control congestion at intermediary devices.Following research problems are addressed in this thesis; i) The number of congestion detection approaches in IoHT fail early detection since the sender is uninformed of the residual queue size. This proposed scheme introduces a Queue aware priority-based queuing strategy for delaying congestion in the IoHT while dealing with the huge volume of messages exchanged, particularly during emergencies scenario. The proposed approach presents a novel algorithm that employs a queue-aware strategy for early congestion detection, which is beneficial for early prevention of congestion. ii) The Internet of Healthcare Things (IoHT) emerged to help healthcare operations with use of an IoT and sensor networks. Delays in the processing of emergency packets in IoT-based healthcare systems can have fatal repercussions, including patient death. This study addresses the handling of emergent packets in high and low priority queues having high severity level. Depending on the severity level, packets are routed to the appropriate queues for immediate processing. This approach intends to improve data handling efficiency for packets carrying emergent data while also reducing queue congestion. iii) The Internet of Healthcare Things (IoHT) uses sensors attached to patients to continuously monitor health metrics and transfer data to the cloud for further analysis. However, the huge amount of data created creates a challenge, particularly in managing congestion at intermediary devices. To solve this issue, this work provides a queue-aware, priority-based queuing technique for efficiently managing data flow, particularly during emergencies. The suggested method eliminates unnecessary data transmission by providing simply a flag signal when sensor data does not change significantly or when there are no emergencies. This method enables the system to focus on essential, real-time data during emergencies while reducing congestion in normal circumstances..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IoT	-	Internet of Things
IoHT	-	Internet of Helthcare Things
QACA	-	Queue Aware Congestion Avoidance Scheme
DQACA	-	Dual Queue Aware Congestion Avoidance Scheme
WBAN	-	Wireless Body Area Network
IoMT	-	Internet of Medical Things
FIFO	-	First-in-First-out
DD-	-	De- Duplication and Queue-usage Aware Congestion
QACA	-	Avoidance.
WFQ	-	Weighted Fair Queuing
PQ	-	Priority Queuing
EDF	-	Earliest Deadline First
DCCA	-	Distributed Congestion Control Algorithm
PBSCA	-	Priority Based Congestion Avoidance Scheme
PLR	-	Packet Loss Ratio
XAI	-	Explainable Artificial Intelligence
CC	-	Congestion Control
QoS	-	Quality of Service
CHs	-	Cluster Heads
CBWFQ	-	Class-Based Weighted Fair Queuing
WRR	-	Weighted Round Robin
LPQ	-	Low Priority Queue
HPQ	-	High Priority Queue

LIST OF SYMBOLS

INode	-	Intermediate Node
Ack_Ctr	-	Ack counter as per interval
LPQueue	-	Low Priority Queue
HPQueue	-	High Priority Queue
RQueueSize	-	Residual Queue Size
H_Queue_Status	-	High Priority Queue Status
L_Queue_Status	-	Low Priority Queue Status
r_Node	-	Root Node
inode	-	Intermediate Node
Lc_Node	-	Lowest Cost Node
Packet	-	Packet received from the root node
QSBOF	-	Quality-of-service balanced function

ACKNOWLEDGMENT

In the name of Allah, the most Merciful, the most Compassionate all praise be to Allah, the Lord of the worlds; and prayers and peace be upon Muhammad His messenger. I acknowledge my unlimited gratitude to Allah, the Ever-Magnificent; the Ever-Thankful, for His help and blessing. My special praise for the Holy Prophet Hazrat Muhammad (Peace be upon him), hailed as the supreme instructor, the everlasting source of guidance and wisdom for humanity. I express my heartfelt appreciation to my honorable supervisor Dr. Ata Ullah for his invaluable guidance, mentorship and encouraging attitude throughout my research work. I am thankful to the Coordinator Dr. Zia Ur Rehman, Department Head Dr. Fazli Subhan and Dean Dr. Noman Malik, for their support and coaching. Special thanks and lot of prayers to my dearest father Manzoor Ahmad (Late), whose love has been my source of strength and inspiration.

I shall also acknowledge the extended assistance from the administration of the Department of Computer Science, who supported me all through my research experience and simplified the challenges I faced. For all whom I did not mention but shall not neglect their significant contribution, thanks for everything.

DEDICATION

This thesis work is dedicated to my parents, family, and my teachers throughout my education career who have not only loved me unconditionally but whose good examples have taught me to work hard for the things that I aspire to achieve.

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

In the recent past years, the Internet of Things (IoT) has gained the huge momentum. In 1999, Kevin Ashton introduced the term "IoT" for use in supply chain management [1, 2]. The term "Internet of Things" refers to the concept that everything is connected to the internet, but it also implies that every object communicates with every other device via the internet. Simply IoT consists of devices, ranging from simple sensors to smartphones and wearable devices, that are connected together. When these devices are connected to an automated system, information can be collected and action taken on a specific task to assist someone in a particular situation. An IoT system comprises smart devices that utilize embedded systems, including processors, sensors, and communication hardware, to collect, transmit, and respond to data received from the environment. The IoT sends the sensed data to the cloud through a gateway or, in some cases, sends it directly to the cloud for necessary action. Different types of protocols are used for the interconnectivity, networking, and communication. Mostly, the types of protocol used depend on the nature of the application or environment [3].

The implementation of IoT is based on a 5-layered architecture. The first layer, known as the business layer, is responsible for user privacy, IoT application management, and research activities related to IoT. The name of the second layer is the application layer. This layer monitors which application will be used for IoT. The third layer is known as the processing

layer. This layer handles the information collected by the perception layer. The working of this layer is very complex because a huge amount of data is collected by the perception layer, which is to be processed by this layer. The transport layer performs the transportation function by transferring information between the processing layer and the perception layer, and vice versa, using established network standards. The fifth layer is called the perception layer, and the main job of this layer is to collect information from the environment and pass this information to the next layer [4]. The most important requirement of the Internet of Things is that things must be connected, which is used to fill the gap between the physical and virtual worlds. Many factors are involved in the design of the Internet of Things, including networking, communication, security, business models, and processes. Today, IoT is widely used in supply chain management, manufacturing, environmental monitoring, retailing, smart shelf operations, healthcare, food and restaurant industry, logistic industry, travel and tourism industry, library services, and many other areas [5] .

1.2 Internet of Healthcare Things

In modern society, improving healthcare infrastructure and providing quality care to patients, as well as addressing the shortage of medical staff, are primary issues [6]. Medical care and healthcare are the most important application areas of IoT. IoT can be very useful in many applications of healthcare, like the detection of chronic diseases, health monitoring, and elderly care. By detecting and preventing illnesses and hazardous situations, remote patient monitoring offers significant potential to reduce healthcare expenses while simultaneously enhancing healthcare quality.By virtue of IoHT, it is possible to remotely track patient physiological parameters, automate diagnostics, and initiate interventions without the need for the patient to visit a clinical facility physically. This becomes particularly beneficial for patients requiring long-term care, elderly individuals with chronic illnesses. Its beneficial for all those who are residing in remote areas and having limited access to healthcare facilities.

IoHT provides a real-time data stream from various patient devices, aiding in clinical decision-making, enhancing patient engagement, optimizing hospitalization resource efficiency, and, most importantly, minimizing healthcare costs. The home-based provision of healthcare facilities is the prominent feature of an IoT. IoT-based healthcare services aim to provide affordable

treatment, which enhances quality of life, and also improves the overall experience for end users [7]. A positive trend is also the establishment of cost-effective and secure connectivity among patients, clinics, and healthcare organizations [8].

Internet of Healthcare Things (IoHT) is that area of the IoT that views remote health monitoring systems [9, 10]. In the last few years, rapid development has been seen in making medical sensors that are compact, energy-efficient, and inexpensive, which has lately fueled the development of IoHT-based solutions. These solutions include wearable ECG monitors for patients having heart problems, continuous glucose monitoring systems for diabetes patients, smart pulse oximeters to monitor oxygen saturation, and motion sensors for fall detection in elderly care. All of these devices collect data in entirely different forms, ranging from continuous, high-volume waveform signals to periodic, single-value measurements. IoHT is comprised of wearable and non-wearable devices, sensor/actuator circuits, wireless communication technologies, remote health monitoring, and IoT tools and applications [11]. The main element of IoHT is the sensors, which are lightweight and intelligent . The other main advantage of wearable sensors or IoHT is that, before appearing for a physical checkup, the doctor has some data or information about the patient, allowing them to make a more informed decision [12]. Choosing the right sensors for IoHT presents several challenges, including ensuring sufficient data rates for medical applications [13], ensuring secure data transmission, reducing power usage to extend node lifetime, and filtering out noise from human movement and surrounding environments [14].

IoT provides an ideal platform for the healthcare system, which facilitates interaction between the patient and the hospital (doctor). Due to the increase in population and the need for remote monitoring, as well as the rising cost of medical facilities and difficulty in accessing medical resources, IoT becomes an important part of the healthcare platform. In IoT-based healthcare, patients can be monitored remotely over the internet, allowing for the timely detection of critical situations and enabling the correct action to be taken at the right time, thereby saving the patient's life. IoT can also support the collection of patient data, which is used to compute information regarding the patient's health [15]. IoHT enables human-machine interaction and real-time health monitoring systems, hence increasing patient engagement in decision-making. Significant benefits of IoHT applications include cheaper costs in healthcare systems, real-time emergency response, and remote monitoring in pandemic conditions [16]. The architecture of an IoHT is shown in Fig.1.1 Developing a flexible and adjustable architecture is crucial because the Internet of Things must be capable of connecting diverse objects [17]. The structure of an IoHT systems

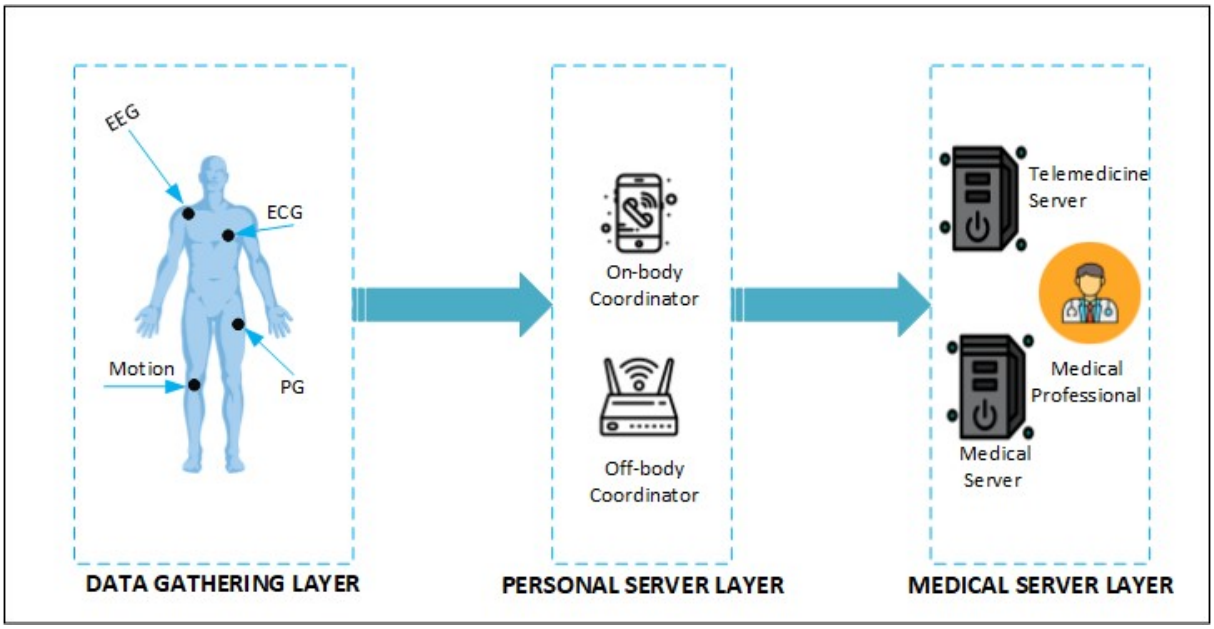


Figure 1.1: Architecture of an IoHT

consist of three levels: the medical layer, the personal layer, and the data gathering layer [18]. IoHT is using advanced technologies, which give healthcare to ensure continuous monitoring and management of the patient in real time situations. Data from the patient is collected through sensors or wearable devices. These devices include heart rate monitors, ECG sensors, glucose meters, and wearable fitness trackers, which display physiological parameters such as body temperature, oxygen levels, and heart rate, among other vital indications. The data can be collected seamlessly and unobtrusively through these smart devices, which are worn, implanted, or embedded within objects used daily.

Upon collection, the data is forwarded to the communication layer through networks, such as cellular networks, or more specialized IoT communication protocols, like Zigbee. However, this layer ensures that data flows from the sensors to the corresponding gateways, which are crucial for efficiently routing the data to cloud servers or local storage systems for further processing. Local data-based preprocessing at these gateways optimizes bandwidth constraints and reduces latency, particularly in cases of immediate action for emergency health issues. At the cloud computing and data processing levels, most computational activities occur. These stores and processes vast amounts of health data using big data analytics and machine learning techniques. This predictive capability is a powerful tool for early intervention and prevention, making patients feel more secure and in control of their health [19]. In practical deployments, the proposed schemes are particularly suited for ward-level patient monitoring systems, emergency triage units,

and remote healthcare environments, where multiple wearable sensors generate heterogeneous traffic streams. Routine physiological data (e.g., temperature, blood pressure) produce periodic low-rate traffic, whereas emergency events such as cardiac arrest or oxygen desaturation generate sudden bursts of high-priority packets. These traffic assumptions are explicitly considered in the simulation setup to reflect realistic IoHT operational conditions.

1.3 Congestion in IoHT Network

Congestion in IoHT (Internet of Healthcare Things) networks occurs when a data packet arrives at a network node at a speed exceeding the rate at which it is processed, forwarded, or stored. It leads to unwanted buildup of a queue above the buffering capacity, increased transmission delay, and ultimately results in packets being lost. While traditional computer networks have extensively studied congestion, its effects in IoHT bear far more dire consequences, as the data transmitted is time-bound and safety-critical. In non-medical IoT cases, such as environmental monitoring or industrial automation, the operational impact of a few seconds of delay or random packet loss is usually insignificant. In contrast, in IoHT, a few seconds' delay for delivering an emergency packet, such as a warning of cardiac arrest or oxygen desaturation, could result in irreversible injury or fatal consequences.

The main reasons for the congestion in an IoHT network are very complex and extensive. First and foremost, the hardware capacities are limited in many IoHT devices, which operate with some buffer constraints, modest processing power, and low-bandwidth wireless links. For instance, a typical wearable sensor node may be able to accommodate just a few dozen packets at a time; thus, when a sudden burst occurs during an emergency event, it can be quickly filled. Second, IoHT traffic is heterogeneous in itself because the data streams differ in size, frequency, and urgency. For example, an ECG monitor is a high-priority sensor that continuously generates data streams. Low-frequency sensors, such as temperature monitors, transmit their updates only once every few minutes. Both data types then flow into the same intermediate node, competing for the same buffer and bandwidth resources [20].

Moreover, the multihop nature of many IoHT networks worsens the congestion issue. The data in such networks needs to go through many other intermediate nodes (these are called cluster heads, relay nodes, or sink gateways) before it gets to the hospital server or the cloud system.

Every intermediate node aggregates traffic from several upstream devices, so congestion in one part of the network can cause flooding in the next downstream, triggering buffer overflows and packet loss. This is an issue that arises when emergency data needs to pass through congested nodes, as it will likely be delayed by other non-urgent traffic [21].

This adds complexity since many congestion control protocols already exist for IoT that operate in this manner. Traditional feedback-based schemes detect congestion reactively, only when packet losses have already accumulated. For example, with a sink node, a particular sender would only have the data rate adjusted when such an event occurs, such as a drop in delivery ratio or an increase in round-trip time. For IoHT, this late detection is intolerable, for it would lead to critical medical data loss. On the same lines, schemes based on static priority burden high-priority classes, such as ECG, with preference over low-priority classes, like temperature updates at periodic intervals, but even these ignore the intra-class urgencies of specific events. An emergency packet could still be delayed when it is behind other high-priority packets that are not time-critical [22].

Moreover, the multihop nature of many IoHT networks worsens the congestion issue. The data in such networks needs to go through many other intermediate nodes (these are called cluster heads, relay nodes, or sink gateways) before it gets to the hospital server or the cloud system. Every intermediate node aggregates traffic from several upstream devices, so congestion in one part of the network can cause flooding in the next downstream, triggering buffer overflows and packet loss. This issue arises when emergency data needs to pass through congested nodes, as it will likely be delayed by other non-urgent traffic [23].

This adds complexity since many congestion control protocols already exist for IoT that operate in this manner. Traditional feedback-based schemes detect congestion reactively, only when packet loss occurs. For example, with a sink node, a particular sender would only have the data rate adjusted when such an event occurs, such as a drop in delivery ratio or an increase in round-trip time. For IoHT, this late detection is intolerable, for it would lead to critical medical data loss. An emergency packet could still be delayed when it is behind other high-priority packets that are not time-critical [24].

1.4 Limitations of Existing Approaches

Moreover, various congestion control schemes have emerged within the broader research field of the Internet of Things (IoT). However, many of these schemes are not well-suited for the very stringent conditions of the Internet of Healthcare Things (IoHT). Indeed, most traditional protocols for congestion control in the IoT are designed for other applications, such as environmental monitoring, industrial automation, or smart cities, where an acceptable level of delay or packet loss is tolerable. However, for the case of IoHT, it needs to carry time-critical safety data, whereby a delay or loss in life-critical packets can have severe clinical consequences [25, 26].

A feature of most available control schemes for congestion is that they operate re-actively. Most ignore the onset of congestion until performance deteriorates, evidenced by packet loss, increased round-trip time, or reduced throughput [27]. For healthcare applications, such an approach should be satisfactory because it allows buffer overflow and dropping of essential data that should be salvaged when mitigation is attempted. Failure to counteract the initial losses rendered these reactive methodologies inadequate in IoHT environments, as care could be compromised by an individual packet loss that contained urgent patient information, thereby hindering timely intervention [28].

The pre-designed protocols have some control overheads that are likely to complicate the setup of IoHT. Any regime that causes active feedback messages, heavy control signaling, or complex negotiation mechanisms consumes a share of the few available resources in resource-constrained devices. Reduced effective throughput further for medical data consumption and increases power consumption in battery-operated healthcare devices that are usually scheduled over long periods of maintenance-free action [29, 30]. The excessive control load can also increase congestion during periods of high load, creating a feedback loop that further delays urgent transmissions. Furthermore, most IoT congestion control mechanisms are not fairness-aware and, therefore, fail to equitably allocate resources between high-priority and low-priority data deliveries during congestion. Prioritizing emergency data is essential; however, starving lower-priority medical data for prolonged periods can harm patient monitoring over the long term and compromise the accuracy of predictive modeling [31].

The most critical research gap is the absence of lightweight, adaptive queue management mechanisms for the IoHT. Most threshold-based systems tend to be static and cannot adjust to

sudden bursts of emergency traffic or changes in background traffic conditions. Also, computationally complex adaptive schemes are impractical for small, resource-limited medical sensor nodes. The lack of a scalable, adaptive, and clinically aware congestion control solution is another testament to the emergence of new techniques, which blend proactive detection, intra-class emergency prioritization, traffic isolation, fairness, and minimal overhead [27, 28].

1.5 Problem Statement

Despite existing congestion control mechanisms, current IoHT networks lack sender-side queue awareness, fail to provide intra-class emergency prioritization, and inefficiently utilize bandwidth due to redundant data transmission, leading to increased packet loss and delay under high traffic load. Following research problems are addressed in this research:

1.5.1 Excessive Acknowledgment Problem

The rapid development of IoT technologies in healthcare has enabled the deployment of wearable and implantable sensors in large numbers, allowing the continuous monitoring of the patient's vital signs, including heart rate, blood pressure, temperature, and oxygen saturation. These devices produce continuous data, which is then transmitted over intermediate nodes for processing and analysis. During emergencies, the number of packets passing through the network can increase rapidly. If the sending node is unaware of the queue status of the receiving or neighboring node, this sudden burst of data results in severe congestion, buffer overflow, and high packet loss. Making matters worse, there is no effective acknowledgement mechanism in place, so the sender is unable to modify its rate of transmission based on circumstances within the network. In fact, delays in packet transmission are worsened. Such transmission failures are potentially fatal in healthcare applications, as the loss or delay of critical emergency packets can realistically lead to late diagnosis, intervention, or patient death.

1.5.2 Classification of the Emergency Packets

Moreover, there is a related challenge in the queue management techniques currently available; the first type of intermediate node assumed to manage traffic within high- and low-priority queues usually processes packets strictly according to the FIFO rule of queue treatment. Such approaches treat all packets in the same class as equally urgent, while ignoring the fact that within the same priority class, emergency levels can differ. For instance, ECG data are routine, and cardiac arrest alerts are classified as both high-priority queues; however, FIFO processing could mean that the alert waits behind a number of non-urgent packets. Such an absence of differentiation is an intra-class emergency, as it adds extra time to critical packets in the waiting time queue, thus increasing the chances of packet loss and delay during real-time congestion. Such limitations will hinder the timely delivery of life-critical information while also reducing the reliability of healthcare IoT networks.

1.5.3 Data Duplication

Sensors frequently transmit identical readings to the intermediate node. Even when an item exhibits no discernible change, these sensors send every measured value in full 16- or 32-bit precision. Such states fall largely within a narrow margin of normalcy, where a person can be seen to do well or even have minor conditions. This redundant transmission wastes sensor energy, shortens the useful life of the device, and increases the communication overhead. Furthermore, such packets, being repetitive, may occupy a slot in the high- and low-priority queues, consuming space in buffers, and thus slowing down the processing of urgent data. At the end of it all, these problems—the absence of awareness of queue status, FIFO processing only without emergency sensitivity, and redundant data transmission—trigger an urgent need for an intelligent, queue-aware congestion avoidance mechanism that dynamically prioritizes emergency packets, reduces unnecessary transmission, and guarantees critical medical information is transmitted reliably and in a timely manner.

1.6 Research Questions

1. How to reduce the communication cost in priority based queuing in healthcare applications?
2. How to effectively prioritize the emergency messages in the high and low priority queues in healthcare applications?
3. How can an IoHT data transmission scheme be designed to intelligently detect and suppress redundant physiological measurements at the sensor level, thereby reducing energy consumption, minimising communication overhead, and ensuring the timely delivery of emergency medical data?

1.7 Research Objective

The objective of this study is to develop a priority-based queuing technique to control congestion in an IoT-based health monitoring system, which will ultimately mitigate congestion, especially during emergencies when numerous messages need to be exchanged between nodes and the gateway. The study aims to achieve the following goals in order to reach its research objective.

1. To minimize the communication cost due to excessive messaging in queue based solution.
2. To prioritize the emergent messages as per the severity level.
3. To enhance throughput by implementing redundant packet suppression mechanism while reducing congestion, delay, and energy consumption.

1.8 Aim of Research

The purpose of this research study is to design and develop intelligent congestion avoidance mechanisms for the implementation of the IoHT, ensuring the reliable and timely delivery of patient-state data, as well as energy efficiency. The scope of this study is to focus on three critical

challenges with existing healthcare IoT networks: the lack of queue status awareness at sender nodes, the inability of FIFO-based queue processing to handle intra-class emergency levels, and the excessive transmission of redundant physiological measurements that do not meet the patient's primary needs. By integrating queue-aware acknowledgement feedback, emergency-sensitive packet prioritization, and redundant packet suppression, the proposed research aims to achieve significant reductions in packet loss, latency, and network congestion especially in high-traffic emergency scenarios while maintaining the operational lifespan of resource-constrained medical sensors.

In addition, this research aims to develop and validate three new schemes: Queue-Aware Congestion Avoidance Scheme (QACA), Dual-Queue Aware Congestion Aviodance Scheme (D-QACA), and De-Duplication and Queue-usage Aware Congestion Avoidance Scheme (DD-QACA), which will collectively address the aforementioned inefficiencies. The schemes are designed to operate within the real-life operational constraints of IoHT systems, considering low computational overhead, compatibility with current network architectures, and adaptability to dynamic healthcare traffic behavior. By achieving these aims, the research aims to take the lead in the field of IoHT congestion control, promoting increased patient safety, reduced energy consumption, and improved overall quality of service in healthcare monitoring systems.

1.9 Scope of the Research Work

This research focuses on the design, development, and evaluation of mechanisms for mitigating congestion in the Internet of Healthcare Things (IoHT), it will specifically cover queue-aware forwarding and redundant packet suppression. The research mainely focuses on the transmission of physiological data collected from wearable and implantable sensors in healthcare monitoring systems, where reliability, low latency, and energy efficiency are of significant importance. The schemes recommended, QACA, D-QACA, and DD-QACA, are designed to operate with minimal computational and memory overhead while maintaining high levels of network performance.

The focus is limited to packet-level congestion control with no scope for physical-layer optimisations, hardware design, or security-specific solutions. It is assumed in the research that sensor nodes and intermediate devices operate within a trusted healthcare network, and the primary challenge lies in optimizing packet flow and priorities, as well as redundancy control,

to minimize delays and packet loss. The performance evaluation of the proposed schemes will be conducted through simulation-based evaluation using metrics such as throughput, latency, energy consumption, and packet delivery ratio to compare the existing approaches under varying network loads, traffic patterns, and emergency scenarios.

1.10 Summary

This chapter provides an overview of the research context, motivation, and problem statement, highlighting the major challenges concerning IoHT networks that undermine the timely, reliable, and energy-efficient transfer of patient data. Current systems suffer from three major problems: (i) sender nodes are not aware of the queue status; thus, packets get congested and are dropped; (ii) FIFO-based queue processing disregards emergency priority considerations within the same class, which delays packets that are life-critical; (iii) transmission of unchanged physiological measurements floods the network, wasting precious energy and sensorial life. These challenges were observed in the context of healthcare monitoring environments, where even slight delays or losses of emergency signaling may result in severe clinical consequences.

The chapter also outlines the purpose of the research, which is to design intelligent congestion avoidance mechanisms that integrate queue-aware acknowledgement feedback, emergency-sensitive packet prioritization, and redundant packet suppression. The boundaries of the study were delineated to address packet-level congestion control in IoHT environments, including the design and evaluation of the three proposed schemes QACA, D-QACA, and DD-QACA in simulated healthcare scenarios. By presenting the motivation, existing challenges, problem statement, aim, and scope, this chapter lays a firm basis for the forthcoming methodological and experimental design, which will serve to plug the existing gaps, thereby contributing to enhanced network performance, patient safety, and energy efficiency.

1.11 Thesis Organization

This thesis is organized into seven chapters, which contain all parts of the study. These chapters deal with specific aspects of analysis, discussion of the problem, methodology for finding solutions, experimental evaluation, and conclusion.

- Chapter 1 – Introduction

  Introduces the study's background, motivation, and significance. The patient will learn more about the challenges in Internet of Healthcare Things (IoHT) networks, such as network congestion, emergency data prioritization, and redundant packet transmission. This chapter discusses the research problem, its aim, scope, and objectives, which lay the foundation for the rest of the thesis..

- Chapter 2 – Literature Review

  An exhaustive review of the existing congestion control mechanisms for IoT and IoHT, coupled with queue management strategies and redundancy suppression techniques, is undertaken. The limitations of the existing schemes are critically discussed with a view to establishing a need for the proposed research.

- Chapter 3 – Research Methodology

  Describing the research framework, design approach, and simulation environment, which details the data flow architecture, traffic modeling, and profiling evaluation metrics. Additionally, the same applies to the evaluated parameters and tools used in the experiential setup.

- Chapter 4 – Queue-Aware Congestion Avoidance (QACA)

  Presents the first of the proposed schemes, which incorporates queue status feedback into the congestion control process. The chapter describes the problem of acknowledgments, the design principles behind QACA, and its operational workflow.

- Chapter 5 – Dual Queue-Aware Congestion Avoidance (D-QACA)

  The second proposed scheme, which refines packet scheduling by considering intra-class emergency levels for both high-priority and low-priority queues. It explains how D-QACA enables urgent packets to reach their destinations more quickly and how it reduces packet loss during periods of congestion.

- Chapter 6 – De-Duplication and Queue-usage Aware Congestion Avoidance Scheme (DD-QACA)

  Describes the third suggested scheme which integrates queue-aware message forwarding with suppression of redundant packets at the sensor level. In addition, DD-QACA improves energy consumption, reduces communication overhead, and guarantees timely delivery of important data.

- Chapter 7 – Results, Comparison, Conclusions, and Future Work

  Presents the simulated results for all three proposed methods and compares their performance with that of existing approaches. The chapter then concludes with a deliberation on the contributions, as well as the practical implications, limitations of the study, and possible future directions for research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The Internet of Healthcare Things (IoHT) enables a paradigm shift in patient management, allowing continuous monitoring from anywhere, remote diagnosis, and real-time medical intervention. It is typically set up with a variety of sensors for physiological data collection that must be delivered quickly and reliably, via low-latency networks, to support actual clinical decision-making. However, volumes of data generated by these sensors prove to be a real challenge in maintaining network performance across multiple patients, particularly regarding monitoring frequency. The congestion within the IoHT establishment can result in packet loss, high delay, and reduced throughput conditions, which can be catastrophic in a healthcare environment, depending on how it hinders emergency data delivery.

This chapter provides a review of current strategies for congestion management in the Internet of Health Things (IoHT), with a primary focus on queue-based approaches. The issues are explored in how the available solutions attempt to resolve the challenges of massive sensor data stream delivery while ensuring the timely delivery of life-critical information. Following this, various queue management strategies with respect to contrasting prioritization mechanisms and optimizations for data transmissions will provide the background for this study, which will address their strengths and weaknesses. The literature review critically reviewed all such techniques, highlighting important research gaps and opportunities to channel maximum efforts

toward improving the congestion control mechanisms of IoHT systems. This analysis lays the groundwork for the proposed study, which aims to address the shortcomings of existing approaches by introducing innovative queue-awareness and redundancy suppression.

2.2 Congestion in IoT Networks

When the volume of data that a node or communication link is required to handle exceeds the amount it can process or transmit, congestion occurs. It has a direct impact on the end-to-end Quality of Service (quality of service). Increased queueing time, packet loss, blocked connections, longer response times, and reduced throughput are general examples of such phenomena. In an IoT networks devices are typically highly constrained in terms of resources and energy, this presents a significant challenge for providing reliable and time-constrained data delivery, especially in mission-critical applications such as health monitoring. Therefore, this section highlights the current state of developed congestion control schemes in IoT networks and the architecture-agnostic strategies for congestion detection, avoidance, and mitigation while considering QoS-level applications.

2.2.1 Congestion Detection in IoT Networks

In IoT networks, congestion detection can be achieved through the identification of irregularities in the normal flow of traffic. this light, the Congestion-Aware Routing protocol was proposed, which can detect congestion through queue occupancy monitoring of both the current traffic load and the previous traffic load [32]. CoAR uses an adaptive buffer threshold mechanism to manage high packet traffic volumes. When the queue occupancy at a node reaches 50 percent of the node's total queue size, the parent's queue size is increased; otherwise, it is decreased [33]. The CoAP-R, on the other hand, improves network performance during heavy traffic conditions by detecting congestion through current and past measurements of channel load and buffer occupancy. When CoAP-R detects that the sending rate of a source node has exceeded a certain allowed maximum. In that case, it will control the packet-sending rate to curb congestion, essentially enforcing a maximum (min fair allocation of the available bandwidth) to determine the proper data transmission rate [22].

The new congestion notification mechanism discussed in [34] for low-power IoT devices works in resource-poor environments while relieving congestion. The scheme specifies the Congestion Info Block (CIB), which encodes the congestion status of nodes along a routing path into binary status, allowing for efficient communication of congestion conditions with a network manager at lower costs compared to standard ECN. Another work in [35] addresses continuous object detection, which tends to perpetuate congestion, leading to packet loss and constant energy consumption. The CDCAPC algorithm, in conjunction with PCCS, is used to manage incoming traffic. The receiver accepts packets based on priority; if there are no high-priority packets, acceptance is determined by a threshold value. If congestion persists, RBNACC constructs clusters and updates the border node's parent for load balancing.

With the rapid growth of numerous devices within a single IoT platform interconnected via the Internet, a maximum number of different networks will result, leading to congestion as more devices join. One of the techniques under consideration is measuring demand on a single link by monitoring the quantity and size of transmitted packets. The effect of load measurement on data transmission rate adjustment, based on the status of network conditions, is another aspect of research. A network threshold increase, therefore, causes an immediate modification in some transmission rates, making it particularly suitable for lightweight IoT devices [36, 37].

In such heterogeneous networks, it is clear that congestion has an indirect impact on the performance of IoT applications. The naïve Bayesian TCP (NB-TCP) mechanism is designed to monitor packet loss rates across different types of wired and wireless networks, and classify all packets into high- and low-priority categories. In congestion, lower-priority packets are served first, with high-priority packets being dropped only when necessary. On the receiver side, the indication of a congested state is indicated by missing packet sequences, which serve as an indicator of congestion. The throughput for NB-TCP is an improvement over several methods that achieve higher fairness and friendliness as compared to earlier methods [38].

The Adaptive Congestion Management-Carrier Sense Multiple Access (ACC-CSMA) protocol has been developed to sense and measure the level of congestion using a backoff index, which evaluates the channel's status. The network is subdivided into primary and secondary segments, assigning appropriate devices to each segment. Any instance of severe congestion would then require the secondary devices to reroute towards secondary networks from their present bottlenecks and help in a quicker reach to their destinations. One application of this scheme is in cellular uplinks, where drones do scheduling to prevent routing through congested IoT

devices, thus reducing channel load. This scheme additionally integrates an online congestion management mechanism for dynamic adaptation [39, **?**]. Other illustrations are from a smart city where many users submit requests for synchronously required information (e.g., parking availability, traffic updates, weather reports), leading to access congestion at the gateway due to high cache-hit rates. To improve connectivity and reliability while reducing packet drop, a multi-gateway architecture integrated with a resource caching strategy and a load balancing approach based on Multi-Criteria Decision Making (MCDM) is proposed [40].

Sometimes, IoT systems interact with devices continuously through control and status messages with one central server. Therefore, as more devices connect to a server, the probability of buffer overflow increases because more retransmissions occur due to lost packets, adding to congestion and thus increasing delay. This is most critical for the delivery of health applications in IoT technology, which requires timely delivery. Furthermore, in Mobile Ad Hoc Networks (MANETs) supporting the IoT, such phenomena generate larger volumes of messages. The new approach of Bandwidth-Aware Routing Strategy (BARS) considers both residual bandwidth along network paths for transmission and available cache space to store data temporarily. Simulations in the NS-2.35 show that BARS improvements in the delivery ratio, overall delay, throughput, and packet loss [41].

In this paper [42], Anitha et al. present a comprehensive survey of detection signals and locations for congestion within low-power Internet of Things (IoT) networks. This paper categorizes detectors into queue-centric, MAC-centric (channel busy/CCA), delay-based (one-way or round-trip time, RTT), and loss indicators, and discusses the impact of granularity (node/local/edge) on responsiveness and overhead. A good contribution found in this paper is the taxonomy mapping detection metrics to signalling strategies (piggybacking versus dedicated control). The survey concludes that queue-aware hints are the most applicable in constrained motes. However, it indicates that static thresholds sometimes fail in bursty workloads such as those mostly seen in emergencies. In IoHT, authors recommend hybrid detectors that combine local light signals with occasional edge verifications to minimize false positives.

Lim limits his study to congestion detection within RPL-based networks for 6LoWPAN, usually deployed in the IoT. Among the mechanisms upon which the paper bases its review for RPL extensions are queue occupancy monitoring, parent-level load indicators, packet loss patterns, and inter-arrival versus service-rate comparisons. It weighs up the costs of putting congestion hints into DIO/DAO messages against a separate signalling method. According to

Lim, these costs should ideally be routed compatibly so that they do not increase the overhead in existing control-plane churn in which topologies are maintained in RPL. The survey is vital when considering IoHT gateways that use RPL because it shows the extent to which routing control channels can double as timely congestion indicators when well-designed [43].

The proposed metric would become part of RPL's objective function, allowing nodes to choose less loaded parents. All detection is local and piggybacked into existing routing beacons and reducing overhead. This design is particularly appealing for resource-constrained deployments, as it requires no additional control frames and allows traffic to be redirected early from hotspots. Two major limitations are likely routing oscillations when many nodes change parents simultaneously and must assume that alternative parents have sufficient capacity, both of which require damping or hysteresis in practice. In healthcare scenarios, the methodology adopted by QU-RPL tends to provide preliminary alerts of imminent overload conditions at no signalling cost [44].

This enhances the routing purpose of RPL by formulating a congestion-aware objective, where traditional metrics (such as ETX) are complemented by information from buffer occupancy. The detection function is threshold-based—if occupancy exceeds the pre-set watermarks, the node advertises its load and influences the parent selection process. The strong point of CA-OF is that it detects congestion directly in routing decisions, eliminating the need for separate feedback channels. However, CA-OF is based on careful weight choices between reliability and congestion metrics; misconfiguration can lead to suboptimal routes or oscillation. Within an IoHT setting, the need for adaptive-weight tuning arises from the diverse clinical traffic profiles but is generally useful for routing by CA-OF [45].

Instead of only probing queue depth, OHCA provides an integrated detector that promptly predicts congestion by comparing the rate of arrival with service rates at the parent. Before the buffers overflow, the system may begin to experience growing stresses as a result of an arrival-vs-service comparison. Lightweight routing updates are used for detection in conjunction with limited mitigation or rerouting. The detection methodology is more sensitive to short bursts than basic queue thresholding techniques. However, this sensitivity itself depends on short-term rate estimations, which might be noisy over limited networks [46].

CoAR synthesizes various signals into a composite congestion score using a multi-criteria decision method (TOPSIS): queue utilization, ETX (link quality), and residual energy. In this way, the detection incorporates both the congestion risk and the link reliability, thus preventing

an overreaction to temporary queue spikes on a poor-quality link. While the multi-metric approach does reduce the chance of false detection compared to single-metric schemes, it introduces calibration difficulty (i.e., how to weight the criteria) and a greater computational burden. Therefore, deploying TOPSIS on tiny devices in the IoHT may require much offloading or simplified variants. However, it will help strike the right balance in the detection, which in turn avoids unnecessary reroutes, thereby saving the emergency packets [47].

PC-RPL demonstrates a marked increase in detection under loss and collision patterns, for it is known that most failures induced by congestion under high-load conditions are due to contention and not by pure buffer overflow. This methodology tracks packet loss trends, including early indicators and MAC-layer retransmissions, and combines them with queue metrics to determine whether to initiate mitigation. Since most contention spikes occur before buffer overloads, the combination is especially advantageous in packed situations. Practically speaking, the drawback is that in lossy wireless channels, it is difficult to distinguish between loss from collisions and connection faults; incorrect identification could result in needless re-navigation. The advantage of using PC-RPL for medical applications is that it is more resilient in situations involving several beds or excessively congested wards, particularly when it comes to collision detection [48].

Marco et al. introduce some MAC-layer indicators (busy-channel probability, growing contention windows) as congestion detectors and factor them into a select reliability metric for selecting a parent. The fundamental idea is that increasing MAC contention will presage a backup in queues and packet loss. Fast reaction times arise from detecting congestion at the MAC layer, as contention rises before queues actually overflow. One constraint is that MAC-layer signals may be local and transitory (like temporary interference), so you must have noise filters to prevent oscillatory responses. In IoHT, such MAC-aware detection would be valuable, as many wireless devices compete for the same medium; however, this detection is susceptible to misinterpretation [49].

Using available bandwidth estimation, this detection metric, combined with ETX, helps avoid parents near service saturation. This allows it to measure the shareable channel capacity and flag nodes with low remaining capacity. Available bandwidth captures actual forwarding capability rather than only buffer state; therefore, this can be useful for low-bandwidth lines, as link bitrates often vary. However, the drawback is that bandwidth estimation imposes probing overhead and can create disturbances on low-power radios, so these probes must be designed properly to be infrequent. Detection considering bandwidth prevents low-capacity congested

links from forwarding alarms in IoHT [50].

GTCC cast the detection as a local, simple metric (generation rate - service rate), and used game-theoretic reasoning to choose whether he should change his parent or adapt his rate. The detection signal is meant to be simple and piggybackable, but the novelty lies in how nodes reason about the likely responses of others to prevent oscillation. In this context, churning is thus reduced even in a high-density topology. However, implementing the complexity of this game model can prove a hurdle; however, the base detection metric is lightweight, making it suitable for IoHT nodes that must avoid heavy computation [51].

2.2.2 Congestion Avoidance in IoT Networks

Wireless mesh technology has become a functional base for the next generation of Internet infrastructure. Connecting a larger number of devices to the IoT network may result in congestion and capacity restrictions. The solution to this is an Integrated Markov State Transition-Open-Loop Smart Caching (MST-OLSC) technique that aims to maintain a constant data flow rate and prevent packet overflow. The method employs a Markov state transition scheduler to learn about transmission patterns, predict future data trends, and optimize schedules. An open-loop smart caching mechanism is additionally integrated, such that each data packet is assigned a token prior to transmission. These tokens are placed in a buffer (bucket). The number of tokens in that bucket represent the transmission rate at which packets are transferred and alleviating traffic congestion. Simulation results demonstrate that MST-OLSC outperformed existing approaches in terms of packet delivery ratio, end-to-end latency, and energy consumption [52].

With the rapid development of today's IoT technology, assuring reliable data delivery across all heterogeneous networks, which differ widely in buffer sizes, is one of the significant challenges. For this, a hybrid protocol combining the Imperialist Competitive Algorithm (ICA), Grey Wolf Optimizer (GWO), and Whale Optimization Algorithm (WOA) has been introduced. It maximizes a fitness function prioritizing nodes with higher residual energy, greater memory capacity, and favorable neighborhood characteristics. For congestion avoidance, dynamic memory allocation of a node is proportioned with respect to its Initial Buffer Size (IBS) and remaining buffer provisioning. Mathematical modeling is used to explain the congestion behavior in the event of a buffer overflow [53].

To avoid NTS (Network Time Slot) collisions, a mechanism based on Backoff is used. Each

node has a Backoff counter defined here. During idle time slots, while the node is waiting, this counter is decremented. The node whose counter first reaches zero is allowed to send its NTS packet, thereby avoiding any transmission conflict [54]. One of the technological challenges in 6G mobile networks is primarily maintaining constant connectivity with latencies that do not approach packet loss. Due to heterogeneous device connectivity, variable internet access speeds generate traffic congestion. In this context, a congestion-avoiding mechanism with slice admission control was proposed. The network is sliced into several independent slices, each with its admission control, which avoids collisions and satisfies quality of service standards [55].

In a constrained network, high retransmission rates and low throughput can result in degraded TCP performance. However, conventional Transmission Control Protocol (TCP) retransmits packets until it receives an acknowledgment, which may cause congestion in specific paths. In this case, an adaptive retransmission timer is introduced, whereby the retransmission timeout of a connection decreases by 0.5 seconds until it reaches 0, and then retransmission is initiated when sufficient bandwidth is available for its use, leading to reduced congestion. With home Wi-Fi networks, small TCP flows, and an increasingly shared, limited medium, performance for IoT devices may be impaired. That is why Adaptive Admission Control (AAC), a queue management policy, and a Restricted Access Window (RAW) scheduling algorithm have been implemented at the Wi-Fi access point to achieve better fairness and reduce congestion. It allocates time windows for groups of devices to access the shared channel one after the other, thereby improving individual response time and reducing contention through Markov regeneration processes [56].

Due to the continuous change in topology, congestion control proves to be quite a challenge. In this work, as presented in [57], the contribution of a trust-aware routing scheme embedded with queue status monitoring for enabling bandwidth-aware routing is demonstrated. This scheme was implemented using the NS2 simulator, which achieves improvement in congestion management in very dynamic MANET environments.

CODA was one of the first congestion-avoidance frameworks designed to provide an end-to-end solution for wireless sensor networks. This two-hop protocol consists of three elements: detection of congestion, determination of hop-by-hop back-pressure that operates in a logically primitive manner (it forwards a request backward, thereby eliminating high-rate data transmission), and concurrent adjustment of aggregated rate, invoked by both the sink and loaded upstream nodes. In principle, CODA is implementation-specific — upstream nodes perform proactive suppression only when downstream buffers are much overloaded from non-

urgent traffic. Although suitable for low-power sensor motes, it is not useful for correlated event-flood-triggered losses [58].

ESRT presents congestion avoidance as a reliability control problem: the sink determines an actionable event reliability, instructing the rate of event transmissions from sources to match this level while controlling congestion. Feasible avoidance utilizes rate adaptation with feedback from the sink to the source, aiming to limit redundant transmissions while maintaining a minimal coverage rate. Strengths: global reliability requirements indeed match application criteria; among the lightest-weight sources. Weakness in IoHT: Latency might be introduced when obtaining sink feedback. ESRT's aggregate reliability metric, if implemented, might only mask an emergency moment on either a per-patient or per-instance basis, thus ruling out the possibility of choosing the highest level of unaltered emergency prioritization [59].

Many avoidance strategies can reduce the amount of traffic through data aggregation, such as in-network aggregation, temporal suppression (sending an event only if the value is considered changed by a significant threshold), or semantic compression. From a healthcare application perspective, edge aggregation can serve as a means to condense waveform data or acknowledge only certain events. Additionally, this arrangement places direct stress on the queuing system and is energy-efficient; however, the disadvantage is that latency and loss of fine granularity can result in loss of service if this data has an emergency priority for immediate treatment [60].

This scheme ensure that critical queues do not empty during periods of light loading while maintaining considerable account capacity for class-2 traffic during times of heavy traffic. Strengths dialogue: easy extensibility and apparent separation; useful for preserving emergencies. Weakness: static sizes, thresholds can and are likely to cause some forms of under utilization (the threshold cannot be set correctly) or starvation due to being out of tune for many scenarios and adjustments [61].

2.2.3 Congestion Control in IoT Networks

Congestion control is defined as the set of strategies, mechanisms, and processes for regulating the flow of data packets entering a network to prevent performance degradation due to congestion. Congestion control may comprise reactive measures to reduce congestion after it occurs or proactive measures to prevent it [62, 63]. In terms of the Internet of Things (IoT), where many devices are constantly exchanging data over restricted and heterogeneous network

infrastructures, effective congestion control is established to retain Quality of Service (quality of service), deliver pertinent data messages on time, and not lose unnecessary packets in environments where very high data loss will not be acceptable. This section presents an analysis of contemporary research efforts focused on congestion control in IoT networks, outlining their core principles, implementation techniques, and performance.

The new value is based on the Information-Centric Networking (ICN) paradigm, which shifts communication from a host-based addressing to a content-based addressing paradigm. The essential goal is to locate and deliver information, rather than establishing direct end-to-end connections. This means that this architectural shift will eventually be beneficial to IoT networks because of the native caching that ICN offers, which enables reduced redundancy in transmissions. One hierarchical model based on ICN has suggested adaptive congestion control mechanisms through chunk-by-chunk partitioning of contents as a strategy for managing network performance. The partitioning of content enables a fine-tuned control mechanism for disseminating content in a network such that chunks could be throttled or prioritized dynamically according to real-time information on congestion levels. Performance gains could be realized from this model, which ran simulation tests using the ndnSIM simulator, as the network would exhibit decreased congestion rates and a lower cache hit ratio, suggestive of more efficient data dissemination [64]

While this network approach to creating opportunities is beneficial to ICN, its operation may generate a high volume of request packets, which can subsequently lead to increased network traffic and congestion. Consequently, a smart congestion control system is designed that utilizes green IoT sensors based on ICN for efficient network operation. This system employs an adaptive scheduling policy for sensors, where data packets are withheld from transmission when the studied level (e.g., urgency of data or triggered value measurement) falls below a specified threshold. Such packets are instead cached locally before being transmitted when conditions are more favorable for the network, thus reducing unnecessary transmissions and lessening congestion [65].

IoT deployments with a high density can produce a large amount of data from many devices, which can cause packet loss and network congestion. Using sophisticated queue management techniques is one method of addressing this congestion. In order to dynamically prioritize network traffic in real time according to the needs of the applications, the Priority Queue-based Token Bucket Algorithm was introduced. Thus, by first forwarding high-priority packets using

its discretionary criteria, PQTBA can operate in either preemptive or non-preemptive mode. The token bucket scheme enforces the data rate by allowing packets to enter based on availability. According to simulation tests, PQTBA outperforms conventional queue management techniques in terms of energy efficiency, packet drops, and throughput. Such a system dynamically modifies the packet transmission rate based on the network quality, traffic buffering capacity, and quality-of-service-related priority requirements. A multi-stage CCE powers this approach. The first step involves checking the recently arriving packets; the second step is completed by the CCE, which uses a Random Access Network (RAN) detection method to determine congestion. Finally, once congestion has been detected, packets are either redirected or cached until they can be sent via another path. This multi-step procedure prevents needless network overburden by enabling dependable packet transfer [66].

A new challenge is emerging with the use of edge computing in IoHT: load balancing and congestion control between edge servers. One of these load-balancing schemes proposes a method that is supposed to be divided into four phases: (1) setting threshold values for load levels, (2) selecting appropriate edge nodes to handle incoming traffic, (3) replacing cached data when necessary, and (4) querying data in an optimized way. When a router becomes overloaded, it forwards excess packets to neighboring edges, thereby refreshing and clearing congestion. Simulation results have shown that this technique has higher efficiency than traditional models of edge computing. However, it mainly targets local load sharing and does not extend to resource sharing with cloud resources [67].

5G IoT networks have high prospects by integrating mobile edge computing (MEC), network function virtualization, and software-defined networking (SDN) under a single operational condition. Consequently, controlling efficient data flow remains a challenge for the highly dynamic and heterogeneous traffic. A container-based virtualization framework for MEC is thus proposed to improve the quality of service at the IoT gateway. The solution comprises two main functions: the Traffic Offloading Function (TOF) and the Radio Network Information Service (RNIS). The combined existence of both functions in a deployable platform enables these flow control algorithms to flexibly modulate traffic flows while ensuring higher service availability [68].

2.3 Congestion Problems in IoHT

In light of population aging and the burden of chronic diseases, this article [69] demonstrates how IoT might transform healthcare. To enable ongoing monitoring and event-driven alerts, they propose an end-to-end architecture that incorporates wearable/implantable sensors, secure communication stacks (including encryption and key management), and cloud storage/analytics. These include emergency response, early deterioration detection, and remote treatment, which lowers hospital readmission rates and expenses. Semantic data models and lightweight, standards-based device middleware are demanded from the standpoint of interoperability design. There is a need for standardization of protocols for device discovery, identity, and quality of service, as well as ML systems for predictive risk scoring and workflow automation.

Although the validation of clinical gaps and regulatory alignment has been noted as an outstanding challenge, the Internet of Things is actually positioned as a foundational enabler. This study examines AAL as a specific implementation of IoT for the elderly and individuals with disabilities. An architecture incorporates wearable technology, in-home sensors, gateways, and caregiver dashboards into a policy model with escalation and human-in-the-loop interventions, mapping AAL use cases such as fall detection, activity recognition, and medication adherence onto sensing modalities and context-aware services. The design contribution places a strong emphasis on exploiting location context and temporal rules to prioritize safety-critical occurrences. The study examines explainability, unobtrusiveness, and user approval as key factors in promoting acceptance. The evaluation is based on scenario-based analyses and small pilots, with findings demonstrating considerably reduced response times and improved general situational awareness for caregivers. Major limitations include personalization overhead, device heterogeneity, and maintaining privacy in shared living spaces. The paper recommends using edge analytics to reduce data exposure and latency [70].

This work builds on AAL by surveying the technology stacks that allow for the creation of assisted living ecosystems, including environmental sensors, wearables, smart appliances, and telepresence. Levels of service are defined, and challenges in the data fusion towards context inference are discussed. The major methodological contribution is a framework for designing risk-aware policies triggering graded interventions (notification $\rightarrow$ confirmation $\rightarrow$ actuation). Usability studies indicate clear benefits such as autonomy and the perception of safety, but also disadvantages such as alarm fatigue and complicated configurations. Proposed privacy-by-design

patterns have been outlined. The authors have contended for interoperable ontologies to avoid vendor lock-in and enable longitudinal care records across providers. [71].

This paper brings together evidence from various studies on how networked ambient technologies improve quality of life, operationalizing it with these three dimensions: mobility, social connectedness, and safety; linked to measurable digital indicators; focusing on reliable sensing under everyday noise conditions (occlusions, lighting, device drift) and graceful degradation: it requires standardized protocols for evaluation that will put together valid combined clinical, human-factors, and cybersecurity endpoints [72].

Extensively put forth is M2M/P2M interaction across the considered sectors within IoT, which minimizes human intervention in programming and service management, with reference to the enabling elements, namely sensors/actuators, RFID/NFC, low-power radios, and cloud/edge integration, in very simple terms and reference workflows with applications right from data acquisition to application services. Touching constraints for safety, adaptability, and energy efficiency must be defined, and advocacy made for context-aware duty cycling as well as dynamic reconfiguration. Big data analytics will, hence, be of paramount importance in extracting value from such hordes of telemetry, although again, caution is very important on matters surrounding data quality and governance. Security must thus be viewed holistically (device hardening, secure boot, PKI, anomaly detection). The roadmap, thus, talks of development in terms of interoperability stacks, self-configuration, and privacy-preserving analytics cite kaur2017.

Bhatia et al.'s research concerns an IoT-based model for telemonitoring conditions inside intensive care units, which is affected by quite a number of intrusions from the environment and also a few limitations in workflow at the level of critical care. The architecture comprises multimodal sensors, fault-tolerant gateways, and alarm rule-based analytics. Deployment in three ICUs showed higher detection sensitivity and lower false alarm rates than the baseline monitors. Quality of service classes prioritize the life-critical data in order to ensure data transport, while buffering and local filtering reduce the network burden during load spikes. Further improvements in response time and documentation accuracy are also reported in the study. Limitations include the combination with existing hospital information systems and a burst-dependent network. Future work addresses adaptive thresholds through ML and building robust low-latency failover routes [73].

This paper discussed the transformative potential of the Internet of Things in healthcare, with a focus on enhancing continuous monitoring, operational efficiency, and supporting patient

outcomes. It articulates the strata of application (chronic disease management, perioperative monitoring, rehabilitation) and inversely maps them onto network requirements (latency, reliability, confidentiality). A major contribution is the discussion of device interoperability and privacy as the foremost barriers, proposing a solution for policy-driven data sharing and standard APIs. Cost models and clinical workflow integration are briefly presented. The results show promising improvements in term of productivity. Edge computing can be utilized for preprocessing sensitive signals to minimize exposure to the cloud and conserve bandwidth [74].

Focusing on congestion in IoT-based medical facilities and WBANs, this review dissects how dense deployments and correlated events overload networks. It classifies congestion control strategies into rate control, queue management, and priority scheduling, analyzing their impact on delay and reliability. A key insight is the need for class-aware handling that protects urgent physiological data under stress. The survey highlights cross-layer cues (MAC contention, queue pressure) as robust early indicators. It notes evaluation gaps: many proposals rely on small-scale simulations without real WBAN dynamics or patient movement models. Recommended directions include hybrid edge-cloud control loops and clinical-grade SLAs for emergency traffic [75].

This presents an energy-efficient congestion control routing protocol for WBAN healthcare applications. Traffic is divided into normal and emergency classifications and routing decisions are based on the remaining energy of each node, the local congestion level, and the reliability of the link. It attempts to direct urgent data along low-latency paths, while regular data traffic is sent along routes that conserve energy. Simulation results demonstrate improved emergency throughput and a longer network lifetime compared to several baseline protocols. This protocol uses lightweight congestion indicators, ensuring feasibility on body-worn medical devices. It has trade-offs, namely, potential for larger route changes under rapid changes of posture, and is sensitive to parameter selection. The authors mention using adaptive hysteresis and a mobility-aware link prediction scheme to reduce oscillations [76].

Structured review (2015-2020) classifying IoT-in-healthcare research under sensor-based, resource-oriented, and security-focused approaches. It incorporates an inventory of platforms, communication stacks, and application archetypes (telehealth, geriatric care, chronic monitoring). Interoperability and scalability emerge as persistent bottlenecks, with homogeneity across devices/protocols impeding integration—security challenges span authentication for constrained nodes and end-to-end confidentiality under intermittent connectivity. Future themes flagged by

the paper include fog computing for time-sensitive tasks, blockchain for auditable sharing, and Internet of Nano-Things (IoNT) for ultra-fine monitoring. It requires benchmark datasets and reproducible evaluation pipelines to advance the field [77].

It is meant for smart health care, where CloudIoT should coordinate with wearable sensors and connect to cloud analytics to provide continuous monitoring of chronic diseases. The whole process involves acquisition, secure transfer, storage, and scalable processing concerning multi-tenant isolation. Privacy threats associated with data movement between cloud services are analyzed, and recommendations are provided for using encryption, access control, and differential auditing. It introduces the idea of fog computing for offloading some preprocessing tasks (e.g., feature extraction and anomaly detection) to the edge, considering latency constraints and bandwidth limitations. A conceptual framework maps tasks to tiers based on latency and energy budgets, which is extremely important for a compliant (HIPAA/GDPR-like) and auditable implementation in practice [78].

It develops an exhaustive study of the merits and demerits of IoT in healthcare analytics under the improved controls, immediate responses, and economic benefit modernizations. It includes the classification of healthcare data like remote monitoring, urgency systems, and assistive technology that correspond with different network service requirements. The safety and security of trust are the primary focus of this study, with a major emphasis on ensuring data confidentiality and privacy, including consent management and data minimization. The interoperability among legacy systems in hospitals is deemed a key driver for scaling the initiative. Furthermore, the future would concentrate on standards-based onboarding of devices, privacy-preserving analytics, and automated compliance checks. The authors welcome techno-social assessments along with technical performance measurements [79].

QQAR is a Q-learning-based QoS-aware routing protocol designed by Arafat et al.in [80] for IoMT and wireless body area networks (WBANs). In QQAR, forwarding nodes are trained via reinforcement learning to avoid congested parents and utilize priority flags at each hop in guiding queue handling. This adaptive mechanism promotes throughput while reducing end-to-end delay under congestion scenarios. A key strength of QQAR is that self-adaptive learning is achieved, as the learning mechanism can adapt to various network conditions. The study focused on routing and parent selection, with queue management as the only secondary mechanism. The protocol also imposes higher computational and energy overhead, which could be a challenge to resource-constrained IoHT devices.

Altowaijri et al.in [81] proposed a healthcare data dissemination protocol in IoMT, incorporating deduplication awareness, along with AODV routing and priority queue management. This solution reduces redundant packet transmissions and distinguishes between high-priority and low-priority data, thereby improving bandwidth utilization and reducing congestion. The proposition achieved improved packet delivery ratio and reduced collision probability. However, since AODV was designed as the routing layer, this approach is limited in scalability for larger IoMT deployments. Handling of intra-class emergency severity was also not addressed. This is primarily network-layer focused, with limited integration of adaptive or explainable intelligence.

Stitini and others in [82] discussed the combination of IoMT with Explainable AI (XAI) in a triage-oriented MQTT broker system. Their design included priority queues added to the patient data streams and integrated XAI techniques to justify why certain data had been prioritized. The proposed method addresses one of the fundamental problems in the field of healthcare AI: trust and transparency. That is why it will give reasons for queue treatment in an interpretable manner. The experimental evaluation confirmed reduced delay for high-priority traffic and thus increased acceptability of automated systems by clinicians. The solution is broker-based and therefore unable to circumvent network-level congestion considerations. Moreover, a dual queue mechanism with intra-queue severity handling was not introduced, thereby hampering its applicability to large-scale IoHT deployments.

Morais Barroca et al. explore how the IoT facilitates a transition to home-centric care. They elaborate on how continuous, longitudinal wearables and ambient-sensor-based monitoring systems could alleviate the burden on hospitals, especially when considering chronic diseases. Tele-consultations and medication management support have a twin focus on mHealth apps and secure backend integration. Data governance has been discussed with regard to the revocation of consent and the limitation of purpose. In addition, integration hurdles with existing clinical workflows and device ecosystems have been pointed out. Usability findings suggest adherence gains but caution against notification fatigue in the absence of personalization. The authors champion standardized APIs and edge-intelligent systems to reduce reliance on cloud [83].

Deep learning enhances big data analytics, yet challenges persist in extracting trustworthy information from large datasets. For that reason, it reduces computational power near the network edge, together with transmission processing and delays. The fog paradigm computerized platforms able to distribute operations are expected to contribute to this reduction of latency in IoHT systems [84].

Bhuiyan et al. analyze the advances in IoT, big data, cloud computing, and WSNs concerning health care. The authors recommend the use of real-time analytics on streaming physiological data for enhancing quality care, but the caveats of scalability, cost, and data veracity come with it. The authors speak concerning the architecture, like lambda/ktheta-style pipeline and tiered storage, that are used to balance latency and retention. Security sections concern their paradigm based on anomaly-based intrusion detection and fine-grained access control in terms of clinical roles. Examples depict dynamic data collection that optimizes triage and chronic care but need further development into scalable solutions that preserve privacy, according to the authors. Recommendations feature standardized schemas, edge filtering, and federated learning for cross-institutional models [85].

Akshatha and others proposed in [86] a priority-enabled MQTT broker that will improve the handling of emergency events in health IoT networks. It incorporated three virtual queues (Urgent, Critical, and FCFS) into the MQTT protocol stack to allow emergency data processing before non-critical updates. The experiments showed improvement in packet delay and jitter, along with a reliable flow for urgent events. The method is practical because MQTT is popularly used in IoT deployments. The solution, however, is broker-centric: no modifications are required by the end devices. It also lacks sensor triage capabilities. Intra-queue severity ordering was not supported, and the system did not provide explainable insights into prioritization.

Ouakasse et al.[87] proposed a QoS-gateway model for the IoMT, which enables pre-queuing emergency packets before they enter the network core. Incorporating lightweight priority queueing in the gateways demonstrated an increased reduction in packet loss probability as well as enhanced throughput as a result of the scheme in simulations. The work concerns one of the principal bottlenecks in IoMT, specifically the points of aggregation, where congestion is most commonly expected. Its main advantage is deployability since it is easier to modify gateways than individual devices. The static classification mechanism was the most significant flaw, as it required codecs to associate packets with predefined priority tags for each message. The study did not include information on emergency severity inside the queue nor adaptive ML-driven classification.

Ahmed and Paulus in [88] investigate the effects of buffer overflows resulting in packet loss in resource-constrained 6LoWPAN-based IoT healthcare networks. The authors discussed the analytical model to compute the packet loss probability due to buffer saturation, thereby enabling a predictive understanding of congestion dynamics. Their performance analysis depict

that conventional congestion management schemes which designed for general IoT applications unable meet the stringent reliability and low-latency requirements of healthcare scenarios. Quantifying both packet drop rates and buffer loss probabilities provides insights into feasible queue management strategies for the timely delivery of critical data. The main contribution is to demonstrate that active monitoring of buffer states would significantly enhance reliability; however, adequate real-time adaptation to changing patient data loads is not thoroughly explored.

In [89] proposes a new Priority Queue-based Token Bucket Algorithm (PQTBA) to address the congestion problem in IoT networks. PQTBA distribute incoming traffic according to priority levels assigned based on the urgency of the application using discretionary laws combined with a hybrid preemptive/non-preemptive scheduling scheme. This allows for the prioritized transfer of critical real-time data, such as emergency healthcare readings or indicators of acute patient deterioration, ahead of the bulk of routine or less urgent data. Simulation results have shown that PQTBA surpasses recently best-performing state-of-the-art methods with respect to throughput, packet loss ratio, and energy efficiency. Notably, its structured token bucket mechanism regulates transmission rates to prevent queue overflows while ensuring fairness among data flows. However, with the added complexity in priority classification, the algorithm also demands accurate and real-time traffic profiling to keep performing well.

Mazloomi et al. in [90] presented a priority-based congestion avoidance scheme (PBCAS) in which incoming data is categorized as critical or non-critical. Packets belonging to the critical group are then routed using a shortest-path SVM classification, while TOPSIS is assigned to route the non-critical data. Other studies addressed the problems of buffer-aware scheduling, fuzzy controllers, and dynamic rate adjustments to reduce packet loss and delay. However, most of these still treat packets within the same category as equal, which is contrary to the ground reality, wherein some emergencies supervene even among those classified as high priority or those classified equally with low priority. This could lead to undue delay for those packets, which is of far greater concern.

There are three traffic classes defined by the scheme proposed by Buenrostro-Mariscal et al in [91] , urgent, important, and best effort, which subsequently have specific queue management policies for these traffic classes. The scheme is specifically designed for Internet of Medical Things traffic and includes QCCP, a prioritization-based congestion control protocol. Simulation-based results show that urgent traffic achieves lower packet loss and higher throughput with less variable latency than the other traditional schemes. QCCP also proved efficient in terms of

fairness, but with a preference for life-critical data. Classification granularity is fixed, and all packets under it receive similar treatment, leaving little room for the kind of nuance required in a more nuanced emergency hierarchy. The process does not leverage adaptive learning or explanation upon which to base decision support.

Tseng et al. in [92] conducted a comparative study of edge-IoMT traffic management, analyzing the scheduling policies FIFO, Priority Queue (PQ), and Weighted Fair Queuing (WFQ) for medical applications. Their experiments at the edge of the network demonstrated that PQ and WFQ reduced jitter and latency for high-priority medical streams compared to the FIFO method. This demonstrated that deploying quality of service scheduling in healthcare edge computing would be an effective approach. The study may provide useful information, but it is primarily evaluative and thus favors a comparison between existing scheduling policies rather than proposing a new algorithm. Additional points concerning the assessment of severity levels within classes of emergencies or the intelligent classification of medical traffic with regard to dynamic quality of service assignment were also not covered. The author in [93] addresses the problem of scheduling in IoT healthcare applications, where even a slight transmission delay could lead to scenarios such as a patient's death. The authors propose a Prioritized Scheduling (PS) scheme as an enhancement to the Earliest Deadline First (EDF) scheduling mechanism. Whereas standard EDF fails in overload scenarios, the PS scheme assigns priority weight adjustments to ensure that life-critical packets are serviced even in situations when the system is starved for resources. Simulation results show a marked improvement in timeliness for high-priority traffic, which consists of delay-sensitive samples, compared to the baseline EDF. However, there are scheduling and classification overheads in utilizing the proposed mechanism, and the implementation itself becomes highly complex with larger deployments. Thus, while PS enhances responsiveness, its computational cost targets limitations to ultra-low power applications in medical sensor networks.

The Distributed Congestion Control Algorithm (DCCA) that combines priority-based routing with distributed congestion monitoring. Differing from centralized approaches, DCCA allows each node to locally detect congestion and adjust its forwarding strategy to ensure the timely delivery of high-priority healthcare data. Simulation studies have validated its efficiency in reducing packet loss and maintaining high delivery ratios even under stressful conditions. The algorithm works to benefit the overall reliability of IoHT networks by clustering congestion management responsibilities at nodes. The only drawback is that the system's dependence on

continuous changes in priority assessment may increase processing overhead, which has yet to be evaluated for scalability in ultra-dense healthcare deployments [94].

This author in this [95] article features an SDN-fog orchestration layer for WBANs, enabling computation offloading and queue and resource-aware scheduling at the edge. Flow control will rely, in this case, on a centralized control mechanism that can change congestion-determined frame number flows and dynamically adjust critical versus routine measurements by controlling queues. The redundancy filters out, for example, constant baselines, which limits the load on backhaul channels as well as reduces occupancy in the queue assigned to gateways. Evaluation results indicate increased throughput, reduced delay, and improved reliability compared to static routing. For IoHT, the contribution is an operationalizable control plane that is realistic for hospital networks in terms of integration cost and practical concerns, such as controller availability during outages.

Nguyen et al. in [96] explored the integration of $M/M/C/K$ priority queuing models within a three-tier IoT-Edge-Fog-Cloud infrastructure specifically for healthcare data. Their research demonstrates that by assigning high priority to critical medical alerts (e.g., sudden heart rate changes) while treating routine telemetry as low priority, the average response time for life-critical events is reduced by 30%. This study emphasizes that static queuing is no longer sufficient for the heterogeneous nature of IoHT, advocating instead for a dynamically adjustable buffer system that scales based on the severity of the incoming medical data packets. The author in [97] conducted an in-depth study on controlling congestion during massive traffic bursts in wireless medical sensor networks. Their work utilizes intentional sequential queuing delays as a feedback mechanism to regulate the data transmission rates of sensor nodes. By applying these queuing constraints at the gateway level, they proved that stability can be maintained in the network even when thousands of medical devices are active simultaneously. This approach prevents the "drop-tail" packet loss common in healthcare facilities where dense sensor deployments often lead to saturated channels.

Nair et al. in [98] investigated the relationship between waiting times and service quality in healthcare through a survey on advanced Queue Management Systems (QMS). Their research models the movement of health data through virtual queues, where priority is determined by a hybrid of medical urgency and available bandwidth. This paper bridges the gap between traditional operations research and IoHT by treating "data packets" and "physical patients" as analogous entities within a shared queuing system, proposing that efficient digital queuing

directly correlates with improved clinical outcomes. The authors in [99] presented a framework for using Artificial Intelligence to mitigate network congestion in IoT-based healthcare systems. The researchers utilized a three-state Markov model to predict periods of high network utilization and applied machine learning to adjust queuing thresholds in real-time. By training the model on historical IoHT traffic patterns, the system can preemptively re-route medical data to underutilized fog nodes, effectively flattening the traffic spikes that typically cause packet loss in emergency departments.

Efrosinin et al. in [100] analyzed the use of machine learning in queuing theory based on a $GI/G/K$ system, which is particularly relevant for the complex, non-exponential arrival rates of IoHT data. Their work suggests that standard $M/M/1$ models fail to capture the "bursty" nature of medical video and imaging data. By using neural networks to estimate the parameters of the $GI/G/K$ queue, they developed a resource allocation strategy that provides upper and lower bounds for performance metrics, allowing hospital IT administrators to guarantee a minimum Level of Service (LoS) for critical devices. Peshkova et al. in [101] addressed the computational challenges of simulating high-volume IoHT networks by introducing splitting-based regenerations for accelerated queuing simulations. Their research provides a mathematical method to more quickly estimate stationary performance metrics, such as mean response time, in complex $GI/M/1$ systems. This allows for the rapid testing of new congestion control protocols in a virtual environment before they are deployed in high-stakes clinical settings, where downtime or failure could have fatal consequences.

The authors in [102] modeled and analyzed performance in hybrid IoHT systems where servers (such as virtual machines for medical data processing) require a non-negligible setup time. By employing a level-dependent quasi-birth-and-death (LDQBD) process, they derived upper and lower bounds for the stationary distribution of data in the system. This research is crucial for cloud-based healthcare services that use auto-scaling, as it provides a queuing-based framework to determine when to "wake up" additional processing power to prevent incoming medical data from backing up. The authors in [103] introduced the SCONE protocol, which enables networks to share throughput advice directly with IoHT endpoints, such as high-definition medical imaging tools. This protocol integrates with L4S (Low Latency, Low Loss, and Scalable throughput) to manage the "application queue" at the sender's side. By allowing medical applications to self-adjust their data rates based on proactive network feedback, the system avoids the delays and interference typically caused by the standard "reactive" congestion control cycles of traditional

TCP.

Pappone et al. in [104] introduced Mutant, an online reinforcement learning (RL) algorithm designed to adapt to the best-performing congestion control schemes in real-time. Unlike traditional ML models that require extensive pre-training on static datasets, Mutant learns from existing protocol behaviors and network fluctuations. In the context of IoHT, this allows the network to maintain an optimal throughput-to-delay ratio even when medical sensors experience abrupt channel fluctuations or varying RTTs (Round-Trip Times), outperforming standard TCP variants in mixed-traffic environments. Oleiwi in [105] proposes a hybrid AI model that combines predictive learning with adaptive decision-making to address congestion in high-demand IoT environments. The research utilizes simulation tools and real-world medical data to demonstrate how hybrid models can forecast "congestion hotspots" before they occur. By integrating these predictions into the network's queuing management, the framework achieved significant improvements in latency and packet loss, specifically tailored for applications requiring real-time feedback like remote patient monitoring.

Ali et al. in presented a Hybrid Cyber Attack Prediction (HCAP) model in [106] that identifies security-induced congestion within IoMT (Internet of Medical Things) environments. While primarily a security paper, it provides a critical link to network performance: it uses ML to distinguish between natural traffic bursts and DDoS-driven congestion. The study highlights that by reducing false-negative rates in threat detection by 20%, the system can maintain stable queues and prevent "malicious" congestion from disrupting the transmission of critical health data. ArXiv presented in [107] Survey on ML-AQM provides an updated taxonomy of machine learning approaches specifically for Active Queue Management. The authors review how Reinforcement Learning (RL) is now being used to discover optimal packet-dropping policies that are "health-aware." Instead of the traditional "tail-drop" method, which can indiscriminately discard vital medical signals, these ML-driven AQM systems learn to drop non-essential background traffic during saturation, ensuring that the health-related queues remain at a predictable and manageable length.

2.4 Critical Analysis of Existing IoHT Congestion Control Approaches

While the preceding discussion has explored a wide range of congestion detection, avoidance, and control mechanisms within the Internet of Healthcare Things (IoHT), it is equally important to evaluate these studies in a structured and comparative manner critically. Such works will enhance awareness of the various design philosophies and implementation strategies, along with their strengths and weaknesses, from which the identification of research gaps for future works within this thesis can be identified. To achieve this, a synthesized comparative analysis for the selected state-of-the-art contribution is provided in Table 2.1. It captures the core aspects of each study, i.e., its basic ideas/methodological focus, the advantages offered within the context of IoHT healthcare applications, and the potential disadvantages or unsolved challenges associated with it. With such a structure in place, an in-depth understanding is facilitated by the way various approaches address congestion in IoHT, with improvements still to be made. The above critical analysis then serves as motivation for the present proposed investigation, focusing on queue-aware congestion avoidance mechanisms for healthcare provision purposes.

Table 2.1: Critical Analysis of the Existing Scheme

Ref	**Basic Idea**	**Advantages**	**Drawbacks**
[69]	Integrated IoT healthcare stack with secure, interoperable, energy-aware design, edge filtering, and QoS.	Continuous monitoring; standards promotion; bandwidth & energy relief.	Validation gaps; security overhead; device heterogeneity.
[70]	AAL with policy-based safety prioritization, multi-modal sensing, and edge analytics.	Faster response; user-centered; reduced latency & exposure.	Personalization cost; privacy in shared spaces; brittle priorities at scale.
[90]	a priority-based congestion avoidance scheme (PBCAS) which categorized data into critical or non-critical. Packets belonging to the critical group are then routed using a shortest-path SVM classification, while TOPSIS is assigned to route the non-critical data.	Dynamic rate adjustments to reduce packet los	No focus on emergency severity and high delay in the processing of emergency packets.
[73]	ICU tele-monitoring with QoS, local filtering, and reduced false alarms.	Higher sensitivity; fewer false alarms; better documentation.	Integration complexity; burst-dependent; needs adaptive ML failover.
[75]	Survey of congestion control in medical IoT/WBANs; hybrid edge-cloud control.	Class-aware handling; early indicators; SLA-based quality care.	Small-scale sims; poor mobility modeling; lacks clinical validation.
[27]	Energy-efficient congestion-aware WBAN routing with emergency/routine classes.	Better emergency throughput; longer network life.	Route oscillations; tuning sensitivity; mobility handling needed.

Ref	Basic Idea	Advantages	Drawbacks
[108]	Analytical model for buffer overflow in 6LoWPAN healthcare networks.	Predicts congestion risk; guides provisioning.	No real-time adaptation; lacks emergency priority; model overhead.
[87]	The gateway has provisions for handling urgent medical traffic on a priority basis; simulations recorded higher throughput and lower loss rates.	offers high-quality service benefits and is very easy to integrate at gateways in hospitals.	NStatic classification, no intra-queue emergency severity, and no ML/XAI justifications.
[89]	Priority-Queue Token Bucket shaping for medical traffic.	Better throughput & PLR; fair rate control.	Needs accurate priority labeling; risk of misclassification.
[80]	RL-based routing, including priority flags and an output priority queue to keep the body-area traffic.	Adapting to the congestion, uses priority forwarding improvements for reliability.	training overhead and more energy consumption.
[93]	EDF-based prioritized scheduling for medical traffic.	Timely delivery under overload; fair & auditable.	High complexity; power-hungry for low-end sensors.
[94]	Distributed congestion control with priority-aware routing.	Reduces loss; avoids bottlenecks; scalable.	Processing overhead; unclear in ultra-dense setups.
[95]	SDN–Fog orchestration for WBAN healthcare.	Higher throughput; low delay; SLA compliance.	Integration cost; controller dependency; needs robust fallback.
[91]	With three traffic classes (urgent, important, best-effort), prioritization works without heavy control overhead.	Improvements in throughput/latency tailored for IoMT, with reduced overhead.	Fixed granularity; no intra-queue severity or ML-driven urgency; queueing not tied to per-packet triage.

Ref	Basic Idea	Advantages	Drawbacks
[92]	Edge architecture comparison of FIFO, PQ, and WFQ for the provision of congestion control to the medical streams.	PQ/WFQ was observed to reduce delay/jitter for critical flows at the edge, serving as effective quality of service knobs.	Static priorities; no emergency-severity tiers.
[109]	Cloud–Fog healthcare with privacy controls and fog preprocessing.	Privacy & security; latency & congestion relief.	Policy complexity; compliance overhead.
[82]	MQTT-based triage of patient priority handling	XAI expositions on the need to justify priority among patient/message delivery	Limited network-level CC evaluation; not dual-queue

2.5 Challenges in IoHT

The poor performance of IoHT, especially in terms of security, latency, reliability, and efficiency, presents serious challenges to the development of such systems. An advanced IoHT system must embody low latency, high dependability, and low power consumption, while also minimizing packet loss, buffer loss ratio, and security risks. The following section discusses the issues that the different authors face in IoHT:

2.5.1 Latency Issues

Low latency is crucial for H-IoT operations, including real-time data processing, emergency response systems, and remote patient monitoring and telemedicine. Low latency enables medical professionals to make swift decisions and respond promptly to emergencies. However, achieving minimal latency might be difficult, requiring specific hardware and software, as well as the optimization of Network infrastructure. The transmission and interpretation of information in

real-time are significantly challenging in H-IoT, as delays in processing and transmission may hinder the ability of medical professionals to make informed decisions. The time-sensitive nature of an IoHT system requires extremely low latency for end-to-end transmission and data processing. To reduce total delay, communication technology with high availability and bandwidth should be explored and utilized [110]. Fog computing in H-IoT can reduce the adverse effects of cloud delay on systems that respond to emergencies. However, there are still issues regarding response times, latency, and energy consumption that require addressing.

2.5.2 Real Time Data Processing

There have been serious issues for real-time data processing in the IoHT environment at this point, such that it has turned out to be a reluctant issue. Healthcare sensors and devices are sources of continuous, infinite streams of data, such as vital signs, motion patterns, or environmental conditions, which need to be interpreted almost instantaneously if they are to be of any clinical use. Most of the time, data is usually so large that it exceeds the carrying capacity of conventional processing systems, hence causing delays in their application, which can be life-threatening in scenarios requiring immediate intervention, such as cardiac arrest detection or a sudden drop in oxygen saturation levels. Most of the data are fed from multiple and mostly different sources and making synchronization and alignment of the streams quite difficult for accurate interpretation. Network congestion invariably brings buffering and queuing as phenomena that cause decelerated processing and introduce latencies, and hence untrustworthy alerts and decisions. The continuous collection of unprocessed raw data, which overwhelms all available resources and also increased storage utilization. If these problems are not resolved in real-time processing, the IoHT will not function properly, which means that improper diagnoses could occur, as well as delays in treatment and even risk to the patient. Therefore, it has become not just a technical challenge but an immediate necessity in healthcare to solve real-time processing bottlenecks [111].

2.5.3 Energy Consumption

In the Internet of Health Things (IoHT), energy consumption is a prominent challenge, particularly for the wearable devices and implantable ones. Wearable devices include fitness

trackers or smart medical gadgets that can be conveniently charged, ensuring they remain operational with minimal downtime. With implants—such as pacemakers, neurostimulators, or biosensors that are implanted within the body—long-lasting and maintenance-free power sources are necessary to ensure the safety and reliability of their patients. Procedures for replacing or recharging implant batteries tend to be invasive, risky, and costly; therefore, actively optimizing energy is paramount. Optimizing resource management for devices with limited processing capability, low energy reserves, and limited memory is crucial to maximizing network lifetime, especially in light of large-scale IoT deployments. The newly launched fog and edge computing encourage the adoption of short-range power infrastructures with minimal expenditure, allowing data to be processed closer to the data source and reducing energy costs for transmission [112, 113]. The use of these devices needs to be cost-effective, support fast charging, and be safe for continuous human use to be viable. These also need to provide stable output capability without interfering with sensitive medical electronics. If energy constraints continue to prevail, the associated problems of device downtime, data loss, and reliability cessations would compromise patient care and safety. Thus, the development of energy solutions for IoHT, requiring efficiency, longevity, and safety, is not only a challenge but also a priority in healthcare [114].

2.5.4 Bandwidth

Due to the growing number of data-rich applications and the corresponding connected medical devices, bandwidth limitations have emerged one of the major challenge in IoHT. Advancements in health technology require integration with next-generation technologies like 5G, which will saturate most of the currently available bandwidth during high-demand medical events. The IoHT applications include telemedicine consultations, transfers of medical images, and real time patient monitoring. All these applications generate huge volumes of data requiring high-speed and reliable transmission. Such continuously flowing streams can saturate the network bandwidth, resulting in delays, jitter, or packet losses; the quality of service of patient care will be affected. The increase in telemedicine to be held remotely and led to videoconferencing being established as a vital communication platform for modern healthcare systems. Stable and adequate bandwidth is a necessity for the uninterrupted transmission of high-quality audio and video signals, enabling clarity and avoiding interruptions. This demand also intensifies when many patients are being

monitored simultaneously or in conditions where live high-resolution imaging with diagnosis data is transmitted during remote surgical assistance or specialist consultations. Absence of adequate bandwidth in such situations will result in delays, low resolution, or sometimes drops, which may interfere with timely diagnosis and interventions. The adoption of emerging IoT and HMI technologies has compelled governments and organizations to equip their networks with sufficient bandwidth to support real-time and multimedia-rich applications. If bandwidth issues remain unaddressed, they could seriously undermine the reliability, speed, and safety of patient care services. Underscoring the importance of resolving this issue promptly for the future in IoHT [115].

2.5.5 Massive Data

Medical Healthcare 4.0 generates enormous amounts of data from a vast array of distributed medical equipment. Processing huge datasets can strain storage and computational resources. Healthcare data can be structured (e.g., EHRs), unstructured (e.g., medical imaging), or streamed via IoMT devices. Integrating and evaluating various data demands complicated and advanced processing techniques. One of the main challenges is to handle the rapid data generation and real-time processing of older systems. Real-time analysis is crucial for making prompt clinical decisions and administering timely treatments. Obtaining valuable insights from complex healthcare datasets requires advanced data analytics techniques and expertise. Effective data analysis might take a lot of resources and effort [116].

2.5.6 Packet Loss

Packet loss represents a significant challenge in IoHT systems, as even a small amount of missing packets will degrade the delivery quality and reliability of medical services. It is defined as a failure of one or more data packets traveling from one electronic junction to another destination, often caused by congestion, unreliable wireless links, interference, or hardware limitations. Packet loss is especially dangerous in healthcare systems because the majority of information it transfers is life-critical data, such as real-time vital signs, medical imaging, or alerts concerning emergencies. In other words, missing or delayed packets will result in incomplete monitoring and reading faults or diagnosis delays; any of these can seriously jeopardize patient

safety.

In real-time IoHT scenarios, particularly in the above categories—such as telemedicine meetings, surgical interventions at a distance, or continuous patient monitoring—packet loss fundamentally translates into a drop in service quality. For example, it could lead to freezing of frames, distorted audio, or dropped calls during telemedicine calls with video, thereby hampering the clarity of verbal communication between physicians and patients. A loss of packets in wearable or implanted medical devices could mean a failure to record crucial abnormal readings, such as an arrhythmic heartbeat or sudden drops in oxygen levels, which require an immediate response and can be even more serious. In addition, the increase in repeated packet retransmissions raises the network's burden, exacerbating it, especially in battery-limited devices, due to congestion and increased energy consumption.

The problem worsens in healthcare such as multidrop hospital facilities with connected IoHT devices operating concurrently, or in rural/remote areas with insufficient infrastructure for networking. Unconventionally, packet loss may compromise the efficiency of IoHT systems by reducing the levels of confidence among healthcare providers and patients. Therefore, there is a need for the development of robust transmission protocols, as well as strategies for network management and error detection, which are specific features necessary to ensure the accuracy and timeliness of life-critical health data [117, 118].

2.5.7 Data Extraction

Medical devices in IoHT environments are used continuously to record a variety of physiological parameters from patients, including heart rate, blood pressure, oxygen saturation, respiratory rate, and body temperature. For aiding clinical decision-making, Physiological measurements are essential for understanding a patient's current health status. The quick and accurate extraction of such data would also aid in the early detection of potential disease occurrences and would generally help prepare for preventive or emergency interventions. However, raw data-generated sets are large, heterogeneous, and recorded at variable sample rates, making the separation of clinically relevant data very hard. Some physiological signals may also contain noise, artifacts, or irrelevant fluctuations, which can complicate their interpretation and lead to unreliable readings. Patients may exhibit different conditions even though they offer the same measurement, because sometimes the parameters are context-dependent—for example, patient activity or their medical

history. Further complicating the interpretation is that one must examine multiple sensor streams, requiring additional synchronization and alignment opportunities before any meaningful analysis can be conducted. Data extraction in this context is not just about rooting out the technique, but also addressing the critical bottlenecks in active health insights. Without effective extraction, downstream analytics will necessarily be compromised, even for diagnosis and treatment recommendations. Furthermore, since every device manufacturer has a proprietary data format, this also forms a hindrance to standardized accessibility. Privacy and security considerations complicate matters even further, as such sensitive physiological data have to be handled in a confidential and yet clinically useful manner. In high-stakes healthcare emergencies, even minor inaccuracies in the extracted data can translate into delays in intervention, or worse, incorrect medical decisions. Hence, extracting accurate, meaningful, and timely physiological parameters from IoHT devices is crucial for ensuring safe, reliable, and effective healthcare delivery [119, 7].

2.6 Problem Identification

The increasing number of aged people has raised an issue in the sector of healthcare due to the risk of suffering from dangerous diseases because the population of older people is expected to reach 761 million in 2025, which is approximately double compared to 1990 [55] With the increase in the number of older populations, the number of patients who require monitoring will also increase, which will in turn increase the cost of treatment and create burden on the hospital. As TCP congestion control is not used because it is designed for traditional, connection-oriented networks and adds unnecessary overhead, which is inefficient for IoHT systems. IoHT devices often operate with limited bandwidth and low power, making TCP's retransmission and acknowledgment mechanisms unsuitable. Instead, our approach focuses on energy-efficient, real-time congestion control at the application and network layers, which better aligns with the needs of healthcare applications in IoHT environments. Due to this, a proper remote health monitoring system can play an important role in reducing hospitalization, alleviating the burden on hospitals, and reducing treatment costs.

2.6.1 Acknowledgement Problem

Today, IoT devices are widely used in the healthcare field to collect patient data. In such scenarios, different types of sensors are attached to the patient's body for the collection of data regarding vital signs. During the emergency, many packets need to be sent over the IoT network. The large number of packets sent over the IoT network can create congestion if the sender node is unaware of the queue status of the neighboring node. Furthermore, without acknowledgment, the sender node remains uninformed about the queue's condition, which increases packet loss, delay, and congestion. In healthcare, during emergencies, loss of packets can lead to the death of a patient. Therefore, it is essential to minimize network congestion, which can cause delays and packet drops.

The absence of an effective acknowledgment mechanism can lead to worsens the situation [94, 88]. Without acknowledgment feedback, the sender node has no way of knowing the current status of the receiver's queue; therefore, it continues sending packets unboundedly, resulting in queue overflow, packet loss, and an increase in transmission delay. Conversely, such situations are typically unacceptable in a healthcare scenario, as the loss of life-critical data during emergencies can result in a delayed diagnosis, which in turn delays the implementation of treatment; hence, it could lead to a fatal outcome. Therefore, it is necessary to design the congestion avoidance mechanisms that duly incorporate queue-aware acknowledgment feedback. The idea is to enable the sender node to adjust its transmission rate according to the actual queue states of neighboring nodes, thereby minimizing packet loss, reducing delays, preventing congestion, and finally assuring that urgent medical data is reliably and timely delivered.

2.6.2 Classification of Emergency Packets Problem

Most existing IoT and IoHT congestion control schemes process packets in both high-priority queues and low-priority queues in accordance with the First-In-First-Out (FIFO) principle [94, 89]. Although the FIFO technique defines equality in terms of the order of arrival, it is a fact that not all packets in the same priority class have the same level of urgency. Existing mechanisms tend to treat all packets contained in one queue as equally important and process them one after the other without differentiating between routine data and life-critical emergency data. For instance, an emergency cardiac arrest alert from within the high-priority queue will

receive the same treatment as a periodic ECG monitoring packet, even though the latter requires instant transmission. The absence of intra-class differentiation allows delays on emergency packets simply because they arrive later than non-urgent packets; hence, increasing the risk of critical information dropping while congestion builds up.

The queueing limitations mentioned above have enormous implications for health-related applications. FIFO queues can cause critical emergency packets to wait longer because they are delayed behind non-critical packets, resulting in packet loss, transmission delays, or even network congestion, all of which are highly unacceptable in life-critical conditions. In some cases, this kind of delay due to processing order may lead to a lost opportunity for timely medical intervention, such as establishing CPR, adjusting medications, or alerting a physician in a timely manner. Thus, there is a need to progress from merely FIFO-like scheduling within priority queues to emergency-aware processing methods that dynamically identify truly urgent packets and expedite their transmission, regardless of their arrival order. This way, time-critical medical information is delivered with the quickest availability possible, while ensuring an effective network. The time factor is therefore carefully considered.

2.6.3 Packets Duplication Problem

Level 1 sensors are part of healthcare monitoring systems and can either continuously collect or relay physiological data. These can include heartbeats, blood pressure, oxygen saturation, and temperature data from patients to intermediate processing nodes. In this data sharing capacity, the existing system is deficient because most facilities lack filters to weed out duplicate or duplicated measurements before transmission. As a result, the sensors send values that continuously repeat, even when the values read unchanged or fall within the normal range. Thus, to manage the incoming packets at the intermediate node, high-priority and low-priority queues differentiate critical and non-critical data, respectively. However, these queues are often filled with a large volume of non-urgent, repetitive measurements that consume valuable network resources without adding meaningful diagnostic value [94, 88, 89].

Putting aside the transmission of redundant data in its complete 16-bit or 32-bit version means that net waste is compounded: it first involves parting with precious battery resources from these resource-constrained sensors, reducing their operational time and increasing maintenance costs. By using up the bandwidth, this issue increases the cost of communication. These idle packets

in the network can create congestion by occupying both important and unimportant queues. In extreme cases, it can cause some emergency packets to lose or, worse, delayed delivery because the buffers are full, resulting in direct impairment of patient safety. This all makes intelligent data transmission methods a necessity for detecting stable and non-threatening scenarios and replacing redundant with lightweight indicators or flags, thus freeing network capacity for really urgent health care data.

2.7 Summary

This chapter critically reviews the literature addressing fundamental issues regarding IoHT-based healthcare systems, especially relating to congestion. Uncontrolled queues on resource-constrained devices cause buffer overflow and packet losses, affecting the reliability of clinical data the most. Likewise, emergency packet classification ensures that transmission continues even during emergency conditions in the network. The final area is redundant packet identification, which proved to be wastage when consuming valuable bandwidth and processing power, resulting in congestion. Delay in data transmission still counts as one of the fundamental limitations. Because latencies undermine the efficacy of timely medical interventions in remote monitoring and real-time scenarios. The reviewed works have been elaborated broadly to cover a lot of mechanisms to counter these issues. But most of the proposed solutions have trade-offs in terms of complexity, scalability, and energy efficiency. In comparative analysis, the chapter discussed the importance of integrated frameworks that can resolve such critical issues in a much broader context to build dependability in IoHT healthcare applications.

CHAPTER 3

METHODOLOGY

3.1 Overview

This study is based on the recognition that IoHT systems, tailored for supporting life-critical healthcare applications, face unique constraints that are primarily strict and interconnected. In such an environment, network performance is not merely a matter of efficiency, but a measure that determines the safety of patients and the quality of clinical decision-making. Queue awareness, emergency packet classification, redundant packet identification, and minimization of delay are not just desirable features, but operational necessities. The IoHT infrastructure comprises several heterogeneous devices that communicate at varying data rates and levels of criticality. Network conditions become increasingly dynamic, requiring solutions that enable real-time adaptation while ensuring the quality of service and quality of experience for both patients and medical professionals. Chapter 2's literature review reveals that numerous works have been conducted in the area of congestion control and scheduling mechanisms; yet, almost all fail to address these four constraints holistically within an integrated, context-aware framework.

Hence, the Methodology Chapter would serve as the conceptual and operational bridge linking the theoretical bases and identified research gaps of the literature review with the practical execution of a novel, comprehensive solution. It would present a structured step-by-step roadmap that details the research design, system modeling, algorithm development, simulation configuration, and evaluation methods.

3.2 Operational Framework

The operational framework of the research aims to provide a systematic path for further investigating and alleviating some of the critical issues identified in the literature review, including queue awareness, emergency packet classification, identification of redundant packets, and delay minimization in IoHT networks. The functional framework stitches together conceptualization, analytical modeling, simulation design, and performance testing into a single-track methodology, thereby providing traceability from the definition of the problem to the interpretation of the results. The start of the operational framework involves conceptual mapping of the IoHT system and its constituents, including the interaction between wearable/implantable devices, wireless sensor networks, gateways, and cloud/fog computing nodes. This mapping is based on the operational constraints and requirements of life-critical healthcare scenarios, where any degradation in Quality of Service directly jeopardizes patient safety. With the insights from the literature review and identified research gaps, the operational framework proceeds to the abstraction of the problem, where the specific challenges targeted are formally defined in terms of the network parameters, performance metrics, and operational boundaries involved. With the problem definition in place, the framework goes on to the design of proposed mechanisms. As detailed in the research objectives, these mechanisms are intended to work at different layers of the IoHT protocol stack and include strategies for:

1. Queue-aware monitoring to prevent real-time congestion.
2. High-accuracy classifying emergency packets and ensuring their prioritized delivery.
3. Identifying and mitigating the impact of redundancy in packets to reduce congestion, while ensuring that critical data delivery is not impaired.

An operational framework based on simulation-driven evaluation is employed to validate each of the proposed mechanisms. Simulation models that describe realistic channel states, variable traffic loads, patient mobility, and heterogeneous device conditions for healthcare networks are then developed. Multiple simulation scenarios are then set up to assess the mechanisms, both in singular and combined forms, offering an understanding of effectiveness alongside trade-offs. Finally, in this comparative performance assessment, the proposed mechanisms are evaluated against existing solutions from the literature. This step enables the evaluation

to remain relevant, contemporary, and actionable, forming the basis of its comparison. The results are subject to critical analysis and provide recommendations for reflection, highlighting limitations and suggesting areas for further research. All simulation results are averaged over 10 independent runs with different random seeds to ensure statistical reliability and to mitigate stochastic variations in traffic generation and channel behavior. Linking theoretical insights with targeted mechanism design and empirical validation ensures that this operational framework ultimately attains truly scientific and practical relevance, with a view toward the next generation of IoHT systems.

3.3 Simulation Framework

The simulation framework used in this research is designed to evaluate the performance of congestion control and packet prioritization mechanisms within Internet of Healthcare Things (IoHT) networks. This framework models realistic healthcare application demands by simulating traffic loads, varying channel capacities, and node energy consumption, while testing three proposed mechanisms: QACA, D-QACA, and DD-QACA. The simulation operates under two channel configurations: 120 kbps (low bandwidth) and 250 kbps (medium bandwidth), which reflect typical bandwidth conditions in real-world IoHT deployments, such as low-power medical sensors and higher-capacity medical gateways. The purpose of using these different configurations is to highlight the performance differences in terms of congestion levels, packet loss rates, and the ability of the congestion control mechanisms to handle network traffic efficiently under varying bandwidths. The simulation parameters are based on realistic healthcare network scenarios, where traffic is generated periodically and in bursts by emergency sensors such as pulse oximeters, ECG sensors, blood pressure sensors, glucose sensors, and accelerometers. These sensors create diverse network loads, offering a comprehensive view of how the congestion control mechanisms perform under different stress conditions. The framework is designed to replicate both in-hospital and remote patient monitoring environments. To model the network conditions accurately, the channel model incorporates both ideal and impaired conditions. These impairments include free-space path loss, log-normal shadowing caused by obstructions like the human body, multipath fading due to reflections in clinical environments, and interference from other IoT devices. These factors help simulate real-world conditions where environmental factors

such as fading, interference, and obstacles affect network performance. For this simulation, NS-2.35 was utilized, which is a widely used discrete-event network simulator known for its capability to model network protocols and traffic dynamics in realistic wireless and IoT environments.

3.3.1 Channel Model

The channel model used in this research mimics the short-range medical wireless technology operation characteristics of the IEEE 802.15.6 (WBAN) and IEEE 802.15.4 (ZigBee) standards, which are typically implemented in IoHT environments. The model implements both ideal and impaired channel conditions comprising free-space path loss, log-normally distributed shadowing due to human body obstruction, and multipath fading caused by reflections in clinical environments. Interference from co-located IoT devices in the same spectrum is created for emulating realistic hospital environments. The simulations are run for two specific capacities: 120 kbps and 250 kbps, representing practical deployment scenarios in the IoHT continuum, ranging from low-energy wearable devices to higher-capacity medical gateways. The capacity influences packet transmission time, buffer occupancy, and, hence, important delays; hence, these rates will be a critical parameter for assessing queue build-up detection, emergency packet handling efficiency, and redundancy suppression efficiency. By incorporating these capacities explicitly into the wireless channel modeling, the assessment will be able to demonstrate how varying data rates affect the Quality of Service in life-critical healthcare communication.

3.3.2 Node energy Model

Adapted to the node energy model within this research, the model considers the operation of low-power IoHT-sensing and IoHT-communication devices, whose energy efficiency directly translates into the lifetime of a network and its reliability. Each node is equipped with a finite energy source and operates according to an energy consumption profile that varies depending on its different operations, as follows: sensing, transmitting, receiving, idle, and sleep. The energy cost associated with packet transmission and reception has been taken into consideration, considering the power draw as a function of data packet size and channel capacity utilization. Control messages consume less energy due to their smaller sizes (100 bits), but they may

accumulate drain under high signaling frequencies. Nodes are deployed at a sensing radius of 20 m and capture physiological data periodically over a wireless link. To account for congestion effects, a maximum queue size of 50 packets is modeled, where prolonged queue retention increases idle listening and processing energy overhead. The two-channel capacities of 150 kbps and 250 kbps used in this study affected not only delay and throughput but also energy expenditure per successfully delivered bit. Cluster Heads (CHs) are modeled with slightly higher energy consumption due to their legitimate aggregation, processing, and transmitting data from associated nodes. The energy model provides a realistic basis for evaluating how queue-aware congestion control and redundancy elimination strategies influence the overall power budget in IoHT deployments.

3.3.3 Simulation Environment

The simulation environment was set up to represent a fully fledged IoHT network scenario within a 400 m × 400 m deployment area, encompassing both in-hospital and extended-range medical monitoring configurations. There are a total of 40 nodes randomly deployed, each with a transmission range of 300 m, ensuring multi-hop connectivity required for dynamic topology changes. Then, out of these 40 nodes, four nodes are designated as Cluster Heads (CHs), which are responsible for gathering data from their respective clusters. The simulations are run for 100 seconds each scenario, with 10 independent runs executed to obtain a statistically significant result. The packets of data were 500 bits in size and were sent during a 10-second data period, along with smaller control messages of 100 bits. The source traffic generated varies between 0 and 40 packets, while the network density ranges from 10 to 100 users, allowing for performance measurements at different load conditions. Two options for channel capacities, namely 150 kbps and 250 kbps, are used to capture the behavior of the network under low and moderate bandwidth constraints. The maximum queue length is set to 50 packets, representing realistic buffer limitations in constrained IoT devices. Time consumption patterns for 0-20 seconds are monitored to study latency, congestion build-up, and packet service rate. This simulation environment, therefore, enables an in-depth analysis of queue build-up detection, emergency packet treatment, and redundancy suppression strategies under realistic medical IoT conditions.

3.3.4 Performance Metrics

The proposed algorithms would be evaluated using performance metrics directly linked to the life-critical operational priorities of IoHT systems. Due to the varied focus of each algorithm, the metrics identified aim to capture their primary functionality objectives while also allowing comparability across scenarios.

Queue-Aware Congestion Avoidance (QACA)

The Queue-Aware Congestion Avoidance (QACA) algorithm is geared towards detecting and responding to the real-time queue occupancy characteristics to improve the service quality of IoHT networks. The evaluation domain consists of three important performance metrics that directly measure the success of the algorithm. Buffer loss probability: This defines the probability at which packets would be discarded due to buffer overflow. Let N_b denote the number of packets dropped due to buffer overflow at intermediate nodes, and N_s denote the total number of generated packets. The buffer loss probability is defined as:

$$P_{\text{buffer}} = \frac{N_b}{N_s} \tag{3.1}$$

. This basically indicates how QACA maintains queue occupancy subject to varying traffic load changes. The lower the buffer loss probability, the more successful it is in congestion avoidance and in utilizing the optimal network. Packet Loss per Second is the packet loss rate over time; that is, it can be used for detecting transient spikes of congestion and prolonged loss conditions. Let N_s denote the total number of packets generated by the source nodes, and N_r denote the number of packets successfully received at the destination. The packet loss probability is defined as:

$$P_{\text{loss}} = \frac{N_s - N_r}{N_s} \tag{3.2}$$

. This measure assumes great value in the real-time domain of healthcare applications, as even short bursts of packet loss would be intolerable for patient monitoring. Packet Delay is the time taken by the packet source and sink to traverse, i.e., queuing, transmission, and propagation delays. Let t_i^{gen} denote the generation time of packet i and t_i^{rec} denote the time at which the same packet is successfully received at the sink node. The end-to-end delay for packet i is defined as:

$$D_i = t_i^{rec} - t_i^{gen} \tag{3.3}$$

The average packet delay over N_r successfully received packets is computed as:

$$\overline{D} = \frac{1}{N_r} \sum_{i=1}^{N_r} \left(t_i^{rec} - t_i^{gen} \right) \tag{3.4}$$

.This parameter assumes more importance in life-critical scenarios of IoHT when there could be a delay in the delivery of physiological data or alerts, leading to a missed timely diagnosis or intervention. Altogether, these three measures would provide an exhaustive evaluation of QACA in ensuring low latency in the reliable transport of data while preventing high-load congestion in healthcare networks.

Dual Queue Aware Congestion Avoidance Scheme for Emergent Packet (D-QACA)

The Dual Queue-Aware Congestion Avoidance algorithm is designed to prioritize life-critical packets during network congestion, ensuring they reach an urgent level to facilitate the timely delivery of emergency data in IoHT scenarios. Buffer Loss Probability, the first performance metric, measures the amount of lost packets due to buffer overflows. It, therefore, reflects how effectively D-QACA stabilizes queues while allocating resources for high-priority traffic. The second metric, High Priority Packet Loss, is a crucial consideration for healthcare applications, as it indicates the number of emergency packets that did not reach their destination, thereby demonstrating the algorithmis ability to safeguard vital physiological signals, alerts, and emergency messages. End-to-end delay is the third in the order of exterior measurements. This represents the total time that a packet, especially the high-priority ones, takes to make its way from the source node to the sink, and it reflects how responsive the system behaves under load. Reducing this delay ensures that urgent medical data reaches medical professionals in a timely manner to support immediate decision-making and intervention. The combination of these three parameters forms an evaluation framework that ensures D-QACA is congestion-avoidant while maintaining clinical reliability through the timely and loss-free delivery of life-critical data.

De-Duplication and Queue-usage Aware Congestion Avoidance Scheme

Redundant packet removal in IoHT networks should be addressed to prevent congestion and performance degradation within the network. Buffer loss probability serves as an important factor in this context, as it is the metric used to quantify the effectiveness of queue-space management in a system after redundant packets are filtered out. High-priority packet-loss served as the second metric that is again vital for the healthcare area since it can define performance measures

based on the ability of the algorithm to protect urgent data during high traffic loads. Finally, collection was end-to-end delay—using this metric we can analyze the time packets take to reach from source to sink and expect overall improvement since it cuts down on redundancies and processing delays, thus not compromising the timeliness.

In addition to these metrics, communication cost is considered as the fourth metric. Communication cost evaluates the total cost of data transmission across the network, factoring in the resource consumption associated with forwarding packets through intermediate nodes. It is especially relevant in IoHT networks where minimizing communication overhead is crucial to optimizing performance, reducing latency, and ensuring efficient data flow in healthcare environments. While energy consumption is critical for IoT devices, including end nodes, we prioritize communication cost for intermediate nodes due to their role in routing and forwarding data. Intermediate nodes handle large amounts of data and are responsible for maintaining efficient flow across the network. By focusing on communication cost, we ensure efficient bandwidth utilization, minimize congestion, and maintain the quality of service for high-priority healthcare data. Energy consumption is still an important consideration, but optimizing communication cost is more directly related to maintaining the network's overall performance and scalability, particularly in high-traffic conditions where energy consumption can be indirectly influenced by communication-related overhead. Considering all these metrics, the evaluation can determine whether the algorithm achieves its dual purpose: relieving network congestion while ensuring the simultaneous flow of high-priority healthcare data, which is essential and non-redundant. This benefit is especially important for IoHT applications, in which patient safety and the quality of medical intervention directly depend on the timing and integrity of data.

3.3.5 Assumptions and Constraints

Assumptions have been specified and strictly enforced in the simulation framework to evaluate the efficacy of QACA, D-QACA, and DD-QACA algorithms, while also providing a uniform basis for experimentation and replicability. The nodes are assumed to be fixed in the area of 400 x 400 m² with a uniform random distribution for the IoHT environment, while maintaining a fixed transmission range of 300 m, which can be assumed to be constant. This ensures the required stable connectivity within the simulation environment. The channel is assumed to be ideal, without interference, fading, and noise beyond the modeled channel capacity limits of

150 kbps and 250 kbps at the physical layer. All nodes have equal energy reserves and are equipped with equivalent hardware configurations, which will render their performance directly comparable across experiments.

The traffic model has been implemented such that every node periodically generates data with predefined sizes of 500 bits and control messages of 100 bits. Each node has a buffer with a maximum capacity of 50 packets; any packets that arrive in excess of this limit will be deleted and will consequently increase the probability of buffer loss. To ensure that the algorithms can handle packets consistently based on priority, it is expected that all emergency packets are pre-classified at the source node, free from mislabeling errors. Furthermore, redundant packet removal is carried out to guarantee the precise identification of duplicates, enabling a theoretical assessment of the amount of congestion that could have been prevented in real-world packet fingerprinting had false positives and rescues not been taken into account.

Nevertheless, under these idealized circumstances, the simulation methodology naturally introduces its own limits. For example, the dynamic patterns of mobility that are present in real-world IoHT scenarios, such as those involving patients or mobile medical equipment, are not accurately depicted by this fixed form of deployment. Multipath fading, interference from integrated modern wireless systems in the same area, and environmental barriers are not included in the idealized channel model. The existence of these components may significantly impact the performance attained in real time. Assumptions for priority categorization and redundancy detection remove some operational variables that may impair real-world performance and understate the algorithm's efficacy. Typically, battery aging and hardware-induced variability are not considered when modeling energy depletion effects at the node level. While bursty or highly correlated event-driven data, which are frequently encountered in emergency medical circumstances, are not explicitly simulated, the network is assessed under conditions of continuous traffic.

These assumptions provide a fairly just and equal setting for comparing various methods, while also enabling the interpretation of simulation results within those parameters. Channel properties, moving objects, and internal environmental interference all have an impact on open settings. As a result, those are suggested for approval in subsequent experiments.

3.4 Summary

These sections elucidate the research methodology employed to address the various challenges encountered in the literature review concerning the Internet of Healthcare Things (IoHT) and the life-critical applications. The methodology was grounded in transparency, reproducibility, and scientific rigor, aligning exactly with the research gap mentioned above. An operational framework was constructed to address four cardinal network performance issues: queue awareness, emergency packet classification, redundant packet identification, and delay minimization. Each objective is crystal clear from gap analysis and leads to designing algorithmic solutions, namely Queue-Aware Congestion Avoidance (QACA), Dynamic Queue-Aware Congestion Avoidance (D-QACA), and the Redundant Packet Removal scheme. Still, practically everything else in research design remained subjective. The simulation framework has painstakingly been built to reflect a realistic IoHT environment, with the basics of the channel model reflecting capacities of 150 kbps and 250 kbps, a node energy model reflective of constrained healthcare devices, and a deployment environment corresponding to operational parameters such as transmission range, sensing radius, and node density. Performance measures like buffer loss probability, packet loss ratio, loss of high-priority packets, and end-to-end delay have emerged as a result of our suggested approach. The assumptions and constraints that underpinned the simulations were also described in this chapter, which improved comprehension of the simulations' scope and relevance. The methodology provides a solid framework for assessing and verifying the suggested approaches, which serves as the basis for the findings and analysis presented in Chapter 4.

CHAPTER 4

QUEUE-AWARE CONGESTION AVOIDANCE SCHEME (QACA)

4.1 Introduction

The rapid evolution of the Internet of Healthcare Things (IoHT) has revolutionized the acquisition, transmission, and real-time processing of medical data. The IoHT enables continuous patient monitoring through the use of interconnected wearable sensors, implantable devices, and stationary medical instruments, which in turn reduces hospital visits, fosters the advancement of preventive care, and catalyzes timely interventions. Nevertheless, it is this metamorphosis that presents a unique set of challenges amidst the management of numerous heterogeneous data streams. Healthcare ecosystems differ from normal IoT systems, where the loss of time or a lost packet might only lead to inconvenience; here, the loss of time or a lost packet could have dire and life-threatening effects. For instance, delays in transmitting ECG anomalies or failing to activate alarms in the ICU could jeopardize patient safety and treatment outcomes [120]. Historically, conventional patient-monitoring systems were unable to track patients in real-time and provide instantaneous assistance. Information was generally gathered periodically, temporarily stored, and examined by the medical staff with minimal delay. The emergence of IoHT has narrowed the gap, making real-time sensing and data transmission possible at any point. Therefore, the healthcare sector has shifted from a reactive model to a proactive and data-driven one, which is particularly relevant to the management of chronic diseases, post-surgery care, and intensive care unit (ICU) monitoring [121].

However, the network topology used in IoHT deployments is heterogeneous, While heterogeneity allows for diversity in devices, it also introduces interoperability and performance issues. However, the standard transport-layer protocols, such as TCP and UDP, were designed to handle traffic in classic Internet environments and would certainly have performed inefficiently in resource-constrained, delay-sensitive environments like those in IoHT's applications. The specific problems caused by the use of such protocols in these scenarios include network congestion, packet loss, lower throughput, and increased energy usage.

All existing congestion control methods-including Weighted Fair Queuing (WFQ), Class-Based Weighted Fair Queuing (CBWFQ), Minimal Latency Queuing (LLQ), Priority Queuing (PQ), Weighted Round Robin (WRR), and Strict Priority (SP)-have been somewhat adapted to the IoHT context but with varying degrees of success. While these strategies enable a more intelligent and prioritized allocation of bandwidth across certain classes of traffic, in some instances, they also guarantee minimum latencies for high-priority data flows. For example, LLQ provides strict Priority queuing for real-time traffic, while WFQ is used for non-critical data, thereby protecting urgent medical transmissions. Similarly, PQ guarantees absolute prioritization on critical packets, while WFQ distributes bandwidth fairly across traffic classes [92].

The repercussions of congestion within healthcare networks are far more severe than those encountered in conventional IoT methods. A drop in a packet on a streaming video application may lead to minor visual artifacts; on the other hand, a lost or delayed packet in an ECG signal alert for cardiac arrest could make the difference between timely intervention and a horrible outcome. During emergencies, massive bursts of high-priority traffic may also be generated by IoHT nodes (e.g., alarm signals, continuous waveform monitoring), placing immense pressure on network buffers and transmission channels. The sending node may continue to transmit without considering the neighboring nodes' queue status at a rate that exacerbates congestion , resulting in unacceptable communication delay and possible loss of life-critical data.

Nonetheless, these mechanisms often fail on two critical aspects:

1. High explosion of medical traffic, especially during emergencies or when multiple patients experience abnormal conditions and send and receive data simultaneously.

2. Queue unawareness: In many existing IoHT systems, sender nodes transmit data without being aware of the queue status of the next-hop or gateway node, which leads to congestion.

3. Resource limitations of WBAN and IoHT gateways, where small buffer sizes and limited

processing capability lead to rapid queue build-up.

4. Simultaneous operation of critical and non-critical traffic, without adequate prioritization mechanisms.

The motivation behind the Queue-Aware Congestion Avoidance (QACA) is to resolve the shortcomings found in existing congestion management schemes applicable to large-scale IoHT deployments. In real-life healthcare scenarios, especially during emergencies, IoHT networks are designed to handle sudden bursts of data packets from sensors and devices. However, the absence of suitable early congestion detection schemes and the failure to share queue state information among nodes often lead to inefficient traffic handling, excessive packet loss, and untimely delivery of life-critical data. This issue not only degrades the system's performance but also jeopardizes patient safety, as delayed or lost packets may carry essential physiological information necessary for timely diagnosis.

Prioritization techniques based on classical queues would normally work well in situations of controlled traffic loads; however, they would uncharacteristically fall short in towering bursts under highly dynamic and unpredictable patterns, such as those of emergency healthcare events. Priority inversion is often observed in such algorithms, where non-critical flows consume network resources, thereby compromising the integrity of high-priority medical data. Such mishandling results in sudden bursts of delays, increased attempts at retransmission, and so forth, ultimately leading to increased energy wastage—a very grave consideration for battery-operated IoHT nodes. Exacerbating the above problem is the lack of visibility, which adds to the ambiguity imposed on sending nodes that are unaware of congestion at later nodes, thereby aggravating overflow conditions at the buffer level and propagating congestion upstream.

To address these shortcomings, the proposed QACA mechanism employs a residual queue-aware strategy, which enhances congestion control at its fundamental level in IoHT networks. In this sense, QACA sets out to achieve two goals: firstly, it provides early detection of congestion before significant packet loss occurs so that the network can realistically resort to prevention strategies rather than react once performance has been severely impaired; secondly, an awareness of the queue states in neighbor nodes is added so that the sender nodes can dynamically alter their transmission rates depending on the actual buffer occupancy downstream and thus avert aggravating downstream congestion. These features contribute directly to resolving the fundamental issues of optimizing throughput, minimizing delay, and maximizing energy efficiency, while also ensuring strict quality of service compliance, all of which cannot be compromised in

assuring the reliability of IoHT systems in clinical practice. By so doing, QACA strengthens the communication infrastructure for healthcare while also providing a scalable framework that can be tailored for other mission-critical IoT settings. The remaining sections of the chapter provide an in-depth examination of the problem space driving the QACA development, an exhaustive description of the system model and algorithm design, and a complete evaluation of its performance through extensive simulation studies, providing comparison benchmarks with numerous leading state-of-the-art techniques.

4.2 System Model and Problem Statement

The proposed system model focuses on a smart health monitoring environment designed to tackle the challenges of network congestion in IoT-based healthcare systems. The primary objective is to develop a priority-based queuing scheme that efficiently manages data transmissions during emergency scenarios, where large volumes of health-related messages must be transmitted from sensor nodes to the gateway in real-time. In such situations, disseminating critical information in a timely manner is crucial because any delay or packet loss will have a direct impact on patients' health. The system model consists of four distinct types of nodes. The leaf node consists of sensor nodes that are responsible for acquiring real-time patient health parameters from sensing units, such as wearables like smartwatches, fitness bands, or health patches, or through untethered equipment. They take direct "samples" from human bodies, including vital signs, physiological measurements, and any other sensor data.

The second type of these nodes is the intermediate node, which may be either a medical device with an additional computing resource or a generic smart gateway that also possesses increased network resources compared to the leaf node. These nodes receive data from leaf nodes, process it based on the predefined priority level, and consequently forward it toward local sink nodes for further processing. Priority assignments in this stage at the affected priority levels also contribute to packet processing and transmission, ensuring that vital information concerning patient intervention reaches the most critical data; hence, the patient is prioritized over all other criteria. Local sink nodes, the third type, are typically represented by hospital servers or edge computing units and assume the preprocessing responsibility of analyzing packet contents to determine the type of medical assistance the patient requires. For example, data suggest that

urgent cardiological interventions may indicate a need for a neurological examination or regular monitoring. The root node, i.e., the cloud infrastructure linked to the hospital management system. It is the cloud infrastructure that serves as a central repository for data storage, advanced analytics, and decision-making processes, providing authorized medical staff with essential access to patient information for diagnosis, treatment planning, and monitoring purposes. The proposed system model is shown in 4.1

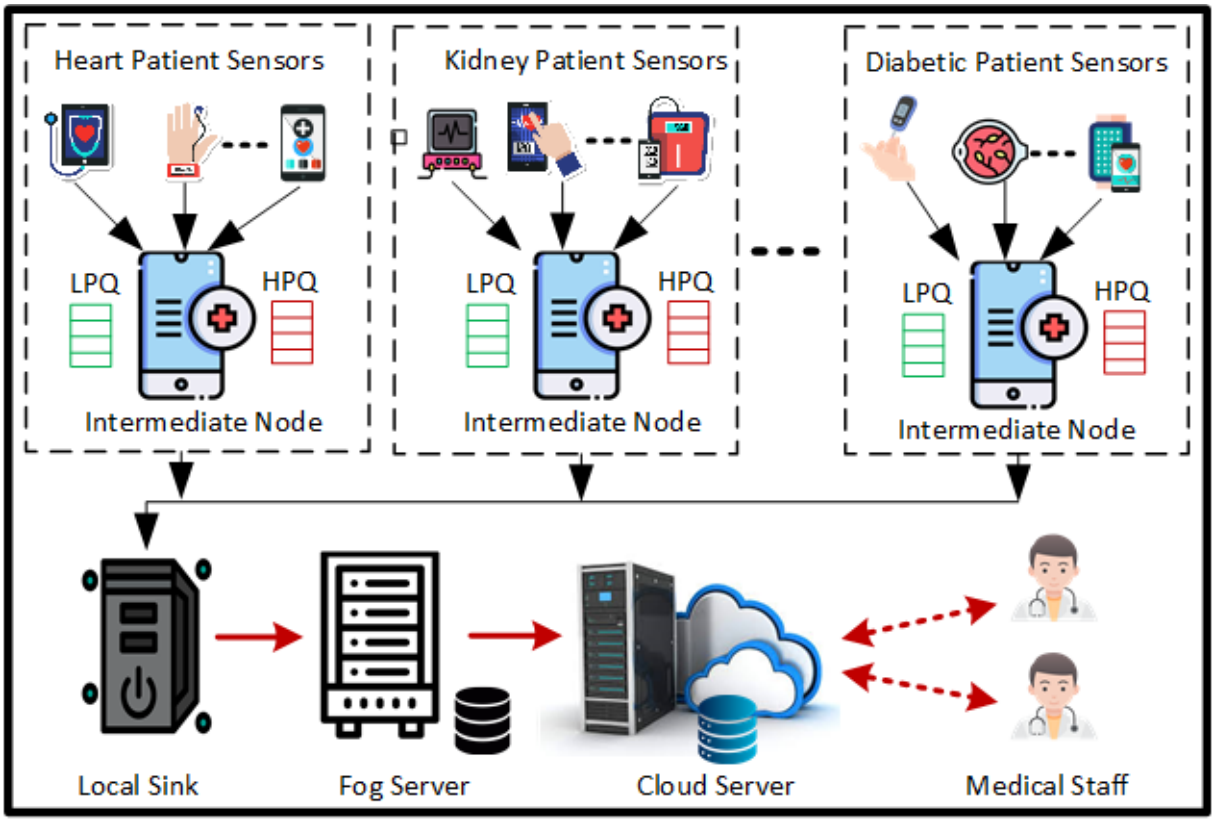


Figure 4.1: System Model

In the presented patient-centric Internet of Healthcare Things (IoHT) architecture, the most convenient devices, such as smartphones, smartwatches, and wearable health sensors, have operated as leaf nodes. That said, aforementioned devices collect data and send it to the intermediate node, which then forwards it to a local sink node. Later, after preprocessing, the refined information proceeds through the network to the cloud (i.e., the root node), allowing for central access and integration into hospital systems. At the intermediate node, two classes of queues are formed for packet organization: high-priority and low-priority. The former is dedicated to the immediate inflow of packets conveying data related to the critical existence of various operations, such as measurements by cardiac stents and brain function, and so forth, without any interruption until professionals at the server receive it. Conversely, low-priority

queues handle data transmission from body nodes that are less critically monitored, like general body temperature readings or daily activities.

However, an extreme challenge arises during an emergency. The consequence of "databurst phenomena" is the sudden creation of a large number of packets, causing delivery congestion within networks due to the limited capacity of the carriers in the entire system. These packets are generated at a very high speed, thereby congesting the entire network. The sender nodes require acknowledgment from the neighbor nodes, which contain the status of their queues. Without acknowledging the receiver, the sender node remains unaware of the receiving node's queue capacity, which can result in packet losses and buffer overflows. This lack of queue awareness leads to an increase in the packet loss ratio, sidetracked transmission delays, and aggravated network congestion, which in turn sacrifices the timely delivery of life-critical data. Therefore, the lack of effective acknowledgment of a queue status under emergency conditions serves as an impediment to IoT-based health monitoring systems. Consequently, resolving this issue will lead to building in valuable system reliability, which eventually ensures the prompt delivery of life-critical information, thereby maintaining the overall efficiency of patient monitoring systems undisturbed; this is particularly important in these critical medical circumstances where seconds count.

4.2.1 Queue Model and Notation

To analytically describe congestion behavior at an intermediate node, we model the node as a single-server queue with a finite buffer. Let λ denote the average packet arrival rate (packets/s) to the node, μ denote the average service (transmission) rate (packets/s), and B denote the buffer capacity measured in packets. Let $Q(t)$ represent the instantaneous queue occupancy (number of packets in the buffer, excluding the packet in service, if any) at time t, where

$$0 \leq Q(t) \leq B. \tag{4.1}$$

Queue Dynamics

Let $A(t)$ and $S(t)$ denote the cumulative number of packet arrivals and successful departures (completions of service) up to time t, respectively. The queue evolution can be expressed as

$$Q(t) = \min\{B,\ \max[0,\ Q(0) + A(t) - S(t)]\}. \tag{4.2}$$

Under stationary conditions, the offered load (utilization) is defined as

$$\rho = \frac{\lambda}{\mu}. \tag{4.3}$$

Congestion is expected when $\rho \geq 1$, meaning that the arrival rate approaches or exceeds the node service capacity, leading to persistent queue build-up and possible buffer overflow.

Buffer Overflow and Drop Condition

An arriving packet at time t is dropped due to buffer overflow if

$$Q(t) = B. \tag{4.4}$$

Equivalently, the number of packets dropped due to buffer limitation depends on how frequently the queue occupancy hits its upper bound B during high-load conditions.

Implication for Congestion Avoidance

The proposed queue-aware mechanism leverages the queue occupancy information $Q(t)$ (or equivalently the residual queue space $B - Q(t)$) to regulate the sender behavior and reduce the probability of operating in the congested regime $\rho \geq 1$.

4.3 QUERYING QUEUE AWARE CONGESTION AVOIDANCE SCHEME

This section deals with the description of the proposed "Queue Aware Congestion Avoidance (QACA)" Scheme. In the patient-centric IoHT congestion control (CC) mechanism, incoming data packets are put into high- and low-level queues as shown in fig. 4.2.

High-priority data packets contain information derived from sensors mounted on critical parts, such as the heart, brain, and others, depending on the patient's disease; low-priority data packets are those from other less critical sensors, which may include those from the feet, hands, legs, etc. The selection scheme uses a priority queue-based scheduler to send data to its respective queue according to its priority. In the event of an emergency, due to the high transmission of packets, the sender node must have the queue status at the receiving node in order to avoid congestion. To solve this problem, the proposed mechanism employs a controlled acknowledgement-based mechanism to share the queue status with the sender nodes. Initially, the intermediate node

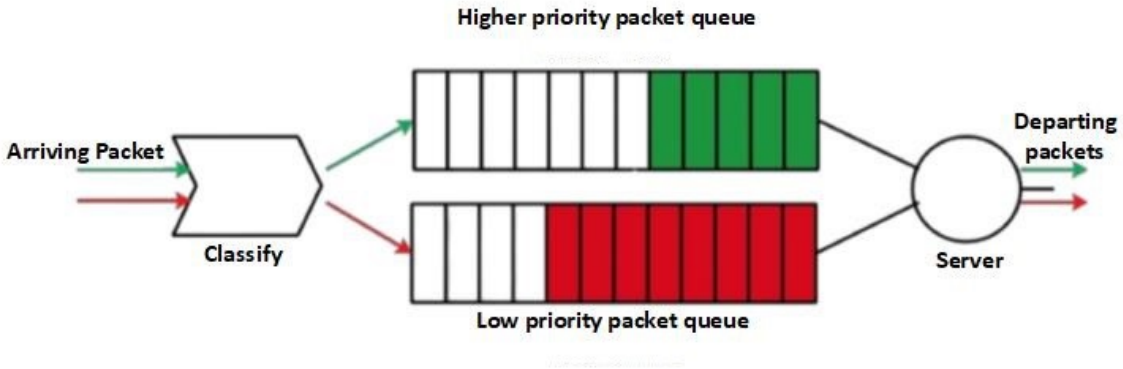


Figure 4.2: Queue Maintained at Intermediate Node

will not send the queue status until it is filled to a defined percentage of more than 50%. The mechanism prescribes three possible rates for sending the acknowledgment packet upon filling the queue at 50, 70, and 90 percent full: after every 10, 5, or 1 packet, respectively. It can enable timely detection of congestion and prevent potential congestion by adjusting the rates of data transmission. The list of notations for this scheme is given in Table 4.1

Table 4.1: List of Notation for QACA

Sr.No	**Notation**	**Description**
1	INode	Intermediate Node
2	P_Interval	Ack count to send Acknowledgement
3	Ack_Ctr	Ack counter as per interval
4	LPQueue	Low Priority Queue
5	HPQueue	High Priority Queue
6	H_Queue_Status	High Priority Queue Status
7	L_Queue_Status	Low Priority Queue Status
8	RQueueSize	Residual Queue Size

Algorithm 1 presents QACA for interval-based acknowledgment processing. At this point, to determine when acknowledgment packets are sent, the system sets the three thresholds for processing queue size (50, 70, 90). According to these criteria, it establishes intervals as (10, 5, 1, 0) for transmitting acknowledgments and sets the counter, Ack_Ctr, to 0 to keep track of the number of acknowledgment packets handled.

In steps (1-10), the Send(Packet Pkt) function processes packets based on their type. If an acknowledgment (Ack) packet is received, it increments the counter Ack_Ctr. The algorithm subsequently assesses the intermediate queue size against established thresholds (50, 70, 90). The common criterion for issuing an acknowledgment then depends on the threshold achieved (i.e., after a certain number of packets (10, 5, or 1) are counted). If none of the thresholds are reached, no acknowledgment will be given. In steps 11 - 16, the decision to send an acknowledgment depends on the intermediate node's LPQ size. It will not send an acknowledgment if none of the size provisions are fulfilled by just using Send_ACK(False, P_Interval). Otherwise, this is not an acknowledgment packet and will be sent immediately to the local sink via Send_Pkt(P). This way, the acknowledgment frequency is adjusted to match the processing load at the node, thereby optimizing network traffic. In steps 17-28, the Send_ACK(Ack_Flag, P_Interval) function determines when to send or mute acknowledgment packets based on the network state. If Ack_Flag is True, the application checks if the value for Ack_Ctr exceeds P_Interval; if so, it

sends the packet to the sensors and resets Ack_Ctr. If not, the acknowledgment is kept until the interval count is exhaustive. If Ack_Flag is False, the acknowledgment is not given until the LPQ size drops below the predetermined threshold, meaning they are only sent when considered necessary for the overall efficiency of the network.

After receiving an ACK from the INode, sensor nodes will adjust the rate at which they send data based on the value of LPQ_Size obtained in the packet from the INodes. The collector node generates one message for all low-priority nodes by aggregating the data values as M1 TempSn1: 98.6, HrtBeatSn2: 70, OxiSn3: 68. In the same message, "," is a delimiter that differentiates the values coming from different sensors. Just like that, another message collects the sensor readings for the high-priority nodes. The one-bit flag indicates whether the sensor is of low or high priority, and then the packets are enqueued to LPQ or HPQ at the INode.

4.4 Complexity and Convergence Analysis

4.4.1 Complexity Analysis

The computational complexity of the proposed X-QACA algorithm is largely determined by its handling of incoming packets and its decision-making process regarding acknowledgment transmission. For each packet, the algorithm executes a constant number of operations, which include checking the packet type, comparing the queue size against predefined thresholds, updating counters, and deciding whether to send or discard an acknowledgment. Each of these steps is performed in constant time, or $\mathcal{O}(1)$, for each packet. Consequently, if n packets are arriving at the intermediate node, the overall time complexity becomes $\mathcal{O}(n)$, as each packet is processed individually and sequentially. This linear complexity ensures that the algorithm scales effectively with the number of packets, making it suitable for real-time, patient-centric IoHT systems. The space complexity is $\mathcal{O}(B_i)$, where B_i represents the maximum buffer size of the LPQ at the intermediate node, which is independent of the number of packets processed.

Algorithm 1: Dynamic Acknowledgment Control at Intermediate Node

```
Input: Ack packet from patient sensors
Output: Decision to send or suppress ACK by intermediate nodes
Threshold.size ← {50, 70, 90}
P-Interval ← {10, 5, 1, 0}
Ack-Ctr ← 0
Function Send(Packet Pkt):
    if Pkt.Type == Ack then
        Increment Ack-Ctr by 1;
        if INode.LPQ-size ≥ Threshold.size[2] then
            Send-ACK(True, P-Interval[2]) ;                    // After 10 packets
        else
            if INode.LPQ-size ≥ Threshold.size[1] then
                Send-ACK(True, P-Interval[1]) ;                 // After 5 packets
            else
                if INode.LPQ-size ≥ Threshold.size[0] then
                    Send-ACK(True, P-Interval[0]) ;              // Every packet
                else
                    Send-ACK(False, P-Interval[3]) ;            // No ACK packet
                end
            end
        end
    end
    else
        Send-Pkt(Pkt) to Local-Sink;
    end
End Function
Function Send-ACK(Ack-Flag, P-Interval):
    if Ack-Flag is True then
        if Ack-Ctr > P-Interval then
            Send-Pkt(Pkt) to Sensors ;                  // contains LPQueue-size
            Ack-Ctr = 0;
        end
        else
            Discard the ACK until interval count is met;
        end
    end
    else
        Discard the ACK until LPQ size reaches threshold;
    end
End Function
```

4.4.2 Convergence Analysis

The X-QACA algorithm achieves convergence by using a controlled acknowledgment mechanism that adjusts dynamically based on the size of the LPQ. As congestion increases, the frequency of acknowledgment messages also rises, prompting sensor nodes to reduce their data transmission rates. This negative feedback loop ensures a swift response to congestion while avoiding overreactions when traffic levels are moderate. The algorithm utilizes queue thresholds

set at 50%, 70%, and 90% capacity as defined trigger points for sending ACKs. This approach facilitates a gradual and stable adjustment process. Over time, the queue size stabilizes within an acceptable operating range, typically between 50% and 70%. This demonstrates the algorithm's ability to converge toward a steady state without causing packet overflow or excessive delays. As a result, the QACA strategy is highly effective in maintaining network stability and reliability, especially in emergencies within patient-centered IoHT systems.

4.5 Results and Analysis

After performing a comparative analysis in this section, the results, including buffer loss probability, number of packets lost, and delay, are presented for the proposed versus the analytical buffer overflow and DCCA models. The parameters used for calling and simulation are encompassed in Table 4.2. To verify the performance of the proposed QACA, this study considers an intermediate domain that takes over the responsibility of sending data to the sink node. Furthermore, three backouts, indicated by the parameter 'm,' and retransmission, denoted by 'n,' are assumed to be the maximum in this study. The packet collision factor is considered to be 0.10, while the number of leaf nodes is varied from 2 to 10. In total, 10 packets are available for transmission, with a packet transmission rate of 32 packets/second. The study assumes a channel capacity of 250 kbps and 120 kbps compared to its base scheme, which includes the analytical buffer scheme [108] and DCCA [94] under analysis.

Table 4.2: Simulation Parameter for Proposed QACA

Sr.No	**Parameter**	**Value**
1	Simulation Time	120 s
2	Deployment Area	450 x 450 m
3	Transmission Range	280 m
4	Number of Nodes	48
5	Data Packet Size	600 bits
6	Number of cluster Head	6
7	Number of simulations	20
8	Density	0 - 60 Packets
8	Control Message Size	128 bits
8	Duration of Data Periods	10 s
8	Channel Capacity	120 kbps & 250 kbps
8	Sensing Radius	20 m
8	Maximum Packets in Queue	50

4.5.1 Buffer Loss Probability

The buffer loss probability is a key performance measurement in patient-centric IoHT applications, indicating the likelihood that an incoming packet will be dropped because there is no buffer space available at the point of reception. The ratio of packets lost due to buffer overflow to the total number of packets that arrived at a node is mathematically expressed in 4.5

$$P_{\text{loss}} = \frac{N_{\text{lost}}}{N_{\text{arrived}}} \tag{4.5}$$

where N_{lost} is the number of packets dropped due to overflow in the buffer and $N_{arrived}$ is the total number of packets that arrived at the node.

Thus, buffer loss implies a potential risk for IoHT applications that are transmitted in a timely and precise manner during medical data transmission. Buffer loss probability takes into account several factors, including packet arrival rate, service rate, buffer size, network congestion,

and queuing strategy. High buffer loss may cause communication delays and, in severe cases, lead to poor data reliability, which can compromise patient safety, making it a key indicator for evaluating the performance of congestion control mechanisms during stimulus time, such as the proposed QACA scheme. QACA-50 shows that it is 50% full, QACA-70 signifies it is actually 70% full, and QACA-90 stands for a buffer level of 90%. Figure 4.3 illustrates buffer loss probability under increasing traffic load. As the load increases beyond the system's service capacity, QACA-70 shows degradation due to higher queue occupancy. However, compared to baseline schemes, QACA maintains significantly lower loss by adjusting transmission behavior through queue-aware acknowledgments. In Figure 4.3 (a), load increases buffer losses that are experienced greatly in the analytical model and by DCCA. In contrast, QACA exhibits comparatively less buffer loss, indicating that it performs better. For example, 12 patients per patient with 250 kbps at the leaf nodes improve the buffer loss probabilities by 64% for the analytical model and by 75% for DCCA.

Intermediate nodes with respect to 250 kbps will experience greater buffer loss than leaf nodes due to increased traffic. Nevertheless, QACA achieved a 67.92% enhancement over the analytical model and a 76.4% better performance over DCCA.

It is depicted that leaf nodes operating at 120 kbps show that $QACA_{50}$ achieves a 70.6% better performance compared to the analytical model, while also yielding a 75.4% improvement over DCCA. At the same time, a 57.14% gain was achieved with $QACA_{90}$ while the queue is 90% full, and a 46.43% gain with $QACA_{70}$. Likewise, Figure 4(d) presents $QACA_{50}$, which shows a 50% improvement over the analytical model at that level, a 62.7% betterment over DCCA, a 28.57% enhancement over $QACA_{90}$, and a 13.8% improvement over $QACA_{70}$. These findings demonstrate that the proposed QACA scheme achieves aachieves a significant reduction in the probability of buffer loss and exhibits strong scalability performance as the number of nodes increases with respect to the number of patients.

4.5.2 Packet Loss Per Second

It is a time during which packets may be lost in every second in the network due to problems such as congestion, buffer overflow, transmission errors, and unstable connections. The formula for packet loss is:

$$\text{PLPS} = \frac{N_{\text{lost}}}{T} \tag{4.6}$$

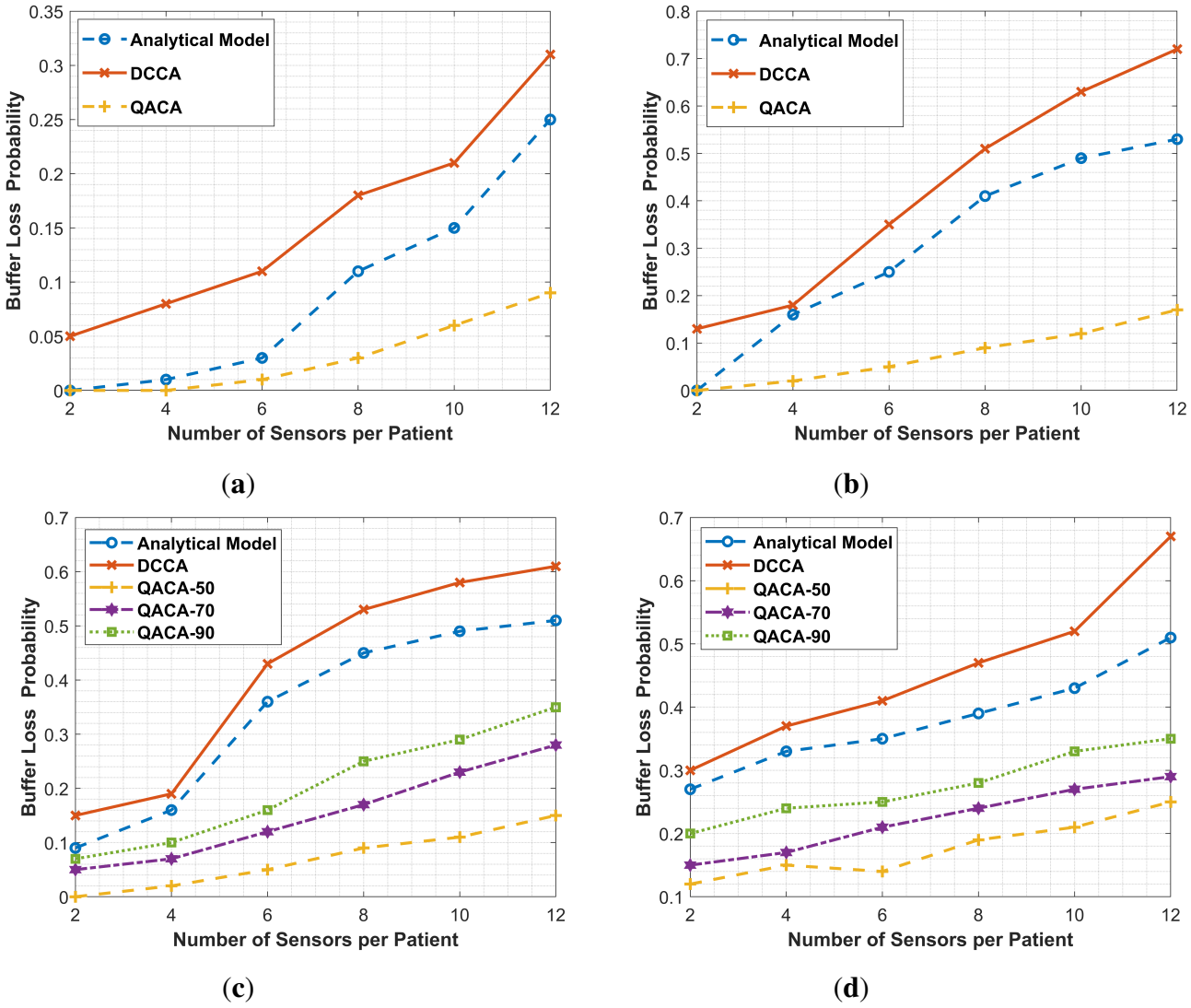


Figure 4.3: Buffer Loss Probability for (**a**) Leaf nodes at 250kbps; (**b**) Intermediate nodes at 250kbps; (**c**) Leaf nodes at 120kbps; (**d**) Intermediate nodes at 120kbps

Where N_lost is the number of packets lost during the time interval, and T is the duration of this interval in seconds.

n a patient-oriented IoHT system, monitoring cannot be separated from continuous availability anywhere, anytime, and packet loss has a direct impact on decisions regarding diagnosis and treatment. Figure 4.4 presents packet loss per second for leaf and intermediate nodes under different channel capacities. As traffic intensity increases, baseline schemes experience rapid loss escalation due to buffer overflow. In contrast, QACA limits packet loss by dynamically regulating the sending rate through queue-aware acknowledgment feedback. For example, Figure 2(a) illustrates PLPS, where each patient is attached to 12 sensors in the leaf node case and operates at a speed of 250 kbps. The results show the significant improvement level achieved by X-QACA, which increased by 73.33% compared to the analytical model and 81% compared to DCCA. In Figure 2(b), X-QACA again achieved 73.68% better performance than the analytical model for intermediate nodes with a 250 kbps speed and 78.26% improvement over DCCA. Figures 2(c) show PLPS at leaf nodes, where the channel capacity is 120 kbps. X-QACA outperforms the other two by achieving a PLPS improvement of 68.42% over the analytical model, while also improving by 73.91% over DCCA. Figure 2(d) shows PLPS for intermediate nodes operating at 120 kbps, where X-QACA achieves a 47.62% improvement over the analytical model and a 56.1% improvement over DCCA. DCCA is even worse in the case of lost packets, since it immediately sends the queue state back to the sender at all levels.

4.5.3 Packet Delay

Packet delay is the total time taken by one packet while traveling from a source to a destination across the network, and it is measured by:

$$\text{Packet delay} = D_{\text{proc}} + D_{\text{queue}} + D_{\text{trans}} + D_{\text{prop}} \tag{4.7}$$

The total packet delay consists of processing delay, queuing delay, transmission delay, and propagation delay. Figure 4.5 illustrates packet delay at intermediate nodes for channel capacities of 120 kbps and 250 kbps. Higher delays are observed at lower bandwidth due to increased queueing time. QACA consistently achieves lower delay by preventing excessive queue buildup through proactive congestion awareness. Figure 3(a) illustrates a scenario of delay under various operating conditions at an intermediate node operating at 120 kbps, with 12 sensor nodes allocated to each patient. The outcome of this assessment indicated that the QACA50 yielded a

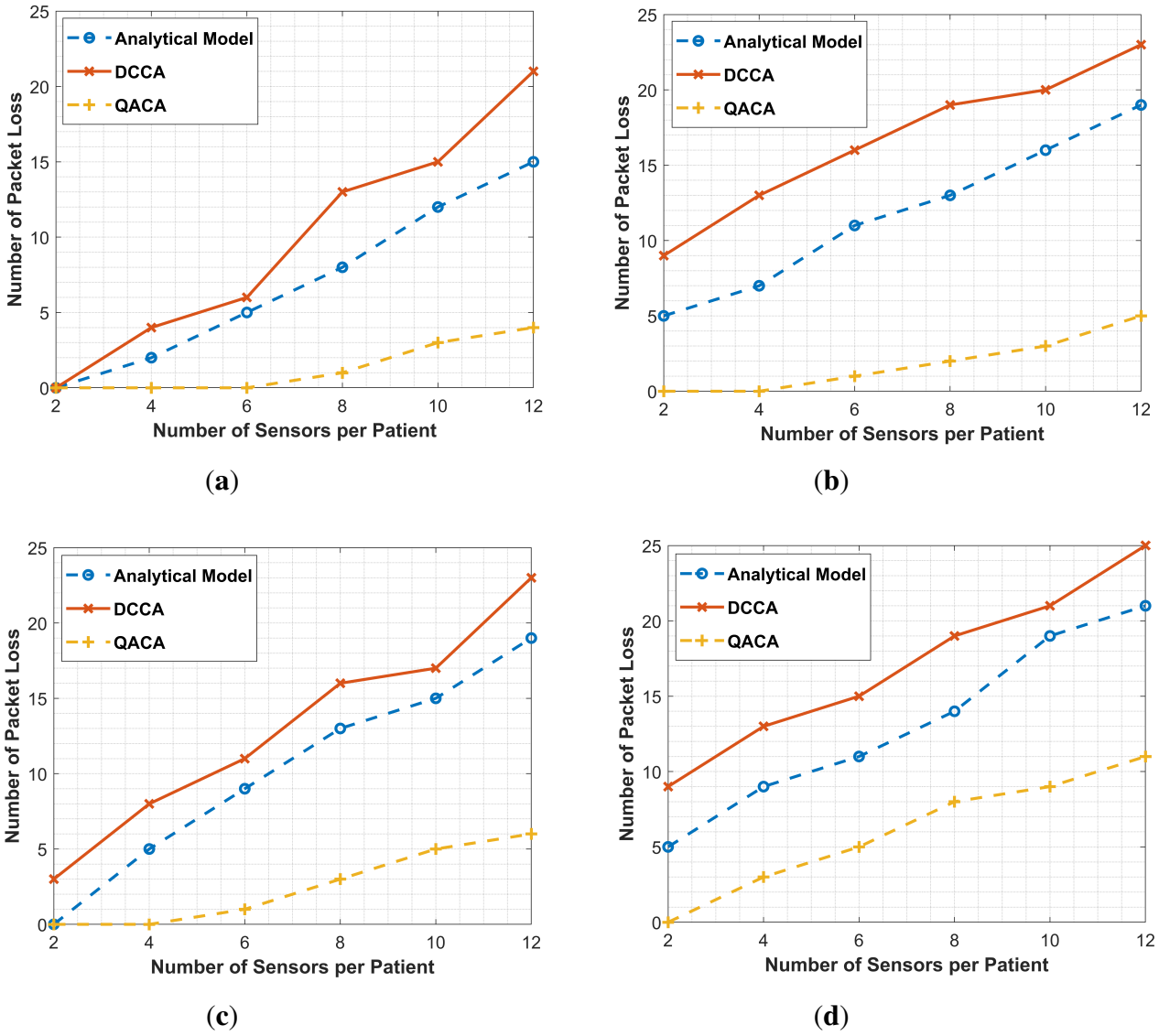


Figure 4.4: Packet loss for (**a**) Leaf nodes at 250kbps; (**b**) Intermediate nodes at 250kbps; (**c**) Leaf nodes at 120kbps; (**d**) Intermediate nodes at 120kbps

46.42% delay reduction compared to the analytical model, a 37.5% reduction against DCCA, a 28.57% reduction versus QACA90, and an 11.87% reduction in the case of QACA70. The data shown in Figure 3(b) refer to the delay for intermediate nodes under operational conditions of 250 kbps, with 12 sensor nodes assigned to each patient. The results, however, show that QACA50 achieves a 54% reduction in packet delay, whereas the benchmark for comparison with DCCA, as per the analytical model, is 48.5%. Further deductions are also noted at present with respect to packet delay and QACA90, which are 32% less and 26% lower, respectively, compared to QACA70. Intermediate nodes were significantly higher in both the analytical model and DCCA for packet delay compared to any other method adopted here. By contrast, QACA had lower delays, which was a result of controlled acknowledgment sent back to the sender node.

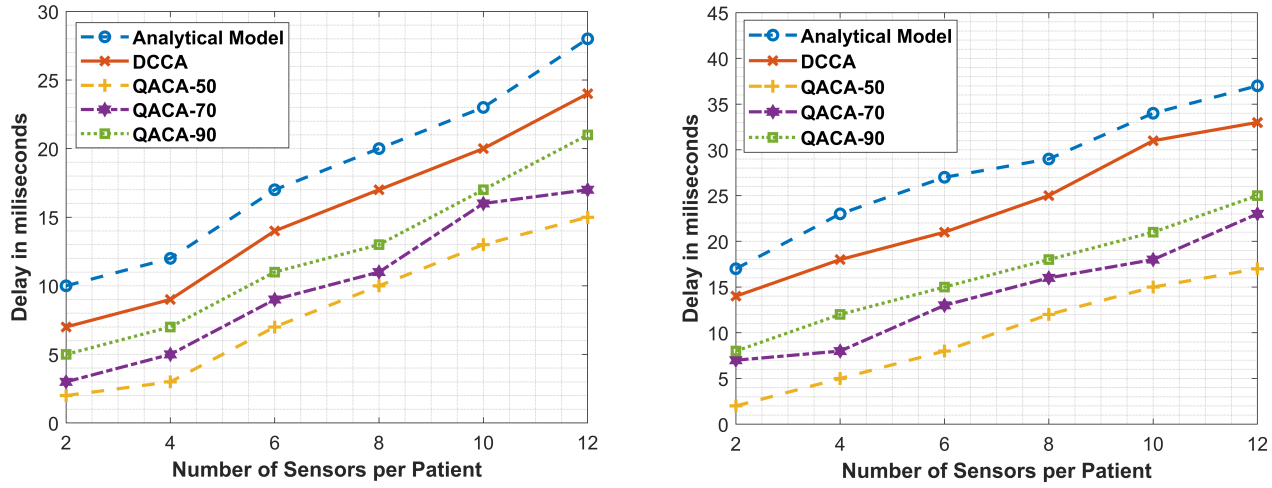


Figure 4.5: Packets delay: (**a**) Intermediate nodes at 120 kbps; (**b**) Intermediate nodes at 250 kbps.

The acknowledgment ratio, as observed in our delay analysis, directly affects it as this ratio increases; queue occupancy and packet delay also increase slowly. This suggests that control over acknowledgment ratios may be one area where some improvement in system performance could be realized.

4.6 Discussion & Implications

The importance of the QACA scheme lies in its ability to handle the critical congestion issues that remain unaddressed by the traditional IoHT framework. In heavy medical monitoring, one of the overseen causes of congestion is the high frequency of ACK messages sent back to the sender, particularly in scenarios of dense body sensor networks or emergencies. While acknowledgments are important from a reliability perspective, an uncontrolled frequency of ACKs wastes bandwidth. It exacerbates packet loss during emergency situations when a large number of messages need to be exchanged. The residual queue-aware nature of QACA addresses this problem by imparting knowledge of the neighboring node's queue state to each node, thus enabling the intelligent handling of data packets and acknowledgments. The proposed scheme ensures a smoother flow control, mitigating packet loss, buffer loss, and delay issues in high-load conditions, while avoiding redundant and excessive ACK traffic and facilitating early detection of queue build-up.

In broader terms, the QACA would also influence the development of the upcoming set of priority-aware IoHT protocols. While this work explicitly provides the basis for residual queue awareness and congestion avoidance, it does not yet incorporate a mechanism that discriminates between emergency-critical packets and ordinary monitoring data. Practically, healthcare traffic ranges from immediate life-saving alerts to routine readings with delays tolerable for clinical support. The present scheme addresses congestion by distinguishing between high-priority emergency queues and low-priority monitoring queues, which will be a limitation to be addressed in the next proposed scheme integrating QACA. This scheme will prioritize emergency packets for absolute precedence of life-critical alerts under congested conditions. These achievements will combine to create a scheme for an IoHT that alleviates congestion and handles the processing of packets according to the nature of the emergency in both high- and low-priority queues. The observed performance trends across the proposed schemes are closely aligned with theoretical expectations derived from queueing behavior and service rate constraints. As traffic load increases, buffer occupancy grows when the packet arrival rate approaches or exceeds the service rate of intermediate nodes, leading to congestion and packet loss in conventional schemes. In QACA, proactive queue-aware acknowledgment feedback enables early regulation of transmission rates, thereby limiting queue buildup and reducing loss probability

4.7 Summary

The QACA is advanced in terms of addressing congestion threats in IoHT networks in this chapter. The research described how poor management of congestion in a healthcare scenario leads to packet loss, buffer overflow, increased delays, and energy wastage, thereby threatening patient safety. One significant problem identified is the considerable number of ACKs generated, which can further exacerbate congestion under heavy traffic conditions and may also lead to performance degradation. QACA is designed to identify congestion early through an active awareness mechanism that dynamically adjusts packet sending based on the condition of the residual queues, and blocks redundant ACK traffic when the queues approach a critical mark. The knowledge of neighbor-queue allows QACA to assist nodes in adjusting their transmission rate in a dynamic manner, thereby preventing buffer overflow and avoiding retransmission

storms. Simulation parameters indicate that QACA offers a significant reduction in packet loss, low buffer occupancy, minimal end-to-end delay, and substantial energy savings compared to classical techniques. Hence, it guarantees a healthier and more reliable delivery of health data, particularly in life-critical telemetry applications, under diverse loading conditions. While QACA addresses congestion and ACK-induced overhead, it has not yet implemented emergency-aware packet prioritization, a limitation that will be addressed in the next stage of research, where emergency packets will be given absolute priority under congestion scenarios. To summarize, the fourth chapter laid the groundwork for further enhancement of IoHT networks, wherein QACA stands as a lightweight and practical scheme that integrates queue-awareness, acknowledgement regulation, and early congestion detection reliability prior to emergency-based prioritization.

CHAPTER 5

DUAL QUEUE-AWARE CONGESTION AVOIDANCE SCHEME FOR EMERGENT PACKET (D-QACA)

5.1 Overview

The Queue-Aware Congestion Avoidance (QACA) scheme, which addresses the underlying problem of early congestion detection in IoHT networks, is discussed in the preceding section. The reliability of networks has improved significantly through the QACA's residual queue awareness and early detection mechanisms, enabling healthcare applications to avoid packet loss, reduce delays, and maintain Quality of Service (quality of service). As discussed in the implications above, one major disadvantage of QACA is its ability to differentiate between emergency and non-emergency traffic. While QACA prioritized healthcare traffic in general, all high-priority data streams were treated uniformly, such that it did not effectively consider the urgency of emergency packets.

This chapter introduces an advancement—the Dual Queue Aware Congestion Avoidance Scheme (D-QACA) — built on the foundation of QACA, but differing in its dual-queue structure compared to single-queue schemes, as it explicitly separates emergency packets from normal priority healthcare packets. Not only does this allow the system to detect and control congestion proactively, but it also ensures that life-critical, emergency data is sent through with ultra-low latency, even during severe network loading. By deploying dual queue management, acknowledgement control, and enhanced scheduling strategies, D-QACA ensures that it alleviates congestion without compromising fairness or exhausting node resources.

The motivation, deeply rooted in matters regarding healthcare communication, would be

even closer to life-and-death considerations. In emergencies, mostly cardiac arrests, paralysis, or sudden respiratory failures, even a few milliseconds of delay can spell the difference between life and death in transferring sensor data. This is why, although queue-aware schemes like QACA still prove better in terms of reliability, they do not prioritize event-driven emergency signals over routine monitoring data. D-QACA, however, precisely compensates for that by ensuring that emergency alerts are accorded strict priority, while still ensuring efficient management of normal physiological data streams.

In summary, this chapter shifts from the broad foundation established by QACA to a more refined and context-sensitive approach applied to real-world healthcare needs in IoHT. The Dual Queue Awareness Congestion Avoidance Scheme would build on QACA, extending its capacity by adding the handling of emergency traffic priority and enhanced queue policy scheduling. These developments make a significant contribution to the development of more dependable, responsive, and energy-efficient IoHT networks, a hallmark of modern healthcare applications.

5.2 System Model and Problem Statement

The system model for this research aims to develop an intelligent health monitoring system that ensures the reliable transfer of health information, particularly under emergencies, where delays and packet loss can have severe impacts on patients. The objective of this model is to implement a priority-based queuing technique that facilitates early congestion detection and efficient packet management, ensuring the timely delivery of emergency data. Introducing congestion-aware mechanisms at several levels in the network will minimize unnecessary packet drops or delays while guaranteeing fairness between normal and emergency traffic flows.

Five different types of nodes are used into the model as shown in the figure 5.1 , which is hierarchically organized into five distinct levels for implementing functions in a specific layer. The nodes interact with each other to form a complete IoHT network, enabling the handling of both emergency and non-emergency medical traffic.

1. Sensing Devices (Leaf Nodes): The first level consists of leaf nodes that represent the sensing devices, which directly interface with the patient's body. Leaf nodes can be wearable devices (such as smartwatches, electronic patches, or fitness trackers) or non-wearable devices, such as implantable sensors or bedside monitoring equipment. Their

main job is to collect real-time physiological parameter data, such as ECG signals, oxygen saturation, blood pressure, body temperature, and activity levels. Smartwatches and smartphones are modern electronic wearables that have taken on this role, as patients widely use them. The leaf nodes serve as gateways for collecting health data and forwarding it to the collector nodes.

2. Collector Nodes: The next level device is the collector node, which may be a mobile device (such as a smartphone) or a dedicated computational device. Its job is to aggregate all incoming data packets from the different leaf nodes and make an initial distribution of these packets into high-priority (emergency) and low-priority (routine monitoring) categories. The collector node is also responsible for indicating the emergency level of the data based on predetermined thresholds. For example, if the heart rate exceeds a critical threshold or oxygen levels drop suddenly, the packet is marked as high-priority. In this manner, it is of great importance for the collector node to filter and classify traffic before this data propagates deeper into the network.

3. Intermediate Nodes: The intermediate nodes are devices that possess high computational power, along with extensive network resources. These nodes perform more extensive processing of data packets received from the collector nodes and evaluate the priority assigned to packets by the collector nodes. Reinforce that priority in subsequent classification, and forward packets to the sink nodes.

4. Fog Servers: Local Sink Nodes: Local sink nodes represent fog servers, which have very high computational and storage capacities—these act as a preprocessing stage for packets coming from intermediate nodes. The major role they play is analyzing data to determine the type of medical service needed. For example, an analysis may reveal crossing the threshold of multiple emergency readings, after which an intervention, even from healthcare professionals, needs to be decided upon. By tide-preprocessing and before forwarding such information to the cloud, sink nodes also minimize their overall transmission latency and bandwidth usage.

5. Root Node (Cloud Server): The root node at the highest level represents such cloud servers through which hospitals connect directly with their central administration system. The cloud node, for example, would be able to store vast amounts of medical data while supporting advanced analytics (such as AI-based diagnosis) and ensuring that the processed

information is available to medical staff in real-time. In this way, the final destination of all patient data is represented in the cloud, where healthcare professionals can access it to make timely decisions.

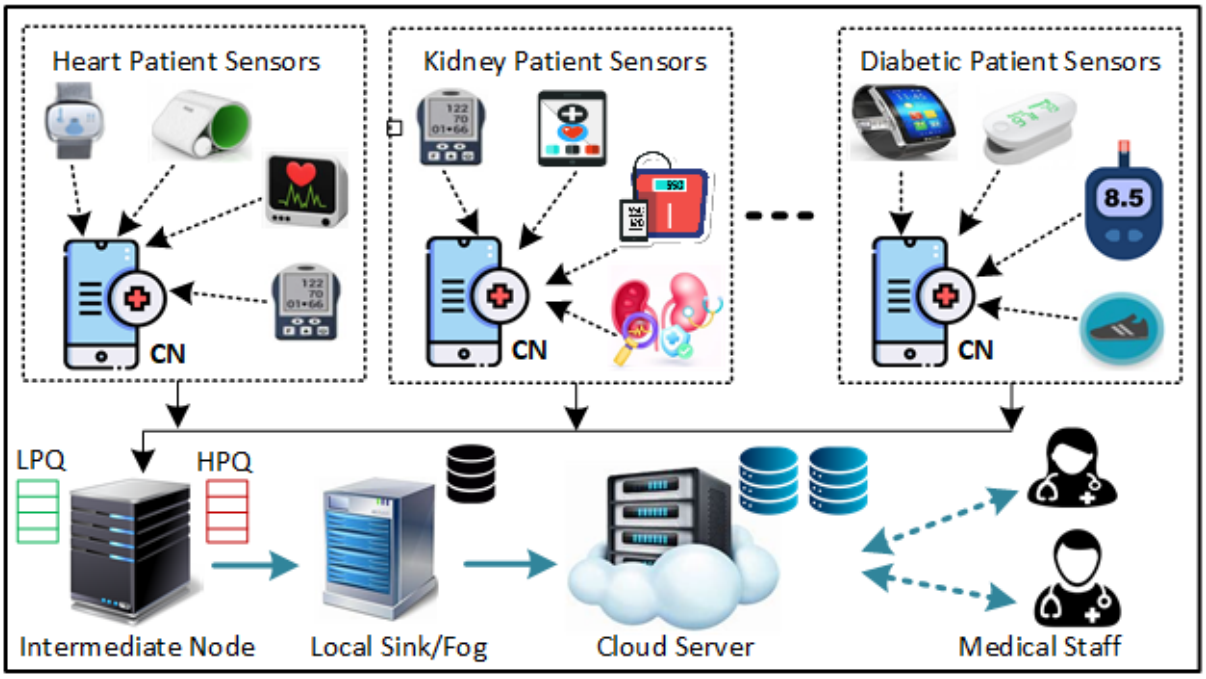


Figure 5.1: Architecture of the System Model

Most of the existing queue management schemes primarily process high-priority and low-priority packets using a First-In-First-Out (FIFO) mechanism. The FIFO approach is indeed fair in the sense that it respects the arrival order of packets, but it does not consider the emergencies that healthcare communication often embodies. Emergency packets might belong to the same priority class, but surely do not share the same. For example, a packet notifying of an emergency of sudden oxygen desaturation is far more urgent than a packet sending a routine periodic heart rate update; nevertheless, both would wait in sequence according to FIFO. Such rigid processing delays critical emergency packets unnecessarily, even when immediate transmission is necessary to avert life-threatening conditions.

The above limitation significantly increases the likelihood of dropped packets, delays, and congestion, especially during emergencies when the network is overwhelmed with numerous health-related messages. Emergency packets that are waiting for less urgent data are at risk of being dropped due to buffer overflow or excessive waiting, which can lead to severe reliability issues for IoHT systems. This loss of packets has not only technical ramifications but, in this case, can have direct clinical implications, since delayed or lost emergency alerts may impede

timely medical action and response. Hence, the absence of an emergency-aware mechanism in the queue setup creates a severe bottleneck, making it impossible to expect any quality of service in life-critical healthcare applications.

Inadvertently, such FIFO-style queuing undermines the performance and responsiveness of health-monitoring systems by treating all packets equally in their respective queues. This will result in congestion, increased retransmissions, delayed acknowledgments, and higher energy and bandwidth consumption. In an ever-constrained IoHT ecosystem, with regard to energy levels on devices and network resources, these effects will have a cascading effect, lowering overall performance. Thus, emergency-based differentiation in queue processing is a significant gap in all existing models, which require more intelligent queue-aware mechanisms to identify and prioritize emergency traffic while maintaining fairness and network stability.

5.3 DUAL QUEUE AWARE CONGESTION AVOIDANCE SCHEME FOR EMERGENCY PACKET

The proposed "Dual Queue Aware Congestion Avoidance (D-QACA) scheme" is explained in this section. The proposed system operates based on priority management of a buffer present at each node, utilizing two buffer queues. With these two buffer queues, each queue has its unique priority. They may be classified as either low-priority or high-priority. Upon the arrival of the input packet, the node sorts and forwards it to the appropriate queue. The packets in the high- and low-priority queues may not have the same level of emergency and may have different emergency levels. The normal queue is replaced with a priority queue in case of an emergency. In the priority queue, each packet is prioritized based on its urgency. The packet containing the emergent data is assigned the highest priority and processed first, if no other packet is being processed. If any packet is currently being handled, it will be placed second. This method would greatly enhance the management of queues in situations where crucial data requires immediate processing. The data packets are classified into high- and low-priority queues based on urgency to ensure that more important information is processed ahead of others. Fairness is achieved by requiring all packets with the same emergency level to be treated on a first-come, first-served (FCFS) or first-in, first-out (FIFO) basis; this prevents a situation that causes bottlenecks for crucial data in processing.

The method also dynamically changes the queue by allowing the arrival of more urgent

packets to replace some of the existing less urgent ones. This again serves to boost the efficiency and responsiveness of the system. The application of this method is crucial for the healthcare category, particularly in real-time patient monitoring and emergency response systems, where it can prioritize patient sensors to ensure timely treatment of vital alerts, such as rapid drops in oxygen and spikes in blood pressure. By arranging data into high- and low-priority queues and then prioritizing according to the emergency level of the packets, the system ensures that the most important cases are executed rapidly. This method further enhances the responsiveness of remote patient monitoring, telemedicine, and automated decision support systems, positively impacting patient outcomes and enabling better targeting of clinical resources in high-stress areas, such as emergency rooms and mass casualty events. The list of notations used in this work is shown in Table 5.1.

Table 5.1: List of Notation for D-QACA

Sr.No	Notation	Description
1	r_Node	Root Node
2	inode	Intermediate Node
3	LPQueue	Low Priority Queue
4	HPQueue	High Priority Queue
5	lc_Node	Lowest Cost Node
6	Packet	Packet Received from the root node

5.4 Queuing at an Intermediate Node

Intermediate nodes have queues in which packets are classified into either high or low priority depending upon the type of sensor. Packets are then processed according to their emergency level and the priorities assigned. High-priority packets are processed based on their emergency level, which contains sensitive data such as heartbeat and brain data. Lesser-priority packets are placed in low-priority queues. The system considers that the sensors generate packets of patient data, which are forwarded sequentially to a collector node; each packet is assigned a priority and an emergency level. The system has two queues: one for high-priority packets and another for low-priority packets; packets with higher urgency are processed first. If there are two packets with the same emergency level, the one arriving first is given priority. The system is assumed to work in a non-preemptive fashion, which means processing does not stop on one packet for a new packet. The proposed queue management model accommodates high-priority packets by placing them at an index next to that of the packet currently in process, as illustrated in 5.2. This model has a finite buffer size, which effectively prevents the eternal loss of service. These assumptions provide a realistic and robust basis for ensuring the behavior of the system in healthcare.

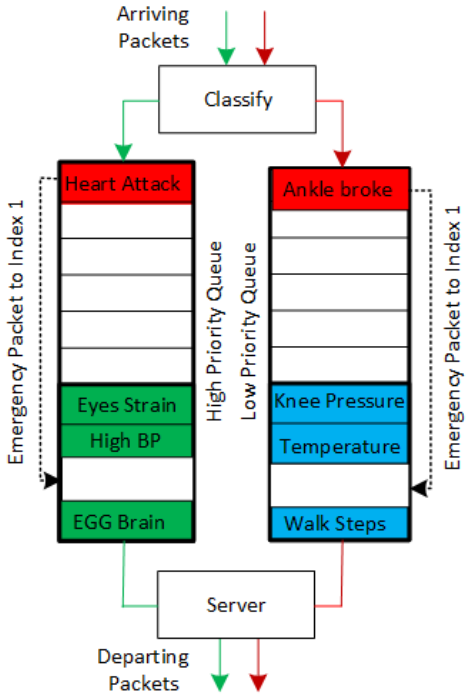


Figure 5.2: Queue maintained at Intermediate Node

5.5 Emergency-aware packet placement algorithm

The funtionality of the D-QACA is presented in Algorithm 1, which is called the Emergency-aware Packet Placement Algorithm . Its main task is to manage emergency packets by placing them at the front of the queue, ensuring timely deliveries without incurring excessive time delays. Steps (1-7) include the Receive Packet function, which is meant to manage all the incoming packets by checking their priorities and inserting them into their relevant queue. The function would initially confirm a newly accepted packet to establish the packet's priority status. If the priority status indicates that it is higher-priority traffic, the packet shall indeed be inserted into the High Priority Queue (HPQueue). For packets that do not carry a high-priority flag, treated as lower-priority packets, they shall be placed at the Low Priority Queue (LPQueue). After assigning the packet to its relevant queues according to its priority, the function ends. As such, each packet gets served according to its priority. Steps 8 to 18: The insertion queue function is set to insert a packet into the queue, which can be either a high-priority or a low-priority queue. An index was initialized to zero before checking the size of the queue. If the queue is empty or has only one packet, this incoming packet is directly added to the queue, regardless of its emergency flag (Eflag). If the queue contains more than two packets and the packet has an Eflag of 0, the packet gets appended to the bottom of the queue. However, if the queue contains more

packets and the incoming packet's Eflag is set to 1, indicating high urgency, it will then be stored in the second position in the queue. Thus, packets having high levels of urgency will then be prioritized within respective queues.

Algorithm 2: Priority-Based Packet Reception and Queueing with ACK

Input: Packet p received from **root node**
Output: Acknowledgment (**ACK**) sent by each node to its parent
p.flag $\in \{\texttt{HighPriority}, \texttt{LowPriority}\}$;
p.Eflag $\in \{0, 1\}$;
Function `ReceivePacket(`p`)`
  if $p.flag = HighPriority$ **then**
    `InsertInQueue(`$HPQueue, p$`)`;
  else
    `InsertInQueue(`$LPQueue, p$`)`;
  end
  `SendACK(`$P_{node}, p.id$`)`;
Function `InsertInQueue(`$Queue$, p`)`
  $n \leftarrow |Queue|$;
  if $n \leq 1$ **then**
    PushBack(Queue, p);
    ; `// Queue empty or one element`
  else if $p.Eflag = 0$ **then**
    PushBack(Queue, p);
    ; `// Routine packet` $\rightarrow$ `append`
  else
    InsertAt(Queue, 2, p);
    ; `// Emergency packet: insert after head`
  end
Function `SendACK(`$P_{node}, pktID$`)`
  Create `ACK` $\leftarrow \langle pktID,\ status = queued,\ queueDepth = |HPQueue| + |LPQueue| \rangle$;
  Transmit `ACK` to P_{node};

5.6 Computational and Time Complexity Analysis of D-QACA

5.6.1 Computational Complexity Analysis

The computational complexity of the Dual Queue-Aware Congestion Avoidance Scheme (D-QACA) is analyzed by examining the operations performed at each intermediate IoHT node during packet processing. For each incoming packet, the node executes packet classification to determine whether the packet is emergency or normal, followed by insertion into the appropriate queue (High Priority Queue or Low Priority Queue). Both classification and enqueue operations are simple logical checks and FIFO insertions, which require constant time. Queue occupancy is tracked using counters that are updated upon each enqueue and dequeue operation, introducing

no additional computational burden. During scheduling, the node checks the availability of packets in the High Priority Queue and serves it preferentially; otherwise, packets from the Low Priority Queue are transmitted. Since no sorting, searching, or reordering of packets is required, the computational cost per packet remains constant. Therefore, the computational complexity per packet can be expressed as:

$$C_{\text{D-QACA}} = O(1). \tag{5.1}$$

5.6.2 Time Complexity

Let N denote the total number of packets processed at a node. As each packet undergoes a fixed sequence of constant-time operations, the total processing time grows linearly with the number of packets. Hence, the overall time complexity for processing N packets is given by:

$$T_{\text{D-QACA}}(N) = O(N). \tag{5.2}$$

In high-traffic scenarios where all packets are classified as emergency packets, the same processing steps are applied without additional overhead. Consequently, the worst-case, average-case, and best-case time complexities of D-QACA remain identical. This linear time behavior ensures that D-QACA scales efficiently with increasing traffic load while maintaining suitability for resource-constrained IoHT devices.

5.7 RESULTS AND DISCUSSION

The simulation results presented here are obtained for two different channel capacities: 120 kbps and 250 kbps. The residual queue is set to 70 for this analysis, meaning that acknowledgment will be sent when the queue is 70% full. The list of simulation parameters is shown in Table 5.2. The performance of the proposed D-QACA was then validated through extensive simulation using NS2.35, where node configuration, node deployment, message initiation, and trace annotation were included in the TCL file. The assumption in this study was that the intermediate node would be responsible for forwarding the data to the sink node. The second assumption is that a maximum of three backouts, represented by m, by and retransmission n, are allowed. The packet collision ratio is 0.10, and the number of leaf nodes is 2 to 10, while the number of packets that can be sent is set to 10 at a packet transmission rate of 32 packets/second. All

specifications are referenced from [108]. It studies packet drop due to low buffer capacity. Since low-priority packets are arriving in the buffer, there is no space to accommodate high-priority packets. The number of packets dropped due to buffer overflow, compared to the total number of packets reaching the local sink node, is treated as a measure of the high-priority packets. The base schemes for this study are the Analytical Model [108], DCCA [94], and QACA, which performed the simulations using channel capacities of 250 Kbps and 120 Kbps.

Table 5.2: Simulation Parameter for D-QACA

Sr.No	**Parameter**	**Value**
1	Simulation Time	110 s
2	Deployment Area	500x500 m
3	Transmission Range	350 m
4	Number of Nodes	50
5	Data Packet Size	500 Bits
6	No. of Cluster Heads	4
7	No. of Simulations	10
8	Time Consumed	0 - 20 seconds
9	No. of Packets	0 - 40 PAckets
10	Density	10 - 110 Users
11	Control Message Size	100 bits
12	Duration of Data Periods	10 s
13	Cahnnel Capacity	120 kbps & 250 kbps
12	Sensing Radius	20 m
12	Maximum Packets in Queue	50

5.7.1 Buffer Loss Probability

Figure 5.3 shows buffer loss probability for leaf and intermediate nodes under the D-QACA scheme. As traffic load increases, baseline approaches exhibit early saturation of buffers. D-QACA significantly reduces buffer loss by isolating emergency packets and prioritizing their placement within dedicated queues. In figure 5.3 (a), whereas the behavior of the Analytical Model and DCCA in terms of buffer loss is severely affected when the load is increased, QACA-70 experiences a sharp deterioration at load 6 and performs well otherwise. On the contrary, D-QACA-70 exhibits the minimum buffer loss for all loads, resulting in excellent performance for leaf nodes. With 6 sensors per patient, the buffer loss probabilities were 0.11, 0.09, 0.07, and

0.05 for the Analytical Model, DCCA, QACA-70, and D-QACA-70, respectively. In Figure 5.3 (b), the increase in uplink buffer loss as compared to leaf nodes is specifically assigned to heavy traffic at the intermediate node for 250 kbps. The buffer loss values for the Analytical Model and DCCA are 0.36 and 0.46, respectively, for 6 sensors per patient. In contrast, QACA-70 reduces these values to 0.15, while D-QACA-70 achieves the lowest value of 0.09. In the second scenario with a reduced channel capacity of 120 kbps, Figure 5.3(c) shows the buffer loss probability at the leaf node, where the Analytical Model and DCCA present an increase of 0.25 and 0.35, respectively, with 6 sensors per patient. QACA-70 achieves a better performance of 0.09, but it is D-QACA-70 that surpasses this by reaching a low of 0.05. Figure 5.3 (d) explains how the buffer loss probabilities at intermediate nodes are faring with a channel capacity of 120 kbps. The Analytical Model and DCCA yield considerable increases in the buffer loss probability, opposite to those obtained with QACA-70 and D-QACA-70. Again, at a higher rate, D-QACA-70 shows the least buffer loss. It did not show any significant buffer loss probability for the proposed scheme for emergency packets, even in extreme fullness of 90 percent in queues at intermediate and leaf nodes. The packets are dropped from the queue tail, while the emergency packets enter directly at the first index or immediately adjacent to the current packet being executed.

5.7.2 High-Priority Packet Lost Per Second

The most significant amount of packet loss experienced by the Analytical Model and DCCA, especially in leaf nodes with a channel capacity of 250 kbps, is 11% and 13%, respectively, when 6 sensors per patient are considered, as illustrated in Figure 5.4. With increasing load, these schemes increase rapidly in number. Poor congestion handling results in more packets being lost because they are not dynamically prioritized under high load. There will be figures for performance comparisons of these two: QACA-70 loses more packets at 8, whereas D-QACA-70 loses only 5 packets. The difference lies in the dynamic priority queuing employed for D-QACA, which effectively prioritizes high-priority packets more efficiently and handles congestion more effectively. In Figure 5.4 (b) for Intermediate Nodes with 250 kbps, the trend is the same. From the DCCA and Analytical Model, very high packet loss is observed, while QACA-70 exhibits moderate to low performance. In contrast, D-QACA-70 experiences less packet loss, demonstrating adaptability to behavioral changes in various traffic situations. At 120 kbps, as shown in Figure 5.4(c), the channel capacity decreases to leaf nodes, accompanied by increased

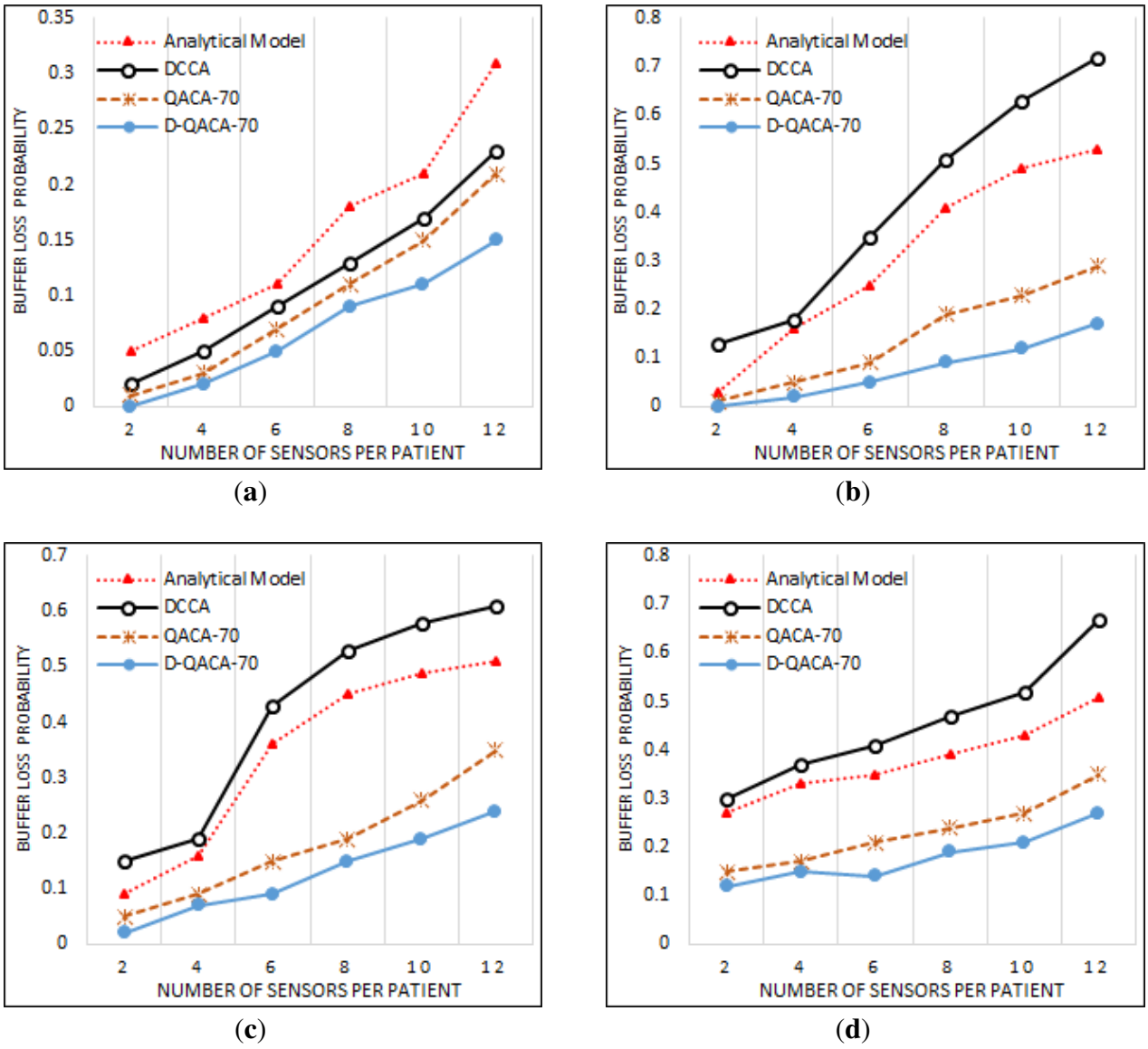


Figure 5.3: Buffer Loss Probability for (**a**) Leaf nodes at 250kbps; (**b**) Intermediate nodes at 250kbps; (**c**) Leaf nodes at 120kbps; (**d**) Intermediate nodes at 120kbps

packet loss in all models, due to bandwidth limitation. However, among the other schemes, D-QACA-70 still performs significantly better, as it exhibits better control of high-priority traffic due to its dynamic adjustment to buffer conditions and load. The proposed scheme, D-QACA-70, classifies packets according to their selection procedure and passes them to the high- or low-priority queue designated for immediate processing, depending on their urgency. The incoming packet is placed in the next index of the queue, which reduces the chances of high-priority packet loss due to "drop tail" being almost zero. When 10 sensors are attached to each patient, high-priority packet loss is 19, 21, and 17, with D-QACA-70 outperforming other schemes by providing 9 fewer high-priority packet losses, as shown in Figure 5.4(d).

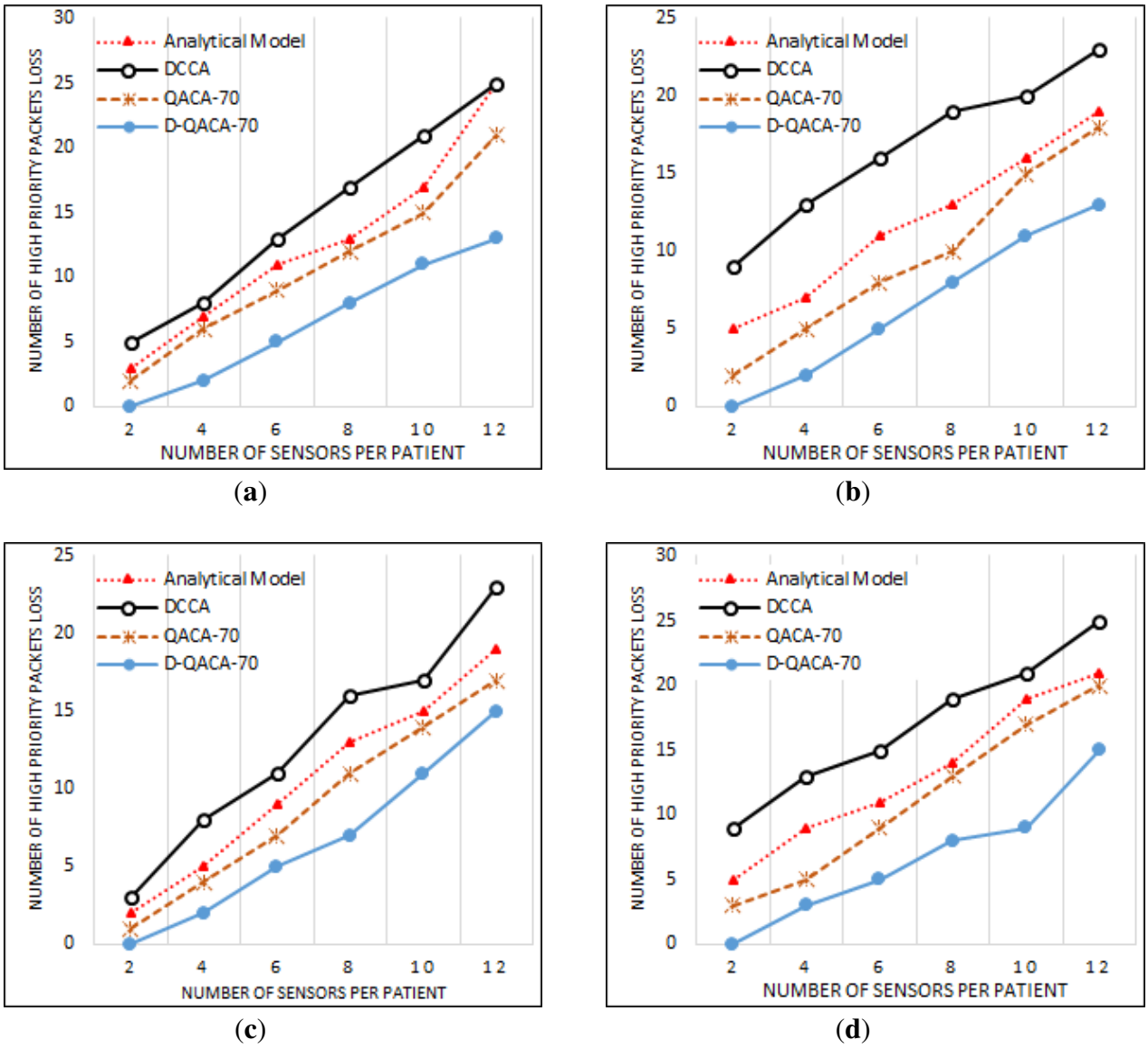


Figure 5.4: High Priority Packet Loss (**a**) Leaf nodes at 250kbps; (**b**) Intermediate nodes at 250kbps; (**c**) Leaf nodes at 120kbps; (**d**) Intermediate nodes at 120kbps

5.7.3 Packets Delay

D-QACA-70 (Emergency) exhibits a packet delay of 3 ms, 20 ms, 17 ms, and 11 ms for the Analytical model, DCCA, QACA-70, and D-QACA-70, respectively, as shown in Figure 5.5. The scenario considers 8 sensors per patient within a channel capacity of 250 kbps. D-QACA-70 has a higher value since it incurs the least computational cost to carry out emergency checks and swap at index 1 or higher incrementally if there are already existing emergency packets. However, in total, it has the least delay in processing emergency packets compared to these schemes. Figure 5.5 (b) depicts packet delay for a channel capacity of 250 kbps as the number of sensors per patient varies from 2 to 12. For 8 sensors per patient, the packet delays for the

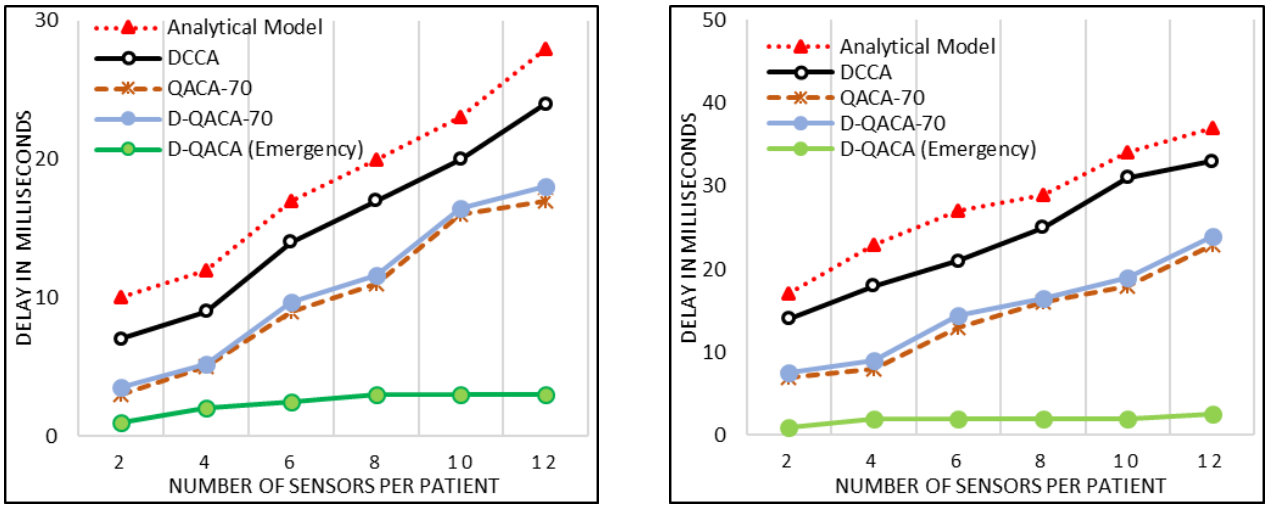


Figure 5.5: Packets delay: (**a**) Intermediate nodes at 120 kbps; (**b**) Intermediate nodes at 250 kbps.

Analytical model, DCCA, D-QACA-70, and QACA-70 are 29 ms, 25 ms, 16.5 ms, and 16 ms, respectively. In contrast, D-QACA-70 (Emergency) requires just 2 milliseconds to handle urgent scenarios, which is significantly less.

5.8 Summary

This research presents the design and analysis of the Dual Queue-Aware Congestion Avoidance (D-QACA) methodology, which is particularly well-suited for the Patient-Centric Internet of Health Things environment. The core issue of this study was prioritizing emergency health data packets in a limited, constrained network to address bottleneck situations. In IoHT systems, medical data are continuously generated from wearable or implantable medical devices using heterogeneous sensors. The majority of the information sent would be in line with normal physiological readings; only a small percentage constitutes life-critical or emergency events, for example, abnormal heart signals or neurological fluctuations. Time-critical is the transmission of these packets, as even a slight delay or loss could risk patient lives.

The proposed D-QACA framework ensures that prioritized packets are processed first, thereby preventing them from suffering loss or delay during congestion scenarios. D-QACA differs from traditional queuing strategies that treat all packets equally or subscribe to their arrival sequence in that D-QACA is queue-aware: emergency packets bypass standard queuing if any resources

are available, and when the queue is already active, such packets are strategically inserted in the second position. The implementation of this mechanism results in a reduction of buffer loss probability for emergency packets by almost 80% and a reduction in emergency packet loss probability by 90%. This justifies the robustness of the scheme in delivering life-theriving data without placing intolerable costs on normal traffic processing.

With the number of IoHT devices and users ever increasing, any healthcare monitoring system must not only be reliable but also scalable. By minimizing critical packet loss and ensuring timely transmission, D-QACA enhances the trustworthiness of real-time monitoring and the utilization of IoHT-based healthcare in both clinical and remote environments. The scheme also ensures that all life-critical data is delivered smoothly, thereby facilitating faster diagnosis and timely medical intervention, and enhancing the link between IoHT and hospital management systems, as well as cloud-enabled infrastructures. The limitation in this research is that QACA and D-QACA do not account for redundant information. Sensors typically convey continuous streams of information, and emergency events are rare. Transmission of complete readings, including redundant normal values, may create heavy loads on the network, thereby enhancing congestion. Although D-QACA has proven to be effective in prioritizing emergency packets, the most challenging issue remains managing redundancy. An optimized mechanism would allow normal values to be represented by compressed or Boolean indicators, thereby reducing unnecessary load on the network. This mechanism would limit competitive congestion and improve system efficiency without compromising the reliability of critical information delivery. The improvements observed in D-QACA are primarily attributed to emergency packet isolation, which prevents life-critical packets from being delayed behind non-urgent traffic within the same priority class. Furthermore, under high-load conditions, service rate limitations become the dominant factor influencing delay, explaining the gradual performance degradation at extreme traffic intensities.

This limitation also affects the broader perspective of the IoHT ecosystem. With the expansion of IoHT deployment, managing redundancy will become extremely critical for energy savings, bandwidth optimization, and scalability. This way, future developments need to incorporate redundancy detection and suppression into the D-QACA framework. Overall, the D-QACA represents a significant leap forward in congestion avoidance for patient-centered IoHT applications, considering the relative importance of emergency packet prioritization. Improvements in buffer loss, packet delay, and packet loss have been shown to provide a good indication of

applicability for real-time healthcare monitoring. To expand its applicability, however, attention will have to be paid to redundant data transmission. This condition will be critical for a holistic and sustainable approach to congestion management in large-scale IoHT deployments. This next chapter will be addressed subsequently through the study of redundant minimization strategies, which can serve as a complementary enhancement for D-QACA.

CHAPTER 6

DE-DUPLICATION AND QUEUE-USAGE AWARE CONGESTION AVOIDANCE SCHEME

6.1 Overview

Through an exponential increase in connected healthcare devices within the Internet of Health Things ecosystem, various opportunities for continuous patient monitoring and proactive healthcare are emerging like never before. However, this advancement has presented unprecedented challenges to network reliability, especially in emergency scenarios where the timely delivery of high-priority medical data could be a matter of life and death. The D-QACA scheme was presented in the previous chapter, where inputting timid medical packets could reduce cases of buffer overflow by placing emergency packets deliberately at the front of the queue. While the demonstration of D-QACA significantly improved the reduction of packet loss for critical and heavily overloaded data, it struggled to proactively mitigate the onslaught of redundant transmissions that exist within the IoHT environment in real-life scenarios.

Medical sensors, in practice, continuously generate data at a fixed rate, largely oblivious to whether a parameter has changed significantly or remained relatively stable. For instance, those wearable devices that monitor heart rate, blood pressure, or temperature often transmit consecutive values that are the same or almost identical. Continuous alerting about redundant information consumes a significant amount of precious network bandwidth, unnecessarily reduces buffer loss in queues, and wastes resources at the sensor nodes. More importantly, during periods of network congestion, the redundancy of working data aggravates buffer occupancy and results in dropped packets, increased delays, and sometimes starvation of critical medical alerts. This

point of inefficiency creates another gap in the D-QACA framework: While it addresses the packet loss of emergency packets, nothing in D-QACA aims to reduce redundancy.

To counter the aforementioned challenges, this chapter proposes DD-QACA, i.e., De-Duplication and Queue-usage Aware Congestion Avoidance. The DD-QACA scheme implements a twofold strategy, that is: (1) to incorporate a de-duplication mechanism that filters out repetitive sensor readings, allowing unique and truly useful data to flow into the upper layers of the network, and (2) to take into consideration queue-usage information that dynamically adjusts packet scheduling and congestion control decisions in accordance with real-time buffer occupancy. Hence, by enforcing the above-mentioned strategies, the scheme effectively utilizes available network resources by rejecting duplicate data, thereby increasing the chances of avoiding congestion and further enhancing the timely processing of emergency packets.

The DD-QACA reduces redundancy and adaptively monitors queues, enhancing network efficiency and information reliability. This method not only protects the critical medical data flow during congested scenarios but also optimizes the use of available bandwidth, thereby conserving energy on the part of the sensors and contributing to the overall improvement of quality of service with the increased data. This chapter describes the system design, algorithmic framework, and operational flow of the DD-QACA scheme, which will be the subject of an extensive performance analysis. It will be demonstrated that DD-QACA outperforms classical congestion avoidance schemes in balancing priority packet handling with intelligent data filtering.

6.2 System Model & Problem Statement

The section highlights system models shown in that incorporate a smart health monitoring system for the early detection of congestion in network queues and the prevention of packet loss in emergencies during message transfer. The study aims to investigate priority queueing techniques in IoT-based health monitoring systems for handling emergency data. In emergency conditions, it creates a message exchange between the nodes and the gateway, and this is when congestion occurs. So far, the congestion has increased; the model assures that emergency packets will reach their recipients on time. There are four sorts of nodes that construct this model: i) leaf nodes, sensor devices consisting of both wearable and non-wearable sensors. The sensing devices mentioned above are used to collect data from patients. ii) The collector node

(mobile or dedicated computational equipment) collects the packets that come from the network and classifies them as high or low priority, checks for redundancy in terms of duplication, and determines their severity. iii) Intermediate nodes refer to medical objects having large computing and network resources; these intermediate nodes process data collected from sensing devices according to packet priority, then send it to a local sink node for further processing. iv) Local sink nodes are fog servers with huge computational power. The primary function of the sink node is to preprocess packets received from intermediate nodes. In other words, it attempts to determine the type of healthcare services required by a patient based on packet data. iv) the root node, which is described as the cloud. The cloud connects to the hospital administration systems that doctors use. The use of IoHT sensing devices, such as smartwatches and smartphones, among other gadgets, forms a leaf node. Currently, electronic watches and smartphones often function as leaf nodes through which sensing data can be transferred to local sink nodes via collector and intermediary nodes. There, at that root node, the cloud servers would provide transformation on the information received from the network.

The major problem is that the sensors placed at Level 1 in health monitoring systems do not filter out duplicate or redundant information sent to the intermediate nodes by patients. The incoming data at the intermediate node can be managed through high- and low-priority queues that filter out important from non-essential data. A very large number of these readings fall within normal limits, representing steady, non-threatening conditions. Sending such redundant information in its complete 16- or 32-bit form wastes more energy from the sensors and increases communication costs throughout the network. This, in itself, not only lowers the operating lifespan of the sensors but also causes traffic congestion, filling high- and low-priority queues, which may even cause delays in transferring critical data.

6.3 De-Duplication and Queue-usage Aware Congestion Avoidance Scheme

This section proposes a "De-duplication and Queue-Usage Aware Congestion Avoidance (DD-QACA) scheme" to solve the identified problem. The proposed system utilizes a priority-based queue, where every node (also referred to as the router) has two buffer queues. Each queue is for different priority levels, namely, low and high priority. When a node receives an input packet, it is categorized and routed to its corresponding queues. The packets in the high-

and low-priority queues are separated into different priority levels and emergencies. Internally complex sensing devices have been developed for patients to collect various data within the healthcare sector. These metrics vary in that they measure part single-number sets, e.g., BPM, which returns 16-bit data for some sensors. Other sensor reading is converted to two values. That is, systolic and diastolic values are recorded from tracking the blood pressure of a sensor. In total, these generate two 16-bit data streams, which sum up to 32 bits. This makes the environment very diverse, which escalates energy usage. Level one of the proposed checks whether the data from sensors lies within a normal range or not. If it falls within a normal range, instead of sending all 16 or 32 bits, only the Boolean digit 0 is sent out, which requires only one bit, as shown in figure 6.1 . Moreover, any threshold change has become a possible example in our

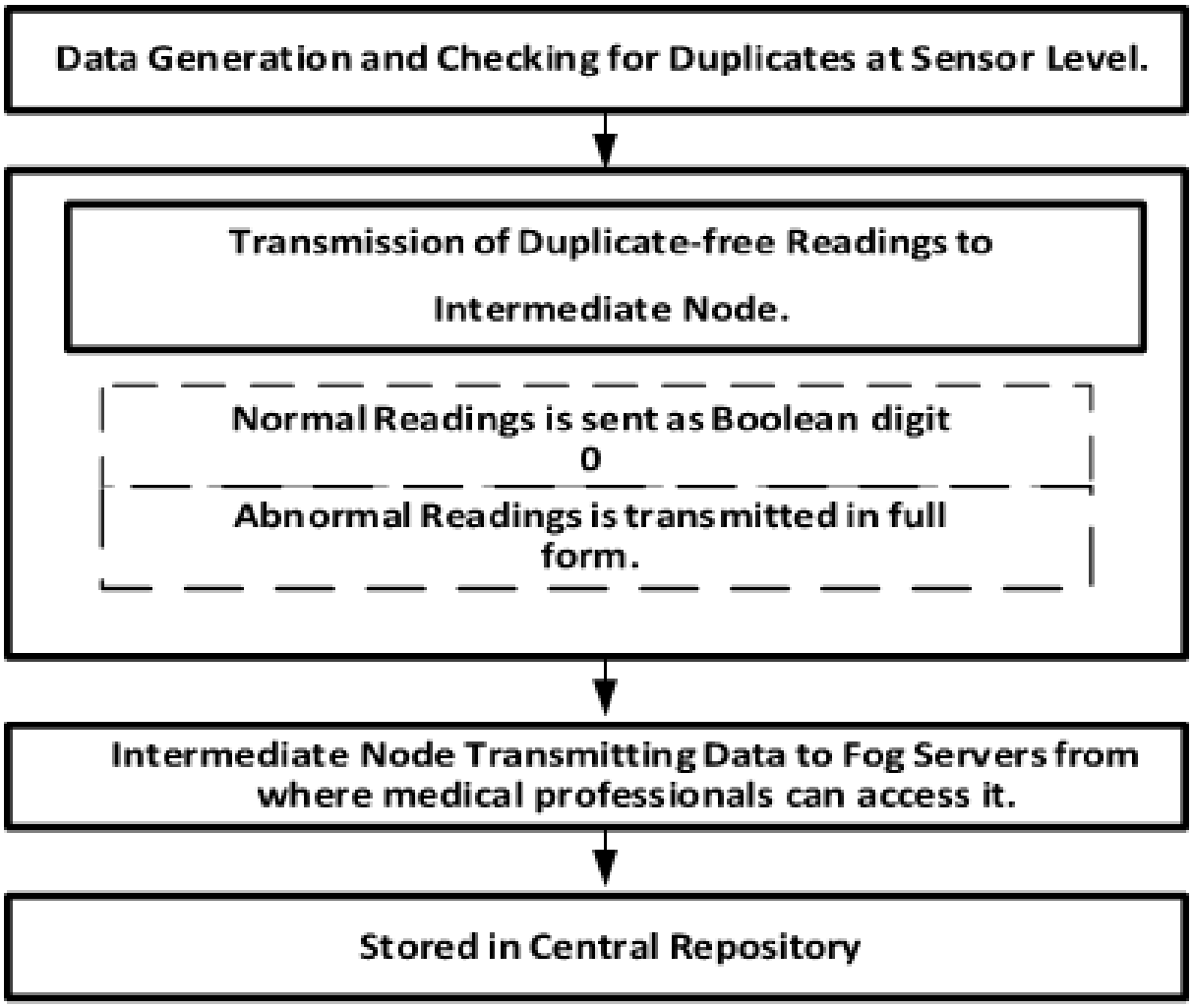


Figure 6.1: Steps of De-Duplication

method, as the mediator node merges and processes the values collected from sensors within designated threshold ranges. These ranges could be easily changed at the intermediary nodes at any time. Sensors record the readings coming from the bodies of these patients in the form of

concatenated strings. For example, one such record made by patient A is: 98.9:75:98:130:180, referring to the body's temperature, pulse, oxygen level, cholesterol levels, and sugar level. The obtained body data are either normal or critical. In a normal average person, such readings of parameters tend to stay the same and fall within the limits of normality. In these cases, the sensor nodes would send a Boolean digit 0 towards the intermediate node. On the contrary, in cases where a person is ill, some irregularities may be visible in one or two sensor readings. At the same time, the rest of the parameters remain unchanged in many cases. The critical valued data are transmitted in their pure format, while the remaining sensors transmit only the Boolean digit. On very rare occasions, all sensors installed may provide erroneous data. In such an event, all values are sent in their original format.

6.3.1 Queuing at the Intermediate Node

The queue on the intermediate node categorizes packets into high and low priority depending on the nature of the sensor. They are processed according to their level of criticality and assigned priority. Any packets carrying sensitive information, such as heartbeat or brain signals, receive an emergency rating and priority for processing. On the contrary, packets containing less critical data are put in a low-priority queue. From a system perspective, when sensors create packets of patient data, these packets are ultimately passed on to a collector node after the level of priority and emergency has been assigned to each packet. There are two queues: high-priority and low-priority. Those considered more urgent are processed immediately. However, if the two packets happen to have the same emergency level, the one which arrived first will be given priority. The system assumes that processing is non-preemptive: in that case, any other packets already processed will be put on hold until the arrival of a new one. The proposed scheme for queue management will dynamically account for the processing of high-priority packets by using an index close to the one containing the packet being processed. This model maintains a finite buffer capacity, thereby ensuring a robust buffer management scheme to prevent overflow. These assumptions provide a reasonable yet sound basis for ensuring system reliability in healthcare.

6.3.2 Mathematical Modeling for Duplicate Reading Removal

In IoHT monitoring, consecutive sensor readings often remain unchanged (or change only marginally) over short intervals. Transmitting such redundant readings increases the effective

traffic load and contributes to congestion. Therefore, the proposed Duplicate Reading Removal Algorithm suppresses readings whose change is below a predefined threshold ε.

Let x_k denote the sensed value at time index k (e.g., temperature, SpO_2, ECG feature), and let x_{k-1} denote the previous reading. The absolute change between consecutive readings is defined as:

$$\Delta_k = |x_k - x_{k-1}|. \tag{6.1}$$

A reading is considered *duplicate* (or redundant) if the change is smaller than a threshold $\varepsilon > 0$, i.e.,

$$\Delta_k < \varepsilon. \tag{6.2}$$

Accordingly, the transmission decision variable $u_k \in \{0,1\}$ is defined as:

$$u_k = \begin{cases} 1, & \text{if } \Delta_k \geq \varepsilon \ \ (\text{transmit}) \\ 0, & \text{if } \Delta_k < \varepsilon \ \ (\text{suppress}) \end{cases} \tag{6.3}$$

Normalized Change (for Multi-Scale Signals)

For sensor signals with different magnitudes, a normalized change can be used:

$$\delta_k = \frac{|x_k - x_{k-1}|}{\max(|x_{k-1}|, \eta)}, \tag{6.4}$$

where η is a small constant to avoid division by zero. In this case, the suppression rule becomes:

$$\delta_k < \varepsilon. \tag{6.5}$$

Extension for Multivariate Readings

If the sensor produces a vector reading $\mathbf{x}_k \in \mathbb{R}^m$, the change can be defined using the Euclidean norm:

$$\Delta_k = \|\mathbf{x}_k - \mathbf{x}_{k-1}\|_2, \tag{6.6}$$

and the same threshold rule $\Delta_k < \varepsilon$ is applied.

This formulation provides a mathematical foundation for redundancy suppression and ensures that only clinically meaningful changes are transmitted, thereby reducing traffic load and improving congestion resilience without impairing emergency data delivery.

6.3.3 Duplicate Reading Removal Algorithm

DD-QACA functionality is referred to as presented algorithm 1, which describes the Duplicate reading removal algorithm. It is designed to manage emergency packets by placing them at the front of the queue, ensuring timely delivery and preventing excessive delays in packet notification. In steps 1-10, algorithm receives as input a set of healthcare sensor nodes which is represented as S={1, n2,...., nk}. Every sensor node collects readings, which are then processed and stored. The output consists of two data strings: dataStringHPQ and dataStringLPQ, which contain processed data from high-priority and low-priority sensors, respectively. Initially, three string variables-aggrString, dataStringHPQ, and dataStringLPQ- are declared empty for consolidated and categorized storage of sensor data. The counter variable, count, is initialized to zero for the purpose of tracking the number of readings or iterations received. Each sensor node $n[i]$ takes measures at intervals of t and stores those readings in $n[i]$.readings[]. The algorithm checks whether the emergency flag $(n[i].emergency\_flag)$ and the full data flag (FDataFlag) are off. If yes, all $n[i]$ sensors have their readings processed. Each sensor node $n[i]$ is checking its priority flag. If the sensor is called High Priority `([i].flag = High_Priority)`, it concatenates its processed readings (obtained by using the AggregateMessage function) into dataStringHPQ, by using a delimiter DL2 to separate values from the different nodes. When the sensor is called Low Priority `([i].flag = Low_Priority)`, its readings are concatenated into dataStringLPQ with the same delimiter.

In steps 11 to 18, the algorithm increments the variable `count` by 1 to keep track of how many sensor readings are processed. When the `count` reaches the specified threshold, the algorithm resets `count` to 0 and sets the flag `FDataFlag` to `true`, ensuring that full sensor values, rather than shortened values, are transmitted. Whenever this criterion is met, the method iterates through all k sensors and appends their full readings to the variable `aggrString`. The values are concatenated using the delimiter `DL1(":")` to distinguish between sensor results for each node. After collecting all sensor readings, the algorithm evaluates the priority level of each sensor. For sensors classified as `High_Priority`, the corresponding `aggrString` containing all readings is concatenated into `dataStringHPQ` using the delimiter `DL2("_")`. The same delimiter `DL2("_")` is applied for `Low_Priority` sensors, ensuring that each sensor's `aggrString` is correctly concatenated into `dataStringLPQ`. This structured concatenation enables accurate classification and efficient data management within the priority queues.

In steps 19 to 26, sensor readings are first processed with the AggregateMessage function before storing them. It has input which corresponds to IN_ARGS or sensor node readings. The value of reading is compared first against the THL, lower threshold, and THU, upper threshold, to check whether it is outside the expected threshold range. The actual value is considered by concatenating in aggrString using the delimiter `DL1 (":")` for correct value separation if the reading is abnormal, remaining below the THL or above THU valued. DL1 concatenates a Boolean digit "0" into aggrString in place of saving the real value-if the reading is within normal range, to reduce redundancy and improve data transmission.

Algorithm 3: Duplicate Reading Removal Algorithm (DRRA)

Input: $S = \{n_1, n_2, \ldots, n_k\}$ (set of healthcare sensors); thresholds TH_L, TH_U; delimiters DL_1 =: (within-node), DL_2 = _ (between-nodes); integer `threshold` (e.g., 20)

Output: `dataStringHPQ`, `dataStringLPQ`

```
1  dataStringHPQ ← ""; dataStringLPQ ← "";
2  count ← 0; FDataFlag ← False;
3  if (∀i) n_i.emergency_flag = OFF and FDataFlag = False then
4      for i ← 1 to k do
5          aggrString ← AggregateMessage(n_i.Readings, TH_L, TH_U, DL_1);
6          if n_i.flag = HighPriority then
7              dataStringHPQ ← Concat(dataStringHPQ, DL_2, aggrString);
8          else
9              dataStringLPQ ← Concat(dataStringLPQ, DL_2, aggrString);
10     count ← count + 1;
11 else if count > threshold or (∃i) n_i.emergency_flag = ON then
12     count ← 0; FDataFlag ← True;
13     for i ← 1 to k do
14         aggrString ← Join( n_i.Readings, DL_1 ) ;      // send full readings
15         if n_i.flag = HighPriority then
16             dataStringHPQ ← Concat(dataStringHPQ, DL_2, aggrString);
17         else
18             dataStringLPQ ← Concat(dataStringLPQ, DL_2, aggrString);
19 Function AggregateMessage(vals, TH_L, TH_U, DL_1):
20     out ← "";
21     foreach v ∈ vals do
22         if v < TH_L or v > TH_U then
23             out ← Concat(out, DL_1, v) ;          // abnormal → keep value
24         else
25             out ← Concat(out, DL_1, 0) ;              // normal → mark 0
26     return out ;
```

6.4 Sensor Energy Savings with and without De-duplication

In resource-constrained IoHT systems, radio transmission is typically the dominant source of energy consumption. The proposed de-duplication mechanism reduces the number of packet transmissions by suppressing redundant readings, which directly lowers the energy consumed by sensor nodes.

Let E_{tx} denote the energy required to transmit one packet, E_{rx} denote the energy required to receive one packet (e.g., acknowledgments or control packets), and E_{proc} denote the local processing energy per sampled reading (classification and de-duplication decision). Let N_s denote the total number of sampled readings generated during an observation interval, and let N_{tx} denote the total number of transmitted packets.

6.4.1 Energy Consumption without De-duplication

Without de-duplication, every sampled reading is transmitted, i.e., $N_{tx}^{(0)} = N_s$. The total energy consumption over the interval can be expressed as:

$$E_{\text{total}}^{(0)} = N_s E_{tx} + N_{rx}^{(0)} E_{rx} + N_s E_{proc}, \tag{6.7}$$

where $N_{rx}^{(0)}$ is the number of received packets (e.g., ACKs) in the baseline case.

6.4.2 Energy Consumption with De-duplication

With de-duplication, only non-redundant readings are transmitted. Let $u_k \in \{0,1\}$ be the transmission decision variable defined in Section (Duplicate Reading Removal), where $u_k = 1$ indicates transmission and $u_k = 0$ indicates suppression. The number of transmitted packets becomes:

$$N_{tx}^{(d)} = \sum_{k=1}^{N_s} u_k. \tag{6.8}$$

Hence, the total energy consumption with de-duplication is:

$$E_{\text{total}}^{(d)} = N_{tx}^{(d)} E_{tx} + N_{rx}^{(d)} E_{rx} + N_s E_{proc}. \tag{6.9}$$

6.4.3 Energy Savings and Saving Ratio

The absolute energy saving due to de-duplication is:

$$\Delta E = E_{\text{total}}^{(0)} - E_{\text{total}}^{(d)}. \tag{6.10}$$

The energy saving ratio (percentage reduction) is computed as:

$$\mathscr{S}_E(\%) = \frac{E_{\text{total}}^{(0)} - E_{\text{total}}^{(d)}}{E_{\text{total}}^{(0)}} \times 100. \tag{6.11}$$

6.4.4 Interpretation

Since $N_{tx}^{(d)} \leq N_s$, the transmission energy term is reduced by a factor proportional to the suppression rate. Therefore, the proposed de-duplication mechanism decreases the overall sensor energy consumption and extends node lifetime, while preserving clinically meaningful changes in the transmitted healthcare data.

6.5 RESULTS AND DISCUSSION

To demonstrate the effectiveness of the proposed DD-QACA method, extensive simulations are conducted using NS-2.35. Sensor nodes are deployed in a network of 500 × 500 m in size. The transmission range among these nodes is 30 m, which is the same for all kinds of features. The sensor nodes then send data packets every 30 seconds. The TCL file contains the complete information for configuring and setting up network nodes. The sending and receiving of data packets are implemented using separate C classes. The processed data results are then retrieved using AWK files. In a traditional IoT-based hospital monitoring system, sensors such as ECGs, blood pressure monitors, pulse oximeters, temperature sensors, accelerometers, and EEGs continuously gather information from patients and forward that to a collector node (CN). The CN collects this information and transmits it to an intermediate node (IN), which then sends it to a centralized server for further processing and analysis. In this arrangement, complete information from the sensor is transmitted in the event of an emergency, where each patient generates a large amount of data and sends it to the intermediate node every 15 seconds. For instance, the data collected by an ECG sensor can be 250 Hz with 16-bit resolution, which generates about

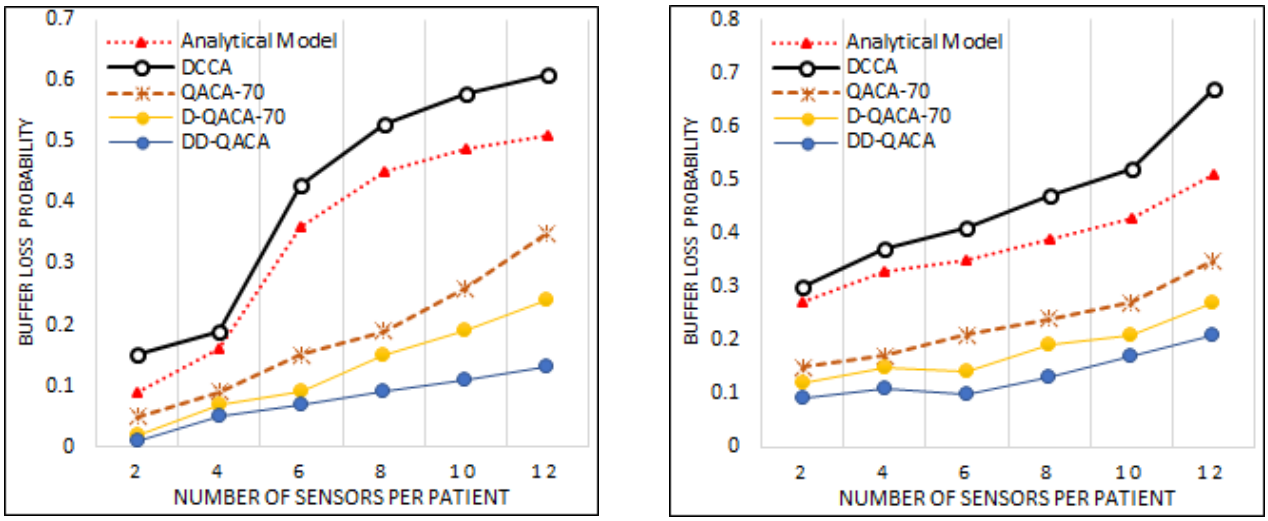
500 bytes per second (4 kbps). A blood pressure monitor generates 2 bytes per second (16 bits per second) by taking one reading of 16 bits per second. A pulse oximeter for measuring oxygen levels and pulse rate generates 4 bytes per second (32 bps) at 1 Hz sampling and 16-bit resolution. A temperature sensor generates 2 bytes per second (16 bits per second) at a sampling frequency of 1 Hz and uses 12 bits of resolution. Accelerometers/gyroscopes capture motion data with a resolution of 16 bits across three axes at a frequency of 50 Hz, generating 300 bytes of data every second (2.4 kbps). The EEG sensor, sampling at 500 Hz with a 24-bit resolution on 8 channels, is the most data-intensive, generating around 12,000 bytes per second (96 kbps). Thus, when all vital signs are combined, a single patient transmits approximately 106.4 kbps of data to the CN. The performance of the proposed model is compared with others. The list of simulation parameter is shown in 6.1

Table 6.1: Simulation Parameter for DD-QACA

Sr.No	**Parameter**	**Value**
1	Simulation Duration	1500 s
2	Deployment Area	500x500 m
3	Transmission Range	350 m
4	Number of Nodes	60
5	Data Packet Size	500 Bits
6	No. of Cluster Heads	4
7	No. of Simulations	10
8	Time Consumed	0 - 30 seconds
9	No. of Packets	0 - 50 PAckets
10	Density	10 - 110 Users
11	Control Message Size	100 bits
12	Duration of Data Periods	10 s
13	Cahnnel Capacity	120 kbps & 250 kbps
12	Sensing Radius	30 m
12	Maximum Packets in Queue	50

6.5.1 Buffer Loss Probability

Figures 6.2 show the performance of buffer loss probability under channel capacities of 250 kbps and 120 kbps mentioned above. By suppressing redundant sensor readings, DD-QACA reduces queue occupancy and delays buffer saturation, thereby achieving lower packet loss compared to congestion-unaware schemes. It minimizes buffer loss probability while prioritizing emergency packets. The additional 6.2 (a) shows the values of 36%, 43%, 15%, 9%, and 7% for the Analytical Model, DCCA, QACA-70, D-QACA-70, and the proposed DD-QACA, respectively, with 6 sensors attached per patient. In this case, 10 CNs per patient yield data from 60 sensors, resulting in increased packet size and a higher buffer loss ratio. In turn, 6.2 (b) depicts



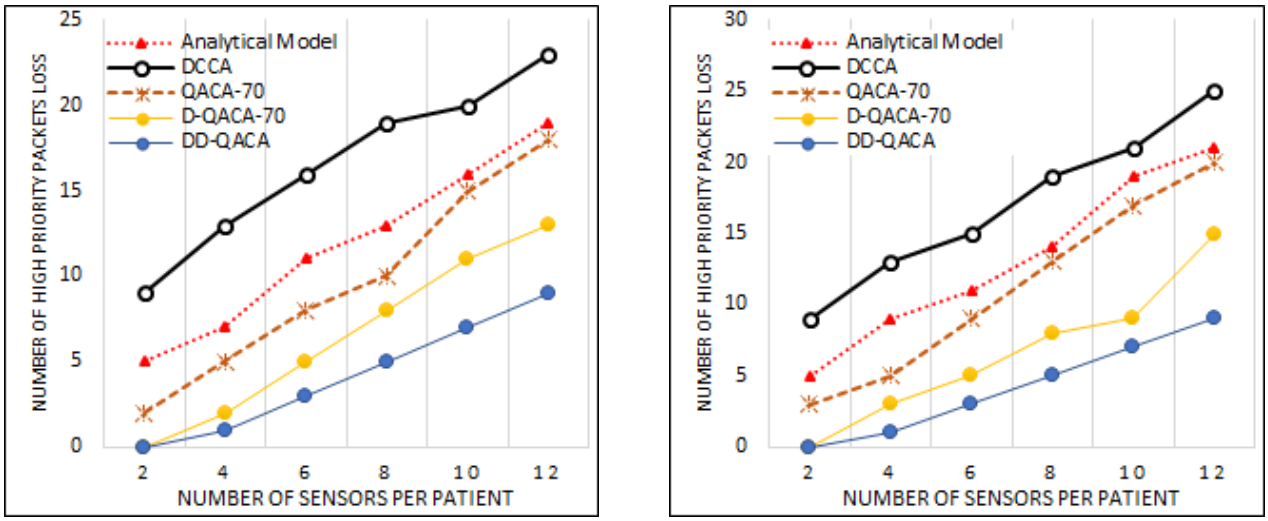
(**a**) (**b**)

Figure 6.2: Buffer Loss Probability for: (**a**) Intermediate nodes having channel capacity at 250 kbps; (**b**) represents the same for 120 kbps.

the buffer loss probability for intermediate nodes for a channel capacity of 120 kbps. Buffer loss probabilities in the Analytical Model, DCCA, QACA-70, and D-QACA-70 remain high due to the continuous transmission of duplicate sensor information to intermediate nodes. This constraint leads to excessive data accumulation in network buffers, thereby increasing the probability of packet loss. The proposed DD-QACA manages buffer filling by controlling the amount of packets. Full quantities are sent only in emergencies, and Boolean flag quantities are sent for unmodified data. Emergency messages are also sent to higher indexes or processed directly whenever queues become idle at intermediate nodes, restoring timeliness in the proposed system. When examining intermediate and leaf nodes, the removal of emergency packets becomes almost nil, although the queue is 90% full. Emergency packets are immediately assigned to the first index next to the current packet being processed, while packets are discarded from the tail of the queue.

6.5.2 High priority packet lost per second

Figure 6.3 shows high-priority packet loss under DD-QACA. The results indicate that redundancy suppression combined with queue-aware forwarding significantly improves the delivery reliability of emergency packets, especially during periods of heavy traffic. As seen in Figure 6.3(a), the Analytical Model is characterized by maximum packet loss, ranging from 5 packets at 2 nodes to 19 packets at 8 nodes and up to 23 packets at 12 nodes, owing to overheads



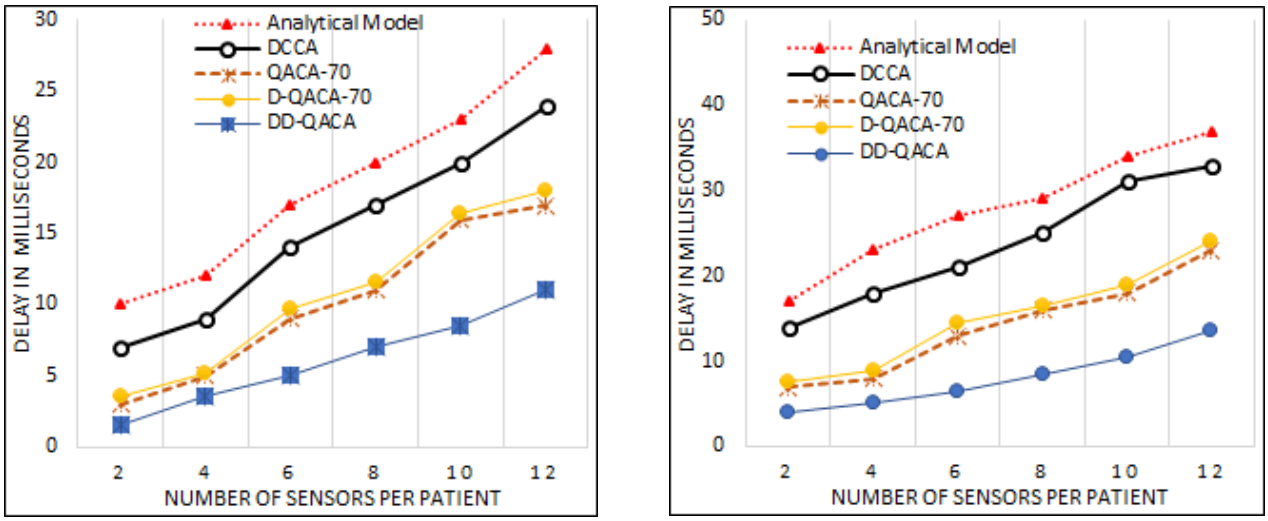
(**a**) (**b**)

Figure 6.3: High Priority Packet Loss: (**a**) Intermediate nodes having channel capacity at 250 kbps; (**b**) represents the same for 120 kbps.

from continuous transmission of full data and acknowledgments. Alongside this, DCCA also encounters considerable packet loss, starting from 9 packets at 2 nodes, increasing to 19 packets at 8 nodes, and peaking at 23 packets at 12 nodes. This happens because DCCA broadcasts full readings and sends acknowledgment only upon congestion. QACA-70, which optimizes acknowledgments, has lower packet losses than these models. Nevertheless, it loses 2 packets at 2 nodes, 10 packets at 8 nodes, and 18 packets at 12 nodes. D-QACA-70 reduces packet loss by granting priority to important data, resulting in 0 packet losses at 2 nodes, 8 packet losses at 8 nodes, and 13 packet losses at 12 nodes. The flag-based transmission mechanism, used in DD-QACA, makes it superior to others, reducing congestion and thereby keeping packet loss minimal: 0 packets at 2 nodes, 5 packets at 8 nodes, and 9 packets at 12 nodes. Similarly, 6.3(b) shows high-priority packet losses at intermediate nodes with a channel capacity of 120 kbps.

6.5.3 Packets Delay

Figure 6.4 presents packet delay at intermediate nodes for DD-QACA. Reduced traffic volume due to duplicate suppression leads to shorter queuing delays, allowing critical packets to be delivered with minimal latency even under constrained bandwidth conditions, at the intermediate nodes with the channel capacities of 250 kbps and 120 kbps. According to Figure 6.3(a), it can be concluded that the values for the Analytical Model, DCCA, QACA-70, D-QACA-70, and DD-QACA are 17, 14, 9, 9.7, and 5, respectively, when six sensors are attached to the patient's



(**a**) (**b**)

Figure 6.4: Packets Delay: (**a**) Intermediate nodes having channel capacity at 120 kbps; (**b**) represents the same for 250 kbps.

body. DD-QACA has consistently demonstrated a low level of delay, proving its efficiency. Figure 6.3 (b), on the other hand, depicts the same but with a channel capacity of 120 kbps at the intermediate node. The reduction in packet size proposed in the DD-QACA schemes would assist in the minimization of the transmission delays. Packet length is dynamically varied in DD-QACA compared to other methods, including analytical, DCCA, QACA-70, and D-QACA-70, by replacing common data with Boolean digits (0) and transmitting valid values only in emergencies. The above has great potential to reduce packet size, thereby speeding up data transmission and reducing queuing overhead time in the network. There is a relationship between the rate at which packets are reduced in size and the number of readings that change into Boolean values. If 50% of readings, for instance, are altered, then the overall packet size reduces to approximately 48%. A figure of 70% of readings replaced will result in 67%. The packets are therefore reduced in size to nearly 95%, where all normal readings have been replaced with Boolean values, thereby cutting communication overhead and significantly reducing packet delays. For instance, a normal method would require 32 bits (4 bytes) for each full data reading before transferring 10 sensor readings, which would amount to 320 bits in total (40 bytes). Replacing it with 50 percent reading reduced it to 165 bits (21), while 70 percent would further reduce it to 103 bits (13). Sending only Boolean values has the potential to bring the packet size down to 10 bits (2 bytes), hence having a great impact on reducing the transmission overhead.

6.5.4 Communication Cost

Communication cost quantifies the transmission overhead generated by the network during the simulation interval. Let N_{tx} denote the total number of packets transmitted by all nodes (including data and control packets, if applicable). The communication cost is defined as:

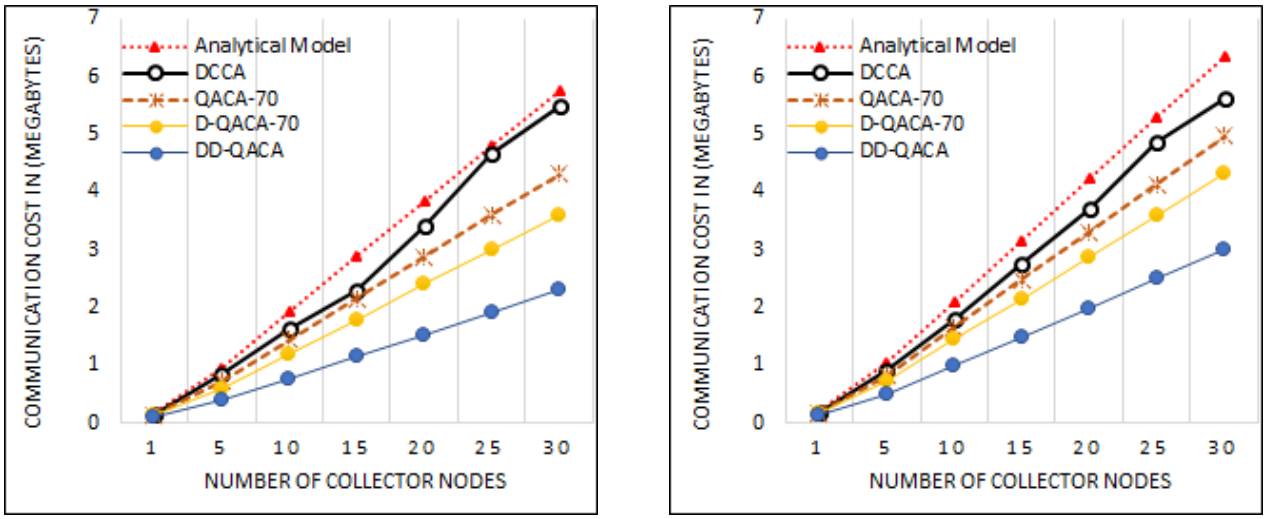
$$C_{\text{comm}} = N_{tx}. \tag{6.12}$$

If packet sizes vary, communication cost can be expressed in bytes. Let L_i denote the size (in bytes) of the i-th transmitted packet and let N_{tx} be the total number of transmitted packets. Then:

$$C_{\text{comm}}^{(\text{bytes})} = \sum_{i=1}^{N_{tx}} L_i. \tag{6.13}$$

A lower value of C_{comm} indicates reduced transmission overhead, which is especially important for energy efficiency and congestion mitigation in resource-constrained IoHT networks.

Figure 6.5 illustrates the communication cost in bytes for different schemes as the number of collector nodes is varied, with channel capacities of 250 kbps and 120 kbps. Figure 6.5(a) illustrates the intermediate node for the 250 kbps scenario, where lower costs are incurred during communication due to higher bandwidth, smoother delivery, and better data quality. The reason for the comparatively greater communication cost is that the analytical model broadcasts every piece of data, as well as the acknowledgment and status of the queues. For instance, it has a cost of 2.8 MB at 15 collector nodes for analysis models. DCCA also transmits all readings at a lower cost, which amounts to 2.3 MB. However, QACA-70 eliminates some acknowledgments without compromising full readings, at a reduced cost of 2.16 MB. D-QACA-70 indeed prioritizes high-priority packets, thereby reducing the cost to 1.8 MB. DD-QACA is the cheapest option since it sends flag bits during normal conditions and full data in emergency or time-lapsed situations, resulting in a cost of 1.15 MB. Figure 6.5(b) explains the 120 kbps case; the lesser bandwidth causes congestion. All the models, therefore, incur increased communication costs. 3.16 MB costs to the analytical model while DCCA goes behind it at a close distance of 2.78 MB. Further reducing the value of expenses is QACA-70, with just 2.48 MB, while D-QACA-70 continues to improve data transfer, having 2.16 MB. In fact, DD-QACA is the farthest and most accomplished, with its transmission cost amounting to just 1.49 MB. The proposed method thus reduces communication cost by simply sending flag bits during normal conditions, while sending all data, along with predetermined intervals, or only during emergencies, thereby reducing unnecessary transmissions. It is clear, however, that most methods, such as the Analytical Model



(**a**) (**b**)

Figure 6.5: Communication Cost: (**a**) for channel capacity at 120 kbps; (**b**) represents the same for 250 kbps.

and DCCA, transmit the entire data every time, exacerbating the costs through more frequent transmissions and acknowledgments. Besides them, QACA-70 and D-QACA-70 confirm the reduction in the number of acknowledgments while still giving high priority to full readings of important data, at heightened communication costs, compared to DD-QACA.

6.6 Discussion & Implications

The proposed study presents the DD-QACA scheme as an effective method for enhancing the efficiency of transmitting high-priority data in IoHT networks. This scheme significantly reduces buffer loss by replacing redundant sensor readings with flag values whenever measurements fall within the normal range. At the same time, it gives priority treatment to emergency packets either by placing them at the next available index in the queue or processing them immediately when no packet is under transmission. This design ensures that time-critical information reaches its destination without being delayed unnecessarily.

The DD-QACA approach saves bandwidth by suppressing the transmission of redundant data, thereby providing savings in storage at the collector node. Instead of sending full sensor values continuously, only a Boolean flag is sent under normal operating conditions. This approach becomes increasingly effective with the addition of more sensors. Compared to conventional models that suffer from congestion and higher packet loss, the proposed scheme exhibits a

remarkable gain in reliability. Results from simulations carried out on NS-2.35, with the TCL configuration defining node deployment, message initiation, and annotation of traces, demonstrate that the buffer loss probability for emergency packets is reduced by almost 90% compared to the baseline schemes. Compared to the previous method, it also results in a 95% reduction in the likelihood of emergency packet loss, thereby ensuring the secure transmission of vital health information to the collector node and enhancing the system's credibility in handling emergencies. In DD-QACA, suppression of redundant sensor readings reduces effective arrival rates, resulting in lower queue occupancy, reduced contention, and improved delay and reliability. These results confirm that the proposed mechanisms operate consistently with queueing theory principles while providing practical benefits for emergency-driven IoHT traffic.

Some limitations remain notwithstanding the aforementioned merits. Full sensor readings that need to be sent at preset intervals consume excessive power. While replacing duplicate values with Boolean flags greatly alleviates communication costs and energy bills, a certain amount of computation must be performed to create and manage such flags. However, the cost of this computation is negligible compared to the savings in communication resources. In general, thus, the DD-QACA mechanism not only enhances the performance of the network but also provides a truly dependable and resource-effective means to support patient-centric emergency data delivery in IoHT systems.

6.7 Summary

This chapter discusses the DD-QACA scheme, which aims to provide a faster and more secure transmission of highly prioritized healthcare data in IoHT networks. The proposed scheme implements a de-duplication mechanism to replace sensor readings bearing the same data with Boolean flags, thereby reducing communication overhead, conserving bandwidth, and minimizing storage requirements on the collector nodes. Emergency packets are prioritized by inserting them within the queue at higher indices or by processing them immediately if the channel is free, thereby assuring time-bound delivery of critical health information. The results indicated that the proposed scheme was able to reduce the buffer loss probability for emergency packets by 90% and the loss probability for emergency packets by 95% compared to the baseline schemes. These enhancements significantly enhance the overall reliability of the system in

emergencies. The method indeed incurs a minimal extra energy expenditure due to computation in flagging the Boolean replacement; however, this is almost negligible compared to the huge savings in communication resources. In brief, DD-QACA guarantees higher reliability, reduced congestion, and enhanced efficiency in patient-centered IoHT systems.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Overview

The research presented in this dissertation has targeted congestion control as a major problem in Internet of Health Things (IoHT) networks. The transmission of data in the IoHT networks must be reliable and timely, especially when monitoring patients' health conditions and responding to emergencies. With the introduction of IoT technologies in the healthcare field, a communication failure becomes an issue much larger than in other IoT applications. A delay or a loss of medical packets could mean failure in health, thus making research in congestion avoidance life-critical.

Through this research, three novel schemes were proposed and tested to solve the fundamental issues of IoHT data transmission: Queue-Aware Congestion Avoidance (QACA), Dual Queue-Aware Congestion Avoidance (DQACA), and De-Duplication Queue-Aware Congestion Avoidance (DD-QACA). Each of these innovative schemes was developed to address certain limitations of existing methodologies, which can progressively upgrade the reliability, responsiveness, and efficiency of the networks. While QACA itself offered core mechanics by drawing together residual queue capacity monitoring and analysis of acknowledgment frequency to predict and prevent congestion before it occurred, afterwards, this concept is built upon and extended to DQACA with its provision of a dual queue model that bases itself upon differentiating high-priority from low-priority packets, being even more concerned about emergency packets that need the speediest possible delivery. Finally, the DD-QACA swung an even wider net by adding level of redundancy, thus assisting with decreasing congestion due to unnecessary double

packets and maximizing queue effectiveness. The simulation results demonstrate that QACA reduces buffer loss by proactively regulating traffic, D-QACA significantly lowers emergency packet delay through severity-aware prioritization, and DD-QACA achieves additional gains by suppressing redundant transmissions, resulting in improved reliability, reduced latency, and enhanced energy efficiency in IoHT networks. These form a complete framework for making intelligent congestion control in IoHT. They push the state of the art by showing how the combination of acknowledgment control, queue differentiation, and duplicate elimination would offer more resilient and life-critical communication systems. Extensive simulation are performed to validate the performance of the proposed schemes.

7.2 Summary of Contributions

Followings are the main contributions of this research:

1. **Introduction of Queue-Aware Congestion Avoidance (QACA):** QACA is designed to identify congestion early through an active awareness mechanism that dynamically adjusts packet sending rate based on the condition of the residual queues, and send the controlled ACks. The knowledge of neighbor-queue allows QACA to assist nodes in adjusting their transmission rate in a dynamic manner, thereby preventing buffer overflow and avoiding retransmission storms. Simulation results show that QACA reduces average packet delay by up to 30% and buffer loss probability by approximately 40% under high traffic load conditions.

2. **Dual Queue-Aware Congestion Avoidance Scheme (D-QACA):** One of the limitations of QACA was that, although it managed congestion better, it treated all packets within one queue the same. This mechanism was inherited by DQACA, which introduced a dual-queue structure that distinguished high-priority emergency packets from normal packets. It ensures that emergency data always gets expedited. Unlike FIFO-based strategies, which delay even urgent packets behind routine transmissions, DQACA ensured that emergency data was always expedited. Simulation results show that this approach significantly reduces end-to-end delays for emergency traffic while maintaining fairness in low-priority traffic. Compared to conventional schemes, D-QACA achieves up to 50% reduction in emergency packet delay and approximately 60% reduction in emergency packet loss, ensuring timely

delivery of life-critical healthcare data during congestion.

3. **De-Duplication and Queue-usage Aware Congestion Avoidance Scheme (DD-QACA):** The proposed study presents the DD-QACA scheme as an effective method for enhancing the efficiency of transmitting high-priority data in IoHT networks. This scheme significantly reduces buffer loss by replacing redundant sensor readings with flag values whenever measurements fall within the normal range. At the same time, it gives priority treatment to emergency packets either by placing them at the next available index in the queue or processing them immediately when no packet is under transmission. Results indicate that DD-QACA reduces overall communication cost by up to 45%, while achieving nearly 70% reduction in buffer loss probability and improving energy efficiency of sensor nodes. Collectively, these quantitative improvements confirm the effectiveness of the proposed framework in enhancing reliability, latency, and scalability for Internet of Healthcare Things networks.

7.3 Limitation

The proposed congestion avoidance schemes are evaluated through extensive simulation under fixed node deployment and idealized channel conditions in order to ensure controlled and reproducible performance analysis. While this setup enables clear isolation of queueing and congestion effects, system behavior may vary in real-world environments characterized by node mobility, dynamic topology changes, and unpredictable wireless interference. In addition, the simulations assume stable sensor placement and synchronized reporting intervals, which may not always hold in practical hospital or home-care scenarios. The performance of the proposed schemes is also influenced by parameter settings such as queue threshold values, acknowledgment ratios, and redundancy suppression thresholds. Suboptimal tuning of these parameters could impact responsiveness and efficiency under extreme traffic conditions. Furthermore, the current evaluation does not explicitly model hardware-level constraints, such as sensor processing delays or battery degradation over time. Scalability under very large network sizes and heterogeneous device capabilities also requires further investigation. These limitations motivate future work toward adaptive, learning-based optimization mechanisms that can dynamically tune system parameters based on observed network and traffic conditions.

7.4 Summary of Proposed Schemes and Contributions

Table 7.1 presents a concise comparison of the proposed congestion avoidance schemes, highlighting their core mechanisms, primary contributions, and expected benefits. The progressive evolution from QACA to D-QACA and DD-QACA demonstrates how queue awareness is systematically extended to support emergency-sensitive prioritization and redundant data suppression, resulting in improved reliability, reduced latency, and enhanced energy efficiency in IoHT networks.

Table 7.1: Comparison of Proposed Congestion Avoidance Schemes

Scheme	**Key Mechanism**	**Main Contribution**	**Expected Benefit**
QACA	Queue-aware acknowledgment (ACK) feedback	Early congestion detection using residual queue information	Reduced buffer overflow and packet loss
D-QACA	Dual-queue architecture with emergency sensitivity	Intra-class emergency packet prioritization based on severity level	Lower delay and loss for emergency packets
DD-QACA	Redundant data suppression combined with queue awareness	Reduced communication overhead through duplicate reading elimination	Improved energy efficiency and scalability

7.5 Future Work and Open Challenges

7.5.1 Open Challenges

Following are the open challenges:

1. **Dynamic and Context-Aware Priority Assignment By Using ML Classifiers:** The present DQACA model has productively solved the congestion problem through dual queues for emergency and non-emergency packets, thereby not letting emergent information delay or get behind processed routine traffic. However, the existing design considers all emergency packets equal in urgency, which is practically untrue in many health care settings. Medical emergencies vary in degrees of severity; treating all of them the same may still result in critical delays. For instance, a packet reporting a case of cardiac arrest should be transmitted and processed with far greater immediacy compared to one reporting a moderate fever or a non-life-threatening condition. By treating both packets as

high-priority but equal, the system may risk making inefficient use of its emergency queue, where life-critical and less critical emergency data may still get delayed. Future work should extend the DQACA model to incorporate a dynamic and context-aware priority assignment framework to address this limitation. This framework would not only separate emergency traffic from non-emergency traffic but also categorize emergency packets into multiple levels of criticality, ensuring that life-threatening data is always processed first during transmission.

This development can be accomplished through integrating machine learning algorithms into the health data context, which will learn from patient vitals, historical medical records, and environmental conditions. In doing so, the system will automatically assess the severity of each packet in real-time, assign a priority score, and then enable dynamic reordering of packets within an emergency queue. That means those most urgent medical conditions would receive the fastest response, while lower priority emergency and non-emergency traffic are still treated fairly. Thus, this approach represents a significant advancement towards making IoHT congestion control not only queue-aware but also context-aware, aiming to match the real-world criticality of medical events as closely as possible, thereby improving network performance and reducing the risk of delayed interventions in life-threatening scenarios.

2. **Privacy-Preserving Queue Status Sharing through Federated Learning:** Another vital area for consideration in future work is how to facilitate the secure sharing of queue status and acknowledgment information among QACA. The recent work has already discussed some aspects of acknowledgment frequency, in order to avoid flooding with redundant ACK traffic; however, this still leaves a challenging open question of how best to design such exchanges in terms of content and confidentiality. In the application domains of IoHT systems where acknowledgment processes and queue status indicators augment reliability, these very signals, at the same time, act as tacit carriers of sensitive medical and network-related information. For example, a sudden flood of ACKs or queue updates from a certain patient device would reveal information about a possible medical emergency that is unwarrantedly sensitive to privacy. Therefore, enhancements to QACA could entail the future addition of federated learning (FL)-based frameworks that allow queue dynamics and acknowledgment strategy learning in a distributed manner, without exposing raw data in an exchange. Rather than explicitly sharing queue states or acknowledgment histories,

each node could train a lightweight predictive model locally using its traffic and buffer observations and then transmit only encrypted model updates to a global aggregator. Such an arrangement would ensure secrecy since no raw queue or ACK information would ever leave the local device. Such a federated approach would also instill an adaptive intelligence in acknowledgment handling. By learning from distributed patterns, nodes could distinguish when ACKs were warranted, when congestion could arise from queue conditions, and when acknowledgment scenarios could be handled by silence, such as aggregated or delayed confirmation. This would provide an additional channel for minimizing unnecessary signaling overhead while ensuring reliability. Furthermore, embedding privacy-preserving protocols, such as secure aggregation or differential privacy, into the federated updates can provide formal guarantees that no individual patient information and no information about any traffic behavior of a specific node can be gleaned through the shared updates. Thus, QACA queue-aware congestion control, combined with the open prospect of federated learning, offers a promising approach towards privacy-preserving reliability in IoHT. Future implementations could thus investigate fine-tuning the trade-off between acknowledgment minimalism, predictive intelligence, and secrecy preservation, so that IoHT networks remain robust and clinically trustworthy even under large-scale deployment.

3. **Advancing Redundancy Elimination through ML and XAI Beyond DD-QACA:** The redundancy of packets has already been resloved in the DD-QACA scheme, where de-duplication techniques limit unnecessary replication, ensuring that identical or irrelevant packets are not sent multiple times in the network. This has achieved a substantial reduction in buffer congestion and performance enhancements across the system. However, although DD-QACA has promising capabilities at the fundamental level, its redundancy elimination method largely comprises mere rules without efforts to tap the power of predictive and adaptive features brought forth by current Artificial Intelligence techniques.

7.5.2 Future Work

Future advancements are possible in its extension through Machine Learning (ML) methods, which will make redundancy detection more dynamic and context-sensitive. Unlike fixed heuristics, ML algorithms—such as clustering, anomaly detection, or deep learning

models—continuously learn patterns from real-time sensor data streams, detecting not only obvious duplicates but also semantically redundant data (e.g., repetitive readings with negligible variations that do not impact clinical decision-making). This will enhance efficiency in cases where multiple heterogeneous sensors transmit correlated signals that may not be identical byte for byte but redundantly represent the same clinical information (e.g., ECG patches, smartwatches, chest straps). Therefore, ML can be supplemented with Explainable Artificial Intelligence (XAI) frameworks in healthcare. XAI can clearly explain to clinicians and administrators why a packet was dropped, merged, or de-prioritized. For example, an XAI module can explain that two ECG packets were 95% similar in waveform patterns, and forwarding both would not affect the diagnosis. This interoperability produce trust in the system where accountability in data handling is a crucial consideration. As DD-QACA already regulates redundancy effectively, future integration with ML and XAI can transform redundancy management from a static, rules-driven approach to a smart, adaptive, and explainable process. It will ensure efficiency in the network while maintaining clinical dependability, ensuring that the most critical health records can be transferred without creating queues and compromising patient safety. Future work may also explore interoperability of the proposed schemes with existing IoT protocol stacks such as RPL and 6LoWPAN, as well as deployment constraints in hospital environments including regulatory compliance, device heterogeneity, and robustness to parameter sensitivity.

BIBLIOGRAPHY

[1] M. C. Domingo, “An overview of the internet of things for people with disabilities,” *journal of Network and Computer Applications*, vol. 35, no. 2, pp. 584–596, 2012. 1

[2] K. Ashton *et al.*, “That ‘internet of things’ thing,” *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009. 1

[3] F. Wortmann and K. Flüchter, “Internet of things: technology and value added,” *Business & information systems engineering*, vol. 57, no. 3, pp. 221–224, 2015. 1

[4] R. Parashar, A. Khan, and A. Neha, “A survey: The internet of things,” *International Journal of Technical Research and Applications*, vol. 4, no. 3, pp. 251–257, 2016. 2

[5] S. Li, “"the internet of things: a survey,” *" Information systems frontiers 17: 243-259.*, 2015. 2

[6] L. Catarinucci, D. De Donno, L. Mainetti, L. Palano, L. Patrono, M. L. Stefanizzi, and L. Tarricone, “An iot-aware architecture for smart healthcare systems,” *IEEE internet of things journal*, vol. 2, no. 6, pp. 515–526, 2015. 2

[7] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, “The internet of things for health care: a comprehensive survey,” *IEEE access*, vol. 3, pp. 678–708, 2015. 3, 45

[8] M. Wazid, A. K. Das, J. J. Rodrigues, S. Shetty, and Y. Park, “Iomt malware detection approaches: analysis and research challenges,” *IEEE access*, vol. 7, pp. 182 459–182 476, 2019. 3

[9] F. Alsubaei, “"iomt-saf: Internet of medical things security assessment framework,” *" Internet of Things 8: 100123.*, 2019. 3

[10] S. M. R. Al Masud, “"study and analysis of scientific scopes, issues and challenges towards developing a righteous wireless body area network,” *" International Journal Of Soft Computing And Engineering (IJSCE) 3(2).*, 2013. 3

[11] M. Chen, “"body area networks: A survey," *" Mobile networks and applications 16: 171-193.*, 2011. 3

[12] A. M. Ahmed and R. Paulus, “"congestion detection technique for multipath routing and load balancing in wsn," *" Wireless Networks 23: 881-888.*, 2017. 3

[13] L. Filipe, “"wireless body area networks for healthcare applications: Protocol stack review," *" International Journal of Distributed Sensor Networks 11(10): 213705.*, 2015. 3

[14] L. Song, Y. Wang, J.-J. Yang, and J. Li, “Health sensing by wearable sensors and mobile phones: A survey,” in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2014, pp. 453–459. 3

[15] S. Beitelspacher, M. Mubashir, K. M. Besher, and M. Z. Ali, “Prioritizing health care data traffic in a congested iot cloud network,” in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2020, pp. 1–6. 3

[16] L. Haoyu, “"an iomt cloud-based real time sleep apnea detection scheme by using the spo2 estimation supported by heart rate variability," *" Future Generation Computer Systems 98: 69-77.*, 2019. 3

[17] D. Hemalatha and B. E. Afreen, “"development in rfid (radio frequency identification) technology in internet of things (iot)," *" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 4(11): 4030-4038.*, 2015. 3

[18] A. Firdausi, “"overview the internet of things (iot) system security, applications, architecture and business models," *" Indonesia: Universitas of Electrical Engineering: 1-8.*, 2016. 4

[19] N. Yaakob and I. Khalil, “A novel congestion avoidance technique for simultaneous real-time medical data transmission,” *IEEE journal of biomedical and health informatics*, vol. 20, no. 2, pp. 669–681, 2015. 4

[20] R. Bhalerao, S. S. Subramanian, and J. Pasquale, “An analysis and improvement of congestion control in the coap internet-of-things protocol,” in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2016, pp. 889–894. 5

[21] M. Quwaider and Y. Shatnawi, “"neural network model as internet of things congestion control using pid controller and immune-hill-climbing algorithm," *" Simulation modelling practice and theory 101: 102022.*, 2020. 6

[22] A. Betzler, C. Gomez, I. Demirkol, and J. Paradells, “Coap congestion control for the internet of things,” *IEEE Communications Magazine*, vol. 54, no. 7, pp. 154–160, 2016. 6, 16

[23] R. Hashemzehi, “"congestion in wireless sensor networks and mechanisms for controlling congestion," *" Indian Journal of Computer Science and Engineering (IJCSE) 4(3).*, 2013. 6

[24] A. Bohloulzadeh and M. Rajaei, “"a survey on congestion control protocols in wireless sensor networks," *" International Journal of Wireless Information Networks 27: 365-384.*, 2020. 6

[25] S. H. Han and Y.-B. Ko, “A survey on internet of things and fog computing for healthcare,” *Future Generation Computer Systems*, vol. 78, pp. 641–658, 2019. 7

[26] M. C. Domingo, “An overview of the internet of things for people with disabilities,” *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 584–596, 2012. 7

[27] X. Liu, W. Yu, T. Dillon, and W. Jia, “A priority-based congestion control protocol for critical iot services,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3376–3385, 2018. 7, 8, 38

[28] A. Yousef, K. A. Darabkh, and R. R. Darabkh, “An energy-efficient congestion control protocol for iot healthcare applications,” *IEEE Access*, vol. 8, pp. 123 202–123 222, 2020. 7, 8

[29] M. A. Rahman and R. Buyya, “Fog computing for iot-based healthcare: Promises, challenges, and future directions,” *IEEE Access*, vol. 7, pp. 45 936–45 956, 2019. 7

[30] A. H. Sodhro, A. K. Sangaiah, S. Pirbhulal, and Z. Luo, “Energy-efficient adaptive transmission in wireless body area networks for healthcare applications,” *Computers and Electrical Engineering*, vol. 67, pp. 362–375, 2018. 7

[31] O. Alrawi and C. Chio, “Iot in healthcare: Security and privacy considerations,” *IEEE Internet of Things Magazine*, vol. 2, no. 2, pp. 66–71, 2019. 7

[32] A. M. Ahmed and R. Paulus, “Congestion detection technique for multipath routing and load balancing in wsn,” *Wireless Networks*, vol. 23, pp. 881–888, 2017. 16

[33] K. S. Bhandari, “"coar: Congestion-aware routing protocol for low power and lossy networks for iot applications,” *" Sensors 18(11): 3838.*, 2018. 16

[34] M. Aboubakar, P. Roux, M. Kellil, and A. Bouabdallah, “A novel scheme for congestion notification in iot low power networks,” in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 932–937. 17

[35] T. Rahman, X. Yao, and G. Tao, “Consistent data collection and assortment in the progression of continuous objects in iot,” *IEEE Access*, vol. 6, pp. 51 875–51 885, 2018. 17

[36] L. P. Verma, “"an adaptive congestion control algorithm,” *" MMC_A 92(1): 30-36.*, 2019. 17

[37] L. P. Verma and M. Kumar, “"an iot based congestion control algorithm,” *" Internet of Things 9: 100157.*, 2020. 17

[38] Y. Chen, “"adaptive method for packet loss types in iot: an naive bayes distinguisher,” *" Electronics 8(2): 134.*, 2019. 17

[39] S. Zhuo, H. Shokri-Ghadikolaei, C. Fischione, and Z. Wang, “Online congestion measurement and control in cognitive wireless sensor networks,” *IEEE Access*, vol. 7, pp. 137 704–137 719, 2019. 18

[40] F. Banaie, M. H. Yaghmaee, S. A. Hosseini, and F. Tashtarian, “Load-balancing algorithm for multiple gateways in fog-based internet of things,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7043–7053, 2020. 18

[41] N. Akhtar, “"congestion avoidance for smart devices by caching information in manets and iot,” 2019. 18

[42] R. Anitha, S. Kumar, and P. Reddy, “A comprehensive survey on congestion detection signals and locations in low-power iot networks,” *Journal of Network and Computer Applications*, vol. 235, p. 103723, 2024. 18

[43] J. Lim, “A survey of congestion detection mechanisms for rpl-based 6lowpan iot networks,” *Sensors*, vol. 19, no. 21, p. 4698, 2019. 19

[44] J. Kim, S. Park, and H. Lee, “Qu-rpl: Queue utilization based congestion detection for rpl networks,” in *Proceedings of the IEEE International Conference on Embedded and Ubiquitous Computing*, 2015, pp. 144–151. 19

[45] H. Al-Kashoash and A. Kemp, “Congestion-aware objective function for rpl-based low-power and lossy networks,” *IEEE Sensors Journal*, vol. 18, no. 1, pp. 378–389, 2018. 19

[46] R. Sukjaimuk, “"adaptive congestion control in information-centric networking for the iot sensor network,” *" Journal of Advanced Simulation in Science and Engineering 5(1): 17-28.*, 2018. 19

[47] S. Bhandari and S. Moh, “Coar: Congestion avoidance and detection using multi-metric fusion in iot,” *IEEE Access*, vol. 6, pp. 85 053–85 067, 2018. 20

[48] H. Kim, J. Lee, and S. Park, “Pc-rpl: Loss and collision aware congestion detection for rpl,” *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5128–5137, 2018. 20

[49] A. Marco, M. Rossi, and A. Zanella, “Mac-layer busy channel probability for congestion detection in iot,” *Ad Hoc Networks*, vol. 94, p. 101942, 2019. 20

[50] M. Farooq and D. Pesch, “Available bandwidth estimation for congestion detection in wireless sensor networks,” in *Proceedings of the IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–6. 21

[51] H. Al-Kashoash and A. Kemp, “Gtcc: Game-theoretic congestion control and detection for low-power networks,” *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 760–771, 2017. 21

[52] N. Yuvaraj and G. Saravanan, “"markov transition and smart cache congestion control for iot enabled wireless mesh networks,” *" Peer-To-Peer Networking and Applications 14(1): 58-68.*, 2021. 21

[53] R. Hamidouche, “"an efficient clustering strategy avoiding buffer overflow in iot sensors: A bio-inspired based approach,” 2019. 21

[54] M. Shafiq, “"handshake sense multiple access control for cognitive radio-based iot networks,” *" Sensors 19(2): 241.*, 2019. 22

[55] P. Borole, “"6g network access and edge-assisted congestion rule mechanism using software-defined networking,” *" International Journal of Future Generation Communication and Networking 13(1s): 107-112.*, 2020. 22, 45

[56] S. R. Pokhrel, “"adaptive admission control for iot applications in home wifi networks,” 2019. 22

[57] B. Pillai, “A model for early detection and avoidance of congestion in manet, easychair.” 2020. 22

[58] C. Wan, S. Eisenman, and A. Campbell, “Coda: Congestion detection and avoidance in sensor networks,” *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2003, early cooperative congestion avoidance for WSNs; backpressure + sink control. 23

[59] R. Sankar *et al.*, “Esrt: Event-to-sink reliable transport for wireless sensor networks,” *ACM/IEEE IPSN*, 2003, sink-driven reliability control with source rate adjustment. 23

[60] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: A scalable and robust communication paradigm for sensor networks,” *ACM/ACM Transactions on Sensor Networks / SIGCOMM venues*, 2002, aggregation and in-network processing to reduce traffic. 23

[61] V. Authors, “Priority-based queueing mechanisms for critical iot applications,” *Journal of Network and Computer Applications*, 2016, survey and implementations of class-based service to protect critical flows. 23

[62] S. Borasia and V. Raisinghani, “A review of congestion control mechanisms for wireless sensor networks,” 2011. 23

[63] C. Sergiou, “"a comprehensive survey of congestion control protocols in wireless sensor networks,” 2014. 23

[64] C. Pruthvi, “"a systematic survey on content caching in icn and icn-iot: Challenges, approaches and strategies,” *" Computer Networks 233: 109896.*, 2023. 24

[65] R. Sukjaimuk, “"a smart congestion control mechanism for the green iot sensor-enabled information-centric networking,” *" Sensors 18(9): 2889.*, 2018. 24

[66] M. Nasimi, “Edge-assisted congestion control mechanism for 5g network using software-defined networking,” 2018. 25

[67] R. Mogi, “Load balancing method for iot sensor system using multi-access edge computing,” 2018. 25

[68] H.-C. Hsieh, “"mobile edge computing platform with container-based virtualization technology for iot applications,” *" Wireless Personal Communications 102: 527-542.*, 2018. 25

[69] S. B. Baker, “"internet of things for smart healthcare: Technologies, challenges, and opportunities,” 2017. 26, 38

[70] D. Singh, “Ambient assisted living technologies from the perspectives of older people and professionals,” 2017. 26, 38

[71] T. Kleinberger, “"ambient intelligence in assisted living: Enable elderly people to handle future interfaces,” *"*, 2007. 27

[72] D. Calvaresi, “"exploring the ambient assisted living domain: a systematic review,” *" Journal of Ambient Intelligence and Humanized Computing 8: 239-257.*, 2017. 27

[73] M. Bhatia and S. K. Sood, “"temporal informative analysis in smart-icu monitoring: M-healthcare perspective,” *" Journal of medical systems 40: 1-15.*, 2016. 27, 38

[74] A. Rghioui and A. Oumnad, “"challenges and opportunities of internet of things in healthcare,” *" International Journal of Electrical & Computer Engineering (2088-8708) 8(5).*, 2018. 28

[75] S. I. Ahmad, S. K. Kuri, C. K. Garhwal, and C. K. Garhwal, “A review of different congestion management mechanisms for remote healthcare monitoring system in iot domain.” 28, 38

[76] K. M. Awan, “"a priority-based congestion-avoidance routing protocol using iot-based heterogeneous medical sensors for energy efficiency in healthcare wireless body area networks," *" International Journal of Distributed Sensor Networks 15(6): 1550147719853980.*, 2019. 28

[77] M. H. Kashani, “"a systematic review of iot in healthcare: Applications, techniques, and trends," *" Journal of Network and Computer Applications 192: 103164.*, 2021. 29

[78] J. L. Shah and H. F. Bhat, “"cloudiot for smart healthcare: architecture, issues, and challenges," *" Internet of things use cases for the healthcare industry: 87-126.*, 2020. 29

[79] Z. N. Aghdam, “"the role of the internet of things in healthcare: Future trends and challenges," *" Computer methods and programs in biomedicine 199: 105903.*, 2021. 29

[80] M. Arafat *et al.*, “Qqar: A q-learning based qos-aware routing protocol for iomt and wbans,” *IEEE Transactions on Network Science and Engineering*, 2024. 29, 39

[81] S. Altowaijri *et al.*, “Deduplication-aware healthcare data distribution in iomt,” *Future Generation Computer Systems*, vol. 152, pp. 355–367, 2024. 30

[82] A. Stitini *et al.*, “Combining iomt and explainable ai for enhanced triage: An mqtt broker approach,” in *2024 IEEE International Conference on Smart Healthcare*. IEEE, 2024, pp. 88–95. 30, 40

[83] I. de Morais Barroca Filho and G. S. de Aquino Junior, “Iot-based healthcare applications: a review,” 2017. 30

[84] K. Alatoun, “"a novel low-latency and energy-efficient task scheduling framework for internet of medical things in an edge fog cloud system," *" Sensors 22(14): 5327.*, 2022. 30

[85] M. N. Bhuiyan, “"internet of things (iot): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities," 2021. 31

[86] N. Akshatha *et al.*, “Priority-enabled mqtt protocol for emergency event handling in healthcare iot,” *International Journal of Communication Systems*, 2024. 31

[87] S. Ouakasse *et al.*, “Qos-gateway: Prioritizing emergency data in iomt networks,” *Computer Networks*, vol. 237, p. 110009, 2024. 31, 39

[88] H. Verma, N. Chauhan, and L. K. Awasthi, "Modelling buffer-overflow in 6lowpan-based resource-constraint iot-healthcare network," *Wireless Personal Communications*, vol. 129, no. 2, pp. 1113–1128, 2023. 31, 46, 47

[89] P. Anitha, "Pqtba: Priority queue based token bucket algorithm for congestion control in iot network," 2023. 32, 39, 46, 47

[90] H. Mazloomi *et al.*, "A priority-based congestion avoidance scheme for healthcare wireless sensor networks," *IET Wireless Sensor Systems*, vol. 12, no. 3, pp. 163–172, 2022. 32, 38

[91] R. Buenrostro-Mariscal *et al.*, "Qccp: A prioritization-driven congestion control protocol for internet of medical things," *IEEE Access*, vol. 11, pp. 145 987–145 998, 2023. 32, 39

[92] L.-M. Tseng, P.-F. Chen, and C.-Y. Wen, "Design of edge-iomt network architecture with weight-based scheduling," *Sensors*, vol. 23, no. 20, p. 8553, 2023. 33, 40, 60

[93] B. K. Asingwire, "Performance analysis of delay and size-dependent scheduling for iot-based healthcare traffic using heterogeneous multi-server priority queueing system," 2024. 33, 39

[94] P. Chanak and I. Banerjee, "Congestion free routing mechanism for iot-enabled wireless sensor networks for smart healthcare applications," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 3, pp. 223–232, 2020. 34, 39, 46, 47, 70, 89

[95] S. Tripathy *et al.*, "Sdn-fog orchestration for queue- and resource-aware scheduling in wban healthcare," *Internet of Things*, vol. 25, p. 100–118, 2024, fog-assisted scheduling to reduce congestion. 34, 39

[96] T. Nguyen *et al.*, "Leveraging priority queueing in iot-edge-fog-cloud infrastructures for real-time healthcare," *Journal of Network and Computer Applications*, vol. 234, pp. 103–120, 2025. 34

[97] R. Al-azzawi and A. Al-mahdawi, "An in-depth study on controlling congestion during massive traffic in wireless sensor networks," *International Journal of Communication Systems*, vol. 38, no. 2, p. e5412, 2025. 34

[98] S. Nair and P. Kumar, "Patient satisfaction and health service queuing: A virtual ioht perspective," *Health Policy and Technology*, vol. 14, p. 100821, 2025. 34

[99] A.-E. R. Group, “Ai-driven framework for mitigating network congestion in iot-based healthcare systems,” in *2025 International Conference on Advanced Computing (ICAC)*, 2025, pp. 210–215. 35

[100] D. Efrosinin *et al.*, “Analysis of machine learning in gi/g/k queuing systems for ioht data bursts,” *Operations Research Letters*, vol. 53, pp. 107–115, 2025. 35

[101] I. Peshkova *et al.*, “Accelerated simulation of medical queues using splitting-based regenerations,” *Simulation Modelling Practice and Theory*, vol. 138, p. 102850, 2025. 35

[102] K. Kawanishi and Y. Ino, “Performance analysis of hybrid ioht systems with setup time and quasi-birth-and-death processes,” *Journal of Computational and Applied Mathematics*, vol. 452, p. 115890, 2025. 35

[103] E. Research, “The scone protocol: Throughput advice for low-latency ioht video traffic,” Ericsson Research Labs, Tech. Rep. TR-2025-09, 2025. 35

[104] L. Pappone, A. Sacco, and F. Esposito, “Mutant: Learning congestion control from existing protocols via online reinforcement learning,” in *Proceedings of the 22nd USENIX Symposium on Networked Systems Design and Implementation (NSDI ’25)*, Philadelphia, PA, USA, April 2025. 36

[105] A. H. Oleiwi, “Artificial intelligence approaches to mitigating network congestion in iot systems,” *Babylonian Journal of Artificial Intelligence*, vol. 2025, pp. 117–127, 2025. 36

[106] M. F. Ali, M. S. Mohmood, B. S. Shukur *et al.*, “Hcap: Hybrid cyber attack prediction model for securing healthcare applications,” *PLOS One*, vol. 20, no. 5, p. e0321941, 2025. 36

[107] N. R. Collective, “Machine learning approaches for active queue management: A survey, taxonomy, and future directions,” *arXiv preprint*, 2025, arXiv:2410.02563 [cs.NI]. 36

[108] H. Verma, “"modelling buffer-overflow in 6lowpan-based resource-constraint iot-healthcare network,” *" Wireless Personal Communications 129(2): 1113-1128.*, 2023. 39, 70, 89

[109] T. Saikumari and G. V. Sudha, “"an improved congestion handling in blockchain secured cloud based healthcare system,” *" International Journal of Intelligent Engineering & Systems 17(2).*, 2024. 40

[110] G. Lee, W. Saad, and M. Bennis, “An online optimization framework for distributed fog network formation with minimal latency,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2244–2258, 2019. 41

[111] G. Ghosh, “"secure surveillance systems using partial-regeneration-based non-dominated optimization and 5d-chaotic map,” *" Symmetry 13(8): 1447.*, 2021. 41

[112] M. Kumar, “"a modified ga-based load balanced clustering algorithm for wsn: Mgalbc,” *" International Journal of Embedded and Real-Time Communication Systems (IJERTCS) 12(1): 44-63.*, 2021. 42

[113] P. Rani, “"robust and secure data transmission using artificial intelligence techniques in ad-hoc networks,” *" Sensors 22(1): 251.*, 2021. 42

[114] S. El-Sappagh, “"multimodal multitask deep learning model for alzheimer’s disease progression detection based on time series data,” *" Neurocomputing 412: 197-215.*, 2020. 42

[115] A. Haleem, “"medical 4,” *0 technologies for healthcare: Features, capabilities, and applications." Internet of Things and Cyber-Physical Systems 2: 12-30.*, 2022. 43

[116] G. L. Tortorella, “"healthcare 4,” *0: trends, challenges and research directions." Production Planning & Control 31(15): 1245-1260.*, 2020. 43

[117] A. Chandrasekaran *et al.*, “Average delay-based early congestion detection in internet of health things systems,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2023. 44

[118] K. Sharma, “Internet of healthcare things security vulnerabilities and jamming attack analysis,” *Expert Systems*, vol. 39, no. 1, p. e12853, 2022, investigates packet-loss ratios under jamming conditions in IoHT networks. 44

[119] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): A vision, architectural elements, and future directions,” *Future Generation Computer Systems*,

vol. 29, no. 7, pp. 1645–1660, 2013, discusses IoT architectures including medical data acquisition and the challenges of extracting useful physiological information. 45

[120] C. Li, J. Wang, S. Wang, and Y. Zhang, "A review of iot applications in healthcare," *Neurocomputing*, vol. 565, p. 127017, 2024. 59

[121] M. S. H. Talpur *et al.*, "Illuminating healthcare management: A comprehensive review of iot-enabled chronic disease monitoring," *IEEE Access*, 2024. 59

APPENDIX

PUBLICATION DETAILS

The following table presents a summary of the publications, including the paper title, authors, status (Published, Accepted, or Revision Submitted), Impact Factor (IF), and HEC category.

Table A.1: List of Publications

Sr#	Authors	Paper Title	Journal Name	Status	HEC Category	Impact Factor (IF)
1	Muhammad Zafarullah, Ata Ullah, Fazli Subhan, Sajjad A. Ghauri	Queue-Aware Congestion Avoidance in IoHT: Enabling Future Integration With Large Models for Transmission Optimization	Internet Technology Letter	Published	X	0.9
2	Muhammad Zafarullah, Ata Ullah, Zohaib Ahmad, Fazli Subhan	LIGHTWEIGHT ML-DRIVEN QUEUE-AWARE CONGESTION AVOIDANCE FOR REAL-TIME HEALTHCARE DATA SHARING IN IOT	Computing and Informatics	Accepted	W	1.7
3	Muhammad Zafarullah, Ata Ullah, Sajjad A. Ghauri, Nauman Anwar Baig	Redundant Packets Removal Queue-Aware Congestion Avoidance Scheme for IoT Based Healthcare Applications	Cost Effectiveness and Resource Allocation	Revision Submitted	W	2.5
4	Muhammad Zafarullah, Ata Ullah, Sajjad A. Ghauri, Nauman Anwar Baig	Emergency Packets Handling in Queue-Aware Congestion Avoidance Schemes in IoHT	PLoS One	Revision Submitted	W	2.6