IntechOpen

Quality of Service (QoS)

Challenges and Solutions

Edited by Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily

Quality of Service (QoS) - Challenges and Solutions

Edited by Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily

Published in London, United Kingdom

Quality of Service (QoS) - Challenges and Solutions
http://dx.doi.org/10.5772/intechopen.1002087
Edited by Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily

Contributors
Abdulmajeed Al-Jumaily, Bakhytzhan Kulambayev, Deon P. Du Plessis, Jose Costa-Requena, Khomotso C. Tsoane, Mika Skarp, Mirzakulova Sharafat, Mohamadou Ndiaye, Puseletso Sebothoma, Rajesh K. Yadav, Samiullah Mehraban, Tahira Mahboob, Topside Ehleketani Mathonsi, Umair Shafiq Khan, Yaseein Soubhi Hussein

First published in London, United Kingdom, 2025 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Quality of Service (QoS) - Challenges and Solutions
Edited by Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily
p. cm.
Print ISBN 978-0-85014-918-0
Online ISBN 978-0-85014-917-3
eBook (PDF) ISBN 978-0-85014-919-7

If disposing of this product, please recycle the paper responsibly.

Meet the editor

Dr. Yaseein Soubhi Hussein is currently an associate professor in the Department of Information Systems and Computer Science at Ahmed Bin Mohammed Military College (ABMMC) in Qatar. He obtained his Ph.D. in Communication and Networks Engineering from the Universiti Putra Malaysia (UPM), Malaysia, in 2014; an M.Eng. in Telecommunications Engineering from the University of Malaya (UM), Malaysia, in 2010; and a BSc in Electrical Engineering from University of Baghdad (UOB), Iraq, in 1999. He previously worked as a senior lecturer in the Faculty of Computing, Engineering and Technology at the Asia Pacific University of Technology & Innovation (APU) in Malaysia. Dr. Hussein was a consultant for a project on visible light communication (VLC) with the Malaysian telecommunications company Telekom Malaysia Berhad and a postdoctoral research fellow in the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. He is also a senior member of the Institute of Electrical and Electronics Engineers (IEEE). His main research interests are 5G and 6G wireless networks, network optimization, VLC (known as Li-Fi), and resource allocation.

Abdulmajeed Al-Jumaily is a distinguished wireless communications professional with extensive mobile and satellite communications expertise. He earned his BSc in Electronics and Communication Engineering in 2003, an MSc in Wireless Communications Engineering from Universiti Putra Malaysia (UPM) in 2014, and a Ph.D. in Wireless Communication and Networks Engineering from UPM in 2022. Following his Ph.D., Dr. Al-Jumaily undertook a research fellowship through the Erasmus+ KA107 exchange program at Universidad Carlos III de Madrid (UC3M), Spain. He is a postdoctoral researcher in the Signal Theory and Communications Department (DTSC) at UC3M. In parallel to this role, he was a research engineer at DEKRA, Spain, where he worked on advanced projects in wireless communications. Dr. Al-Jumaily is a senior member of the Institute of Electrical and Electronics Engineers (IEEE).

Contents

Preface

The relentless growth of digital technologies and the increasing reliance on network connectivity have placed unprecedented demands on network performance. Ensuring Quality of Service (QoS) has become a concern for network operators and service providers worldwide. Modern networks must accommodate various applications and services, from real-time streaming to mission-critical data transmission, all while ensuring optimal performance, security, and reliability. At the heart of addressing these demands lies QoS—a critical concept that ensures networks meet performance requirements across diverse environments.

Quality of Service (QoS) – Challenges and Solutions delves into the multifaceted aspects of QoS, offering readers an in-depth exploration of QoS, addressing its theoretical foundations, implementation strategies, and practical applications. It is organized into two sections offering insights into the challenges and solutions for optimizing network performance.

Section 1: "Challenges and Fundamentals in Quality of Service (QoS)"

This section provides a deep dive into QoS's core concepts, principles, and challenges. It addresses the architectural complexities, management strategies, and design principles that underpin effective QoS implementation. Readers will gain an understanding of the challenges posed by hybrid networks, virtualization, and the layered nature of network architectures.

- Chapter 1: "Introductory Chapter: Challenges and Solutions in Quality of Service (QoS) – Optimizing Network Performance"

This chapter introduces the concept of QoS, exploring its importance and the challenges faced in maintaining performance in modern networks. It sets the stage for the discussions in the subsequent chapters.

- Chapter 2: "Network Softwarization and Virtualization: Management of QoS in Wireless and Mobile Networks"

This chapter examines the impact of network softwarization and virtualization on QoS, highlighting innovative management approaches for wireless and mobile networks.

- Chapter 3: "Quality of Service Challenges in Hybrid IP/Software-Defined Networks"

Focusing on hybrid networks, this chapter discusses how QoS can be effectively managed in environments combining traditional IP networks with software-defined networks (SDNs).

- Chapter 4: "The Role of QoS at the OSI Model Layers"

This chapter analyzes the role of QoS across the OSI model, offering insights into how QoS mechanisms are implemented at each layer to ensure end-to-end performance.

Section 2: "Technological Solutions and Implementations"

This section highlights cutting-edge solutions and technologies that address QoS challenges. It emphasizes practical implementations and strategies to optimize performance, improve reliability, and ensure security across diverse network scenarios.

- Chapter 5: "Implementation of an Enhanced Security Algorithm for Wireless Sensor Networks"

This chapter introduces a novel security algorithm designed to enhance QoS in wireless sensor networks, addressing the critical need for secure and efficient data transmission.

- Chapter 6: "Omnidirectional Antenna Design for Underground Wireless Sensor Network QoS Improvement"

Here, the focus is on designing and implementing omnidirectional antennas to improve QoS in challenging underground wireless sensor network environments.

- Chapter 7: "Quality of Service (QoS) for the Future 6G Network Data Transport"

This chapter explores QoS considerations for 6G networks, examining how emerging technologies will influence QoS and proposing solutions to meet the demands of future applications.

- Chapter 8: "Quality of Service (QoS): In Sliced Network"

This chapter investigates network slicing as a tool for achieving QoS, particularly in multi-tenant and multi-service environments, demonstrating its potential to optimize performance and resource allocation.

This book is designed to be a valuable resource for researchers, industry professionals, and students in networking and telecommunications. By covering both the theoretical and practical dimensions of QoS, it bridges the gap between understanding fundamental challenges and implementing effective solutions.

As we move toward a more interconnected and data-driven world, we hope this book serves as a guide and an inspiration for further innovation in QoS.

Yaseein Soubhi Hussein
Associate Professor,
Department of Information Systems and Computer Science,
Ahmed Bin Mohammed Military College (ABMMC),
Doha, Qatar

Abdulmajeed Al-Jumaily
Postdoctoral Researcher,
Department of Signal Theory and Communications,
Charles III University of Madrid (UC3M),
Madrid, Spain

Section 1

Challenges and Fundamentals in Quality of Service (QoS)

Chapter 1

Introductory Chapter: Challenges and Solutions in Quality of Service (QoS) – Optimizing Network Performance

Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily

1. Introduction

This chapter introduces the ebook by exploring modern solutions, including their theoretical frameworks and sensible implementations. As community infrastructures become increasingly complex and interconnected, the risk of cyber threats grows, making it essential to include superior security features to guard against those threats while maintaining brilliant service. The dependent innovation provides a comprehensive review of the chapter. Each phase addresses a distinct QoS challenge, proposing a precise and prepared technique for the subject. The book aims to give readers in-depth knowledge of the modern advancements in community slicing, side computing, machine-gaining knowledge, and service orchestration. It will spotlight the combination of more advantageous QoS with advanced technologies, including those past 5G and beyond, by discussing methods to optimize assets via QoS-aware gadget learning models, dynamic resource allocation, the usage of predictive analytics, and effective QoS management.

Additionally, the book will examine dynamic traffic routing algorithms for real-time QoS optimization and improving a security-conscious community infrastructure using virtualization and cloud-primarily based sources to improve network performance and resilience.

In the modern landscape of statistics and communication technology, Quality of Service (QoS) emerges as a vital idea, ensuring that networks meet the performance expectations of diverse programs and services. QoS is a cornerstone of network control, vital for ensuring that verbal exchange networks function correctly while meeting the diverse wishes of customers. As networks preserve to adapt, becoming increasingly complicated and assisting a wider variety of devices and applications, the importance of QoS has grown correspondingly. QoS encompasses the control of community sources to assure the performance metrics necessary for the highest quality operation of programs, which includes bandwidth, latency, jitter, and packet loss [1]. QoS is essential in optimizing community performance in numerous domains, including telecommunications, laptop networks, and rising technologies. Understanding the core principles of QoS is critical as we explore the challenges and solutions vital for current network environments, where performance, reliability, and subjective satisfaction are essential. As digital connectivity continues to suffuse all aspects of lifestyles, the function of QoS in providing reliable and efficient network services is essential [2]. This chapter provides a

first-level view of the significance of QoS, the demanding situations it presents, and the progressive answers that are advanced in dealing with these demanding situations.

1.1 Fundamental concepts of QoS

QoS is described by several key performance indicators (KPIs) that determine the quality and reliability of network offerings. These KPIs encompass bandwidth, latency, jitter, and packet loss, all of which play a crucial role in the overall performance of different packages. Maintaining high QoS requirements in modern-day networks creates numerous enormously demanding situations. One of the number one demanding situations is the developing complexity of networks, which must now support an extensive selection of gadgets, applications, and offerings, each with precise overall performance requirements. Bandwidth refers to a community's maximum facts transfer rate, which is, in particular, essential for applications including video streaming and big record transfers [3]. Latency refers to the put-off before a transfer of statistics begins following an education, an essential parameter for real-time applications such as VoIP and online gaming [4]. Jitter measures the variability in packet arrival instances, affecting the best of voice and video communications [5]. Packet loss, denoting the quantity of data packets that fail to attain their vacation spot, can seriously degrade the performance of any network carrier, in particular, people who rely on continuous records streams [6]. This complexity leads to demanding situations, including network congestion, fluctuating site visitor patterns, and mixing legacy structures with new technologies. These issues are similarly compounded by the call for actual-time responsiveness in programs like streaming, gaming, and critical infrastructure structures, where even minor performance degradations may have reliable-sized influences. Understanding these standards is essential for powerful QoS management, specifically in heterogeneous network environments where exclusive packages have various QoS necessities [7]. Addressing these demanding situations calls for deep information on the elements that affect QoS and the improvement of effective strategies to make certain continuous performance throughout diverse and worrying network environments.

1.2 Challenges in QoS management

Managing QoS in current networks is a significant challenge. One of the first challenges is the dynamic nature of network environments, where conditions can change rapidly because of fluctuating site visitors masses, regularly leading to inconsistent carrier satisfaction [8]. In reaction to these demanding situations, modern-day solutions have evolved to optimize QoS in present-day networks. Machine learning and synthetic intelligence are increasingly applied to allow predictive analytics and dynamic resource allocation, allowing networks to evolve to changing situations in real time. For instance, during peak utilization times, networks can also be congested, resulting in multiplied latency and packet loss. Addressing these fluctuations requires sophisticated monitoring equipment and adaptive management mechanisms that can reply in real-time [9]. Another venture lies in the allocation of constrained network assets. Network directors need to prioritize site visitors to ensure that critical offerings receive the necessary assets, which may be challenging given the finite availability of bandwidth and different sources [10].

Network slicing and facet computing have also emerged as critical strategies for reinforcing QoS via improving resource performance and decreasing latency. These improvements represent cutting-edge studies in network control, presenting promising avenues for addressing the QoS challenges posed by modern-day and destiny

networks. Security also poses a widespread venture, as ensuring QoS in the face of potential cyber threats calls for solid infrastructure and security measures [11]. For example, distributed Denial of Service (DDoS) attacks can weigh down network assets, extensively decreasing QoS [12]. Leveraging those technologies makes it viable to maintain or beautify QoS despite network environments' growing complexity.

1.3 Innovative solutions for QoS

Several progressive solutions have been developed to manage these demanding situations; numerous extreme solutions have been developed. One such solution is network reduction, particularly beneficial in subsequent-era networks. Network slicing allows for the introduction of more than one digital network on an available physical infrastructure, each optimized to satisfy unique QoS necessities [13]. In response to these challenges, cutting-edge answers, gadget learning, and synthetic intelligence are increasingly utilized to enable predictive analytics and dynamic resource allocation, permitting networks to conform to changing conditions in real-time. This technique enhances resource usage and isolates network capabilities, enhancing standard QoS. Edge computing additionally enhances QoS via processing data closer to the supply, decreasing latency and improving reaction instances for actual-time packages [14]. Furthermore, community cutting and area computing have emerged as essential strategies for enhancing QoS by improving resource performance and decreasing latency. These advancements represent the slicing edge of studies in network control, presenting promising avenues for addressing QoS-demanding situations in current and destiny networks. Machine learning (ML) and artificial intelligence (AI) are increasingly integrated into QoS management systems. This technology enables predictive QoS control by analyzing sizeable quantities of community statistics to forecast visitor patterns and dynamically modify QoS parameters in real time [15]. Service orchestration, which involves the dynamic coordination and control of network assets, guarantees that the best sources are allotted when and in which they are needed, preserving the preferred level of QoS [16]. This approach enhances QoS and reduces the load on centralized statistics facilities, mainly for extra-efficient community operations [17]. In industrial Internet of Things (IoT) applications, side computing can ensure timely record processing and choice-making, which is important for maintaining operational efficiency and protection [18]. Machine Learning (ML) and Artificial Intelligence (AI) are also playing increasingly important roles in QoS management. ML algorithms can analyze considerable quantities of network records to expect visitor patterns and alter QoS parameters dynamically [19]. This predictive functionality allows for proactive QoS control, ensuring optimal community performance even under various situations [20]. AI-driven community management structures can dynamically reroute site visitors to avoid congestion and keep high QoS ranges [21]. Service orchestration entails the dynamic coordination and control of network resources to meet specific provider requirements [22]. This procedure guarantees that the proper assets are allotted at the proper time, retaining the desired stage of QoS. Service orchestration is mainly useful in complicated network environments where several offerings with distinct QoS necessities operate concurrently [23].

1.4 QoS within the 5G era

The 5G era represents a sizable soar in conversation capabilities, bringing forth new possibilities and challenges for QoS control. 5G networks are designed to assist a large

number of connected devices, extremely dependable Ultra-Reliable and Low Latency Communications (URLLC), and more desirable cellular broadband, Enhanced Mobile Broadband (eMBB) [24]. These advancements necessitate re-evaluating traditional QoS procedures to house the diverse and stringent necessities of 5G applications [25]. Network slicing, as stated earlier, is a fundamental factor of 5G QoS. By growing devoted virtual networks for one-of-a-kind use cases, community reduction ensures that every application receives the ideal resources and QoS ensures [26]. A self-reliant car network slice might require extremely low latency and excessive reliability, whereas a slice for video streaming may prioritize bandwidth and throughput [7].

Edge computing is also essential to 5G QoS. The allotted architecture of 5G, blended with part computing, permits the processing of records at the threshold of the community in the direction of the cease-customers [27, 28]. This appreciably reduces latency and enhances the performance of latency-touchy applications, including AR and VR [29]. AR packages in retail, for example, can offer more excellent purchaser experiences by processing information in actual time at the threshold [30]. ML and AI are anticipated to play even more significant distinguished roles in 5G networks. The complexity of dealing with QoS in 5G requires advanced algorithms capable of actual-time selection-making and self-optimization [31].

AI-pushed network management structures can dynamically modify network parameters to maintain QoS, expect and mitigate congestion, and optimize aid allocation primarily based on actual-time facts analytics [32]. AI can help in the predictive preservation of community infrastructure, preventing ability disasters that might degrade QoS [19].

1.5 Future prospects and beyond 5G

Future solutions may also include the development of dynamic traffic routing algorithms that optimize site visitor flows in real-time, decreasing latency and packet loss while balancing community loads. Additionally, the combination of superior safety features, which include blockchain generation and AI-pushed chance detection, might be critical to shield against cyber threats while retaining excellent providers; looking past 5G, the destiny of QoS management may involve even greater sophisticated technology and methodologies. Emerging tendencies, which include the Internet of Things (IoT), Industry 4.0, and intelligent towns, will area additional needs on network infrastructures, necessitating similar advancements in QoS [33, 34]. One promising area is the improvement of dynamic visitor routing algorithms. These algorithms can optimize site visitor flows in actual time, making sure that information packets take the most green paths via the network [35]. This dynamic approach can notably enhance QoS by reducing latency, minimizing packet loss, and balancing network loads [33]. Security will continue to be a paramount situation in future QoS management. As community infrastructures become more complicated and interconnected, the capacity for cyber threats will increase [36].

Future QoS solutions will want to combine superior security measures to protect against those threats while retaining excellent service. This may also involve using the blockchain era for stable and prominent resource management and AI-pushed threat detection and response systems [37]. Blockchain can ensure steady transactions and useful resource allocations, preventing fraudulent activities that might affect QoS [37]. Virtualization and cloud-primarily based resources also play vital roles in QoS's destiny. Virtual network functions (VNFs) and Software-Defined Networking (SDN) allow for extra bendy and efficient resource control [38, 39]. By decoupling network

functions from bodily hardware, these technologies enable dynamic aid allocation and scaling, which might be essential for preserving QoS in an increasing number of complicated community environments [40]. SDN can dynamically modify community configurations to optimize overall performance primarily based on actual-time call-for, improving basic QoS [40].

Overall, the QoS – Challenges and Solutions affords complete know-how for modern-day improvements in QoS. Through an in-intensity exploration of fundamental standards, demanding situations, and progressive answers, valuable insights into QoS's dynamic and multifaceted nature are offered. Whether a researcher, practitioner, or scholar, this book serves as an in-depth manual to the techniques and technology with the purpose of shaping the destiny of communique networks [10]. By analyzing the effect of 5G and beyond 5G technologies on QoS, the book highlights the importance of adaptive methodologies and advanced technology, including community cutting, aspect computing, gadget learning, and service orchestration [41]. It additionally addresses the vital function of safety in preserving QoS and explores future innovations' capability, including dynamic traffic routing and virtualization [33].

In an era where virtual conversation is integral to every component of existence, know-how and dealing with QoS is more critical than ever. This book equips readers with the information and tools essential to navigate the complexities of QoS, ensuring the shipping of first-rate network offerings in an ever-evolving technological panorama [42]. Through this exploration, a robust expertise in coping with the demanding situations of QoS, optimizing network performance, and enhancing user satisfaction is gained. "Quality of Service (QoS) – Challenges and Solutions" is not just a technical guide but a comprehensive, useful resource for each person looking to recognize and improve first-class communication networks in the virtual age. This book affords an in-depth roadmap for the future of QoS, highlighting the progressive solutions and emerging traits that will be an excellent way to form the panorama of verbal exchange networks for years yet to come.

Author details

Yaseein Soubhi Hussein[1*] and Abdulmajeed Al-Jumaily[2]

1 Department of Information Systems and Computer Science, Ahmed Bin Mohammed Military College, Qatar

2 Department of Signal Theory and Communications (DTSC), University Charles III of Madrid, UC3M, Spain

*Address all correspondence to: dr.yaseein@abmmc.edu.qa

IntechOpen

References

[1] Santhi J, Prabha K. QOS aware vertical handover process in heterogeneous wireless network. Measurement: Sensors. 2023;**26**:100710

[2] Banafaa M et al. 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. Alexandria Engineering Journal. 2023;**64**:245-274

[3] Nourildean SW, Mohammed YA, Attallah HA. Virtual local area network performance improvement using ad hoc routing protocols in a wireless network. Computers. 2023;**12**(2):28

[4] Mast N, Khan S, Uddin MI, Ghadi YY, Alkahtani HK, Mostafa SM. A cross-layer solution for contention control to enhance TCP performance in wireless *ad-hoc* networks. IEEE Access. 2023;**11**:24875-24893

[5] Mazhar T et al. Quality of service (QoS) performance analysis in a traffic engineering model for next-generation wireless sensor networks. Symmetry. 2023;**15**(2):513

[6] Murty DSKN. Information flow identification using the package marking system. Journal of Informatics Education and Research. 2023;**3**(2):654-660

[7] Li X et al. Network slicing for 5G: Challenges and opportunities. IEEE Internet Computing. 2017;**21**(5):20-27

[8] Zhang N, Liu Y-F, Farmanbar H, Chang T-H, Hong M, Luo Z-Q. Network slicing for service-oriented networks under resource constraints. IEEE Journal on Selected Areas in Communications. 2017;**35**(11):2512-2521

[9] Sachs J et al. Adaptive 5G low-latency communication for tactile internet services. Proceedings of the IEEE. 2018;**107**(2):325-349

[10] Zeb S, Rathore MA, Hassan SA, Raza S, Dev K, Fortino G. Toward AI-enabled nextG networks with edge intelligence-assisted microservice orchestration. IEEE Wireless Communications. 2023;**30**(3):148-156

[11] Kazmi SHA, Qamar F, Hassan R, Nisar K. Routing-based interference mitigation in SDN enabled beyond 5G communication networks: A comprehensive survey. IEEE Access. 2023;**11**:4023-4041

[12] Ahmed E et al. The role of big data analytics in internet of things. Computer Networks. 2017;**129**:459-471

[13] Samadi R, Nazari A, Seitz J. Intelligent energy-aware routing protocol in mobile IoT networks based on SDN. IEEE Transactions on Green Communications and Networking. 2023;**7**:2093-2103

[14] Sallam G, Zheng Z, Ji B. Placement and allocation of virtual network functions: Multi-dimensional case. IEEE Transactions on Mobile Computing. 2022;**22**(7):4195-4211

[15] Foukas X, Patounas G, Elmokashfi A, Marina MK. Network slicing in 5G: Survey and challenges. IEEE Communications Magazine. 2017;**55**(5):94-100

[16] Gill SS et al. Modern computing: Vision and challenges. Telematics and Informatics Reports. 2024;**13**:100116

[17] Satyanarayanan M. The emergence of edge computing. Computer. 2017;**50**(1):30-39

[18] Ghosh P, Biswas A, Ghosh S. Fundamentals and technicalities of big data and analytics. In: Intelligent Systems in Healthcare and Disease Identification Using Data Science. United Kingdom: Chapman and Hall/CRC; 2023. pp. 51-106

[19] Zhang H, Liu N, Chu X, Long K, Aghvami A-H, Leung VC. Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine. 2017;**55**(8):138-145

[20] Tayyaba SK et al. 5G vehicular network resource management for improving radio access through machine learning. IEEE Access. 2020;**8**:6792-6800

[21] Castro OEL, Deng X, Park JH. Comprehensive survey on AI-based technologies for enhancing IoT privacy and security: Trends, challenges, and solutions. Human-centric Computing and Information Sciences. 2023;**13**:39

[22] Chiang Y et al. Management and orchestration of edge computing for IoT: A comprehensive survey. IEEE Internet of Things Journal. 2023;**10**(16):14307-14331

[23] Yousaf FZ, Bredel M, Schaller S, Schneider F. NFV and SDN—Key technology enablers for 5G networks. IEEE Journal on Selected Areas in Communications. 2017;**35**(11):2468-2478

[24] Siddiqui MUA, Abumarshoud H, Bariah L, Muhaidat S, Imran MA, Mohjazi L. Urllc in beyond 5g and 6g networks: An interference management perspective. IEEE Access. 2023;**11**: 54639-54663

[25] Kabalci Y, Mutlu U. Emerging communication technologies for V2X: Standards and protocols. In: Smart Grid 3.0: Computational and Communication Technologies. Cham, Switzerland: Springer; 2023. pp. 301-329

[26] Salahdine F, Han T, Zhang N. 5G, 6G, and beyond: Recent advances and future challenges. Annals of Telecommunications. 2023;**78**(9):525-549

[27] Al-Jumaily A, Alrubaee SH, Jiménez VPG, Al-Saegh AM, Mahmood AA, Al-Jumeily D. Architecture design of B5G and 6G millimeter-wave radio access network: Using wireless communications to increase coverage, capacity and performance. In: 2023 16th International Conference on Developments in eSystems Engineering (DeSE). Istanbul, Turkey: IEEE; 2023. pp. 212-217

[28] Raeisi-Varzaneh M, Dakkak O, Habbal A, Kim B-S. Resource scheduling in edge computing: Architecture, taxonomy, open issues and future research directions. IEEE Access. 2023;**11**:25329-25350

[29] Nandy P. The intersection of 5G and blockchain technology: A paradigm shift in connectivity and security. Scientific Voyage. 2023;**3**(3):1-5

[30] Zhang Y, Azizui MFAB, Yan C. The application of 6G and augmented reality technology in education and training. Wireless Personal Communications. 2024;**15**,:1-19

[31] Nassar A, Yilmaz Y. Deep reinforcement learning for adaptive network slicing in 5G for intelligent vehicular systems and smart cities. IEEE Internet of Things Journal. 2021;**9**(1):222-235

[32] Alwahedi F, Aldhaheri A, Ferrag MA, Battah A, Tihanyi N. Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models.

Internet of Things and Cyber-Physical Systems. 2024;**4**:167-185

[33] Liyanage M, Ahmad I, Abro AB, Gurtov A, Ylianttila M. A Comprehensive Guide to 5G Security. Wiley Online Library; 2018

[34] Hussein YS, Jiménez VPG, Al-Jumaily A. Emerging technology for 6G networks. International Journal of Information and Communication Engineering. 2024;**18**(3):143-149

[35] Aruna R et al. Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless *ad-hoc* network using craft protocol. Wireless Networks. 2024;**30**(2):711-735

[36] Frías N, Johnson F, Valle C. Hybrid algorithms for energy minimizing vehicle routing problem: Integrating clusterization and ant colony optimization. IEEE Access. 2023;**11**:125800-125821

[37] Kouicem DE, Bouabdallah A, Lakhlef H. Internet of things security: A top-down survey. Computer Networks. 2018;**141**:199-221

[38] Gil Jiménez V, Al-Jumaily A, Sali A, Al-Jumeily D. Hybrid chaos particle swarm optimization algorithm for smart cloud service system based on optimization resource scheduling and allocation. Journal of Autonomous Intelligence. 2023;**6**(2):652-652

[39] Nawaz H, Ali MA, Rai SI, Maqsood M. Comparative analysis of cloud based SDN and NFV in 5g networks. The Asian Bulletin of Big Data Management. 2024;**4**(1):206-216

[40] Pillai SEVS, Polimetla K. Integrating network security into software defined networking (SDN) architectures. In: 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS). Bangalore Section, India: IEEE; 2024. pp. 1-6

[41] Al-Ansi A, Al-Ansi AM, Muthanna A, Elgendy IA, Koucheryavy A. Survey on intelligence edge computing in 6G: Characteristics, challenges, potential use cases, and market drivers. Future Internet. 2021;**13**(5):118

[42] Beshley M, Kryvinska N, Beshley H. Quality of service management method in a heterogeneous wireless network using big data technology and mobile QoE application. Simulation Modelling Practice and Theory. 2023;**127**:102771

Chapter 2

Network Softwarization and Virtualization: Management of QoS in Wireless and Mobile Networks

Umair Shafiq Khan and Tahira Mahboob

Abstract

The significant evolution of QoS management methods in wireless and mobile networks has relied on developments in various technologies and frameworks based on network softwarization and virtualization. This work details technologies, namely network functions virtualization (NFV) and software-defined networking (SDN) that encourage a dynamic, efficient, and flexible network architecture. Furthermore, the developments in these domains that benefited efficient resource utilization together with providing QoS management solutions required to guarantee optimal network performance have been discussed in detail. The architectural advancements of futuristic networks and their implementations, such as the separation of the control and data planes in SDN and resource virtualization in NFV to achieve desired QoS, the evolution of OpenFlow protocol, QoS mechanisms, and the related works done by standardizing organization, have also been covered. This work also includes use case scenarios for these technologies and their utilization in wireless and mobile networks. Managing quality-of-service (QoS) concerns in complex networks to enhance the network's agility and security has also been discussed. It further details the implementations supported by edge computing and artificial intelligence applications.

Keywords: mobile networks, network function virtualization (NFV), quality of service (QoS), software-defined networking (SDN), wireless networks

1. Introduction

The massive growth in the number of mobile devices and the development of Internet of things (IoT) technologies increase the resource and traffic demands for wireless and mobile networks. The guaranteed quality-of-service (QoS) and ubiquitous connectivity are essential. Their applications include automotive, autonomous vehicles, and personal devices [1]. Consequently, meeting user and network demands is challenging for network engineers, operators, and researchers, which has changed its focus to modern approaches such as network softwarization

and virtualization. These approaches have provided a paradigm shift in network management from traditional methodologies by adding flexibility and dynamicity, which is not possible while using commodity servers. Network softwarization, supported by SDN, separates the control plane from the data plane. The network orchestration and configuration become centralized where resources can be managed dynamically [2]. This abstraction enables the network operators with flexibility to provide solutions that efficiently utilize network resources while meeting individual user's resource demands.

Another concept significant for today's networks is network virtualization, which entails disassociating network functions from the physical. It allows the formation of many virtual networks on the same physical hardware, and each of these virtual networks can be customized to different service delivery requirements or customers. This split also helps scalability and optimizes usage, which is necessary for dealing with the complex and, at times, conflicting needs and expectations of today's networked applications [3].

Therefore, QoS management has necessitated service delivery demands in wireless and mobile networks. Apart from conventional traffic classification and prioritization for providing QoS in softwarized and virtualized environments, it is necessary to have intelligent control over the application of network functions concerning dynamic modifications to traffic and network conditions. Sophisticated concepts like artificial intelligence are now incorporated to forecast traffic patterns and allocate resources intelligently to meet the demands and ensure the required service level [4].

Based on mobile and wireless topologies, user mobility, dynamic network conditions, and operational requirements create additional challenges. Not only do networks have to handle the density of connections being much higher than in fixed environments, but continuity of service must also be possible when a user crosses from one network to another or when signals are weak. Network virtualizations, tenant-based network slicing, and edge computing can be potential solutions. Network slicing makes it possible to establish micro-connections for services or traffic intensity. At the same time, edge computing is a process of functioning as many micro-data centers are located close to the consumer to minimize delays [5]. Therefore, ultra-reliable low latency communications (URLLC) and enhanced mobile broadband (eMBB) can be addressed by leveraging these technologies.

In this chapter, the reader will find a comprehensive evolution in next-generation networks emanating from network softwarization and virtualization, focusing on QoS management. It goes as far as to discuss what these technologies are based on, the problems targeted, and the complex solutions created to ensure that the growing amount of stress on network structures is not directly related to quality degradation. Understanding the conceptual backgrounds and the real-life applications of the discussed technologies will emphasize the importance of designing network management's further evolution.

2. SDN and NFV fundamentals

This section delves into the foundational technologies reshaping the network infrastructure landscape: software-defined networking (SDN) and network functions virtualization (NFV).

2.1 Software-defined networking (SDN)

SDN decouples separate the data and control planes enabling network managers to appropriately manage and route traffic and deploy service at whatever point they desire in the whole network supported by a global perspective view of the network [2].

SDN has several layers that are managed and operated as different structures of the networking stack. The application plane is at the top of the architecture and deals with network applications that enforce policies such as routing, QoS, access control lists (ACLs), and other network services. These applications are connected to the network by the northbound interface, which contains a set of interfaces or API, such as restful API (REST), directing communication with the control plane.

The control plane plays a significant role in the SDN architecture as the network is centrally controlled by an SDN controller or a hypervisor, providing real-time network control based on the applications and related policies and service-level agreements (SLAs) defined earlier among service providers/operators and users [6]. It translates policy information from the application plane into operational data for the data plane. This translation and communication occur through the southbound interface (SBI) using protocols, such as OpenFlow where the controller can program the network elements/node/switches, dynamically, bringing programmability into the network. Last but not least, the data plane is made of physical and virtual devices, including switches, routers, etc., controlled by the instructions from the control plane [2, 6]. Also, SDN empowers refined security measures to be engaged in the new threat approaches, which have become crucial in the modern networked world.

Additionally, SDN has been used to enable the management of cloud service integration into enterprise and carrier networks. SDN makes traffic management within cloud networks more precise and efficient, supporting the delivery of cloud services and improving business outcomes for organizations [7].

2.2 Network functions virtualization (NFV)

NFV alters how services are implemented because network functions are handed off from hardware-specific devices and implemented in software to run on vendor-agnostic commodity servers. This enables a reduction in capital and operational (CAPEX) expenditures for provisioning and managing network services, depending on the current network of user demand [3].

NFV solves the issues of accommodation and flexibility for the networks by eliminating functions such as firewalls, load balancers, and intrusion systems from the traditional hardware units. It enables functions to be executed on virtual machines, which may be relocated or cloned to other physical servers on demand. This minimizes delay in provisioning the newer services [3, 8].

Figure 1 illustrates the architecture of virtual network functions (VNFs), a significant component of NFV. The primary emphasis in improving network management is separating the control and data planes; however, the NFV architecture is centred on dissociating network functions from equipment. VNFs are executed as applications on the VMs distributed over the NFVI physical hosts; these comprise compute nodes, storage assets, and network appliances. Such a setup allows for better modularity of network services and their expansion plans depending on the layer 3 (L3's) architecture.

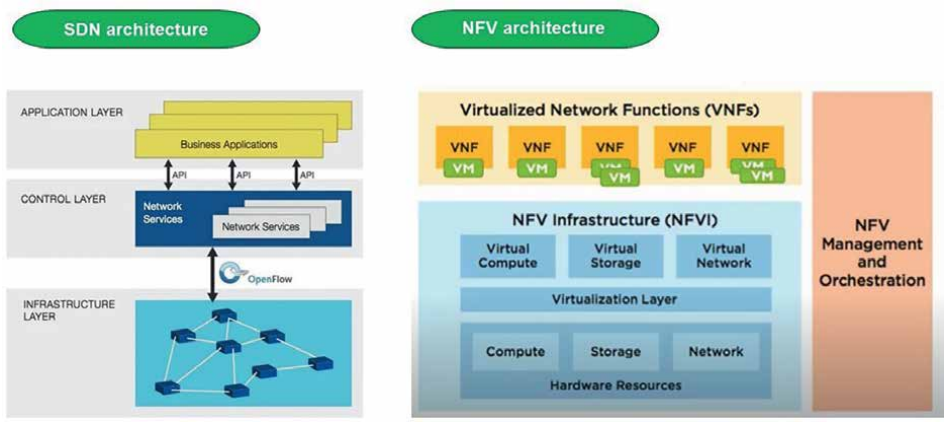

Figure 1.
Architecture of software-defined networks (SDNs) vs. network functions virtualization.

At the structure of NFV, the highest level is the NFV management and orchestration layer, which controls the deployment and tasking of NFV network functions along with VMs. It corrects the use of resources and guarantees the proper state of VNFs. This is different from what SDN has consistently advocated for in the manner it coordinates traffic direction through control interfaces that can be programmed. Both architectures attempt to improve network programmability and decrease operational expenses; however, NFV concentrates on infrastructure management and the flexibility of services, whereas SDN deals with centralized traffic flow.

Moreover, NFV helps make network services more flexible. It permits the fast introduction of new services and is highly responsive to changed business or customer needs. NFV uses standard IT virtualization; therefore, network operators are in a position to introduce new networking services early [3, 8]. This flexibility is more significant than other circumstances that take a long time to deploy services, as seen in the mobile network instances and events.

Further, NFV can be combined with other cloud techniques; this allows such solutions to create cloud-based network functions, which are embedded, managed, and used based on cloud methodologies. Such an approach also helps build network services that can be easily scalable and adaptable simultaneously, as it forms a strong foundation that can continue to advance in network innovation [9].

2.3 Synergy between SDN and NFV

While SDN and NFV are independently powerful, their combined use can unleash significant synergies that amplify their benefits, particularly in complex network environments. SDN's ability to provide centralized dynamic traffic management complements NFV's flexibility in service deployment [10]. The integration of SDN and NFV facilitates the creation of highly automated and flexible network platforms capable of supporting high-bandwidth applications and services. SDN can control the network resources in real time to perform NFV functions and achieve the function of perfect VNF matching with the network resources. This synchronization requires network management and service control to deliver reliable, scalable, and optimized networks [10, 11].

SDN and NFV facilitate network architecture and functioning complications, augmenting the capacity to effectively provision network services across different domains and satisfy constantly rising consumer expectations [10, 11]. SDN and NFV are not only the enabling technologies to implement missing network functionalities, but they also create a basis for the further evolution of the technology, including new ideas like software-based network slicing for dynamically segmenting the network resources and providing the appropriate services and capabilities to the different classes of users and applications. This optimizes the use of network resources and improves the functions of service individuality and protection [12].

The Venn diagram in **Figure 2** comparing NFV, SDN, and NV shows how these technologies collectively boost and redesign network operations and functions. In its simplest form, NFV mainly targets separating network functions from the appliances on which they run to be executed in software contexts. This abstraction enhances the scalability and topology of networks and smoothly brings easier and cheaper solutions to deploy, manage, and grow the network services. SDN enables the control of the networks' actions centrally, and it offers potent tools to the administrators to manage the traffic flow in the network and make operation decisions dynamically. This centralization exact regulation of network resources, hence boosting performance and flexibility. On the other hand, NV broadens the concept by admitting several virtual networks on a single physical platform that can be operated and administrated apart.

The intersections within the diagram elucidate deeper synergies: The intersections within the diagram elucidate deeper synergies:

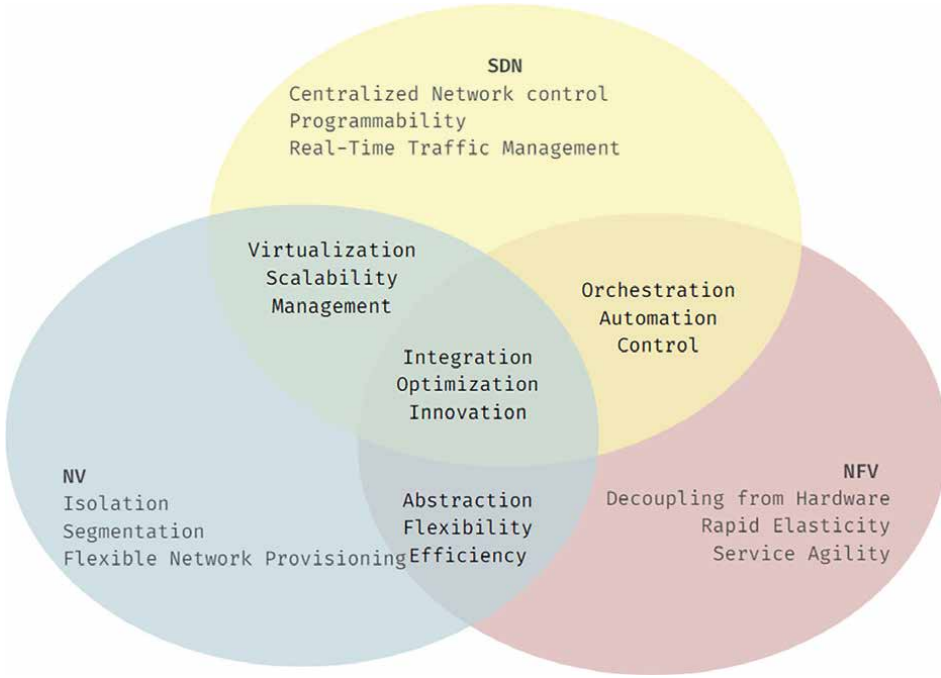

Figure 2.
Relationship between SDN, NFV, and NV.

- Between NFV and SDN: They both manage and control one aspect of the network using complicated processes, often automated—integrating these two results in effective resource optimization optimization workings of diverse networks.

- Between NFV and SDN: They both manage and control one aspect of the network using complicated processes, often auto-integrating these two results in effective resource optimization workings of diverse networks.

- Between NFV and NV: They help enhance network abstraction and flexibility, enabling efficient execution of network services on different networks and platforms.

In the middle of NFV, SDN, and NV, there is a synergy that builds on the best capabilities of the three, enhancing the overall network design. This integrated concept further improves the network's functionality and expandability for the end client. It encourages service leaders to offer increasingly complex network services and structures by providing a firm foundation on which to build. When visualizing relationships and functionalities using a Venn diagram, it is possible to see how NFV, SDN, and NV complete each other and build up a new network environment that is much more resilient, flexible, and efficient.

2.4 Quality-of-service in SDN and NFV-based networks

The SDN and NFV inundated networks require high QoS as they incorporate complex traffic management and resource allocation schemes to meet SLAs. The best real-world example of its practice is the utilization protocols in SDN structures to shape the network traffic flow and to enable the different flows to be managed in line with specific QoS requirements. OpenFlow, by the open network foundation (ONF), founding protocol supporting SDN implementations, allows fine-grain traffic policies that can forward, throttle, or sculpt traffic to the needs of specific applications in terms of bandwidth and latency [6, 13].

In SDN, advanced dynamic QoS policies are implemented through programmable network controllers, which adapt rules based on real-time network conditions to optimize, even under varying loads [6, 13]. OpenFlow switches, OvS-switch, integral to SDN architecture, leverage a robust switch fabric with queues and metering functions to manage network traffic effectively, depicted in **Figure 2**. These switches maintain flexible tables that can be programmed to match various packet fields, enabling administrators to implement granular QoS policies.

Within the switching fabric, the metering functions within these switches measure and control the rate at which packets traverse the network, ensuring no application or service exceeds its allocated bandwidth. Meanwhile, queuing mechanisms manage congestion and prioritize traffic during high traffic volumes, applying actions such as packet forwarding, dropping, and rate-limiting based on established metrics [6].

Figure 3 of an OvS switch configuration illustrates the pivotal role of the central controller in QoS management. All the commands are transmitted through a secure channel, where the controller establishes a dynamic setup with the switch. The flow table is a priority tool in QoS management since it addresses the incoming packets from the network and executes the corresponding rules by unchanged QoS protocols.

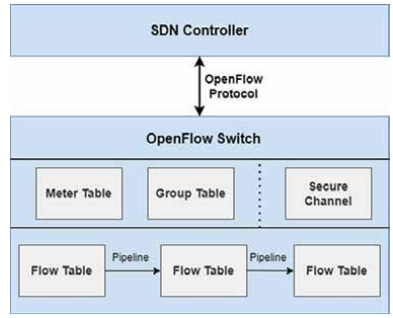

Figure 3.
Components of an OpenFlow switch.

This helps to guarantee that overly critical services remain a high priority over low-importance ones. From the flow table, the packets move to the group table to ensure QoS through replicating packets for multicast—essential for equal data distribution required for the application of broadcast and multicast services. That is why the meter table is subsequently used for rate limiting, monitoring, and controlling the relative traffic flow according to the bandwidth allocation so that a single flow does not become dominant and bring down the whole network.

The ports are responsible for the entry and exit of all network traffic managed directly by the controller. The controller determines the traffic flow according to QoS requirements in real time. Subordinate to these, the data path transmits and forwards data packets according to the QoS rules predetermined by network upstream layers. This setting emphasizes the OpenFlow switch's ability to efficiently manage a precise and demanding QoS policy. It ensures that all the existing resources are distributed in the network according to business-critical functions while preserving top-notch network performance and availability [6, 13]. Here queues can be set up, where each queue could have a predefined rate limited to control the traffic flow.

Also, the NFV benefits the SDN by enabling the faster deployment of the network services through a conceptually flexible structure. For example, virtual network functions (VNFs) and virtual routers, firewalls, and WAN accelerators can be instantiated to improve service provisioning and quality. This dynamic capability is vital within this thought as it entails creating processes for customizing offered within the network to address the existing demand [3, 8].

SDN and NFV integration also enables the resource orchestration process, including the QoS for vital services. SDN is a centralized technique that allows for better traffic and resource management, and NFV enables service flexibility; together, both systems help ensure modern network applications that rely on video streaming, cloud computing, and large data transmissions are provided with the necessary QoS [14].

Table 1 represents the timeline of OpenFlow versions regarding its QoS and the generational advancement of mobile and wireless networks from 2008 to 2015. Each version from 1.0 to 1.5 presents new features, such as improved QoS for multicast, rate-limiting of metering, and optical ports. These features make OpenFlow useful in promising solutions for network management and support network events in mobile and wireless networks. Consequently, by using SDN's OpenFlow protocol and engaging NFV properly, networks can have a dynamic and proactive solution for managing QoS that can address the conditions and challenges of current and future networks.

Year	OpenFlow (OF) ver.	QoS improvements	Event in mobile network	Event in wireless network
2008	1.0	Basic QoS settings	Initial OF specification released	Basic OF features tested on Wi-Fi
2010	1.1	Enhanced QoS for groups and multicast	Experimental LTE setups using OF	Development of wireless mesh networks
2012	1.2	Introduction of metering for rate-limiting	First trials of OF for network slicing in 4G	OF applied managing heterogeneous wireless networks
2013	1.3	Advanced QoS with multiple tables	Integration of SDN controllers with OF in mobile core networks	Enhanced QoS management in Wi-Fi networks via OF
2014	1.4	Egress processing for better QoS handling	dynamic spectrum management in Mobile networks	Deployment of OF in large-scale public Wi-Fi systems
2015	1.5	Optical port enhancements for diverse traffic management	OF assists in the transition to 5G, supporting network slicing and orchestration	Introduction of OF in wireless ad-hoc networks for improved routing

Table 1.
Timeline of documented OpenFlow versions and corresponding developments in QoS and network events.

2.5 QoS provisioning in wireless and mobile networks: Trends

QoS provision in wireless and mobile networks faces unique challenges due to these environments' inherent variability and unpredictability. However, the challenges above are effectively solved by SDN and NFV. These technologies' mobility characteristics underline the possibility of offering QoS that quickly adapts to mobile environments [7].

The mobile network has been improved by using SDN that separates the control and the forwarding planes where the controller is used. This centralization provides the network operators with the network's overall topology and control, thus allowing them to manage the network and the different resources. SDN has the strategic advantage of changing the network policies based on users' and conditions' mobility, which is common in, for instance, mobile environments. For example, whenever mobile users switch from one cell tower to another, the SDN controller can identify input to modify the network topology and quality. On traffic management, for instance, SDN controllers employ efficient algorithms in handling data depending on the type of traffic and QoS demands and thus enhance the efficiency of the network. For example, suppose the business uses a video conference call. In that case, the controller might prefer this traffic over the other less sensitive applications, providing further bandwidth to ensure the call does not drop [7, 15].

Table 2 [16–18] compares three significant networking technologies: P4, OpenFlow, and extended berkeley packet filter (eBPF). These three tools/applications differ in design, flexibility, and applicability. P4 provides much higher flexibility in programming data planes with the possibility of programming PISA chips and supporting any protocol suitable for the current mixed network topologies. OpenFlow, primarily applied in SDN to manage the forwarding tables, has less flexibility

Aspect	P4	OpenFlow	eBPF[a]
Purpose	Program data plane	Forwarding tables	Kernel ext.[b]
Flexibility	High	Limited	High
Programming model	Top-down	Bottom-up	Flexible
Chip dependency	PISA chips	Fixed ASICs	Software
Protocol support	Any protocol	Fixed headers	Any data
Interoperability	Common language	TTP standards	Universal
Control plane	SDN control	SDN separation	Via API
API generation	Autogenerates	Fixed API	Dynamic
Evolution	Software updates	Hardware updates	Kernel updates
Use Case	IPv4 steps	Data centers	Monitoring
Future	Mixed networks	Fixed switches	Expanded use
Community	P4.org	ONF	Linux Foundation

[a]*eBPF allows for high flexibility to extend and customize the functionalities of the Linux kernel.*
[b]*Kernel ext. let developers load code directly into the kernel communications dynamically.*

Table 2.
Comparison of P4, OpenFlow, and eBPF technologies [16–18].

and relies on fixed application specific integrated circuits (ASICs) and hardware upgrades, which works best in data centres where topology remains constant and well-defined.

On the other hand, eBPF is visible because it can augment and improve the linux kernel functionality more flexibly online, which can cover any data. This supports numerous usages, especially in performance analysis and security measures through intercommunication in the kernel space. This is somewhat true since eBPF and P4 offer more dynamic and autogenerated APIs compared with the fixed API of OpenFlow. Thus, the further development of these technologies, backed up by communities such as the P4, continues smoothly. It is discerned from the involvement of critical associations such as Open Networking Foundation |org, ONF, and linux foundation and observed that they hold a significant position and have evoked certain unique usage areas in contemporary network situations.

NFV enables mobile operators to quickly deploy virtualization across geographically dispersed environments, significantly improving the dynamic management of network traffic flows and service quality. SDN and NFV integration in mobile networks enhances applications and services such as virtual reality (VR) and augmented reality (AR) since they require high bandwidth and low latency [7, 15].

Thus, based on the current advancements in wireless networks, the functions of SDN and NFV in QoS delivery are gaining more importance. Mobile applications' programmability and capability to operate in a versatile and scalable fashion are required to satisfy the increased requirements posed by the current mobile application and services environment [7, 15]. 5G and IoT were introduced utilizing SDN and NFV architecture to enhance functionality in wireless networks regarding scalability, data handling, and the provision of diverse quality services [19].

The current work focuses on how SDN and NFV transform network management and QoS. These technologies empower networks with the capabilities they need to address the challenges of today's network landscape and define the

evolution of further advances in network technologies. With the help of SDN and NFV technology concepts, networks can introduce the ability to slice the networks to utilize different sections of the networks for various types of traffic that require varying QoS levels. For instance, a network slice intended for emergency services can be designed to reflect the main parameters as reliability and low delay. In contrast, a video-stream slice should enhance the goals, such as bandwidth and throughput [11, 19].

Multi-access edge computing (MEC) is another area of interest for SDN, which deals with traffic analysis and decision-making that can happen nearer to the mobile users in the network. It minimizes latency and enhances the performance of real-time processing applications such as augmented reality or self-driving cars [19, 20]. The computing and services are offered closer to the user, thus reducing the delays significantly. This is supported by the deployments of 5G networks providing deeper penetrations into the networks. Therefore, heterogeneous network deployments based on MEC, SDN, and NFV can offer a significant improvement in service quality.

NFV cooperates by enabling vRAN to optimize resources by user traffic and the required QoS. This flexibility is essential in ensuring that mobile network users receive a quality experience, especially during peak hours or in areas with high traffic density [3, 8]. Machine learning and artificial intelligence are also incorporated into SDN and NFV to support QoS and other relevant aspects. These technologies facilitate the creation of models that predict traffic flow and network incidences, addressing QoS concerns when service levels are affected by changes in network conditions [15, 21].

Table 3 offers more specifically developed descriptions of the evolution of networking tools such as NetConf/YANG, OpenFlow, Open vSwitch, P4, Ryu SDN Framework, and the ONOS platform related to QoS provisioning. Technologies, frameworks, protocols, and algorithms have evolved by extending advanced QoS features to encompass a variety of functionalities, including device configuration and rate limitations. The table depicts a steady transition of current wireless and mobile networks from 3G settings to optimal 5G configurations. It introduces the developments in various technologies, focusing on resource scaling, ubiquitous connectivity, and particularly highlights the QoS requirements and optimal network resources allocations.

3. Challenges in managing QoS in softwarized and virtualized networks

Managing quality-of-service (QoS) in networks that leverage softwarization and virtualization presents unique challenges. These challenges stem from the dynamic nature of these networks, the variability in network conditions, and the complexity of resource allocation.

Dynamic of network traffic: The new paradigm of virtualized networks consequently brings dynamism into network traffic and conditions compared to traditional ones. This causes fluctuations in the capacity to safeguard the QoS because the network has to adjust to emerging requests and traffic requirements [22].

Scalability and elasticity problem: Although virtualization enables the scalability of a system, managing this scalability to maintain QoS requires innovative solutions. Scalability, the ability to adjust resources in a network guaranteeing quality-of-service, is challenging. Even with this, the mechanical method of resource allocations has

Tool	Org.	Year	Ver.	QoS Features	Wireless Impact	Mobile Networks Impact
NetConf /YANG	IETF	2006	1.0	Initial device configuration capabilities	Basic device config. management	Early 3G configurations
NetConf /YANG	IETF	2011	1.1	Enhanced datastores support	Enhanced device management	Improved LTE configurations
OpenFlow	ONF	2008	1.0	Basic rate-limiting and VLAN tagging	Basic QoS in Wi-Fi networks	Traffic management trials in 4G
OpenFlow	ONF	2009	1.1	Fine-grained flow control	Improved Wi-Fi network management	Enhanced mobile broadband control
OpenFlow	ONF	2011	1.2	Support for MPLS labels and group tables	Better traffic segregation in Wi-Fi	Enhanced network slicing in LTE
OpenFlow	ONF	2012	1.3	Robust metering features	Advanced Wi-Fi management	Advanced 4G traffic management
OpenFlow	ONF	2013	1.4	Per-flow metering and egress queue management	Real-time QoS management in Wi-Fi	High-speed mobile data QoS
OpenFlow	ONF	2015	1.5	Scalable flow aggregation	Large-scale Wi-Fi QoS management	5G network management
Open vSwitch	Open vSwitch Community	2009	1.0	Basic virtual switching	Support for Wi-Fi virtual networks	Data center support for mobile networks
Open vSwitch	Open vSwitch Community	2012	1.4	Traffic shaping and QoS rate limiting	Wi-Fi hot spot traffic management	Cloud services for mobile operators
P4	P4.org	2013	1.0	Basic packet processing	Flexible packet handling in Wi-Fi	Limited mobile networks adoption
P4	P4.org	2016	2.0	Real-time packet processing	Customizable network behaviors in 5G	5G testing and rollouts
Ryu SDN Framework	OSRG	2011	1.0	Basic OpenFlow support for QoS	Experimental wireless setups	Mobile network experimentation

Tool	Org.	Year	Ver.	QoS Features	Wireless Impact	Mobile Networks Impact
Ryu SDN Framework	OSRG	2014	3.0	Better OpenFlow 1.3 support	Dynamic QoS in metropolitan Wi-Fi	4G network enhancements
ONOS	ON.Lab	2014	1.0	Basic flow control	Large-scale Wi-Fi management	City-wide mobile networks
ONOS	ON.Lab	2015	1.1 Blackbird	Performance enhancements	High-density Wi-Fi handling	Mobile backhaul management
ONOS	ON.Lab	2015	1.4 Drake	Intent-based networking	Simplified Wi-Fi operations	Automated mobile operations
ONOS	ON.Lab	2016	1.9 Falcon	Scalability and protocol support	Complex Wi-Fi deployments	Evolving LTE support
ONOS	ON.Lab	2017	1.8	Bandwidth calendaring	Next-gen Wi-Fi QoS	5G bandwidth management
ONOS	ON.Lab	2017	2.0 Hummingbird	High availability and security	High-demand wireless scenarios	Large-scale mobile network QoS
ONOS	ON.Lab	2019	2.2	Adaptive streaming and resource allocation	Enhanced wireless streaming	Improved mobile network resource allocation

Table 3.
Evolution and impact of networking tools on wireless and mobile networks.

to be accurate to avoid underutilization, leading to resource wastage and ultimately substandard services, under provision leading to degraded service [9].

Resource allocation complexity: Managing resources within virtualized networks is more challenging because they are abstracted, making it difficult to access them directly. Managing and organizing physical resources must maintain the service's QoS. This is not just about implementing network or computation resources but also about near availability, responsiveness, and performance as perceived by the user and quality-of-experience (QoE) has evolved [12].

Interference and congestion: Based on the levels of virtualization distributed systems, resources are shared at different levels of service and tenant layers, which results in interference and congestion. Such problems can negatively impact the quality of delivered services, especially if the pertinent services have high priorities and are offered alongside those of lower priorities. Proper isolation measures and traffic prioritization to reduce these risks [14].

Maintaining QoS because of different network slices: Network slicing is utilized in softwarized networks provisioning multiple logically separated virtual networks over a common physical substrate. Every slice can be tailored for a particular type of service; however, achieving a uniform QoS on these slices is difficult due to different services' different requirements and SLAs [19].

Resource limitation constraints in switching fabric:

In the context of network softwarization and virtualization, the virtualization fabric plays quite an essential role in handling the network's data flow; however, the characteristic resource constraint in the switching fabric presents serious issues primarily as the network expands. This section analyzes the issues related to the switching fabric's ability to process multiple data streams and buffer management and the effect of these issues on the network's performance and QoS.

The general disadvantages of switching fabric are always finite bandwidth, a smaller buffer size, and limited processing capability. However, its disadvantages appear when the network load is high, which results in packet loss and high latency. Since traffic and the number of connected gadgets increase, such problems intensify and influence the network's capability to guarantee constant and high-quality service [23].

Network designers and operators use strategies to overcome these difficulties. One practical construct applies sophisticated traffic control techniques catering to the specific QoS requirements. For instance, traffic shaping and load balancing assist in controlling the load on the network resources necessary for priorities [24]. Furthermore, buffer management techniques are critical in optimizing limited space. Algorithms for dynamic buffer allocation, which adjust the buffering strategies based on real-time traffic conditions and network congestion levels, are particularly effective. This adaptive buffering helps smooth out burst traffic and reduces the risk of buffer overflow, thereby minimizing packaging and maintaining QoS integrity [25].

Another strategy involves architectural enhancements in the switching fabric design. Implementing scalable fabric architectures, such as close networks or hyperscale network architectures, provides pathways for expanding capacity and routing flexibility to handle increasing traffic loads without a corresponding increase in packet loss or service degradation [26]. These innovative strategies are crucial for network architects and engineers as they design systems to support high-density, high-throughput network environments without compromising service quality. Understanding these limitations and employing appropriate mitigation techniques are essential for maintaining network infrastructures.

4. Advanced techniques for QoS management in softwarized and virtualized networks

In response to the challenges highlighted in the previous section, several advanced techniques have been developed to improve the management of QoS in softwarized and virtualized networks. These techniques leverage innovations in algorithmic solutions, machine learning, and policy-driven frameworks to manage network resources dynamically.

Machine learning for predictive QoS management: Predictive QoS management is another outgrowth of network awareness that has used machine learning (ML) to effectively predict traffic flow and start preparing for change before it happens. Doing some work to acquire ML models that foresee high-demanding situations and, consequently, order additional resources to maintain QoS goals without disruption. For example, deep learning strategies have been widely applied to predict network loads with a subsequent efficient allocation of resources [27].

Policy-driven management frameworks: Managing frameworks is essential to setting policy reforms on utilizing these frameworks, which focus on the resources available, the business objectives, and the QoS. Thus, through automation of the enforcement of these policies, it is possible to manage the fluctuations in service requirements efficiently and continue offering the required QoS [28].

Network function virtualization: NFV orchestration plays a significant role in controlling the lowly virtualized network life cycle. It also orchestrates the scaling and healing of VNF across the network segments. In other words, ideal NFV orchestration guarantees the evolution of a network to meet service demands without compromising the QoS at the end-user level, as well as efficient resource utilization to avoid bottlenecks [11].

Adaptive traffic management (ATM): ATM has been defined as coordinating the transmission of data packets in a network according to the prevailing conditions and appropriate QoS specifications. Methodologies like traffic prioritization, prioritization distribution, and priority-controlled queuing make allocating bandwidth and latency to vital services possible even when the rest of the traffic is heavy [13].

AI-driven decision-making: AI goes with ML to supplement decision-making features that allow responsiveness to a situation and proactivity. Automated systems can quickly determine priorities for resource allocation and traffic control within required key performance indicators (KPIs) and apply new strategies based on the dynamics of the network. This leads to highly adaptive and enduring network infrastructure [15].

5. Case studies and real-world applications of network softwarization and virtualization

The application of network softwarization and virtualization has proven effective in various real-world scenarios, demonstrating their potential to revolutionize work management and improve QoS. This section presents several case studies highlighting successful deployments and the lessons learned from these experiences.

5.1 Case study: Virtualization in mobile networks

Virtual network functions are placed closer to the user's end, eliminating delays and damaging tasks needing immediate response. Also, using software, networks'

flexibility and scalability are enhanced through software-defined networking, which ensures that the networks can handle varying loads. Outsourcing of all functions in a network implies that several tasks are guaranteed to be fulfilled without common human errors, thus improving the network services. Tasks such as safeguarding that are central to ensuring the delivery of services with security are also elaborated. Collectively, these enhancements are possible and provide a solid foundation and flexibility necessary for a network infrastructure in 5G technologies [29].

5.2 Case study: Softwarization for IoT integration

A use case in a smart city involves the possibilities of employing SDNs in convergence with the Internet of things (IoT), where considerable enhancements in managing the traffic networks and improving the QoS have been evident (**Figure 4**). Through SDN, the city's administration could control a now dynamically constructed web of IoT devices such as traffic lights, surveillance systems, or public Wi-Fi systems. The SDN controller uses real-time data to prioritize certain types of traffic, to control resource usage during a specific time of the day, and thus never compromise essential services like emergency responses. This helped avoid statistics of traffic jams and improved the time taken in critical communication. Also, incorporated with edge computing, the author's proposal ensured that data computation was carried out closer to the data collecting centre, reducing time for general operational responses. This was particularly helpful for the real-time applications that are strategic to the running of the city, traffic, and security [30].

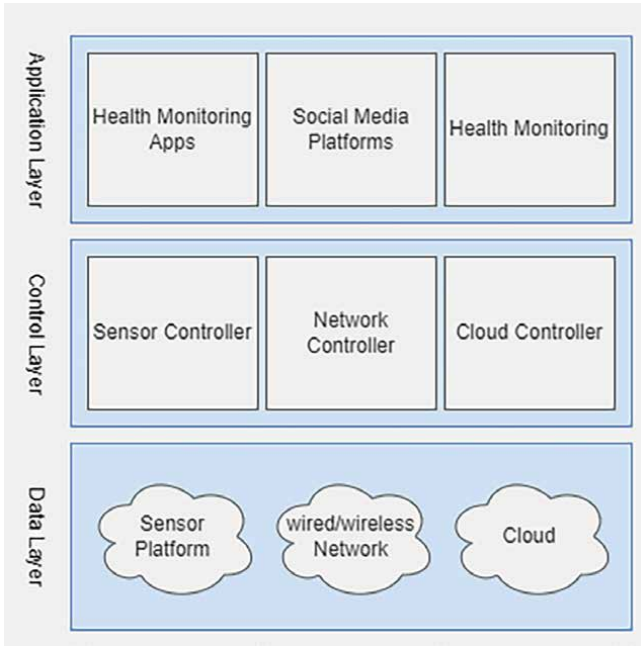

Figure 4.
Architecture of SDN for IoT devices to improve the QoS.

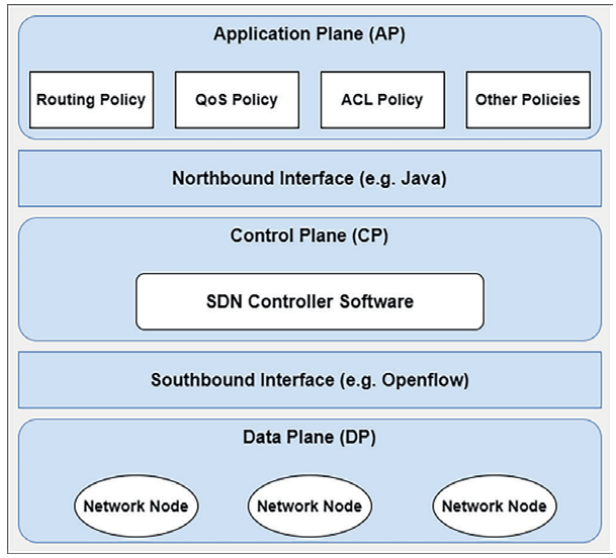

Figure 5.
The general SDN architecture.

Figure 5 also demonstrates the layers of infrastructure, control, and application; this shows the connectivity of the different network components for smooth traffic flow. The architecture focuses on the aspects of open and flexible management of resources needed for QoS through monitoring and controlling them in real time with the help of SDN, which contributes to QoS enhancement due to dynamic and flexible management of resources required for QoS through the use of real-time monitoring and controls options available in SDN.

5.3 Case study: Enhancing connectivity via open RAN: Leveraging SDN for optimal QoS management

The case study proposes a QoE enhancement function (QoE^2F) xApp for 5G Open RAN coupled with an adaptive genetic algorithm that deals with the issues of user association, resources, and power in HetNets. Through simulations, this approach raises the quality-of-experience for users when accessing high-resolution video services by a large margin. This is due to optimizing Near-RT RIC within the Open RAN environment, which leads to a better provision of video services regardless of the density of the user and network scenario [31, 32].

As depicted in **Figure 6**, the standardized data from the distributed unit (DU) and the control unit (CU) through the A1 interface is used by the non-RT remote intelligent controller (RIC) to improve the management of network applications such as policies and QoS. On the other hand, the Near-RT RIC utilizes the E2 interface to orchestrate a change in policies and management of the RAN resources and, therefore, have a direct impact on QoE by allocating available resources to areas of demand in quasi-real-time. These are in addition to the already opened F1 and fronthaul interfaces that enhance multi-vendor operation within the distributed unit remote resource unit (DU-RRU) link and diverseness of networks, which are mandatory for efficient service provision.

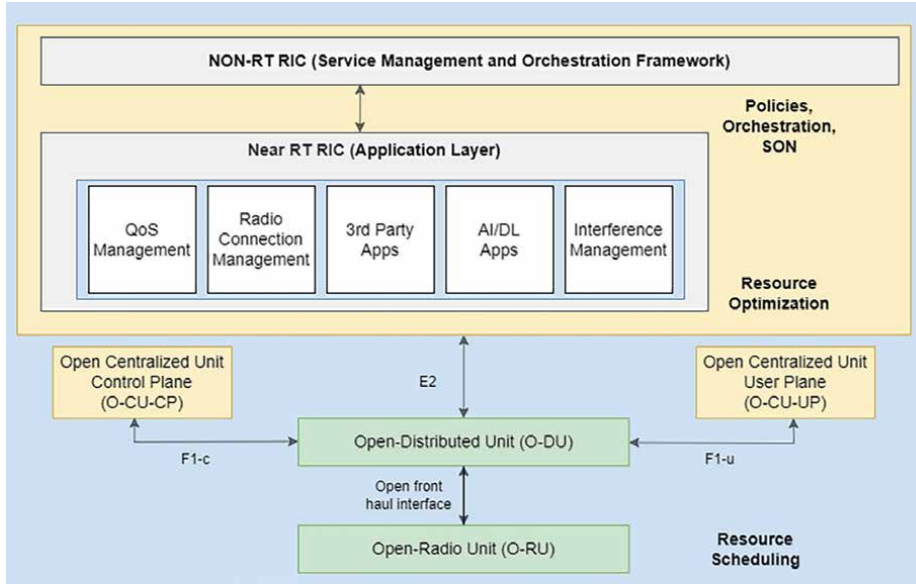

Figure 6.
Architecture of Open RAN to improve the QoS.

6. Future trends and innovations in network software and virtualization

As network technologies continue to evolve, the landscape of network softwarization and virtualization significantly changes with the advent of new technologies and

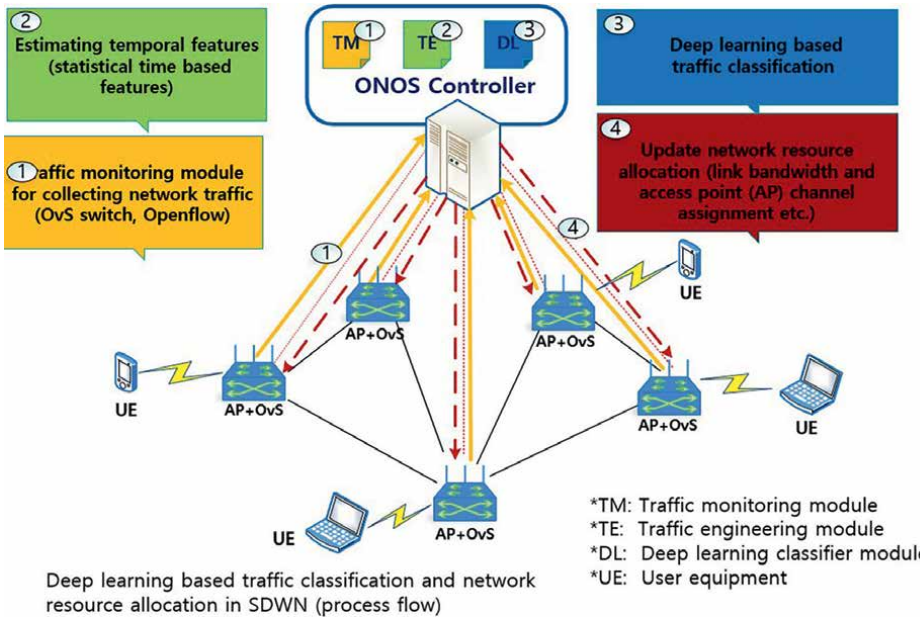

Figure 7.
Process flow of deep learning-based traffic classification and network resource allocation in SDWN.

methodologies. This section explores future trends and innovations that could shape the management of wireless and mobile networks, highlighting the potential for continued QoS advancements.

Compatibility with 5G and beyond: Today, the process of creating popular 5G networks has started, and the work on building 6G is being discussed. They envisage a next-generation network with higher bit rates, low delay, and enhanced availability. Software is also crucial to attaining these advantages because it integrates softwarization and virtualization. It is foreseen that future improvements will incorporate AI-based methods to improve functions in a network, consequently improving the ability to maintain and control QoS on a dynamic level [22].

Roles of artificial intelligence: Undoubtedly, AI integration is supposed to enhance the networks' performance and adaptability to respondents' needs with high levels of performance and security [21]. In the case of incorporating artificial intelligence in software-defined wireless network (SDWN), as illustrated in **Figure 7**, the system has several specific modules managed by the ONOS controller. The traffic monitoring (TM) module: Real-time data of the existing network structure is collected by implementing OvS switches within the commodity platforms, such as Raspberry Pi, and configuring them as an access point. The southbound API utilized could be OpenFlow. The data collected is retrieved by the controller via the TM module. The data is then fed to the traffic engineering module (TE). This data is significant because temporal features' assessment is made by statistical analyses considering network traffic characteristics over time. After that, a deep learning (DL) classifier module utilizes this analyzed data to sort the traffic based on predefined criteria such as type and priority. This classification is beneficial for the next stage of dynamic resource allocation wherein the link bandwidth and the access point channel assignment will be changed to improve the network's performance. The entire process is represented, starting from data collection and then resource management, resulting in the optimization of the SDWN network, and hence, a workable solution is presented to attain the required in the wireless network.

Expansion of edge computing: Another sub-segment that is likely to experience immense growth is edge computing. However, as data generation rises, mainly because of IoT devices, edge computing provides a model to process data near the source rather than in data centres. This lowers latency and bandwidth usage, which is essential for uses that demand real-time processing [20].

Enhanced security protocols: Security remains paramount, especially as networks become more open and interconnected. Future trends in softwarization and virtualization development of advanced security protocols that can adapt to emerging threats dynamically. These protocols will likely integrate machine learning techniques to predict and mitigate risks before they impact network performance or compromise data integrity [33]. Applying the multi-layer knowledge graphs in the context of threat perception in cybersecurity makes one appreciate how advanced semantic networks might be used to improve the threat detection and analysis implemented in cyberspace by employing the restricted Boltzmann machine (RBM) [34].

Integration with blockchain: To do so, focus on the risks introduced by integrating communication technologies and IoT devices, and how blockchain can strengthen it. It assesses blockchain performance with AI solutions, improving the means of identifying and protecting from threats within cyberspace effectively [35].

7. Conclusion

The contemporary progress of network softwarization and virtualization in administering wireless and mobile virtualization, an exponential traffic load, and complicated network structure aim to satisfy the ever-rising expectations for QoS; more precisely, it increases the flexibility and efficacy of network management, thus opening the way to new and advanced solutions. This work focused on detailing the conceptual models of softwarization and virtualization in wireless and mobile networks. Aspects such as software-defined networking (SDN) and network functions virtualization (NFV) advancements have led to current networks' intelligence and enabled flexibility in responding to traffic and user characteristics. The work details the evolution of quality-of-service (QoS) provisioning and the related progress in the academic, industry, and research community. The work comprehensively states the significant contributions of SDN and NFV in achieving today's networks' service performance, leading to promising solutions and paving the way for futuristic networks.

The dynamics of the technological structure of networks with intelligent service and artificial intelligence remain a significant opportunity and challenge to guarantee quality service delivery. This work emphasized the need to employ algorithmic prescription, edge computing, and policy-based architectures to operate resources and networks proactively. The synergy of SDN and NFV has been paramount in delivering required service performance to network operators and users, thus recognizing softwarization and virtualization as an established core of future network management as the network environments evolve.

Author details

Umair Shafiq Khan[1†] and Tahira Mahboob[2*†]

1 Department of Electrical Engineering, Information Technology University of the Punjab, Lahore, Pakistan

2 Department of Computer and Software Engineering, Information Technology University of the Punjab, Lahore, Pakistan

*Address all correspondence to: tahira.mahboob@itu.edu.pk

†These authors contributed equally.

References

[1] Cisco Annual Internet Report (2018-2023) White Paper. Cisco. 2020. Available from: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html [Accessed: June 25, 2024]

[2] Nadeau TD, Gray K. SDN: Software-Defined Networks. Sebastopol (CA): O'Reilly Media, Inc.; 2013, 384 p. ISBN 978-1-4493-4230-2.

[3] Moreno V, Reddy K. Network Virtualization. 1st ed. USA: Cisco Press; 2006

[4] Vinayakumar R, Soman KP, Poornachandran P. Applying deep learning approaches for network traffic prediction. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2017. Udupi, India: IEEE; 2017. pp. 2353-2358. DOI: 10.1109/ICACCI.2017.8126198

[5] Barakabitze AA, Ahmad A, Mijumbi R, Hines A. 5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges. Computer Networks. 2020;**167**:106984. DOI: 10.1016/j.comnet.2019.106984

[6] McKeown N et al. OpenFlow: Enabling innovation in campus networks. ACM SIGCOMM Computer Communication Review. 2008;**38**(2):69-74

[7] Kim H, Feamster N. Improving network management with software-defined networking. IEEE Communications Magazine. 2013;**51**(2):114-119. DOI: 10.1109/MCOM.2013.6461195

[8] Han B, Gopalakrishnan V, Ji L, Lee S. Network function virtualization: Challenges and opportunities for innovations. Vol. 53. no. 2. IEEE communications magazine; 2015. pp. 90-97

[9] Fischer A et al. Virtual network embedding: A survey. IEEE Communications Surveys & Tutorials. 2013;**15**(4):1888-1906. DOI: 10.1109/SURV.2013.013013.00155

[10] Jarschel M et al. Interfaces, attributes, and use cases: A compass for SDN. IEEE Communications Magazine. 2014;**52**(6):210-217. DOI: 10.1109/MCOM.2014.6829966

[11] Homma S et al. NFV orchestration with dynamic QoS adjustment for network slicing in 5G networks. IEEE Network. 2019;**33**(2):164-171

[12] Chandra A, Gong W, Shenoy P. Dynamic resource allocation for shared data centres using online measurements. ACM SIGMETRICS Perform Evaluation Review. 2003;**31**(1):300-301. DOI: 10.1145/885651.781067

[13] Irazabal M, Lopez-Aguilera E, Demirkol I. Active queue management as quality of service enabler for 5G networks. In: 2019 European Conference on Networks and Communications (EuCNC). Valencia, Spain: IEEE; 2019. pp. 421-426. DOI: 10.1109/EuCNC.2019.8802027

[14] Zhang H, Liu N, Chu X, Long K, Aghvami A-H, Leung VCM. Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine. 2017;**55**(8):138-145. DOI: 10.1109/MCOM.2017.1600940

[15] Chen A, Law J, Aibin M. A survey on traffic prediction techniques using artificial intelligence for communication networks. Telecom. 2021;**2**(4):518-535. DOI: 10.3390/telecom2040029

[16] Open Networking Foundation (ONF). Available from: https://www.opennetworking.org/ [Accessed: June 28, 2024]

[17] eBPF Documentation. Available from: https://ebpf.io/what-is-ebpf/ [Accessed: June 30, 2024]

[18] P4 Open Source Programming Language. Available from: https://p4.org/ [Accessed: June 30, 2024]

[19] Ordonez-Lucena J et al. Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges. IEEE Communications Magazine. 2017;**55**(5):80-87. DOI: 10.1109/MCOM.2017.1600935

[20] Tan L et al. Edge computing in next-generation networks: SDN and NFV as enablers. IEEE Communications Magazine. 2019;**57**(2):42-48

[21] Turner N et al. AI in network management: From theory to practice. IEEE Network. 2019;**33**(6):204-211

[22] Zhao M et al. The impact of 5G on network management: Opportunities and challenges. IEEE Wireless Communications. 2021;**27**(4):70-77

[23] Docker for the Virtualization. 2016. Available from: https:/virtualizationm/what-container [Accessed: June 28, 2024]

[24] Patel A, Smith J. Challenges in network switch design for scalable fabrics. Journal of Network Engineering. 2020;**45**(2):134-145

[25] Liu Y, Zhang H. Traffic management algorithms for improved network QoS. IEEE Transactions on Networking. 2021;**38**(6):1652-1667

[26] Intelligent Buffer Management on Cisco Nexus 9000 Series Switches [White Paper]. Available from: https://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-seriesswitches/white-paper-c11-738488.html [Accessed: July 30, 2024]

[27] Zhang C et al. Deep learning for network traffic control in SDN-enabled networks. IEEE Access. 2019;**7**:137522-137535

[28] Giotis K, Kryftis Y, Maglaris V. Policy-based orchestration of NFV services in software-defined networks. In: Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft). London, UK: IEEE; 2015. pp. 1-5. DOI: 10.1109/NETSOFT.2015.7116145

[29] Little AD. Mobile Network Architecture. Available from: https://www.adlittle.com/sites/default/files/reports/adl_mobile_network_architecture.pdf [Accessed: July 25, 2024]

[30] Ja'afreh M, Adhami H, Alchalabi AE, et al. Toward integrating software-defined networks with the internet of things: A review. Cluster Computing. 2022;**25**:1619-1636. DOI: 10.1007/s10586-021-03402-4

[31] Agarwal B, Togou MA, Ruffini M, Muntean G-M. QoE-driven optimization in 5G O-RAN-enabled HetNets for enhanced video service quality. IEEE Communications Magazine. 2023;**61**(1):56-62. DOI: 10.1109/MCOM.003.2200229

[32] Ndikumana A, Nguyen KK, Cheriet M. Renewable energy powered and open RAN-based architecture for 5G fixed wireless

access provisioning in rural areas. Vol. 8. no. 3. IEEE Transactions on Green Communications and Networking. Sep 2024. pp. 994-1007. DOI: 10.1109/TGCN.2024.3431989

[33] Molina, Zarca A, Bernal Bernabe J, Farris I, Khettab Y, Taleb T, Skarmeta A. Enhancing IoT security through network softwarization and virtual security appliances. International Journal of Network Management. 2018;**28**(5):1-18, e2038. DOI: 10.1002/nem.2038

[34] Sali A, Al-Jumaily A, Jiménez VPG, Al-Jumeily D. Cybersecurity threat perception technology based on knowledge graph. Journal of Autonomous Intelligence. 2023;**6**(3):1-8

[35] Mohammed SH, Al-Jumaily A, Singh MSJ, Jimenez VPG, Jaber AS, Hussein YS, et al. A Review on the Evaluation of Feature Selection Using Machine Learning for Cyber-Attack Detection in Smart Grid. IEEE Access. 2024;**12**:44023-44042

Chapter 3

Quality of Service Challenges in Hybrid IP/Software-Defined Networks

Samiullah Mehraban and Rajesh K. Yadav

Abstract

In recent decades, there has been a tremendous growth in the network traffic volume, which causes substantial challenges for effective traffic management. Traffic engineering (TE) is an efficient strategy for minimizing network congestion and improving network performance. In today's network, supporting end-to-end quality of service is an ongoing problem in the existing network infrastructures. Software-defined networking (SDN) designs were introduced to solve these challenges by separating the control and data layers. However, migrating directly to a full SDN is challenging, and the optimal option is to migrate incrementally to a hybrid SDN where conventional network devices coexist with the SDN nodes to take the benefits of both networks. This study investigates challenges with hybrid SDN traffic engineering and quality of service. Quality of Service (QoS) in a hybrid SDN is defined by network parameters such as jitter, latency, loss, bandwidth, and network link utilization, measured using industry standards and procedures. We can witness considerable gains in overall network performance and user experience by implementing advanced QoS regulations in the hybrid environment. Using machine learning techniques for adaptive QoS and traffic prediction can increase the network's ability to deal with diverse and unpredictable traffic patterns.

Keywords: SDN, hSDN, TE, MLU, QoS

1. Introduction

In today's network, one of the main challenges that professionals and industries face is the complexity of the network, which is challenging for network providers to manage such a complex network. The Software-Defined Networking (SDN) concept was initiated in response to these challenges. SDN works on the fundamental principle of separating the data and control planes. In recent years, SDN has been one of the most well-discussed subjects in telecommunication systems. SDN is an essential network architecture developed alongside virtualization, mobility, the Internet of Things, and other technologies. SDN offers dependable and adaptable network administration; it enables intelligent flow scheduling systems to enhance network

performance and link utilization, and it supports advanced Traffic Engineering (TE) because of its global perspective and centralized network control.

Similar to the rapid growth of the Internet, traffic volume has increased significantly during the past few decades, as managing tremendous traffic is challenging. TE is a practical way to reduce network congestion and improve network performance; it has drawn much interest from academics and industry in recent years. Traffic engineering is a potential real-time approach that measures and analyzes network traffic to enhance traffic routing and balance network flow. Traffic engineering is a technique that looks at data transmission behavior over communication networks to improve network performance.

Optimizing network performance is the primary goal of traffic engineering to increase network reliability [1]. It can be achieved by strengthening the network's resilience to faults, reducing network congestion by evenly distributing the load among the links, etc. In contrast, a centrally controlled controller in an SDN network communicates with forwarding components to maintain a global perspective, traffic demand, network topology, and link condition information. The SDN network's direct routing channels set it apart and make it a top choice for many traffic engineering solutions.

SDN offers a valuable and adaptable method of routing optimization; with the separation of the control and data layers and the logical centralization of the control layer, traffic can be dynamically divided and delivered to different channels using SDN switches since the controller controls flow routing in the control plane [2]. Programmability and globally centralized network management are compelling SDN network features that enable traffic engineering. With all the significant advantages and futures SDN provides, directly migrating to a fully SDN is challenging, and the optimal solution is to migrate to a hybrid SDN first, then to a fully SDN environment. Employing particular complete TE approaches in a hybrid SDN network is possible because the flow splitting ratio on SDN switches is flexible. To increase centralized control's profitability, network providers who want to migrate to hybrid SDN must choose an appropriate migration sequence considering various optimization techniques. These methods can be used with other network optimization strategies and hybrid network deployment. In a hybrid SDN network, traffic engineering is more complex; solutions based on static routes or access control lists (ACLS) offer limited capabilities [3].

TE in a hybrid SDN is challenging since an SDN controller does not entirely control traditional devices. The controller in the hybrid SDN can only manage SDN nodes to enable TE; conventional legacy devices will be controlled through legacy devices themselves. The most widely used traffic engineering techniques in conventional legacy networks are MPLS (Multi-Protocol Label Switching), which uses labels, and Generalized MPLS, which extends MPLS to handle new switching technologies and interfaces. Regarding optimizing routing, this traffic routing provides greater flexibility than the ECMP (Equal-Cost Multipath) utilized in the conventional OSPF (Open Shortest Path First) protocol.

The traffic engineering method recommended for the hybrid SDN network will be examined in this article. A traffic engineering technique called load balancing seeks to maximize the distribution of traffic loads across various resources following a predetermined performance threshold. Since load balancing looks for the average link rate usage for all network links, it may achieve the TE aim of minimizing maximum link utilization in the hybrid SDN network by making network congestion accessible.

The OSPF protocol is one of the most used IGP (Internet Gateway Protocol) in conventional networks. After assigning it a cost or weight, it chooses the shortest path with negligible forwarding costs for each connection. Equal-cost multipath forwarding (ECMP) routes traffic between the source and destination addresses and distributes it evenly in cases with numerous shortest pathways. In a hybrid SDN environment, OpenFlow, flow abstraction, comprehensive packet matching, and packet filtering techniques are not supported by all nodes in the network; hence, TE employs many optimization strategies to obtain the best possible network performance characteristics.

Data on network state must be gathered, measured, and tracked to perform traffic measurement. The study of managing network traffic based on network status information obtained from traffic measurement is known as traffic management. Traffic measurement is challenging in a hybrid SDN network because all the devices in the network environment are not SDN-enabled, so few devices in the hybrid SDN environment are SDN nodes.

2. Traffic engineering

See **Table 1**.

Reference	Aspect	Objective	Limitation
[4]	IP-TE	To accomplish load balancing, maximize the weight of OSPF links.	The shortest path is used to route traffic, which is a drawback of IP network routing optimization.
[5]	IP-TE	They address the limitation of equitably splitting traffic on an equal-cost multipath to achieve a nearly optimal solution.	Every route prefix necessitates a complex configuration.
[6]	IP-TE	Designed unique link-state routing protocols to maximize traffic engineering efficiency.	It can operate at peak efficiency, but most common devices cannot support it.
[7]	IP-TE	Solve the multi-commodity flow maximization problem to represent the ideal intradomain TE.	They add a second weight to each complex and error-prone link in OSPF routing to allow flexible traffic splitting over the shortest path.
Reference	**Aspect**	**Proposed work**	
[8]	MPLS-TE	They develop a hybrid network with MPLS/OSPF and use a linear programming module to define the traffic engineering problem, which is a multi-commodity flow problem.	
[9]	SDN-MPLS	With an emphasis on an SDN/OSPF hybrid network scenario, they propose a novel TE architecture for SDN/MPLS hybrid networks.	
[10]	SDN-TE	Provide a crowd collaboration solution for SDN vehicular networks designed for SINET (smart identifier networking).	
[11]	SDN -TE	Controlled co-flows by offering a network architecture with performance close to ideal.	
[12]	SDN -TE	Provide an effective resource reallocation system for software-defined data centers.	

[13]	hSDN -TE	Provide a framework for the gradual deployment of hybrid networks and offer a variety of heuristic methods for reducing the maximum link utilization in TE.
[14]	hSDN-TE	SDN was implemented at the border routers, and OSPF was split into various subdomains to allow for more precise traffic control between subdomains.
[3]	hSDN-TE	They studied two hybrid modes: barrier mode, where two forms of traffic are routed in various capacity spaces, and hybrid mode, where any form of traffic can fully utilize each link.
[15]	hSDN-TE	They proposed a novel hybrid switching design that integrates conventional and SDN switching to achieve both scalability and optimal performance.
[16]	hSDN-TE	Present fibbing is a technique for centrally managing the link status of the routing protocol by creating fictitious nodes to provide more effective routing.
[17]	hSDN-TE	Provide an energy-efficient routing technique to reduce energy usage.
[18]	hSDN-TE	They proposed a heuristic algorithm that jointly optimizes the OSPF weight configuration and traffic splitting ratio of SDN nodes.
[19]	hSDN-TE	Studied the problem of TE over multiple traffic matrices and proposed a heuristic approach for optimizing routing over multiple traffic matrices by combining offline weight configuration and online splitting ratio optimization.

Table 1.
Traffic engineering.

3. Quality of services (QoS) in hybrid SDN

Overall, suppose we define Quality of Service for a network. In that case, evaluating user satisfaction with specific services using a composite factor: bandwidth, delay, loss, and jitter. All technologies that manage data flow inside a network to reduce packet loss, latency, and jitter are called Quality of Service (QoS) technologies. Providing specific data transmission in a particular network aspect is what we refer to as Quality of Service.

Considerations for Quality of Service include data accuracy, packet loss, bandwidth, and service level. For instance, QoS ensures bandwidth delivery for bandwidth-sensitive data if the network has exceptional bandwidth; if we have an ideal link, QoS is unnecessary. QoS denotes low delay if the network provides low latency transmission service for time-sensitive data delivery. Since application flows compete for available network resources, they must be differentiated to achieve QoS. Network resources must be assigned to make the best decision for packet forwarding; occasionally, this process necessitates knowledge of the present network conditions. With a broad, dynamic range of Quality of Service features, the architecture supports various communication applications, including those for people and traditional services [20].

In a network, sufficient bandwidth, low latency, low packet loss rates, and balanced network connection loads are all considered aspects of Quality of Service. As is well known, adjusting the network's link load can result in higher bandwidth use and lower latency. Consequently, the latency and bandwidth QoS requirements can be met by balancing the connection load demand.

Service Level Agreements (SLAs) between service providers and end users largely dictate service quality. Although this approach performs exceptionally well for best-effort service, it does not offer more precise granular traffic management. Specific applications, like VoIP (Voice over Internet Protocol), online gaming,

video conferencing, and many more, have flows sensitive to jitter, latency, and bandwidth, so they require unique handling methods. Setting high-level traffic management strategies and establishing limits on the depth of traffic distinction are not well-established methods.

SDN networks separate data and control layers in a network. This separation gives network controllers more authority over networks. Controllers can view the status of the network globally. The network controller offers a conceptual network perspective, providing network applications under the SDN idea. They are not needed to handle data plane device low-level configuration.

3.1 Quality of service requirements

Quality of service parameters in a network is classified into two categories: policies and metrics, as shown in **Figure 1**.

Quality of service measurements include security, performance, cost, and reliability, which are used to establish QoS characteristics. The primary performance measures are latency, loss, and data rates. Rarely, as traffic increases, the network load needs to be balanced; nevertheless, controlling network loads is challenging. QoS regulations are essential to providing a network management tool for bandwidth management. QoS policies have many advantages, including performance, security, manageability, and security.

The SDN network model's benefits may improve service quality in many network operations. SDN provides a global view of the network, making it possible to preserve relevant states over the whole course of flows. It can also monitor network statistics at other levels, such as port, device, and flow. Furthermore, SDN may help network operators create an automated, robust, and user-friendly QoS management framework for their networks by enabling resource reservation, packet scheduling, and queue management. Because network resources are dynamic, control methods must be well-defined for QoS provisioning in network applications.

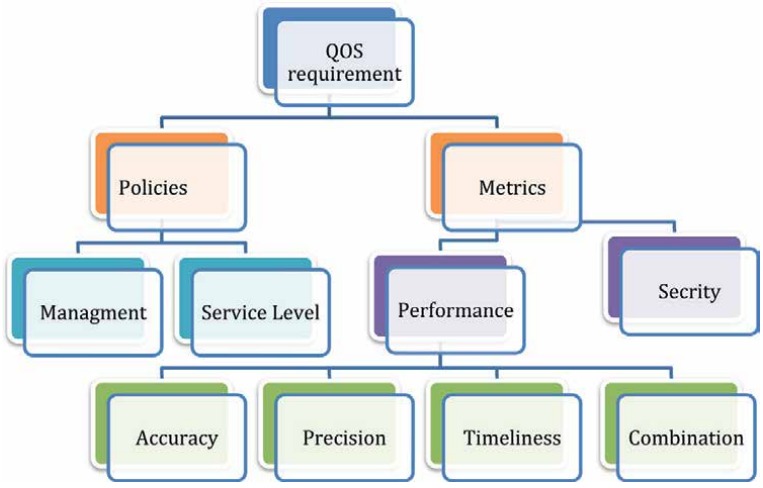

Figure 1.
QoS requirement.

3.2 Quality of service observation

The hybrid network is divided into domains by SDN switches, establishing connections between each domain. Various QoS metrics, including packet loss rate, traffic load, throughput, latency, and so on, must first be gathered to enable QoS-aware forwarding in a hybrid network. After that, these variables are merged to depict network performance precisely.

Traffic load monitoring: The connections to SDN devices may provide load information to the OpenFlow protocol. For a directed link, we have to figure out how many data flows are going through the switch in a certain amount of time. The traffic load on the link can then be determined by computing the number of data flows per unit of time. Flow statistics data is kept in the flow table on an SDN switch. The source/destination IP address, output port, input port, and other matching details for a given routing path are contained in each matching field of a flow entry that corresponds to that path.

Figure 2 demonstrates the hybrid network load monitoring module. In this network, the controller, at each time interval t, sends a flow state request message to the switch S1 over the OpenFlow control channel. It counts the number of data flows that reach port 1 of S2 and compares it with the required flow entries, e11, e12, e13, e2n, etc., using the feedback message it receives.

The SLA message on a link not connected to SDN switches may allow us to obtain information about traffic load. An OSPF router receives SLA broadcasts from SDN nodes and relays the message and its status to the SDN switches. After receiving the SLA from the SDN switch, the controller resolves the message to ascertain the traffic load on that link.

Measuring every flow in a hybrid SDN network would be pointless and prohibitively expensive. The authors proposed a method for gathering the load of specifically designated connections and predicting the remaining amount with a 5% margin of error [21]. An applicable criterion for determining which flows need to be monitored to minimize the estimated error that relies on the flow spread parameter is defined

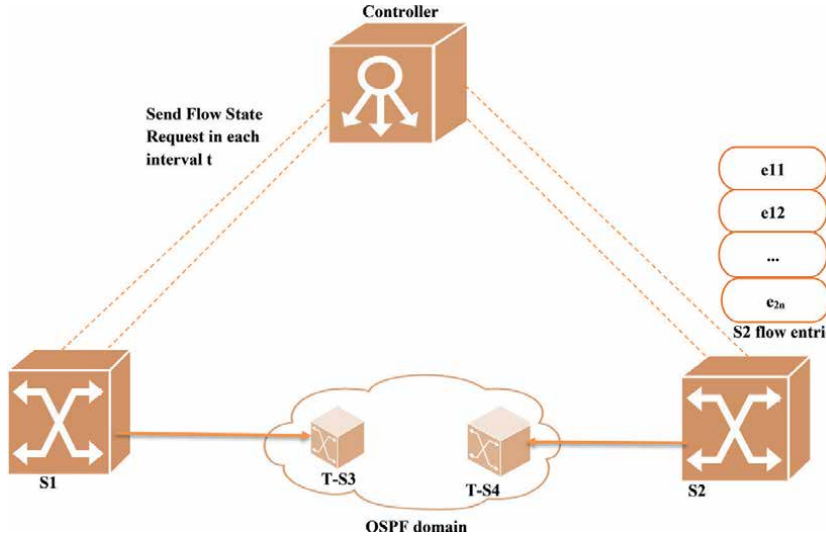

Figure 2.
Traffic load monitoring in a hybrid network.

and assessed by the authors in Ref. [22]. The proposed method may take forwarding table space into consideration and distribute measurement jobs among network devices in an equitable manner.

Figure 3 demonstrates the tunnel construction method in the hybrid network using the penultimate hopping method. In tunnel splicing, the link translator configures the forwarding rule for each device from device n1 to device 1. Depending on the device type, this procedure involves adding OpenFlow entries or assigning MPLS (Multi-Protocol Label Switching) forwarding rules.

Figure 4 depicts a hybrid SDN network that combines SDN with MPLS-based networks. The network topology comprises a mesh of standard MPLS switches at the core and SDN devices at the edges. The network controller injects rules into the MPLS switch routes to push and pop the appropriate labels.

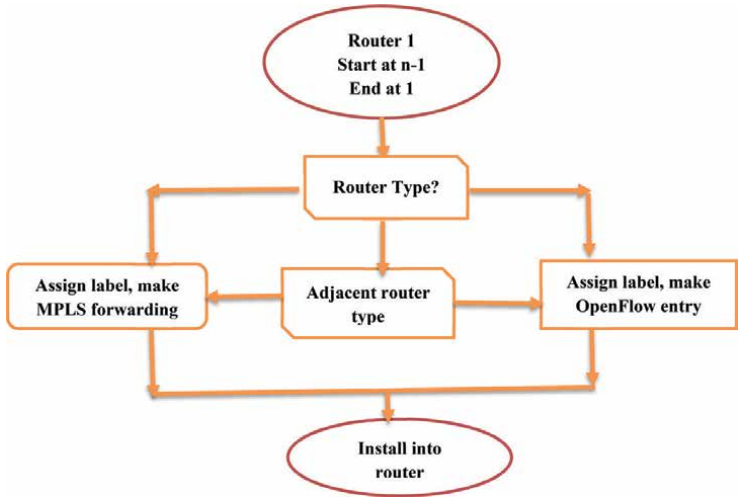

Figure 3.
Tunnel construction procedure.

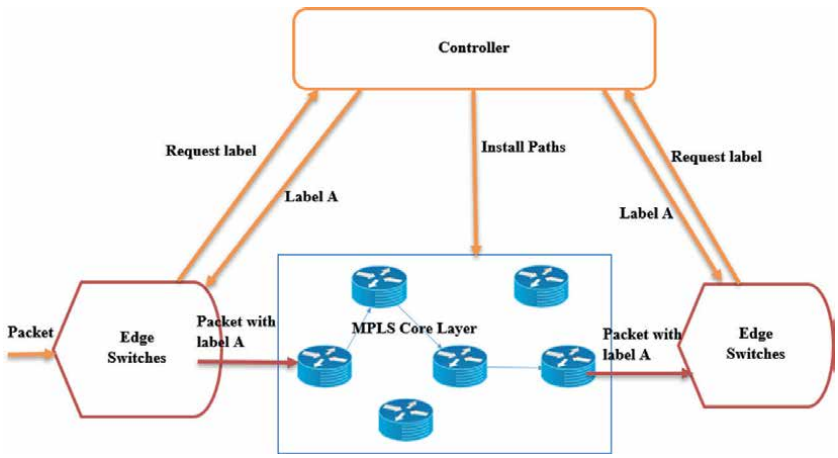

Figure 4.
Installation of flow rules.

Packet loss rate and delay: SDN nodes divide the hybrid network environment into multiple domains in a hybrid SDN environment, frequently linked by SDN switches. The internal domain devices function as OSPF routers; each pair of SDN switches can have several OSPF paths. Determining each link's packet loss rate and latency is problematic [23]. On the other hand, the OpenFlow protocol can be used to calculate the packet loss rate and delay of a channel connecting SDN switches. Although traffic within an OSPF domain does not pass through SDN switches in the hybrid SDN network, a significant portion will pass between multiple OSPF domains, necessitating its passage through one or more SDN switches. As a result, packet loss rate and latency can be estimated.

Figure 5 depicts the delay measurement in a hybrid SDN network; the delay of each path is divided into two components: round trip delay of data flow and transmission delay between SDN switches and controllers. The controller adds the time t1 to the probe data flow before sending the packet-out message to S1. After getting it, S1 sends the message to all ports in compliance with the regulations. This communication arrives at S2 via an OSPF path and is routed as a packet-in message to the controller. The controller receives the message at t2 and uses it to get the transmitting port of S1, the receiving port of S2, and the time at t1.

More investigation has been carried out to enhance the Quality of Service in hybrid SDN networks. They used a variety of traffic management strategies to offer the highest quality of service for their designed topologies, and they proposed numerous ways to optimize single or multiple QoS metrics for hybrid SDN networks.

To improve the QoS of industrial applications, the authors in Ref. [24] proposed a QoS-aware forwarding strategy utilizing the K path partition algorithm and SPMC (Single Path Minimum Cost). Their simulation results show that their proposed technique in an SDN/OSPF hybrid industrial internet effectively balances traffic load using OSPF link bandwidth resources while meeting QoS requirements.

The authors in Ref. [25] presented a comparative optimal traffic control strategy for hybrid SDN QoS optimization. They used public traffic datasets to test their approach. According to their research, this tactic may significantly improve network Quality of Service performance indicators such as latency, jitter, and link utilization.

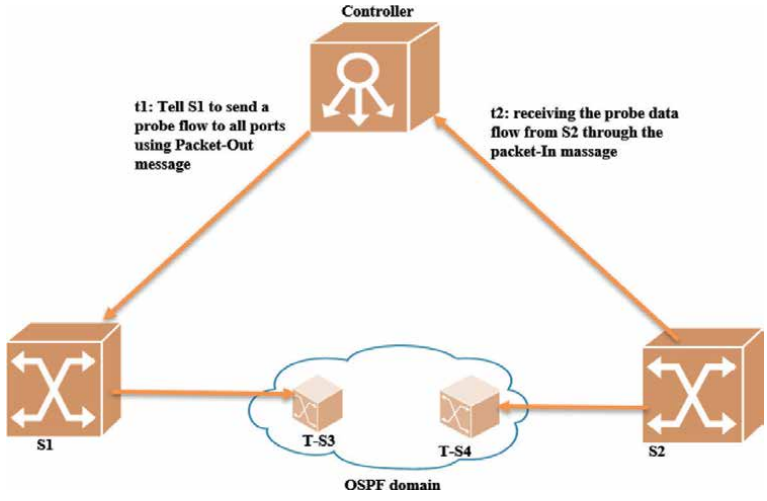

Figure 5.
Hybrid networks delay measurement.

The authors in Ref. [26] proposed a heuristic algorithm for incremental deployment of hybrid SDN under the budget constraint. They employ the depth-first search strategy and a randomized rounding method for throughput maximization routing. They proved that their algorithm can perform much better with the approximation ratio of O (1/log N) than in hypothetical cases. They demonstrated that their recommended incremental deployment could enhance throughput by about 40% compared to the previous deployment approach, and their proposed routing algorithm can improve the throughput by about 31% compared with the ECMP method in the hybrid SDN environment.

The authors in Ref. [27] presented a heuristic approach for deploying SDN devices in hybrid SDN networks. Their strategy functions in two different situations. 1. Improving network control capacity while maintaining the same upgrade spending level. 2. Cutting back on upgrading costs to provide optimal network control performance. The outcomes showed that with only a 10% upgrade cost, this strategy could achieve 95% of the controlled flows.

4. Routing optimization

A novel DRL (Deep Reinforcement Learning)-based architecture that optimizes network QoS performance in hybrid SDN networks is presented by the authors in Ref. [25]. They first search for a suitable SDN migration sequence to maximize total regulated traffic based on past traffic records. Their method emphasizes adaptive algorithms and self-learning routing algorithms under different traffic scenarios, unlike the previous heuristic approach to TE in hybrid SDN networks. The findings show that, compared to other methods, their recommended strategy significantly reduces network latency and jitter and optimizes maximum link utilization with increasing ratios in the deployment of SDN. In traditional distributed networks, increasing OSPF link weight is the primary goal of routing optimization.

Like SDN network development, network operators can choose their routes in real-time networks and control the network and traffic flow much more quickly. Previous studies have concentrated on the routing optimization of fully SDN networks, which differs from the hybrid SDN environment; those routing optimizations cannot be directly applied to hybrid SDN networks.

4.1 AI-driven techniques for optimization

The traffic control issue is different from the deployment challenge. Certain instances could go unconsidered even while using different traffic mattresses to display network traffic. Various efforts have been undertaken using artificial intelligence (AI) methods like DRL (Deep Reinforcement Learning) for flexible routing.

To improve end-to-end communication network throughput, the authors in Ref. [28] provide a novel DRL-based method and review DRL's benefits over earlier dynamic control techniques. To optimize the Quality of Experience (QoE) in an SDN network, the authors in Ref. [29] implement DDPG (Deep Deterministic Policy Gradient) to increase network link weights depending on different incentive targets. Artificial intelligence can enhance cybersecurity for smart grids by automating threat detection and response [30].

The authors in Ref. [31] presented a novel approach to reinforcement learning for TE in an SDN network. Automatically identifying which significant flows need to

be diverted is critical to this approach. At the same time, it works well for any traffic matrix, but hybrid SDN applications are not the best fit. Once deployed, SDN devices can only control a fixed set of flows and limited selection options. The authors in Ref. [32] proposed SINET-based DDPG for SDN routing, in which critic nodes are selected as control nodes, and it constructs routing pathways using its link weights.

4.2 Quality of service routing

The SDN network's open control interface makes it easier to implement a flexible traffic scheduling scheme for the network; it is the ideal architecture for satisfying the Quality of Service requirements of various applications, such as audio and video. Based on Quality of Service routing, the authors in Ref. [33] provide an open QoS concept on an OpenFlow controller with support for end-to-end QoS for multimedia distribution. Traffic pathways are dynamically optimized to fulfill the required QoS. The results of their testing show that open QoS assures transmission, and users report minimal to no visual artifacts when watching the continuous video. They analyze their open QoS across actual work test networks. Moreover, assured services are managed in open QoS without influencing any other network traffic categories, in contrast to current QoS systems.

The authors in Ref. [34] propose a method using SDN nodes to guarantee Quality of Service. Additionally, they use several channels between source, destination, and queueing techniques to offer Quality of Service (QoS) for diverse traffic kinds. According to the experimental results, the suggested HiQoS technique may reduce latency while boosting throughput to guarantee QoS.

The authors present NSAL (Network Service Abstraction Layer) implementation in Ref. [35] at the top of the management and control layers. To guarantee quality service for time-sensitive tasks, the authors offer a consistent data model for traditional and SDN devices. This model enables unified management and configuration of both networks. The authors in Ref. [36] proposed a multi-tenancy network resource allocation strategy based on a hybrid chaos particle swarm optimization algorithm for innovative cloud service systems.

The authors present a QAR (QoS-aware adaptive routing) strategy with multi-layer distributed hierarchical control plan architectures in Ref. [37] for an SDN network. It is primarily used in large-scale SDN networks to reduce signaling delay using three controller design tiers. This QAR technique proposes using a QoS-aware reward function in conjunction with an RL (Reinforcement Learning) technique to provide suitable time-adaptive QoS packet forwarding. Large-scale SDN networks can realistically use this QAR technology because simulation demonstrates its superior performance over the conventional Q-learning method.

To obtain a worldwide perspective on the hybrid SDN network, the authors in Ref. [38] offer the architecture of a hybrid SDN network that may use the spanning tree protocol to identify the existence of classic switches. They can communicate between OpenFlow switches and legacy switches through the Learning Bridge Protocol without requiring any modifications to the legacy switches. By utilizing SDN characteristics, SDN applications can dynamically find routing pathways based on pre-established QoS requirements and the state of the network.

The authors in Ref. [39] developed a revolutionary SUDOI (software-defined ubiquitous data center optical interconnection) architecture in ubiquitous data center optical connectivity to satisfy the extended user access QoS requirement. Their work performance is validated on an OaaS (Optimization as a Service) testbed for

data center services. Their findings demonstrate that SUDOI can efficiently leverage optical networks and application resources in ubiquitous data center optical interconnections without raising the likelihood of blocking.

The authors present a spatial routing system for automobile networks based on hSDN and a clustering technique [40]. When determining which relay to pass the data to, it considers three factors: 1. contact length between cars; 2. each vehicle's load; and 3. a communication failure log integrated with each cluster head. The results of the simulation show that the recommended strategy works well in terms of average routing overhead, packet top rate, and throughput.

4.3 Routing with quality of service awareness

Different QoS measurements, such as packet loss rate, latency, bandwidth, etc., are needed for specific applications. To solve this problem, we might implement QoS-aware routing, which matches QIS (Quality Information System) requirements by creating an SLA for every application that records QoS criteria and assigns routing paths accordingly. The genetic algorithm, or GA, is used in this routing to find the optimal path that gradually satisfies QoS requirements. The evaluation showed that this technique works effectively in networks with various QoS requirements.

Numerous big data applications in data centers need to transfer enormous amounts of data between locations, so it's crucial to consider how these application flows are dispersed across the network to maximize network utilization. The developers of Hedera gave an SA-based method to route flows to other channels based on the current network status [41]. In terms of performance, this study performs better than the average ECMP. The inventors of Long proposed the dynamic rerouting mechanism [42]. In this method, the network controller uses a single-hop or multi-hop algorithm to shift traffic to alternative paths when traffic stress is detected. This technique can improve connection utilization instead of the traditional Round Robin algorithm.

The LARAC (Lagrange Relaxation Based Aggregated Cost) creators [43] proposed a LARAC QoS-aware routing algorithm that uses an aggregated cost technique based on Lagrange relaxation to discover near-optimal pathways in the constrained least-cost problem. For IoT devices, the authors of MINA (Multinetwork Information Architecture) [44] proposed a QoS-aware flow scheduling technique that employs a genetic algorithm to produce the optimal path that satisfies QoS requirements continuously. The evaluation showed that this strategy works effectively in networks with various QoS requirements.

In comparable studies, only MINA addressed a variety of limitations. Some of MINA's main obstacles are static weight in the cost functions, QoS-aware routing, and getting stuck in a local optimum. To solve these problems, the authors in Ref. [13] proposed the QoS-aware routing method, which finds the optimal path that satisfies different QoS requirements by utilizing the simulated annealing (SA) methodology.

4.4 QoS policy administration

SLAs, or service level agreements, are necessary to establish QoS standards for traffic management of various QoS applications, such as online interactive gaming, video streaming, and video conferencing. Every service level agreement consists of a set of goals for producing network-level policies that are converted into primitives at the device level (e.g., packet dropping rate, queue configurations, traffic shaping policies, and forwarding rules). Managing QoS settings in traditional network designs

like MPLS and DiffServ is challenging [45]. On the other hand, SDN offers a global network view and robust northbound API (Application Programming Interface) functionality. It allows network operators to implement different network regulations and start services fast.

The PolicyCop [46] project uses the Northbound SDN API to enable vendor-neutral, flexible, and customizable QoS rule management in SDN/OpenFlow networks. PolicyCop is an appropriate management framework since it takes advantage of OpenFlow characteristics. Both per-flow control and on-demand aggregation are provided. Compared to MPLS and DiffServ, Policy Cop simplifies traffic definition because it does not need to shut down network devices. Due to vendor independence, it reduces operational costs and is easy to deploy in a network. The management, control, and data planes make up the three planes of the Policy Cop design. The management plane is the brains of Policy Cop; the data and control planes are conventional SDN planes. The plane comprises two parts: the policy enforcer and the policy validator. The policy enforcer enforces network policy regulations and responds to policy violations, while the policy validator finds and tracks policy breaches.

4.5 Overhead QoS signaling

All overhead (QoS-related signaling messages) is collected at the network's control mechanism (Controller) because of the logically centralized structure of the SDN design. OpenFlow allows network operators to gather data at different flow levels, including aggregated and individual flows. The per-flow process offers more granularity about QoS-related statuses. Its scalability is an issue. A controller can pool switches in an OpenFlow or SDN network to get information on active flows. The switch may be instructed to report flow statistics at a predefined interval during a flow timeout.

Another essential factor is how often QoS information should be sent from one network device to the controller. While obtaining statistics from the data plan helps the controller maintain a current global picture of network conditions, this process involves trade-offs between signaling overhead, timeliness, and measurement precision speed, and it also causes a controller's control plane to become uneasily scalable [47]. The PayLess [48] solution facilitates various levels of flow aggregation for flow statistics via a REST (Representation State Transfer) API. It gathers statistics adaptively to minimize network overhead and deliver precise real-time information. Compared to the regular polling strategy, this approach reduces massaging overhead by as much as 50% while achieving accuracy that is exceptionally close to the constant periodic polling method.

5. Conclusion

It is well known that SDN is a cutting-edge network architecture that separates the control and data layers to offer exceptional network control, flexibility, and efficiency. However, migrating straight from a conventional network to a fully SDN-based network appears challenging. A contemporary network design, known as hybrid SDN, is the optimal solution as it blends the programmability and adaptability of SDN with the stability and familiarity of conventional network designs. On the other hand, the primary problem with today's network is the scarcity of quality service guarantees. In recent years, this issue has gotten lots of attention worldwide,

and researchers and industry have proposed and implemented different models to solve this problem; however, many of these solutions have either failed or may not deploy and function properly. In this article, we have studied the interconnection of Quality of Service mechanisms within a hybrid SDN environment, emphasizing the balance between conventional network and SDN mechanisms. Our finding indicates hybrid SDN, which provides a more scalable, flexible, and programmable network environment, is the best option for enterprises looking to upgrade their network infrastructure without experiencing service outages or network disruptions. By deploying sophisticated QoS policies in the hybrid environment, we can see significant improvements in the overall network performance, including lower latency, increased bandwidth usage, and enhanced overall user experience. Using machine learning algorithms for adaptive QoS and traffic prediction can improve the network's capacity to deal with diverse and unpredictable traffic patterns.

Author details

Samiullah Mehraban* and Rajesh K. Yadav
Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

*Address all correspondence to: mehraban748@gmail.com

IntechOpen

References

[1] Awduche D et al. Overview and Principles of Internet Traffic Engineering. No. rfc3272. 2002

[2] Mehraban S, Yadav RK. Traffic engineering and quality of service in hybrid software defined networks. China Communications. 2024;**21**(2):96-121

[3] He J, Song W. Achieving near-optimal traffic engineering in hybrid software defined networks. In: 2015 IFIP Networking Conference (IFIP Networking), Toulouse, France. 2015. pp. 1-9. DOI: 10.1109/IFIPNetworking.2015.7145321

[4] Fortz B, Thorup M. Internet traffic engineering by optimizing OSPF weights. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064), Tel Aviv, Israel. Vol. 2. 2000. pp. 519-528. DOI: 10.1109/INFCOM.2000.832225

[5] Sridharan A, Guerin R, Diot C. Achieving near-optimal traffic engineering solutions for current OSPF/IS-IS networks. IEEE/ACM Transactions on Networking. 2005;**13**(2):234-247

[6] Xu D, Chiang M, Rexford J. DEFT: Distributed exponentially-weighted flow splitting. In: IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications, Anchorage, AK, USA. 2007. pp. 71-79. DOI: 10.1109/INFCOM.2007.17

[7] Xu K, Liu H, Liu J, Shen M. One more weight is enough: Toward the optimal traffic engineering with OSPF. In: 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, USA. 2011. pp. 836-846. DOI: 10.1109/ICDCS.2011.49

[8] Zhang M, Liu B, Zhang B. Multi-commodity flow traffic engineering with hybrid MPLS/OSPF routing. In: GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference, Honolulu, HI, USA. 2009. pp. 1-6. DOI: 10.1109/GLOCOM.2009.5426135

[9] Tajiki MM et al. SDN-based resource allocation in MPLS networks: A hybrid approach. Concurrency and Computation: Practice and Experience. 2019;**31**(8):e4728

[10] Quan W et al. Adaptive transmission control for software defined vehicular networks. IEEE Wireless Communications Letters. 2018;**8**(3):653-656

[11] Agarwal S, Rajakrishnan S, Narayan A, Agarwal R, Shmoys D, Vahdat A. Sincronia: Near-optimal network design for coflows. In: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18). New York, NY, USA: Association for Computing Machinery; 2018. pp. 16-29. DOI: 10.1145/3230543.3230569

[12] Tajiki MM et al. CECT: Computationally efficient congestion-avoidance and traffic engineering in software-defined cloud data centers. Cluster Computing. 2018;**21**(4):1881-1897

[13] Vissicchio S et al. Safe update of hybrid SDN networks. IEEE/ACM Transactions on Networking. 2017;**25**(3):1649-1662

[14] Caria M, Das T, Jukan A, Hoffmann M. Divide and conquer:

Partitioning OSPF networks with SDN. In: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ottawa, ON, Canada. 2015. pp. 467-474. DOI: 10.1109/INM.2015.7140324

[15] Xu H, Huang H, Chen S, Zhao G. Scalable software-defined networking through hybrid switching. In: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, GA, USA. 2017. pp. 1-9. DOI: 10.1109/INFOCOM.2017.8057001

[16] Vissicchio S, Vanbever L, Rexford J. Sweet little lies: Fake topologies for flexible routing. In: Proceedings of the 13th ACM Workshop on Hot Topics in Networks (HotNets-XIII). New York, NY, USA: Association for Computing Machinery; 2014. pp. 1-7. DOI: 10.1145/2670518.2673868

[17] Huin N et al. Bringing energy aware routing closer to reality with SDN hybrid networks. IEEE Transactions on Green Communications and Networking. 2018;**2**(4):1128-1139

[18] Guo Y et al. SOTE: Traffic engineering in hybrid software defined networks. Computer Networks. 2019;**154**:60-72

[19] Guo Y et al. Traffic engineering in hybrid SDN networks with multiple traffic matrices. Computer Networks. 2017;**126**:187-199

[20] Al-Jumaily A, Alrubaee SH, Jiménez VPG, Al-Saegh AM, Mahmood AA, Al-Jumeily D. Architecture design of B5G and 6G millimeter-wave radio access network: Using Wireless communications to increase coverage, capacity and performance. In: 2023 16th International Conference on Developments in eSystems Engineering (DeSE), Istanbul, Turkiye. 2023. pp. 212-217. DOI: 10.1109/DeSE60595.2023.10468801

[21] Cheng TY, Jia X. Compressive traffic monitoring in hybrid SDN. IEEE Journal on Selected Areas in Communications. 2018;**36**(12):2731-2743

[22] Polverini M et al. The power of SDN to improve the estimation of the ISP traffic matrix through the flow spread concept. IEEE Journal on Selected Areas in Communications. 2016;**34**(6):1904-1913

[23] Mehraban S, Vora KB, Upadhyay D. Deploy Multi-Protocol Label Switching (MPLS) Using Virtual Routing and Forwarding (VRF). In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India. 2018. pp. 543-548. DOI: 10.1109/ICOEI.2018.8553949

[24] Bi Y et al. Intelligent quality of service aware traffic forwarding for software-defined networking/open shortest path first hybrid industrial internet. IEEE Transactions on Industrial Informatics. 2019;**16**(2):1395-1405

[25] Huang X, Zeng M, Xie K. Intelligent traffic control for QoS optimization in hybrid SDNs. Computer Networks. 2021;**189**:107877

[26] Xu H et al. Incremental deployment and throughput maximization routing for a hybrid SDN. IEEE/ACM Transactions on Networking. 2017;**25**(3):1861-1875

[27] Jia X, Jiang Y, Guo Z. Incremental switch deployment for hybrid software-defined networks. In: 2016 IEEE 41st Conference on Local Computer Networks (LCN), Dubai, United Arab Emirates. 2016. pp. 571-574. DOI: 10.1109/LCN.2016.95

[28] Xu Z et al. Experience-driven networking: A deep reinforcement learning based approach. In: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, HI, USA: Honolulu. 2018. pp. 1871-1879. DOI: 10.1109/INFOCOM.2018.8485853

[29] Huang X et al. Deep reinforcement learning for multimedia traffic control in software defined networking. IEEE Network. 2018;**32**(6):35-41

[30] Mohammed SH et al. A review on the evaluation of feature selection using machine learning for cyber-attack detection in smart grid. IEEE Access. 2024;**12**:44023-44042. DOI: 10.1109/ACCESS.2024.3370911

[31] Zhang J et al. CFR-RL: Traffic engineering with reinforcement learning in SDN. IEEE Journal on Selected Areas in Communications. 2020;**38**(10):2249-2259

[32] Sun P, Lan J, Guo Z, Xu Y, Hu Y. Improving the scalability of deep reinforcement learning-based routing with control on partial nodes. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain. 2020. pp. 3557-3561. DOI: 10.1109/ICASSP40776.2020.9054483

[33] Egilmez HE, Dane ST, Bagci KT, Tekalp AM. OpenQoS: An OpenFlow controller design for multimedia delivery with end-to-end Quality of Service over Software-Defined Networks. In: Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, CA, USA. 2012. pp. 1-8

[34] Yan J et al. HiQoS: An SDN-based multipath QoS solution. China Communications. 2015;**12**(5):123-133

[35] Sieber C et al. Network configuration with quality of service abstractions for SDN and legacy networks. In: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ottawa, ON, Canada. 2015. pp. 1135-1136. DOI: 10.1109/INM.2015.7140446

[36] Gil Jiménez VP et al. Hybrid chaos particle swarm optimization algorithm for smart cloud service system based on optimization resource scheduling and allocation. Journal of Autonomous Intelligence. 2023;**6**(2):652-652

[37] Lin S-C, Akyildiz IF, Wang P, Luo M. QoS-Aware adaptive routing in multi-layer hierarchical software defined networks: A reinforcement learning approach. In: 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA. 2016. pp. 25-33. DOI: 10.1109/SCC.2016.12

[38] Lin C, Wang K, Deng G. A QoS-aware routing in SDN hybrid networks. Procedia Computer Science. 2017;**110**:242-249

[39] Yang H et al. SUDOI: Software defined networking for ubiquitous data center optical interconnection. IEEE Communications Magazine. 2016;**54**(2):86-95

[40] Alouache L et al. HSDN-GRA: A hybrid software-defined networking-based geographic routing protocol with the multi-agent approach. International Journal of Communication Systems. 2020;**33**(15):e4521

[41] Al-Fares M, Radhakrishnan S, Raghavan B, Huang N, Vahdat A. Hedera: dynamic flow scheduling for data center networks. Nsdi. 2010;**10**(8):89-92

[42] Long H, Shen Y, Guo M, Tang F. LABERIO: Dynamic load-balanced Routing in OpenFlow-enabled Networks.

In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), Barcelona, Spain. 2013. pp. 290-297. DOI: 10.1109/AINA.2013.7

[43] Juttner A, Szviatovski B, Mecs I, Rajko Z. Lagrange relaxation based method for the QoS routing problem. In: Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213), Anchorage, AK, USA. Vol. 2. 2001. pp. 859-868. DOI: 10.1109/INFCOM.2001.916277

[44] Qin Z, Denker G, Giannelli C, Bellavista P, Venkatasubramanian N. A Software Defined Networking architecture for the Internet-of-Things. In: 2014 IEEE Network Operations and Management Symposium (NOMS), Krakow, Poland. 2014. pp. 1-9. DOI: 10.1109/NOMS.2014.6838365

[45] Mehraban S, Yadav RK. Quality of services in hybrid SDN (hSDN): A review. In: 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India. 2022. pp. 652-658. DOI: 10.1109/ICCES54183.2022.9835774

[46] Bari MF, Chowdhury SR, Ahmed R, Boutaba R. PolicyCop: An autonomic QoS policy enforcement framework for software defined networks. In: 2013 IEEE SDN for Future Networks and Services (SDN4FNS), Trento, Italy. 2013. pp. 1-7. DOI: 10.1109/SDN4FNS.2013.6702548

[47] Moshref M, Yu M, Govindan R. Resource/accuracy tradeoffs in software-defined measurement. In: Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking (HotSDN '13). New York, NY, USA: Association for Computing Machinery; 2013. pp. 73-78. DOI: 10.1145/2491185.2491196

[48] Chowdhury SR, Bari MF, Ahmed R, Boutaba R. PayLess: A low cost network monitoring framework for Software Defined Networks. In: 2014 IEEE Network Operations and Management Symposium (NOMS), Krakow, Poland. 2014. pp. 1-9. DOI: 10.1109/NOMS.2014.6838227

Chapter 4

The Role of QoS at the OSI Model Layers

Mirzakulova Sharafat and Bakhytzhan Kulambayev

Abstract

The current trend of converging various network types, coupled with the rising traffic volumes and the increasing use of real-time and multimedia applications, has necessitated the transportation of diverse types of traffic, particularly those sensitive to latency. Traditional TCP/IP networks fall short in providing the required Quality of Service (QoS) for these applications, prompting the development of supplementary mechanisms to meet the necessary service levels. Additionally, the advent of new multimedia and cloud services, the expansion of the Internet of Things (IoT), and the integration of optical and wireless communications within the 5G optical network framework demand modifications to network infrastructure. These changes are essential to accommodate scalable traffic growth while simultaneously maintaining high levels of dynamic connectivity, complete flexibility, and enhanced energy efficiency. Within this context, QoS is crucial across all layers of the OSI model, ensuring effective and reliable data transmission within network systems.

Keywords: quality of service, OSI model, bursts, jitter, latency, prioritization, classification

1. Introduction

The Information Age has facilitated rapid global communication and the development of extensive information networks, fundamentally transforming the structure of modern society. Unlike previous eras, this period is characterized by the central role of technology and the emergence of information as a critical commodity. Information has become a valuable intellectual asset and an essential component of society's life-support system, with the proportion of intellectual property, particularly information, steadily increasing, underscoring its growing importance and constant demand.

The ongoing trend of network convergence, particularly through packet-switched IP networks, has led to the integration of various communication channels, services, and networks into a unified framework. This integration allows for the transmission of diverse data types, including voice, video, and other data, over a single network infrastructure. In the early stages of network implementation, the separation of voice and data networks minimized concerns about application diversity. However, as network technologies have evolved, the need for differentiated services within converged IP networks has become apparent. The continuous advancement of technology and

the consolidation of multiple services within a single network have heightened the demand for Quality of Service (QoS) to maintain the required service levels for various applications. As the number of users and the utilization of multi-service networks increase, the demands on bandwidth and network performance grow, making QoS increasingly critical. Modern applications and services necessitate substantial network resources, and users expect these services to function seamlessly at high speeds consistently. Consequently, organizations must adopt processes and technologies that ensure the highest level of service [1].

QoS has thus emerged as a crucial subject, encompassing various technologies that guarantee a specific quality level for different network services.

2. Methodology

This study utilized a systematic approach to investigate the role of Quality of Service (QoS) in contemporary network infrastructures. The methodology comprised several key components:

Literature review: A comprehensive review of existing literature on QoS models, including Best Effort (BE), Integrated Services (IntServ), and Differentiated Services (DiffServ), was undertaken. The review specifically focused on the applicability of these models in modern IP networks and examined essential QoS parameters such as latency, jitter, packet loss, and bandwidth.

Comparative analysis: The study compared various QoS models based on their operational characteristics, scalability, and overall effectiveness. This comparison included an analysis of how each model handles traffic management, data prioritization, and resource allocation.

QoS implementation across OSI layers: The research examined the implementation of QoS across different layers of the OSI model, with a focus on protocols such as TCP, UDP, DiffServ, and MPLS. The study assessed how these technologies interact within the OSI model and their impact on network performance.

Data synthesis: The insights gained from the literature review and comparative analysis were synthesized to develop a comprehensive understanding of QoS in modern networks. This synthesis highlighted key trends and best practices for effective QoS management.

This methodology provided a structured framework for understanding the crucial role of QoS in network systems, leading to significant conclusions regarding its importance in maintaining high-quality network performance.

3. QoS service models

There are three primary models for implementing Quality of Service (QoS) [2]:

Best effort (BE): This is the most basic QoS model where all network packets are treated with equal priority, and no guarantees are provided regarding packet delivery. BE is typically employed when networks lack configured QoS policies or when the infrastructure does not support QoS capabilities. In such cases, all packets, irrespective of their type or importance, are handled equally by the network. During periods of congestion or traffic surges, this model may lead to data loss, increased delays, or instability in data transmission, especially for latency-sensitive applications such as voice or video services. While BE is sufficient for scenarios where QoS is not critical,

it may be inadequate for more complex and demanding applications. Data is transmitted as resources become available, without any assurance of bandwidth, latency, or packet loss.

IntServ: This QoS model offers guarantees for bandwidth and latency by reserving resources across all routers along the data path from the source to the destination. It operates on the principle of resource reservation for each data flow, effectively mimicking dedicated channel services within an IP network. The IntServ model relies on routers that support it, typically utilizing the Resource Reservation Protocol (RSVP) to manage the resource reservation process within the network. Despite its robustness, IntServ is not as prevalent in modern networks due to its limited scalability, complex configuration requirements, and high resource demands, making it less suitable for large or dynamically changing networks.

DiffServ: This QoS architecture provides varying levels of service for different types of traffic by marking packets using the Differentiated Services Code Point (DSCP). Packets are classified and marked according to the type of service they require. A network component, known as the Bandwidth Broker (BB), manages resource allocation for different service classes. The BB maintains state information, such as current network loads and available bandwidth, enabling it to interact with routers and other devices to make informed resource allocation decisions based on the network's real-time status. In DiffServ, state information can either be maintained in the network core or managed by the Bandwidth Broker or edge routers, depending on the network's architecture [3]. Packets with similar QoS requirements are grouped into aggregate flows, which are then managed based on their service class (e.g., voice traffic, video traffic). Network devices handle these packets according to their class, as indicated by the DSCP field in the IP packet header, allowing for appropriate traffic management policies such as prioritization or bandwidth limitations for each class.

A comparative analysis of these QoS models indicates that while the basic BE service remains relevant for non-critical applications such as web browsing or file transfers, more advanced models such as IntServ and DiffServ offer better solutions for applications requiring higher service levels, is indicated in **Table 1**.

In Internet Protocol version 4 (IPv4), the Type of Service (ToS) field enables the implementation of Quality of Service (QoS) by incorporating features that address QoS requirements through route selection criteria such as delay, throughput, and reliability [4].

Best Effort (BE) is typically employed in public and smaller networks due to its simplicity and lack of resource demands. In contrast, Integrated Services (IntServ) is seldom used in modern networks because of its high complexity and the substantial

Parameters	BE	Guaranteed service	Differentiated service
QoS support	Does not support	Assured service	Support
QoS relationships	Non-guaranteed	Per flow	Aggregated
Coverage area	Basic level	Full path	Domain-focused
State data	Stateless	Flow state	Minimal state information
Protocol	Not available	Flow management protocol	DS field
State	Standard	Developed	Implemented

Table 1.
Comparative analysis of QoS models.

resources it requires. Differentiated Services (DiffServ), however, takes full advantage of the flexibility and scalability of IP networks. It does so by converting the information contained in packets into Per-Hop Behavior (PHB) at each network hop, thereby significantly reducing the need for signaling operations. DiffServ is widely recognized as a robust QoS model and is extensively utilized in contemporary networks, particularly in large enterprise and carrier networks.

4. Parameters characterizing the quality of service

Quality of Service (QoS) can be quantitatively evaluated using several key parameters [5]:

Latency: This refers to the time delay required for data to travel from the source to the receiver, typically measured in milliseconds (ms). Latency is a critical metric in QoS, particularly for latency-sensitive applications such as IP telephony and video conferencing.

Jitter: Jitter represents the variations in latency during data transmission. High jitter can degrade the quality of real-time applications, including streaming video and voice communications.

Packet loss: This is the percentage of data packets lost during transmission. Packet loss generally occurs in congested network areas where the volume of incoming packets exceeds the output queue's capacity. It can also result from inadequate input buffer size. Packet loss can significantly affect service quality, especially in multimedia applications.

Bandwidth: Bandwidth measures the maximum amount of data that can be transmitted through the network within a given timeframe, usually expressed in bits per second (bps).

Reliability: This parameter assesses the network's capability to deliver data without loss.

Mean opinion score (MOS): MOS is a metric used to evaluate voice quality, rated on a five-point scale where five represents the highest quality.

The concept of QoS is grounded in the idea of an agreement or guarantee provided by the network to deliver a predefined set of measurable service attributes, including network latency, jitter, available bandwidth, and packet loss probability, among others.

5. Traffic of various applications

Applications are the primary catalyst for network development. The introduction of traffic types such as IP telephony, audio, and video broadcasting has necessitated the need for low packet delay, multicast packet delivery support, and other specific requirements within computer networks.

Three fundamental characteristics have been identified as the main criteria for evaluating network traffic:

1. Relative predictability of data transmission speed.

2. Susceptibility to packet delays.

3. Sensitivity to packet loss and distortions.

Based on the predictability of traffic speed, applications can be categorized into two broad classes:

Stream traffic: This class is characterized by a uniform data flow, maintaining a consistent speed, typically referred to as Constant Bitrate (CBR). Although the speed of the stream may fluctuate, its upper limit is predefined. For example, audio data streams fall into the CBR category, with a simple voice stream having an upper limit of 64 kbps [5].

Burst traffic: This class is marked by irregular traffic patterns, with periods of inactivity followed by bursts of high activity, corresponding to the transmission of large "blocks of data."

The first class of applications is marked by a high degree of predictability in the generated traffic, usually maintaining a more or less constant speed known as *Constant Bitrate (CBR)*. Although the stream's speed may vary, its upper limit is predefined. For instance, audio data streams belong to the CBR class, where the upper limit for a simple voice stream is 64 kbps [6].

The second class is characterized by high unpredictability, where periods of inactivity are followed by bursts of activity, corresponding to the delivery of large "blocks of data" (see **Figure 1**) [7].

This type of traffic is characterized by Variable Bitrate (VBR). For example, in file services, traffic intensity can fluctuate drastically, ranging from zero (when no files are being transferred) to very high levels (when a file location request prompts the application to retrieve data as quickly as possible). In reality, all applications generate some level of irregular traffic, including streaming applications; however, the degree of irregularity (i.e., the ratio of peak instantaneous speed to average speed) varies between these two types of applications.

Another criterion for classifying applications by traffic type is their susceptibility to packet delays. The main types of applications in this context include:

- Asynchronous applications: These impose minimal restrictions on delay times, often referred to as "elastic" traffic, such as email [8].

- Synchronous applications: These are sensitive to delays but tolerate them within certain limits. An example would be VoIP calls, where minor delays may be acceptable.

- Interactive applications: In these, delays may be noticeable to users but do not significantly impair the functionality of the applications, such as working with a remote file in a text editor.

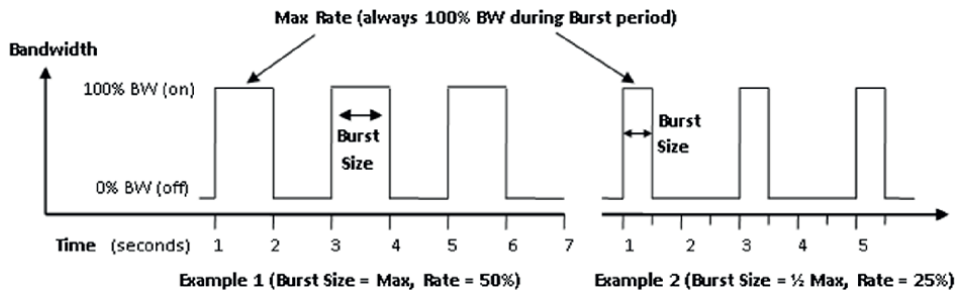

Figure 1.
Burst mode.

- Isochronous applications: These have a specific threshold of delay sensitivity, beyond which the application's performance degrades, such as when delays exceed 100–150 ms, leading to a significant drop in speech quality.

- Ultra-sensitive delay applications: These applications cannot tolerate delays, as they become non-functional with any significant lag. For instance, in real-time control of a technical object, a delay in the control signal could result in an accident.

The third criterion for classifying applications is their sensitivity to packet loss. Based on this criterion, applications are generally divided into two groups—those sensitive to loss [9] and those that are less affected by it. The first group includes almost all applications that transmit alphanumeric data, such as text documents, program code, and numeric arrays. These applications cannot tolerate the loss of even small fragments of data, as such losses can diminish the value of the remaining information, with the absence of a single byte in a program's code rendering it non-functional. Traditional network applications, such as file services, database services, and email, belong to this category.

The second category includes applications transmitting information about inertial physical processes. These applications are more resilient to losses because missing data can be approximated using the received data. For instance, if a packet containing several consecutive voice measurements is lost, the missing measurements can be estimated based on surrounding values. Most audio and video applications fall into this category. However, the resilience to loss has its limits, as the number of lost packets should not exceed 1%. Not all traffic can withstand data loss, particularly compressed voice or video, which are very sensitive to loss and thus belong to the first category of applications.

The primary mathematical model describing traffic in a communication network is a random data stream characterized by statistical properties such as the probability density of data arrival over a period, the probability density of intervals between data arrivals, and the autocorrelation function.

In their 1993 studies, American researchers W. Leland, M. Taqqu, W. Willinger, and D. Wilson [10, 11] discovered that data traffic in both LAN and WAN networks exhibits the properties of self-similarity (scale invariance), long-term dependence (memory effect), and fluctuations not only in short-term intervals but also over longer time periods. Subsequent research by others [12–14] further confirmed these findings. Quality of Service (QoS) is crucial for ensuring the optimal performance of mission-critical applications that require high bandwidth for real-time traffic.

6. Types of real-time applications

The key characteristic of a real-time application is its ability to execute tasks within a specific time constraint. The following services are known for generating real-time traffic [15]:

- Video telephony

- IP telephony

- Internet games
- Online stores
- Data management platforms
- Real-time management systems
- Chat applications
- Collaborative work environments
- Connected device ecosystems
- Data analysis in business
- Location determination services
- Fraud detection
- Online advertising campaign analysis
- Radio diagnostics
- Banking activities

Additionally, various online real estate platforms with advanced search filters facilitate seamless transactions for both sellers and buyers by providing real-time, up-to-date information on available properties, potential buyers, and other relevant data.

7. QoS architecture

Modern public IP networks integrate several advanced technologies such as *Software-Defined Networking (SDN)*, *Dense Wavelength Division Multiplexing (DWDM)*, *Reconfigurable Optical Add-Drop Multiplexer (ROADM)*, and *Quality of Service (QoS)* to ensure high performance, flexibility, and reliability.

Figure 2 illustrates the logical mapping between the basic Open Systems Interconnection (OSI) reference model and the TCP/IP stack (**Figure 1**).

QoS relates to various layers within the OSI model. The QoS parameters associated with service primitives and Protocol Data Unit (PDU) are shown in the illustration 3 [16].

In the context of the OSI network model and protocols related to QoS, Association Control Service Element (ACSE) is used for organizing communication between two applications or application processes, providing the means for establishing, maintaining, and terminating associations. The A-ASSOCIATE primitive is part of the ACSE element and is used to initiate the process of establishing an association between two applications. *ACSE* provides various primitives for managing associations and sessions, presented in **Figure 3**.

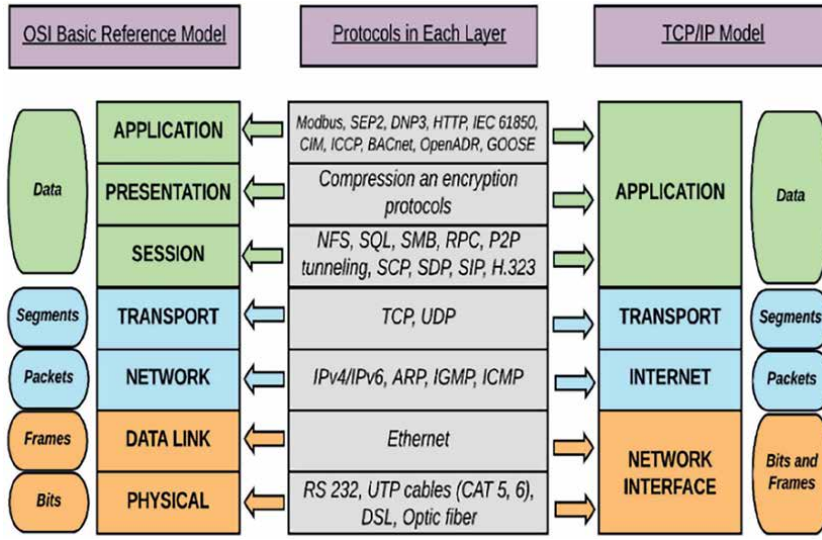

Figure 2.
Logical mapping between the OSI reference model and the TCP/IP stack.

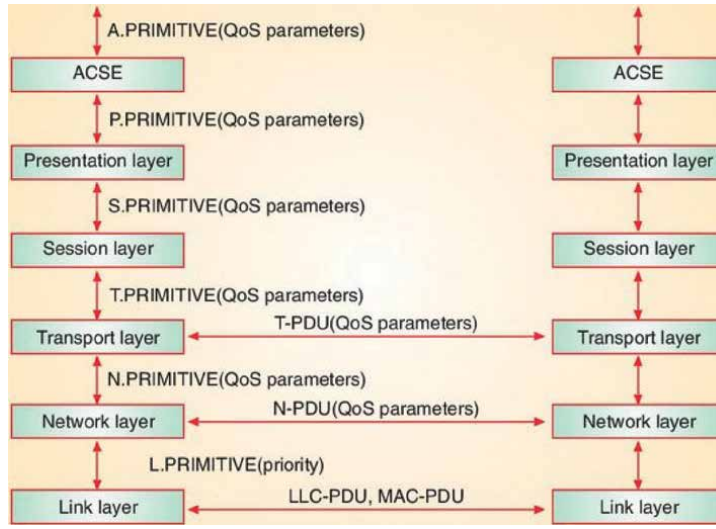

Figure 3.
OSI QoS parameters.

The *A.ASSOCIATE* primitive is used to establish an association between two applications. In this primitive, QoS parameters can be specified, which describe the quality of service requirements for a given session. Different applications may have different QoS requirements (e.g., VoIP, video conferencing, online gaming). These parameters are passed down to the transport layer, which ensures transparent data transfer between transport service users.

Once QoS parameters are established at the application layer and communicated through service primitives, they must be addressed at the lower layers of the OSI model (**Figure 4**) [17]. The application layer is crucial in supporting end-user applications that require specific QoS levels to meet user expectations.

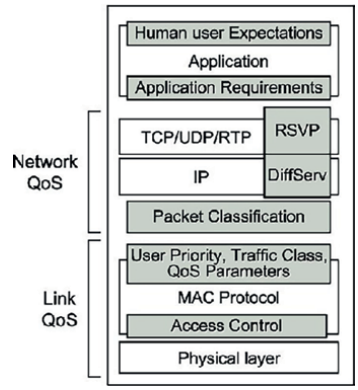

Figure 4.
IP network QoS architecture.

The presentation and session layers play vital roles in managing sessions and transforming data, ensuring that communications remain synchronized and that data is accurately interpreted.

At the Transport Layer (Layer 4 of the OSI model), which is responsible for data transfer between transport service users, QoS parameters are utilized to manage network traffic effectively. Transport layer protocols, such as transmission control protocol (TCP) and user datagram protocol (UDP), take QoS parameters into account to ensure the appropriate level of service quality during data transfer between hosts.

TCP uses a mechanism called flow control to adjust the data transmission rate based on network conditions and the receiver's capacity. TCP also uses a congestion control mechanism to avoid network congestion and packet loss.

In the context of QoS, TCP performs several functions that directly or indirectly impact QoS:

- *Reliable data delivery*: In case of losses or errors, TCP uses sequence numbers in segment headers to track the order of transmitted data. An acknowledgment (ACK) mechanism ensures the correct sequence on the receiving side (**Figure 5**) [18]:

- *congestion control.* The TCP segment header includes a field called the *Receive Window*, which controls the amount of data that can be sent without acknowledgment. This field helps manage the data flow and prevents network congestion. If the network is congested, this window decreases, slowing down data transmission and preventing further losses;

- *flow control.* The *Window Size* field in the TCP header determines the maximum amount of data that can be sent but not yet acknowledged. This helps synchronize the data transmission rate with the receiver's receiving capabilities, ensuring a smooth data flow and avoiding congestion;

- *error control.* The TCP header contains a checksum used to verify data integrity. If the checksum does not match on the receiving side, the data is considered corrupted and is requested again.

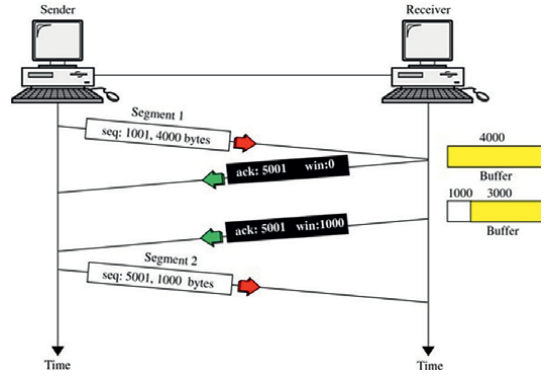

Figure 5.
Reliable data delivery.

UDP is a fast, connectionless protocol that does not guarantee delivery, order, or data integrity. UDP does not use flow control or congestion control and does not require acknowledgments from the receiver. UDP is suitable for applications that require low latency, high bandwidth, and real-time communication, such as streaming, gaming, and Voice over Internet Protocol (VoIP).

UDP does not provide built-in QoS mechanisms, but its lightweight and fast nature makes it suitable for applications where minimal latency is more critical than reliability. QoS for UDP is provided through external mechanisms and network-level configurations that help manage data flow, prioritize critical packets, and minimize data loss under network congestion.

By regulating the flow of data and setting priorities for certain types of traffic, the transport layer helps achieve the required quality of service, ensuring a stable user experience.

8. QoS at the network layer

DiffServ is a Quality of Service (QoS) architecture specifically designed for IP networks, enabling the provision of different service levels for various types of traffic. Unlike other QoS models such as Integrated Services (IntServ), DiffServ emphasizes scalability and efficient traffic management, particularly in large networks. The primary distinction of DiffServ is its focus on classifying services based on traffic classes rather than individual flows. This approach means that packets within the same class are treated uniformly, simplifying QoS management and enhancing the model's scalability.

A significant feature of DiffServ is that the complex tasks of traffic classification and conditioning are performed at the network's edge. DiffServ leverages edge devices, such as ingress and egress routers, to execute these tasks, including traffic classification, policing, rate limiting, and Differentiated Services Code Point (DSCP) marking. This approach reduces the burden on the core network and enhances traffic processing efficiency.

One of the key advantages of DiffServ is that it does not require maintaining the state of each individual traffic flow within the core network. This contrasts with IntServ, where maintaining separate states for each flow is necessary. In DiffServ, the core network simply processes packets based on their DSCP markings, which

significantly enhances scalability. By managing traffic according to classes rather than individual flows and avoiding the need to maintain flow states within the core, DiffServ can effectively support QoS for a large number of users and devices [19].

DiffServ is designed to be compatible with network nodes that do not support DiffServ by utilizing standard IP packets with additional service class information. Nodes that do not support DiffServ treat these packets as standard IP packets, ensuring compatibility.

Another advantage of DiffServ is its incremental deployment capability. It can be introduced into a network gradually, starting with key network nodes and then expanding to other parts of the network. This staged approach minimizes risks and allows for the flexible adaptation of QoS as the network grows and traffic demands evolve.

DiffServ employs a straightforward hierarchy of service levels known as Per-Hop Behavior (PHB), which enables network devices, such as routers and switches, to process traffic based on labels embedded in the packet header. These labels, typically DSCP, determine the priority and handling method for each packet.

Traffic classification within DiffServ involves identifying and categorizing traffic into specific classes. This classification can be based on various parameters, including the incoming interface, source and destination IP addresses, traffic type, port numbers, and other packet attributes. Classification allows for the prioritization of packets by considering multiple factors to determine their priority levels. This process is essential for effective QoS management, ensuring that different types of traffic receive appropriate service levels.

Re-marking of DiffServ codes enables the mapping of traffic to one or more aggregated flows, ensuring the correct QoS within the DiffServ domain. At the network layer, DSCP and IP precedence are typically used. Once traffic is classified, it is assigned a specific label, indicating the level of priority or QoS policy that should be applied as the packet traverses routers and switches along its path through the network.

Examples of operation:

- in a DiffServ-enabled network, packets are classified at network boundaries (e.g., access routers) based on header attributes (e.g., IP addresses, ports). They are then marked with a specific DSCP value;

- in the core network, devices (routers, switches) make decisions about how to handle the packet (e.g., giving it higher priority or applying certain queue management policies) based on its DSCP marking.

Modern applications such as VoIP and video conferencing require high network performance, making prioritizing bandwidth for QoS critical. IPv6 was developed to address key IPv4 issues such as QoS, security, and address shortage. Modern IP networks optimize traffic delivery by offering improved, albeit partial, quality of service guarantees. However, there are services, including voice, with strict requirements for latency and jitter, which required the addition of functionality in IPv4 networks to support such services [20].

Figure 6 shows the headers of IPv4 and IPv6.

In the IPv4 packet header, there is a field called Type of Service (ToS) that is 8 bits long. The first 3 bits are known as IP Precedence bits. The next 4 bits are referred to as Type of Service (TOS) bits. The last bit is not used. The precedence bits form a 3-bit subfield ranging from low 000 (normal packet) to high 111 (control packets). Higher-priority packets are processed first. During overload, low-priority packets may be

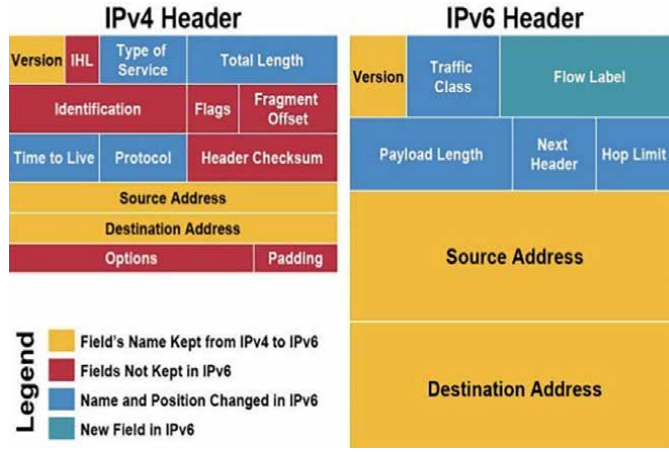

Figure 6.
IPv4 and IPv6 packet headers.

dropped. The Type of Service (TOS) bits are: D—delay requirement, T—throughput requirement, R—reliability requirement, C—cost requirement, with the last bit unused (**Figure 7**) [21].

The differentiated service (DS) byte in DiffServ is standardized and derived from the ToS byte in IPv4 and the Traffic Class byte in IPv6, allowing for consistent QoS mechanisms across both IP versions and ensuring backward compatibility with IP Precedence. The DiffServ field consists of a 6-bit DSCP (Differentiated Services Code Point) and a 2-bit CU (Currently Unused) field that is not used (**Figure 8**) [22].

The first 6 bits of the DS field are used to set the code point, which will affect the Per-Hop Behavior (PHB) at each node. The code point is similar to precedence and is used to set a specific priority. With six bits for code points, a multitude of different priorities can be created, theoretically providing 64 possible values.

The Differentiated Services Code Point (DSCP) codepoint is integral to the Differentiated Services (DiffServ) architecture, as it determines the Per-Hop Behavior (PHB) processing rules that a packet follows at each network node. These

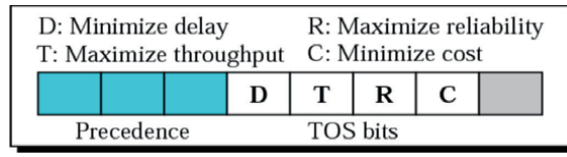

Figure 7.
ToS.

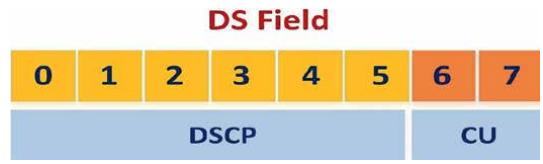

Figure 8.
DS field.

PHB rules are the cornerstone of DiffServ, enabling nodes to reserve resources for specific aggregated traffic flows they manage. This step-by-step resource reservation mechanism allows for the provision of efficient, differentiated services. Service differentiation becomes particularly noticeable when multiple aggregated flows compete for buffer space and node bandwidth [23].

The Default Forwarding (DF) PHB is employed for Best Effort (BE) service, providing basic traffic handling in networks where QoS differentiation is unnecessary. This behavior is standard for traffic that does not require special QoS treatment. A packet marked with the DSCP value 000000, for instance, receives traditional best-effort service, complying with all basic DiffServ requirements.

The Class-Selector (CS) PHB corresponds to a group of DSCP values associated with specific classes of service (CS). This PHB ensures compatibility with legacy systems that utilized the Type of Service (ToS) field in IPv4. In this scenario, only the first three bits of the DSCP value are used, similar to the IP precedence field in older systems (**Figure 9**).

Within the Assured Forwarding (AF) PHB framework, traffic can be classified in various ways depending on the specific network requirements and SLA. An SLA is a formal contract between a service provider and a user that defines the level of QoS, including parameters such as bandwidth, delay, jitter, and packet loss. The AF PHB model uses four service classes (AF1, AF2, AF3, and AF4), and each class can have three levels of drop precedence (DP): low, medium, and high (AFx1, AFx2, and AFx3).

In the case of classification by application type:

- critical—business-critical applications, such as VoIP and video conferencing, receive high priority;
- medium importance—important, but not immediately critical applications, such as email;
- low priority—non-critical applications, such as background tasks or data transfers, may receive lower priority.

In the case of classification by device type:

- traffic from servers—may have higher priority;
- traffic from personal devices—may receive lower priority.

These classification schemes depend on the organization's traffic management policy and can be adapted to meet specific business needs or technical requirements (**Figure 10**).

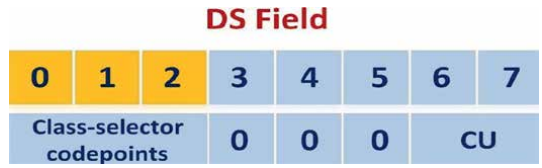

Figure 9.
Class-selector PHB fields.

Figure 10.
AF PHB fields.

The first 3 bits are used to determine the class, each of which is intended to provide a specific level of Quality of Service (QoS), while the next 3 bits are used to determine the drop probability, allowing for even finer control of traffic priorities in the case of network congestion. If Class 4 is assigned the highest priority, then packets in this class will be processed before packets belonging to Class 3.

The Expedited Forwarding (EF) PHB, a key component of DiffServ, provides a reliable service for applications such as Voice over IP (VoIP), video, and online trading programs. EF PHB offers low packet loss, low latency, low jitter, and guaranteed bandwidth. The goal of Expedited Forwarding (EF) is to queue packets where they experience minimal delay and packet loss. To achieve this, a priority queue is used. Packets in the priority queue are sent before those in other queues. However, there is a risk that other queues may not get a chance to send their packets, so a "rate limit" for this queue must be set. This is done through policing. The recommended DSCP value for EF PHB is 101,110.

Mapping CoS to Network Layer QoS is the process of aligning CoS with QoS parameters at the network layer of the OSI model (see **Figure 11**). Class of Service (CoS) is a tag in a network packet used to prioritize different types of traffic.

At the current stage of IP protocol development, DiffServ remains the most effective method for ensuring QoS, especially for real-time applications in IPv4, and its principles continue to be successfully applied in IPv6.

The third field in the IPv6 header, the Flow Label, is necessary for identifying and managing data flows, which helps improve QoS support.

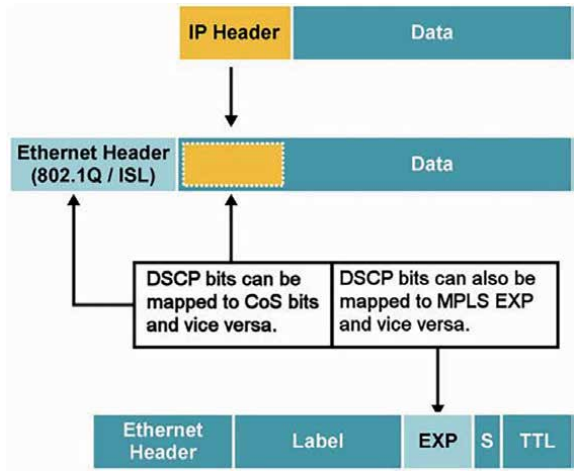

Figure 11.
Mapping CoS to network layer QoS.

9. QoS at the data link layer

Inter-Switch Link (ISL) is a proprietary protocol developed by Cisco used for tagging Virtual Local Area Networks (VLANs) in Ethernet networks. Unlike the IEEE 802.1Q standard, it can only be used between Cisco switches. It allows the transmission of traffic from multiple VLANs over a single physical link between switches. Inter-Switch Link (ISL) encapsulates the original Ethernet frame by placing it inside a new ISL frame, which adds a 26-byte header and a 4-byte trailer. The ISL header contains a VLAN identifier (12 bits) that specifies the VLAN to which the frame belongs.

ISL plays a significant role in ensuring QoS in VLAN networks. The Class of Service (CoS) field exists only within Ethernet frames when using ISL or 802.1Q trunking (see **Figure 12**) [24].

The IEEE 802.1Q standard defines changes to the Ethernet frame structure to allow VLAN information to be transmitted across the network. Ethernet protocol does not support traffic prioritization, so alongside the standard Ethernet protocol IEEE 802.3, the IEEE organization developed the VLAN standard IEEE 802.1Q. The IEEE 802.1Q standard includes the insertion of an additional four-byte VLAN tag into the header of the original Ethernet frame, containing the Class of Service (CoS) priority label IEEE 802.1p [25].

To the Ethernet frame, 32 bits (4 bytes) are added, increasing its size to 1522 bytes. The first 2 bytes (Tag Protocol Identifier, TPID) with a fixed value of 0x8100 indicate that the frame contains an 802.1Q protocol tag. The remaining 2 bytes contain the following information:

- priority—3 bits of the priority field encode up to eight levels of priority (from 0 to 7, where 7 is the highest priority), as specified in the 802.1p standard;

- Canonical Format Indicator (CFI)—1 bit of the canonical format indicator is reserved for identifying frames from other types of networks (such as Token Ring, FDDI) transmitted over Ethernet;

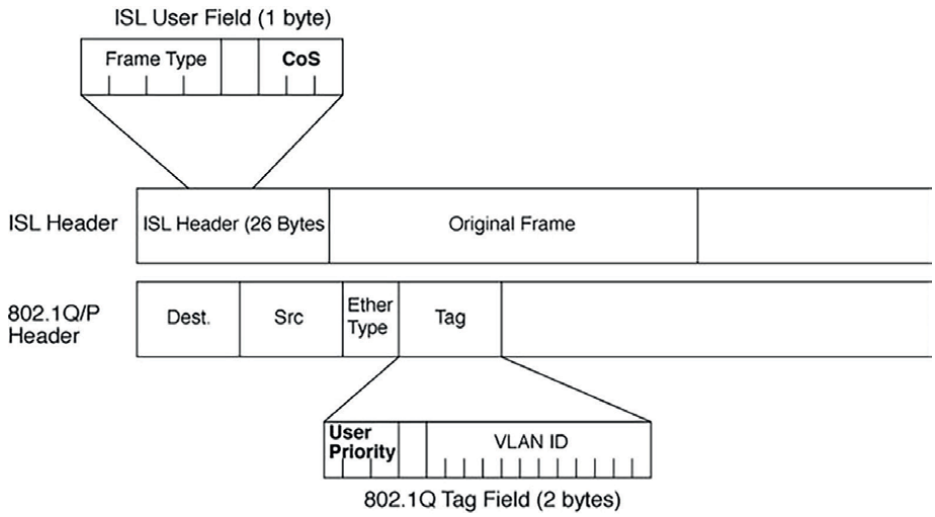

Figure 12.
802.1Q VLAN tag format.

- VID (VLAN ID)—The 12-bit VLAN identifier determines which VLAN the traffic belongs to. Since the VID field is 12 bits wide, up to 4094 unique VLANs can be specified (VID 0 and VID 4095 are reserved).

Switches have separate hardware queues on each physical port to which frames from different applications are mapped and assigned a priority. Each switch port supports between four and eight queues, and the IEEE 802.1P standard specifies three bits for frame priority in its header tag. In the case of a port supporting four queues, an example is shown in **Figure 13**, which illustrates the mapping of priority levels to queues [26].

To implement frame prioritization, it is essential to configure network switches with a queuing algorithm and map 802.1p, Type of Service (ToS), and Differentiated Services Code Point (DSCP) priorities to specific queues. Queuing is a method used to organize and manage data flow in a network, where data packets are sorted into queues to prioritize their processing based on QoS parameters. Some switches allow mapping user-defined 802.1p priorities to either the default four or eight priority queues.

When an untagged frame (a frame header without priority bits) arrives at a switch's input port, it is classified according to the default 802.1p priority value assigned to that port. Network congestion typically occurs where switches connect networks with varying bandwidths. During congestion, frames are buffered and distributed into queues. The order in which these queued data frames are transmitted through the output interface is determined by the queuing mechanism, which helps manage network throughput under congestion conditions.

The congestion management mechanism includes the following queue servicing methods:

First-In, First-Out (FIFO): Frames are processed in the order they arrive, without prioritization.

Priority queuing: Frames are prioritized based on their assigned priority level, with higher-priority frames being processed first.

Weighted round robin (WRR): Frames are processed based on assigned weights, allowing for proportional allocation of bandwidth among different queues.

Custom queuing: Provides a more flexible queuing structure, allowing custom-defined priorities and bandwidth allocation.

Congestion avoidance involves selectively discarding frames to prevent network congestion when output queues reach their maximum capacity. Algorithms such as Tail-Drop, Random Early Detection (RED), and Simple Random Early Detection (SRED) are commonly used for this purpose.

Class of Service (CoS) in Multiprotocol Label Switching (MPLS) works in tandem with the general CoS functions of the routing device. The core of MPLS technology relies on IP tunnels. When a packet enters the virtual network, it is assigned a label,

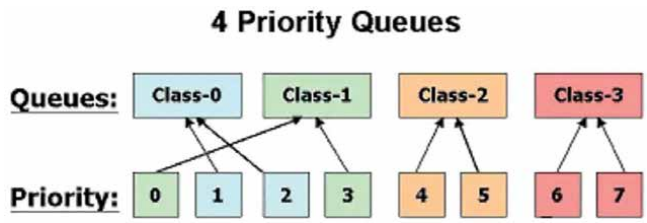

Figure 13.
Classification of frames.

and routers within this network switch packets based on these labels. The label stack format includes a CoS field (also known as EXP or experimental bits). Upon entering a Label Switched Path (LSP) tunnel, the network device tags frames with a CoS value, assigning priority for further forwarding. This label is removed when the packet exits the MPLS domain and enters another. Each physical interface can support up to eight queues [27].

At the physical layer, QoS is concerned with signal fidelity and throughput, ensuring that the hardware can support high standards of data transmission.

10. Conclusion

This chapter has provided an in-depth exploration of the role of Quality of Service within modern network infrastructures, emphasizing its critical importance in maintaining the reliability and performance of diverse applications. The Information Age, characterized by the convergence of various communication channels and services into unified IP networks, has elevated the necessity for robust QoS mechanisms. As network technologies evolve, the demand for differentiated services has grown, making QoS an essential element for ensuring that various applications, particularly those requiring real-time processing, operate seamlessly.

The analysis of QoS models—Best Effort, Integrated Services, and Differentiated Services—has highlighted the trade-offs between simplicity, scalability, and resource guarantees. While BE remains relevant for non-critical applications, IntServ and DiffServ offer more sophisticated approaches to managing network resources, particularly in environments with high demands for latency, jitter, and packet loss control.

The examination of QoS parameters across different layers of the OSI model has reinforced the necessity of implementing QoS at multiple levels within the network stack. From the physical layer to the application layer, ensuring proper QoS management is vital for supporting the growing complexity and volume of network traffic. This is particularly true for mission-critical applications that require high levels of performance and reliability.

In conclusion, the effective implementation of QoS strategies is crucial for optimizing network resources and maintaining high standards of service quality. As networks continue to expand and the demand for real-time services increases, ongoing research and development of advanced QoS mechanisms will be essential. These efforts will help ensure that networks can adapt to the challenges of modern technology, providing users with a stable and efficient communication environment.

Author details

Mirzakulova Sharafat* and Bakhytzhan Kulambayev
Higher School of Telecommunications, Turan University, Kazakhstan

*Address all correspondence to: sh.mirzakulova@turan-edu.kz

IntechOpen

References

[1] Aurrecoechea C, Campbell AT, Hauw L. A survey of QoS architectures. Multimedia Systems. 1998;**6**:138-151

[2] Eisa M, Younas M, Basu K, Awan I. Modelling and simulation of QoS-aware service selection in cloud computing. Simulation Modelling Practice and Theory. 2020;**103**:102108

[3] Kneer H, Zurfluh UE, Dermler G, Stiller B. A business model for charging and accounting of internet services. In: Conference: Electronic Commerce and Web Technologies, First International Conference, EC-Web 2000, London, UK, September 4-6, 2000, Proceedings. 2000. Available from: https://www.researchgate.net/publication/221017276_A_Business_Model_for_Charging_and_Accounting_of_Internet_Services

[4] Tiwari RK, Kumar R, Baranwal C, Buyya R. Decision making framework for heterogeneous QoS information: An application to cloud service selection. Journal of Ambient Intelligence and Humanized Computing. 2023;**14**:2915-2934

[5] Khamosh A, Ahmadi AR, Anwer M, Hamdard J, Aminzai S, Nasrat N. Relationship between IoT service user quality and network QoS factors. Journal of Research in Applied Sciences and Biotechnology. 2023;**2**(2):147-154

[6] Mazhar T, Malik MA, Mohsan SAH, Li Y, Haq I, Ghorashi S, et al. Quality of service (QoS) performance analysis in a traffic engineering model for next-generation wireless sensor networks. Symmetry. 2023;**15**:513. Available from: https://www.mdpi.com/2073-8994/15/2/513

[7] Manavi M. A comprehensive evaluation of the impact of ATM QoS mechanisms on network performance for multimedia and data applications. In: 2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC), Chicago, IL, USA. 2024. pp. 1-7. DOI: 10.1109/INTCEC61833.2024.10602804. Available from: https://ieeexplore.ieee.org/document/10602804

[8] Stepanov SN, Stepanov MS. Approximate method for estimating characteristics jf joint service of real-time traffic and elastic data traffic in multiservice nodes. Stochastic Systems. 2024;**84**:1191-1207

[9] Yang H, Jing F, Wu J, Zukerman M. A study of a loss system with priorities. Open Access. 30 Aug 2024;**10**(6):e36109. Available from: https://www.cell.com/heliyon/fulltext/S2405-8440(24)12140-8

[10] Willinger W, Taqqu MS, Erramilli A. A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks, Stochastic Networks, Theory and Applications. Vol. 4. Royal Statistical Society Lecture Notes Series; 1996. pp. 339-366. Available from: https://www.semanticscholar.org/paper/A-Bibliographical-Guide-to-Self-Similar-Traffic-and-Erramilli-Taqqu/2a3c828e8e53844e47e4470bf338a2efd6f998c1

[11] Willinger W, Taqqu MS, Sherman R, Wilson DV. Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. IEEE ACM Transactions on Networking. 1996;**5**:71-86

[12] Mirzakulova S, Ibrayeva Z, Kuanova S, Mamyrova A, Japparkulov B,

Kamal R. Visual identification of some regularities in packet network traffic. Eastern-European Journal of Enterprise Technologies. 2024;**1/4**(127):32-42

[13] Melo EF, de Olivera HM. An overview of self-similar traffic: Its implications in the network design. Computer Science Networking and Internet Architecture. 6 May 2020. pp. 38-46. Available from: https://www.researchgate.net/publication/341538837_An_Overview_of_Self-Similar_Traffic_Its_Implications_in_the_Network_Design

[14] Serikov T, Zhetpisbayeva A, Mirzakulova S, Zhetpisbayev AK, Ibraeva Z, Tolegenova A, et al. City backbone network traffic forecasting. International Journal of Electronics and Telecommunications. 2021;**67**(3):319-324

[15] Kanagarathinam MR, Sivalingam KM, Choudhary GK. Application prioritization engine for enhancing real-time performance in smartphones. IEEE Transactions on Network and Service Management. Feb 2024;**21**(1):773-788. DOI: 10.1109/TNSM.2023.3291706. Available from: https://ieeexplore.ieee.org/abstract/document/10172036

[16] Sundararajan A, Chavan A, Saleem D, Sarwat AI. A survey of protocol-level challenges and solutions for distributed energy resource cyber-physical security. Energies. 2018;**11**(9):2360. DOI: 10.3390/en11092360

[17] Tian YC, Gao J. Network addressing architecture. In: Network Analysis and Architecture. Singapore: Springer Nature Singapore; 2023. pp. 161-219

[18] QOS Lecture 6 - Classification and Marking. Available from: https://www.slideserve.com/zavad/qos

[19] Review on QoS Provisioning Approaches for Supporting Video Traffic in IEEE802.11e: Challenges and Issues. Available from: https://www.researchgate.net/publication/327949368_Review_on_QoS_Provisioning_Approaches_for_Supporting_Video_Traffic_in_IEEE80211e_Challenges_and_Issues

[20] Adjardjah W, Kumassah F, Abdallah DM, Addor JA. Performance evaluation of VoIP analysis and simulation. Journal of Engineering Research and Reports. 2023;**25**(7):176-191

[21] Saibharath S, Mishra S, Hota C. Joint QoS and energy-efficient resource allocation and scheduling in 5G network slicing. Computer Communications. 2023;**202**:110-123

[22] Lee H, Kim S, Park G. Efficient QoS policy implementation using DSCP redefinition: Towards network load balancing. The Journal of the Convergence on Culture Technology. 2023;**9**(3):715-720

[23] Shudrenko Y, Kuladinithi K, Plöger D, Timm-Giel A. Optimizing data latency for time-critical avionic sensors. In: 2023 33rd International Telecommunication Networks and Applications Conference. IEEE; 2023. pp. 183-189

[24] Yang H, Guo B, Xue X, Deng X, Zhao Y, Cui X, et al. Interruption tolerance strategy for LEO constellation with optical inter-satellite link. IEEE Transactions on Network and Service Management. 2023;**20**(4):4815-4830

[25] Wijaya A, Abdullah A, Windriyani E, Samaeni FC, Romdhan MY, Ardiansah R. Implementasi quality of service (QoS) menggunakan Wireshark pada Jaringan Wireless LAN. Digital Transformation Technology. 2024;**4**(1):296-303

[26] Ramya R, Ramamoorthy S. QoS in multimedia application for IoT devices through edge intelligence. Multimedia Tools and Applications. 2024;**83**(3):9227-9250

[27] Rzym G, Duliński Z, Chołda P. Dynamic link metric selection for traffic aggregation and multipath transmission in software-defined networks. Applied Sciences. 2024;**14**(12):5312

Section 2

Technological Solutions and Implementations

Chapter 5

Implementation of an Enhanced Security Algorithm for Wireless Sensor Networks

Puseletso Sebothoma and Topside Ehleketani Mathonsi

Abstract

Wireless Sensor Networks (WSNs) have become increasingly important due to their widespread accessibility and affordability. However, they face challenges such as limited power resources, high bandwidth requirements, and security threats like Denial-of-Sleep (DoS) attacks. This chapter proposes a new security solution called Dysfunction Sensor Detection-Rivest-Shamir-Adleman Algorithm (DSD-RSA) to address these challenges. DSD-RSA aims to improve network throughput, reduce power consumption, and minimize end-to-end latency in WSNs. This was achieved by incorporating Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA), and Synchronous Medium Access Control (S-MAC) when designing the proposed DSD-RSA algorithm. DSD-RSA outperformed traditional algorithms in terms of energy consumption, end-to-end delay, and throughput. DSD-RSA achieved an average energy consumption of 30%, compared to 50% for WSN-FAHN, 35% for GA-DoSLD, and 45% for GSHMAC. DSD-RSA also demonstrated superior performance in terms of end-to-end delay and throughput.

Keywords: DSD-RSA, DoSL, QoS, WSNs, throughput

1. Introduction

Wireless Sensor Networks (WSNs) are made up of several nodes that are arranged in a specific spatial configuration and are connected by one or more sensors. As Refs. [1, 2] show, these sensors play a crucial role in monitoring a wide range of the physical environment. These nodes, which are often called wireless devices, have small form factors and built-in features including sensing, communication, onboard computing, and storage [3]. The importance of WSNs has increased due to their low-cost network deployment and ubiquitous internet connectivity, which allows them to be flexible in a wide range of applications, such as military, aviation, healthcare, home automation, transportation, fitness, and forestry [4]. The visual delineation of WSN application scenarios across these various domains is vividly portrayed in **Figure 1**.

Sensor nodes in wireless environments face a variety of security issues, which are discussed in Ref. [5]. Malevolent actors in this environment compromise the integrity of the network, causing disruptions to its regular operations. To save energy, nodes often use power-saving techniques that enable sporadic switches to a standby

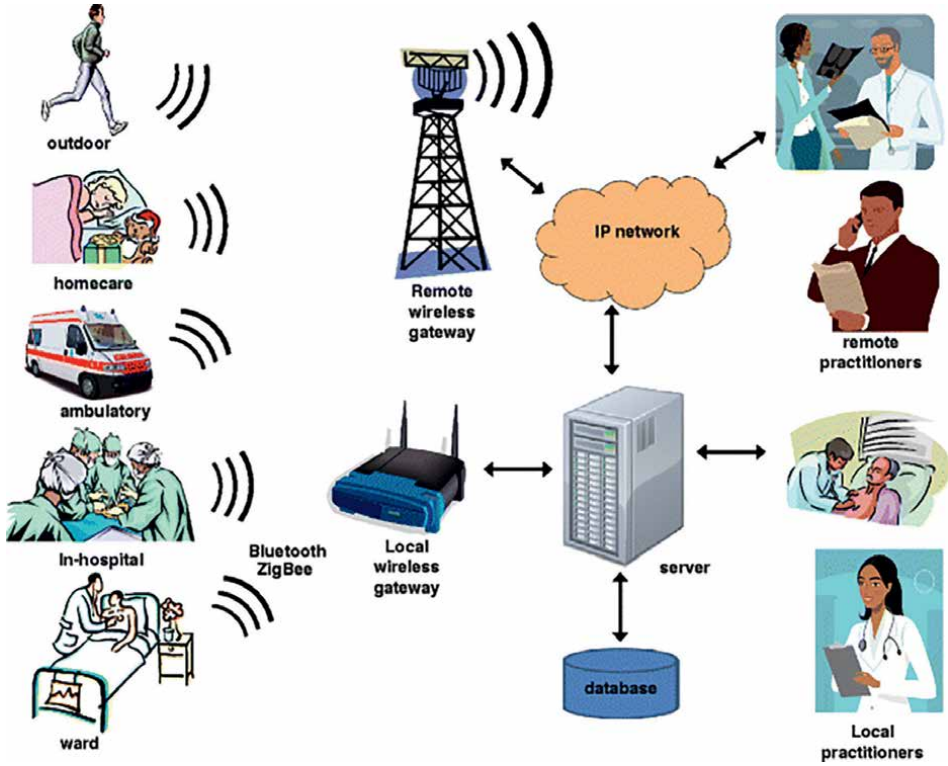

Figure 1.
WSN application scenarios.

state, sometimes referred to as sleep mode [6]. Thus, a node with malicious intent may sneak into the network and prevent nodes that want to go into sleep mode from turning down their communication modules. This malicious activity causes node operational irregularities, which impair network performance [7]. This phenomenon is consistent with the theory of sleep deprivation torture, which is **Figure 2**'s analogue of Denial-of-Sleep attacks (DoSL). DoSL assaults give rise to a series of problems that include increased energy usage, increased end-to-end latency, and decreased network performance. The overall effectiveness and operational efficiency of the network are negatively impacted by this cluster of problems [8].

Prior research has shown significant efforts to protect Wireless Sensor Networks (WSNs) from Denial-of-Sleep (DoSL) attacks and to improve WSN security by means of algorithmic advances and security protocols [9]. Numerous strategies have been proposed by researchers to counter these attacks, most of which focus on energy management [10, 11]. Though useful for energy optimisation, these current techniques often face issues such as higher latency and lower network performance. The Quality of Service (QoS) and overall network performance could be greatly impacted by these constraints. The Dysfunction Sensor Detection-Rivest-Shamir-Adleman (DSD-RSA) algorithm, which is purposefully designed to improve energy efficiency, end-to-end latency, and network throughput within WSNs, is introduced in this study to address these problems. Strengthening network security and operational effectiveness, the DSD-RSA method combines Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA), and the Synchronous Medium Access Control (S-MAC) protocol.

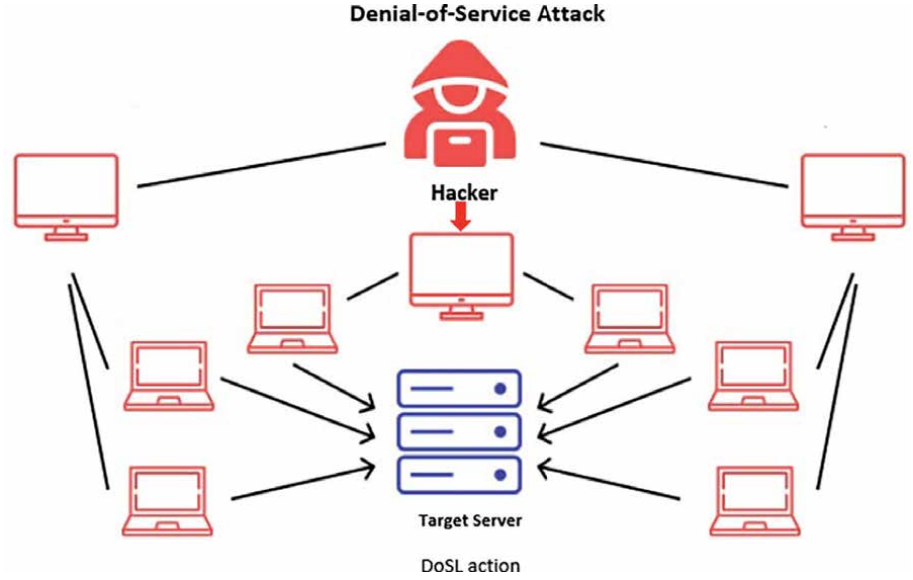

Figure 2.
WSN with DoSL.

The DSD-RSA technique was purposefully designed to address the issues related to energy consumption, network speed, and end-to-end delay in Wireless Sensor Networks (WSNs), all of which are caused by DoSL assaults. AES was utilised in this endeavour to ensure data validity and secrecy, and Rivest-Shamir-Adleman (RSA) enabled safe key exchange. Through the use of the Synchronous Medium Access Control (S-MAC) protocol, nodes were able to meet Quality of Service (QoS) requirements by wisely conserving energy through a listen-and-sleep cycle. Symmetric key encryption was provided by AES, strong encryption was guaranteed by RSA, and nodes may enter sleep mode while idle thanks to S-MAC. The amalgamation of these algorithms under the umbrella of the DSD-RSA framework was orchestrated with the explicit intent of elevating both network security and performance.

This paper's remaining sections are organised as follows: A thorough assessment of earlier studies on systems' defence against Denial-of-Sleep attacks is given in Section II. The suggested DSD-RSA algorithm is presented in Section III. The simulation results that were obtained after the algorithm was applied are shown in Section IV. Lastly, Section V provides an overview of the next steps and wraps up the report with some observations on the results.

2. Related works

The existing literature on DoSL attacks in WSNs reveals that numerous algorithms have been proposed by researchers, each with its own limitations. This section provides a discussion of these existing algorithms and their drawbacks.

Green and Secure Hybrid Medium Access Control (GSHMAC) is a technique devised by Paware et al. [12] to counteract Denial-of-Sleep (DoSL) attacks in Wireless Sensor Networks (WSNs). Using Time Division Multiple Access (TDMA) and Carrier Sense Multiple Access (CSMA) modes for communication within clusters and TDMA

for communication between clusters, GSHMAC is a cluster-based hybrid MAC method. During intra-cluster communication, this mechanism dynamically adjusts modes based on collision thresholds. Additionally, GSHMAC improves security by requiring TDMA-only mode for inter-cluster connections, guaranteeing nodes only exchange messages within allocated time intervals. By altering the underlying MAC mechanism to recognise and prevent replay and collision attacks, the algorithm further improves security by fortifying defences against full-domination attacks. GSHMAC's considerable gains in network throughput and energy usage were confirmed by NS-2 simulations. However, the method that was presented had trouble reducing end-to-end latency since mode switching and TDMA scheduling in a hybrid MAC system take a lot of time.

In contrast, our algorithm is tailored to minimise end-to-end delays in WSNs. While sharing similarities with the approach proposed in the aforementioned study, both schemes deploy hybrid MAC strategies to counter DoSL attacks, thereby augmenting power efficiency and throughput. However, our algorithm advances beyond by incorporating a novel hybrid approach to further fine-tune network performance.

A unique method of countering Denial-of-Sleep (DoSL) assaults in Wireless Sensor Networks (WSNs) was introduced by Fotohi & Bari [13]. This method involved combining the Firefly algorithm with Hopfield Neural Network (HNN) algorithms in a novel way. One noteworthy aspect of their methodology is the employment of a movable sink to augment energy efficiency and prolong network lifetime. In order to prevent DoSL attacks, the Firefly algorithm is deliberately used to coordinate node clustering and authentication at two different levels. Simultaneously, the Hopfield neural network is utilised to ascertain the best route for the mobile sink to travel, guaranteeing efficient data transfer to the Cluster Head (CH). The effectiveness of the suggested WSN-FAHN technique is carefully assessed through large-scale simulations run in the NS-2 environment. The results of these simulations highlight the WSN-FAHN approach's improved performance over existing methods, as seen by several performance indicators. Although the authors claim that their method reduces energy consumption and ensures an impressive average throughput, it is important to note a crucial difference: our proposed algorithm, which is the subject of this study, explicitly addresses the end-to-end delay, which the authors' algorithm fails to take into account.

In response to DoSL attacks, Gunasekaran & Periakaruppan [14] introduced the Genetic technique - Denial-of-Sleep Attack Detection (GA-DoSLD) technique. In order to strengthen the integrity of relay nodes, their methodology combined the AODV routing algorithm with RSA cryptography. Nevertheless, the researchers chose to use the S-MAC protocol in their investigation because they were aware of the energy consumption issues associated with the AODV protocol. We leveraged the S-MAC protocol, a modified RSA algorithm, and AES encryption in our proposed DSD-RSA method to decrease energy usage while thwarting potential assaults and offering strong cryptographic key security. Additionally, our programme divided the network into clusters and adopted an energy-focused strategy, carefully selecting appropriate cluster heads. Through this effort, energy consumption levels were optimised, and overall network performance was improved.

Innovative methods for detecting Denial-of-Sleep (DoSL) assaults in cluster-based sensor networks were presented by Osanaiye et al. [15]. The selection of controller nodes responsible for monitoring and reporting DoSL attack instances was the main focus of their methodology. These assigned controller nodes analysed network traffic in detail and alerted the cluster head right away to any unusual traffic patterns.

This dynamic strategy attempted to increase network lifetime by reducing the amount of energy used by each sensor node while also improving security by preventing attacks. The algorithm was shown to be superior to existing systems such as X-MAC, ZKP, and TE2P through simulations. Similar to this, our DSD-RSA technique aims to reduce energy consumption while enhancing security in Wireless Sensor Networks (WSNs). We present a low-power cyber-security mechanism designed especially for smart grid monitoring applications based on wireless sensor networks. This system maintains exceptional energy efficiency while detecting and isolating a variety of threats, including replay, forgeries, and denial-of-service attacks. The simulated outcomes validate the superiority of our mechanism over the state-of-the-art methods, exhibiting higher power efficiency while maintaining continuously high standards of latency and dependability.

An enhanced DoS detection technique was presented by Suryaprabha & Kumar [16] to strengthen Wireless Sensor Networks (WSNs). Their method combines statistical analysis and machine learning approaches to create a strong defence against Denial-of-Sleep (DoSL) assaults in the network. The algorithm reduces the number of false positives and false negatives, giving energy conservation priority. The programme effectively detects DoSL assaults by utilising a mix of the decision tree algorithm and statistical analysis. The efficiency of the approach in terms of energy efficiency and DoSL detection is demonstrated through the use of MATLAB-based simulations. Compared to current systems, it achieves higher detection accuracy and lower false positive rates. It is noteworthy, however, that the algorithm does not take the network's end-to-end delay into account, which could cause delays in the transfer of data from source nodes to destination nodes.

A novel proactive method was put up by Patil & Chaudhari [17] to address the problem of Denial-of-Sleep (DoSL) assaults in Wireless Sensor Networks (WSNs). By using a reputation-based mechanism, their method seeks to identify and isolate malicious nodes that initiate DoSL assaults. This system creates reputation scores by attentively examining nodes' actions inside the network; nodes with lower scores can be identified as possible harmful actors and removed from the network. Extensive simulations are used to thoroughly test the technique's efficiency, highlighting its ability to prevent DoSL assaults in WSNs beforehand. But it's important to recognise the inherent drawbacks of this strategy. One significant flaw is the absence of performance measurements that are vital for a thorough evaluation of WSN performance, such as throughput and end-to-end delay. Furthermore, the suggested reputation-based method might not be adequate to handle extremely complex attacks that can get past such systems, highlighting the significance of a multi-layered security approach.

This paper's literature review showed that the existing DoSL algorithms fall short of meeting the necessary standards for Quality of Service (QoS). These methods frequently cannot guarantee DoSL detection and prevention, cut down on energy use, minimise latency, or maximise network speed [18].

3. Dysfunction sensor detection-Rivest-Shamir-Adleman

Using a multidisciplinary approach, the Dysfunction Sensor Detection-Rivest-Shamir-Adleman (DSD-RSA) algorithm was carefully designed with data security and energy optimisation as top priorities. The Advanced Encryption Standard (AES), which serves as the foundation for symmetric key encryption, was utilised in this novel technique. The purpose of this strategic implementation was to improve

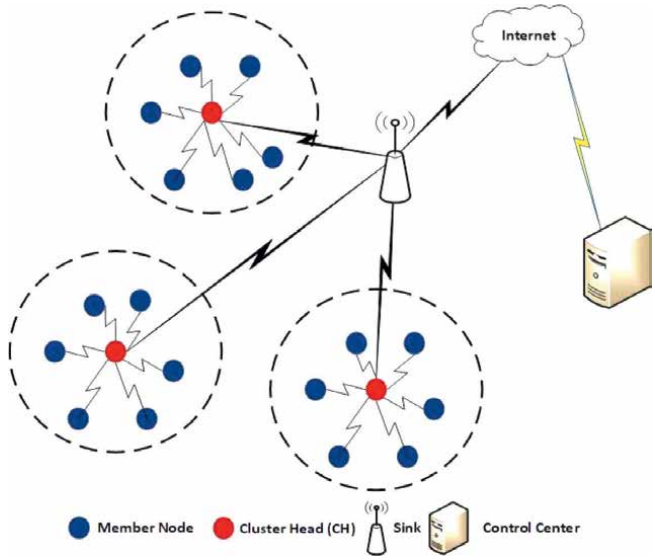

Figure 3.
DoSL WSN systems architecture.

the data integrity and confidentiality of the ecosystem. Simultaneously, the algorithm leveraged the robustness of the Rivest-Shamir-Adleman (RSA) algorithm and employed it for safe key exchange. By strengthening encryption techniques, this protected the channels of communication from possible breaches. Interestingly, the algorithm's framework was able to incorporate the Synchronous Medium Access Control (S-MAC) protocol with ease. Through a cyclical listen-and-sleep cycle, this protocol gave sensors a dynamic technique to save energy. During moments of inactivity, this adaptive functionality allowed sensors to go to a low-power sleep mode while still allowing them to communicate with other network nodes. This energy-saving and communication-continuity symbiosis was carefully designed to meet Quality of Service (QoS) requirements. Through compliance with these specifications, which included assured energy consumption, low latency, and maximum network throughput, the DSD-RSA algorithm was prepared to handle a wide range of efficiency and performance parameters. **Figure 3** graphically illustrates this holistic approach by showing the complex interplay between the algorithm's component parts.

3.1 Advanced encryption standard

A single key is used for both the encryption and decryption phases in the symmetric encryption method known as the AES algorithm. It works by encrypting data in blocks of 128 bits, and it can handle keys of lengths of 128, 192, or 256 bits [19]. This study suggests that in situations where key management is weak or where a key is stolen, digital data should be encrypted using the AES method. The suggested approach uses ciphertext that has been encrypted with a 256-bit key using 14 rounds of the RSA method and the AES technique to carry out another cryptographic operation. Key encryption is done using the AES method, and key transfer activities are carried out via an interlocking protocol to safeguard keys against attackers.

3.2 Rivest-Shamir-Sdleman

Through its dual use in key distribution and message decoding, the RSA algorithm significantly improved data security and strengthened confidentiality measures [20]. In addition to its encryption capabilities, RSA proved to be versatile by including key exchange and node authentication processes, which prevented possible Denial-of-Sleep (DoSL) assaults. The attempt to synchronise the nodes in the network was accomplished by means of an intricate system in which the amount of time that each node would sleep was inferred from the SYNC signals that were transmitted between them. By using the data carried within SYNC packets to construct sleep intervals, this novel method did away with the need for reset sleep intervals. The calculation of the NewSleepTime parameter, an essential component that determines the nodes' sleep periods, is important to this synchronisation. Eq. (1) carefully controlled this computing process, indicating the study's careful attention to technical detail and synchronisation protocols.

$$New_{SleepTime} = \left(Old_{sleepTime} + \frac{\text{Re}ceivedSYN\ PacketSleepTime}{2} \right) \tag{1}$$

where:

NextSyncTime: the time of the next scheduled synchronisation message.

Old sleep time: time that a node has spent sleeping before waking up to receive the next SYNC message.

The section explains how to use SYNC messages to gradually synchronise nodes in a sensor network and how to use key exchange, message decryption, and RSA algorithm-based node authentication to thwart DoSL assaults. Nonetheless, by repeatedly sending SYNC packets, attackers can still initiate DoSL attacks, and sensor nodes' sleep modes can greatly lower energy usage and increase network throughput.

3.3 Sensor medium access protocol

The Sub-MAC (S-MAC) protocol is a carefully designed response to the critical problem of energy efficiency in sensor networks. It accomplishes this by putting into practice a carefully thought-out cyclic listen-sleep system that maximises energy usage for individual nodes. But it's critical to be aware of a potential weakness: the continuous sending of control packets, particularly Request-to-Send (RTS) messages, may unintentionally provide a path for the execution of Denial-of-Sleep (DoSL) attacks. If successful, these attacks can result in needless energy waste. The deployment of Cluster Heads (CH) is a strategic intervention aimed at improving the network's overall resilience and lifespan. This method, which is based on the idea of clustering, reduces communication distances and successfully counteracts the inefficiencies that come with having a centralised washbasin or base. Moreover, the act of combining data prior to sending it to the CH greatly reduces the computing load, promoting increased operational effectiveness. It is important to remember that clustering starts to matter when there are more than five nodes. Within the framework of the novel DSD-RSA method used in this study, a laborious selection procedure guarantees the discovery of a suitable CH for efficient data gathering. This all-encompassing approach is in perfect harmony with the primary goal of improving energy efficiency and overall network performance in the ever-changing domain of sensor networks.

This selection process involves the following steps:

Step 1: Initially, choosing a node that is near the base station is prioritised. For this, the Euclidean Eq. (2) is applied, presuming that every node in the network has Global Positioning System (GPS) capabilities and is aware of its current location.

$$CS = \left(x_i - x_s\right)^2 + \left(x_i - x_s\right)^2 \tag{2}$$

Where: the energy of each selected node is measured using Eq. (2), where C represents the coordinates of each node and S represents the coordinate of the sink.

Step 2: Nodes are deemed suitable as CHs based on their energy efficiency and distance from the base station.

$$Q = \frac{e}{t} \tag{3}$$

This model hinges on two critical parameters: e , the node's initial energy measured in kilojoules, and t , the time elapsed in minutes. These factors play a pivotal role in understanding the system's behaviour.

Step 3: Eq. (3) is used to choose the CH after computing the energy levels of the nodes and figuring out how far apart they are from the base station.

Step 4: The energy used by each node must be determined in order to choose the CH. This makes it possible to determine which node has the most energy left, and that node is then selected to be the cluster head. This is achieved by using Eq. (4):

$$e = Q.t \tag{4}$$

The model takes into consideration the intricacies of the real world by integrating the parameters of the channel, such as its free space or multipath fading susceptibility, as well as the separation distance between the transmitter and receiver. Equation reflects this nuanced approach (5). The writers of this publication have presented this model.

$$E_{tx}\left(T,di\right) = \begin{Bmatrix} T.E_{elec} + T_{efs}.d^2 \ \ di, < di_0 \\ T.E_{elec} + Ke_{mp}.d^4 \ di \geq di_0 \end{Bmatrix} \tag{5}$$

Where:

E_{elec} refers to the amount of energy needed to activate the circuits.

E_{mpi} and E_{fsi} are relay activation energy models for multipath channels and empty space, respectively, di^2 denotes the power loss in free space, while power loss in multipath fading is di^4 . T is time and T_{efs} is the total time taken to activate the circuit.

3.4 Enhanced security prevention DSD-RSA algorithm flowchart

The flowchart of the proposed DSD-RSA algorithm is depicted in **Figure 4**.

3.5 Enhanced security prevention DSD-RSA algorithm

To identify and prevent DoSL attacks while meeting QoS requirements for energy consumption, end-to-end latency, and throughput, the proposed DSD-RSA was created by combining the S-MAC protocol, the AES algorithm, and the RSA algorithm (see Algorithm 1).

Algorithm 1. Pseudo code for DSD-RSA proposed model

1. Step 1: Network Clustering
2. Create a random WSN with sensor nodes.

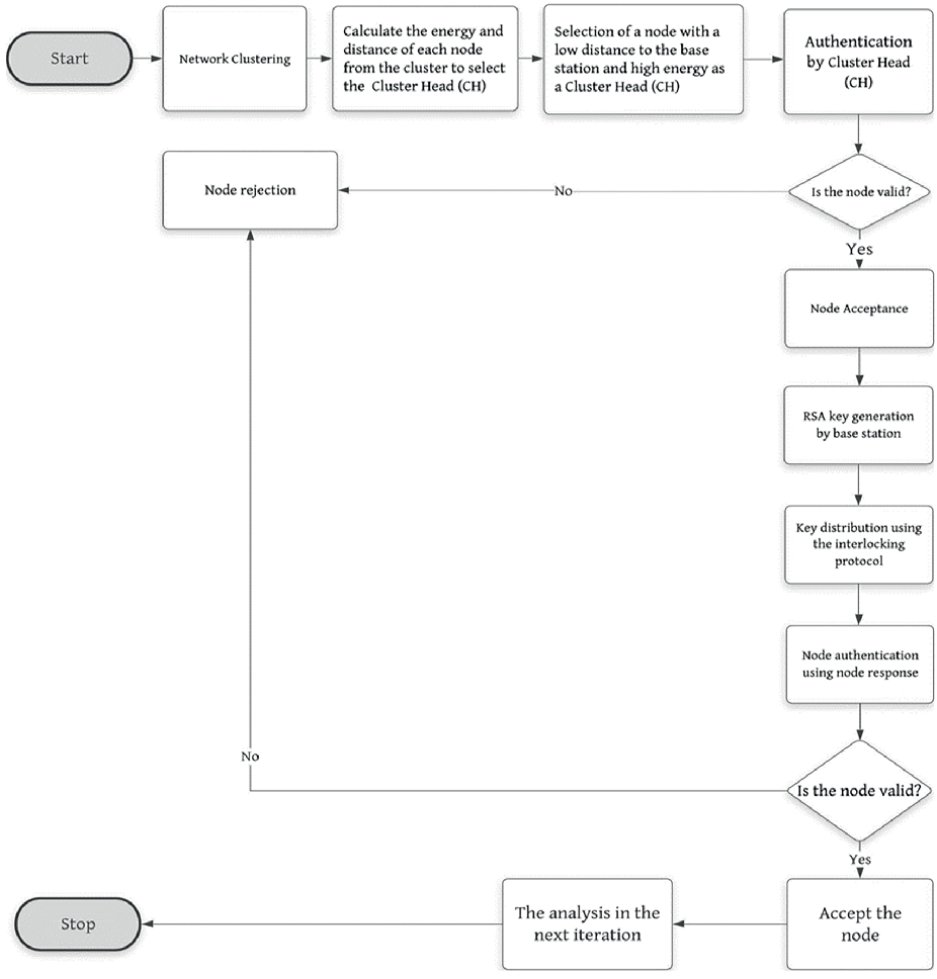

Figure 4.
DSD-RSA algorithm flowchart.

3. Set the parameters for Algorithm 1.
4. Choose the Cluster Head (CH).
5. Compute the energy level of each node.
6. Determine the distance of each node.
7. Select a node with low distance to the base station and high energy as the CH.
8. Node Authentication Procedure
9. Step 2: Cluster-Level Node Authentication
10. Calculate the threshold value.
11. Determine the interval.

12. Repeat the following steps while the interval is greater than the threshold:

13. Authenticate each node at the cluster level.

14. End loop.

15. End repetition.

16. Otherwise, accept the node.

17. End procedure.

18. Step 3: Network-Wide General Authentication

19. Generate RSA keys at the base station (BS).

20. Verify node authenticity by requesting node responses.

21. Reject the node if it is invalid.

22. Accept the node if it is valid.

23. End conditional statement.

24. End procedure.

The DSD-RSA technique, which attempts to identify and mitigate DoSL attacks while guaranteeing QoS requirements including energy consumption, end-to-end delay, and throughput, is introduced in this section. The suggested DSD-RSA algorithm is explained in depth in the section that follows.

4. Simulation results

This research paper presents the implementation of the suggested Dysfunction Sensor Detection-Rivest-Shamir-Adleman (DSD-RSA) algorithm, which was carried out with great care using the IEEE 802.11w model that was included in Network Simulator 2 (NS-2) version 2.35. A Linux Ubuntu 14.04 64-bit operating system with 1 GB of RAM and a generous 10 Gigabytes (GB) of storage capacity was used to carefully set up this simulation environment. The software Oracle Virtual Box Manager (VM) Version 6.1.30, created by Oracle Corporation in 2021, was specifically used for the host platform.

The complex network topology configurations and the DSD-RSA model simulation were implemented using a combination of Tool Command Language (TCL) scripts and C++ source code. The network topology that was defined covered a 400 × 400 m spatial area. It consisted of five Sink Base Stations (SBS1, SBS2, SBS3, SBS4, and SBS5) that were fixedly positioned and entwined with 20 randomly positioned nodes, as **Figure 5** of the study vividly illustrates. Notably, the simulation timeline was carefully calibrated to begin packet transmission at the 20-second interval and end smoothly at the 120-second point. This well-thought-out simulation framework made it easier to carry out the complex evaluation of the DSD-RSA algorithm's performance and how it affects network dynamics in a regulated but realistically set-up environment.

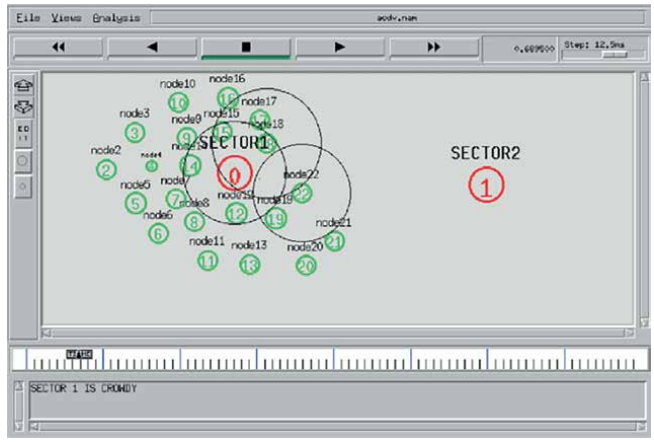

Figure 5.
NAM simulation scenario.

4.1 Average throughput

Data transfer between nodes in a network over a given time period is measured by network throughput. **Figure 6** shows the average throughput of the network. This statistic measures the responsiveness, capacity, and efficiency of the network, all important for achieving Quality of Service (QoS) requirements in a variety of applications and maximising communication.

Examining the results presented in **Figure 6**, the suggested Dysfunction Sensor Detection-Rivest-Shamir-Adleman (DSD-RSA) method demonstrated an impressive 95% average throughput in 120 seconds. Similarly, throughputs of 89, 65, and 60% were achieved by the GA-DoSLD, WSN-FAHN, and GSHMAC algorithms, respectively. This impressive throughput is mostly due to the effective energy conservation measures of the DSD-RSA algorithm, which increased the bandwidth available for effective data transmission. A key component of this success was the

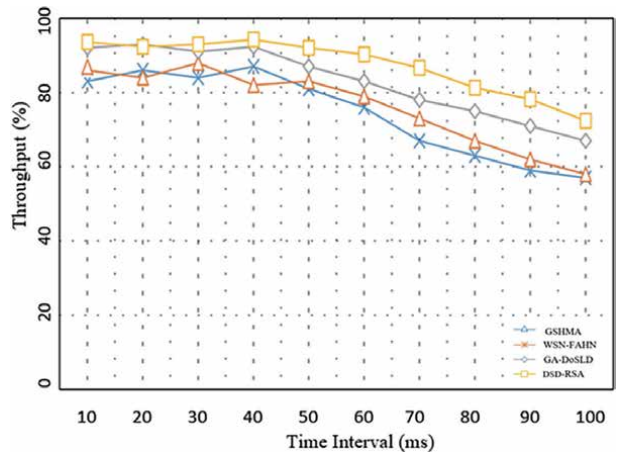

Figure 6.
Average network throughput.

skilful application of the S-MAC protocol. Moreover, the DSD-RSA method's fundamental design included the RSA algorithm, which made it capable of both detecting and preventing Denial-of-Sleep (DoSL) assaults. On the other hand, the algorithms GA-DoSLD, WSN-FAHN, and GSHMAC found it difficult to significantly reduce energy usage while still preventing DoSL attacks. As a result, the overall QoS that these alternative algorithms could provide was compromised because they were unable to allocate large throughputs in an efficient manner. The comparison of these results highlights the DSD-RSA algorithm's strong performance in reducing security risks and improving network performance, which in turn translates to improved quality of service.

4.2 Average end-to-end delay

The end-to-end delay is the average time taken to transmit packets. **Figure 7** portrays the analysis of end-to-end delay concerning packet size. This assessment offers insights into how packet size influences transmission time, critical for optimising network performance and assessing QoS.

The usefulness of the suggested DSD-RSA method in reducing end-to-end delay by preventing pointless transmissions is shown in **Figure 7**. In comparison to the GA-DoSLD, WSN-FAHN, and GSHMAC algorithms, DSD-RSA considerably minimises the overall latency. The DSD-RSA, GA-DoSLD, WSN-FAHN, and GSHMAC algorithms had average delays of 19, 35, 40, and 60%, respectively.

4.3 Energy consumption

The average energy consumption of the proposed DSD-RSA algorithm and the current GA-DoSLD, WSN-FAHN, and GSHMAC algorithms are compared in **Figure 8**. The energy that these algorithms need to transmit data packets is represented by the average energy consumption.

The graph presents unambiguous proof that the study's suggested DSD-RSA technique achieves noticeably lower energy consumption when compared to other

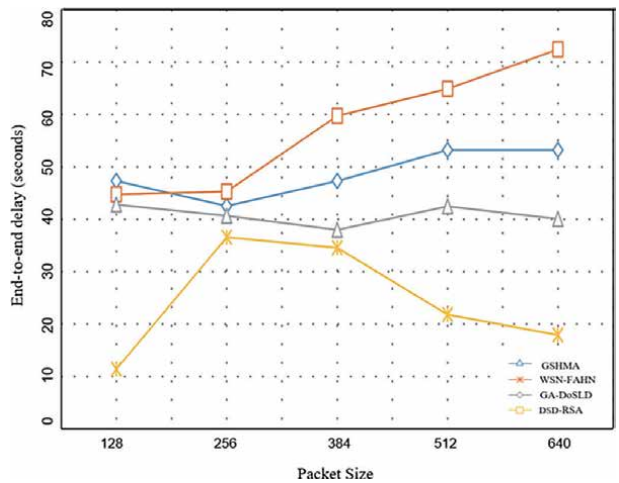

Figure 7.
Average end-to-end delay.

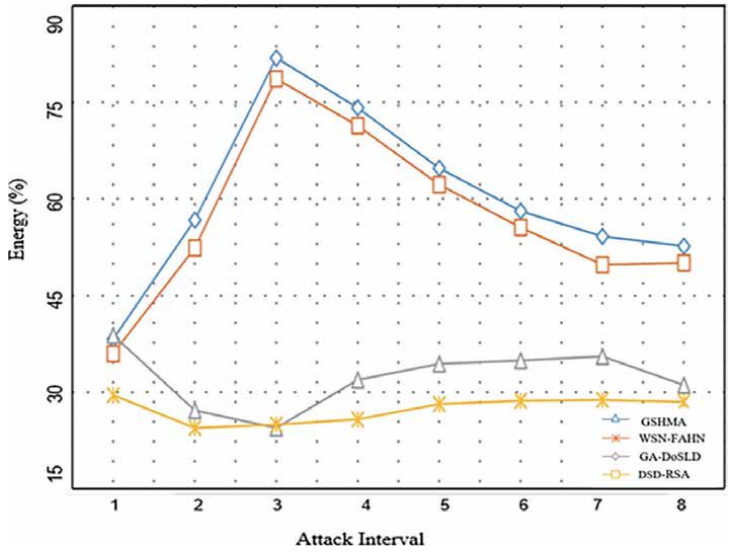

Figure 8.
Energy consumption.

DoSL attack prevention schemes. The suggested DSD-RSA algorithm used 30% of the energy, WSN-FAHN algorithm used 35%, GSHMAC algorithm used 45%, and GA-DoSLD algorithm used 50%.

5. Conclusion and future work

The DSD-RSA algorithm described in this paper is a comprehensive method designed to accomplish multiple important goals, including preventing and identifying DoSL attacks, reducing energy usage, minimising latency, and maximising network performance. In order to create the DSD-RSA algorithm, the S-MAC protocol, the AES algorithm, and the RSA algorithm were strategically combined. This resulted in a strong defence against DoSL attacks in WSNs while maintaining QoS by carefully managing end-to-end delay and throughput.

The S-MAC protocol was chosen because of its effectiveness in energy-rich sensor networks. This protocol performs well throughout extended periods of inactivity and handles delay situations well, making it useful for our needs. Our cryptographic architecture is based on the RSA method, which performs two functions: it keeps attackers at bay and makes secure key exchanges possible. Concurrently, the AES algorithm intervenes to strengthen cryptographic keys against possible compromises.

NS-2 was used to run extensive simulations in order to verify the DSD-RSA algorithm's effectiveness. The results clearly demonstrated how well the system can identify and prevent DoSL attacks, which improves network security. Furthering the field of network performance enhancement, the algorithm also brought about a large reduction in energy usage, minimised end-to-end delays, and noticeably increased network throughput.

Our future plans include incorporating machine learning techniques into the DSD-RSA architecture. By utilising the capabilities of cutting-edge algorithms, this tactical advancement seeks to improve the precision of DoSL detection and

prevention. Concurrently, this project is in line with our objective of minimising energy consumption, reducing end-to-end latency, and increasing network capacity. This upward trajectory demonstrates our steadfast commitment to remaining at the forefront of innovation in strengthening Wireless Sensor Networks' efficiency and resilience against changing security threats.

Acknowledgements

The Tshwane University of Technology is acknowledged by the authors for its financial support of this study. Additionally, the authors attest that the publishing of this work has no conflicts of interest.

Author details

Puseletso Sebothoma and Topside Ehleketani Mathonsi*
Department of Information Technology, Tshwane University of Technology Pretoria, South Africa

*Address all correspondence to: mathonsite@tut.ac.za

References

[1] Raouf MM. Clustering in wireless sensor networks (WSNs). Journal of Baghdad University College of Economic Sciences, Baghdad, Iraq. 2019;**57**(4):1-9

[2] Prasanna S, Rao S. An overview of wireless sensor networks applications and security. International Journal of Soft Computing and Engineering. 2012;**2**(2):2231-2307

[3] Jondhale SR, Maheswar R, Lloret J. Fundamentals of wireless sensor networks. In: Received Signal Strength Based Target Localization and Tracking Using Wireless Sensor Networks. Midtown Manhattan, New York City: Springer; 2022. pp. 1-9

[4] Haldorai A, Ramu A, Murugan S. Smart sensor networking and green technologies in urban areas. In: Computing and Communication Systems in Urban Development: A Detailed Perspective. Midtown Manhattan, New York City: Springer; 2019. pp. 205-224

[5] Bhola J, Soni S, Cheema GK. Recent trends for security applications in wireless sensor networks–a technical review. In: 2019 Sixth International Conference on Computing for Sustainable Global Development (INDIACom). New Delhi, India: IEEE; 2019. pp. 707-712

[6] O'Mahony GD, Curran JT, Harris PJ, Murphy CC. Interference and intrusion in wireless sensor networks. IEEE Aerospace and Electronic Systems Magazine. 2020;**35**(2):4-16

[7] Tamilarasi N. Neutralization of denial of service attack in wireless sensor networks. Pramana Research Journal. 2019;**9**(5):454-461

[8] Udoh E, Getov V. Layered-MAC: An energy-protected and efficient protocol for wireless sensor networks. In: Mobile Wireless Middleware, Operating Systems and Applications: 10th International Conference on Mobile Wireless Middleware, Operating Systems and Applications (MOBILWARE 2021). Cham: Springer International Publishing; 2022. pp. 45-61

[9] Gong J. An application of meta-heuristic and nature-inspired algorithms for designing reliable networks based on the internet of things: A systematic literature review. International Journal of Communication Systems. 2023;**36**(5):e5416

[10] Rajasoundaran S, Prabu AV, Kumar GS, Malla PP, Routray S. Secure opportunistic watchdog production in wireless sensor networks: A review. Wireless Personal Communications. 2021;**120**(2):1895-1919

[11] Amutha J, Sharma S, Sharma SK. Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review, taxonomy, research findings, challenges and future directions. Computer Science Review. 2021;**40**:100376

[12] Pawar PM, Nielsen RH, Prasad NR, Prasad R. GSHMAC: Green and secure hybrid medium access control for wireless sensor network. Wireless Personal Communications. 2018;**100**:267-281

[13] Fotohi R, Firoozi BS. A novel countermeasure technique to protect WSN against denial-of-sleep attacks using firefly and Hopfield neural network (HNN) algorithms.

The Journal of Supercomputing. 2020;**76**(9):6860-6886

[14] Gunasekaran M, Periakaruppan S. GA-DoSLD: Genetic algorithm based denial-of-sleep attack detection in WSN. Security and Communication Networks. 2017;**2017**(1):1-10

[15] Osanaiye OA, Ogundile OO, Aina F. Evaluating DoS jamming attack on reactive routing protocol in wireless sensor networks. African Journal of Science, Technology, Innovation and Development. 2022;**14**(5):1369-1376

[16] Suryaprabha E, Saravana Kumar NM. Enhancement of security using optimized DoS (denial-of-service) detection algorithm for wireless sensor network. Soft Computing. 2020;**24**(14):10681-10691

[17] Patil S, Chaudhari S. DoS attack prevention technique in wireless sensor networks. Procedia Computer Science. 2016;**79**:715-721

[18] Jinhui X, Yang T, Feiyue Y, Leina P, Juan X, Yao H. Intrusion detection system for hybrid DoS attacks using energy trust in wireless sensor networks. Procedia Computer Science. 2018;**131**:1188-1195

[19] Mohammed SJ, Taha DB. Performance evaluation of RSA, ElGamal, and paillier partial homomorphic encryption algorithms. In: 2022 International Conference on Computer Science and Software Engineering (CSASE). Duhok, Iraq: IEEE; 2022. pp. 89-94

[20] Karmakar T, Biswas S, Das I, Nath S. Varibox encryption algorithm: The new generation of hybrid security measure for the era of quantum computation. Journal of Discrete Mathematical Sciences and Cryptography. 2022:**26**(2):478-505

Chapter 6

Omnidirectional Antenna Design for Underground Wireless Sensor Network QoS Improvement

Khomotso C. Tsoane, Topside E. Mathonsi and Deon P. Du Plessis

Abstract

In this chapter, it is proposed to detect intrusions in subterranean Wireless Sensor Networks (WSNs) employing an enhanced omnidirectional antenna. The proposed arrangement offers a 20 x 20 mm compact square shape and is based on the Square Slotted Monopole Antenna (SSMA) idea, leveraging both realistic micro-etching applications on low-cost substrates and Ansys High-Frequency Structure Simulator (HFSS) simulations to achieve superior performance. The omnidirectional antenna significantly increases the Quality of Service (QoS) for subsurface WSN applications by providing exceptional high-frequency depth, signal dispersion, and spreading. The antenna's subsurface penetration and precision intrusion detection are demonstrated by its average fidelity of 0.9 and spread factor of 1.5 at a depth of 30 cm.

Keywords: intrusion detection, subterranean wireless sensor networks, omnidirectional antenna, square slotted monopole antenna, spread factor, quality of service

1. Introduction

As a critical component of the communication chain, antenna design must align with the specific application's requirements. While both ultra-wideband (ULB) and traditional narrowband antennas serve their purposes, ULB antennas distinguish themselves by their significantly broader bandwidth [1–5]. Moreover, ULB technology offers the unique advantage of detecting subsurface irregularities and enabling timely alerts.

According to an assessment of the literature, traditional antenna methods do not perform as well at stopping cable theft. Conversely, connected motion sensors use pyroelectric infrared (PIR) and active ultrasonic sensors to identify any movement

or impact of an object [3–6]. The precision and dependability of connected passive infrared and active ultrasonic sensor systems are well established. The motion sensor can automatically identify approaching disturbances thanks to ULB antenna technology, and it can sound an alert to ward off attackers. Compared to a standard surveillance camera, this motion sensor technology seeks to prevent copper theft more successfully [1–6].

Due to their oscillators and power amplifiers employing high-power voltage to generate continuous signals, Wireless Sensor Networks (WSNs) technologies utilising narrowband sub-GHz systems are not energy efficient [3, 4]. Additionally, since the accuracy of distance estimates increases with increased radio signal bandwidth, localisation capability—which is essential for some applications—is absent from previous WSN technologies [7–9]. These drawbacks impede the widespread deployment of subterranean WSNs, prompting the search for substitute strategies.

This chapter suggests combining an underground WSN with the pulsed ULB method. Barras [10] identifies the following reasons for the suggested solution's use of ULB technology:

1. Smaller antennas: By using this technology, antennas can be made smaller, which lowers the cost and size of modules and makes deployment easier.

2. Limited power consumption: Subterranean sensors have a longer lifespan due to the low gearbox powers permitted by ULB. Additionally, low-power CMOS technology is easily compatible with ULB transceivers.

3. Localisation capability: These systems should be able to achieve good localisation accuracy thanks to the pulsed ULB technique's very broad bandwidth.

4. Higher communication rates: Higher communication rates are achievable when communication systems' bandwidth and capacity grow.

The primary foreseeable constraint on the pulsed ULB approach is the antennas' shallow subterranean depth. The rest of the document is organised as follows: Related work is included in Section 2. The ULB-based omnidirectional antenna design is shown in Section 3. Additionally, simulation findings comparing the new omnidirectional antenna's effectiveness to that of the traditional directional antenna utilised in subterranean WSNs are included in the research.

2. Related works

Numerous experiments have been carried out, including ones employing buried antennas, to examine how the ground affects antennas and signal propagation. Most of these investigations have concentrated on the narrowband frequency region that is lower than 1 GHz. Barras [10], for example, reported on the design and performance analysis of four UHF antennas functioning at 392.5 MHz in a concrete subterranean environment. The antennas that were studied include a 16.7-cm horizontal dipole, a 38.5-cm-diameter ring, a 150-cm-long travelling wave monopole with a 76-cm-diameter disc ground plane, and a 150-cm-long V-shaped travelling wave monopole

with a 43-cm-diameter disc ground plane. These antennas were free space matched in concrete with a dielectric constant of 6.5 to guarantee correct functioning at the intended frequency [11–13].

Numerous studies have examined how soil and snow affect antenna performance; many of these studies concentrate on buried antennas. These investigations have mostly focused on the frequency region below 1 GHz. Das et al. [14] looked at the impact of soil on 145 MHz resonant antennas buried at shallow depths. Their findings indicate a shift in the resonant frequency towards lower frequencies together with a decline in efficiency [15]. Saeidi et al. [16] proposed a 49 x 29 x 3 mm^3 rectangular microstrip antenna that was buried in the ground for dual functionality—moisture monitoring at 1.85 GHz and communication with an external reader at 1.35 GHz. Their findings indicated a 130 MHz shift in the resonance frequency towards lower frequencies when the subterranean antenna was moved from dry to wet conditions.

An Archimedean spiral antenna operating in the 1–3 GHz range and designed to be buried in the ground to study its dielectric properties was proposed by Curiac [17]. The antenna was printed on an 87 x 87 x 1.5 mm^3 square substrate. This ice-buried antenna produced results that showed an increase in directivity, a decrease in efficiency, and a drop in input impedance. In the study by Huang et al. [18], a rectangular broadband antenna measuring 67 x 42.5 x 1.54 mm^3 is described for in-ground broadcasts. This antenna shows matching in the 1.32–3.94 GHz band in free space, and when buried in dry soil, it may function in the 0.869–2 GHz band [19]. When taken as a whole, these studies highlight how much the ground affects buried antenna performance; the most prominent finding is the bandwidth shift towards lower frequencies. This finding is encouraging for the development of small antennas for subterranean uses.

A WSN system was installed in Norway as part of a study on glacier variation monitoring, and sensors were buried up to 100 meters deep in the glacier [20]. A 468 MHz radio link with a strength of 500 mW was used for the long-distance link between the base station and the reference station, and a 433 MHz radio link with a power of 100 mW was used for communication between the base station and the sensors. The results of the study, however, were more concerned with the physical aspects of the glacier than with the features of the propagation channel.

In a study carried out in a snowy setting, the viability of using ULB technology in WSNs for avalanche victim localisation was investigated [21]. By predicting the permittivity of the snow, a propagation model for a snowy environment was presented and experimentally confirmed in the 3.3 to 4.6 GHz region. After that, a theoretical investigation was conducted to estimate attenuation as a function of distance at frequencies of 0.5 GHz and 3.8 GHz using the suggested model. The VWC values that were taken into consideration were 2.3, 6.4, and 12.1%. Both the receiving antenna's height and the transmitting antenna's subterranean depth were set to one meter. The findings showed that there were notable attenuations when the frequency and water content of the snow were increased. For VWC = 6.2%, the attenuation at 3.8 GHz was −131 dB at 10 meters of separation, while for VWC = 2.3%, it was 90.1 dB at 20.2 meters of separation. In the worst situation, with VWC = 12%, the attenuation at 0.7 GHz was 63 dB at 31 meters of separation.

A microprocessor and an antenna that supported global system for mobile communications (GSM) were used in a WSN that was intended to prevent copper wire

theft [22]. An electrical circuit wire connection was incorporated within the antenna so that the microcontroller antenna input port could detect voltage drops in the cable. The microcontroller antenna produced a broad operational band to cover the DVB-H band through the UHF frequency, which was useful for subterranean theft detection and similar to the directional antenna suggested by Alejandro-Granados et al. [22]. An LEC display screen and a cell phone component were added to the antenna to enable thorough processing and visualisation of incoming data in the subterranean setting.

All things considered, the studies stated above draw attention to the difficulties and factors that must be taken into account while designing and operating antennas in subterranean and icy conditions. When constructing antennas for these applications, it is important to take into account how the surrounding medium affects the antenna characteristics, especially frequency shifts and attenuation.

3. Methods

3.1 Numerical measurement

The suggested solution cycle consists of four programmable time intervals. The length of the cycle changes based on the number of coordinators and the time parameters of each activity phase. The unit of time used inside the interval [T1; T3] is the detection activity slot. A basic Wireless Sensor Network (WSN) slot is made up of 60 symbols or 960 μs (3 backoff periods). The dimensioning of activity slot durations is made possible by the slot duration exponent (EDS), which is comparable to the Superframe Order in the IEEE 802.15.4 standard (see Eq. (1)):

$$\mathrm{dslot} = 2\,\mathrm{EDS} \times 960 \times 10 - 3 \tag{1}$$

Synchronisation Period [T0; T1]: The number of nbC coordinators and the duration, in milliseconds, Tb, of a beacon transmission interval determine the length of the synchronisation period. Tb is specified in a way that:

$$\mathrm{Tb} = 0.18 \times \mathrm{nbC} + 2.8 \tag{2}$$

$$\mathrm{d}[\mathrm{T0};\mathrm{T1}] = \mathrm{nbc} \times \mathrm{Tb} \tag{3}$$

Using the following formula (see Eq. (4)), this research calculated the synchronisation period length as a function of the number (*nbC*) of coordinators:

$$\mathrm{d}[\mathrm{T0};\mathrm{T1}] = \mathrm{nbC} \times (0.18 \times \mathrm{nbC} + 2.8). \tag{4}$$

The length of each star's activity period and the total number of stars in the network determine how long [T1; T2] will last. All stars in the network have the same amount of time for their active period, which consists of an intra-star period and a relay interval. The number of activity slots assigned for the intra-star period,

'intra', can be set to *nbS*, which is the duration of the intra-star period in milliseconds (depending on the overall load generated by all stars). Routing Period [T2; T3]: The amount of non-priority traffic determines how long [T2; T3] is measured in milliseconds. Two guard periods are included in [T2; T3], which can be customised by adjusting nbSR, the number of activity slots allotted for the routing time. Thus, we have:

$$d[T2;T3] = nbSR \times dslot + \times dgarde. \tag{5}$$

Sleep Period [T3; T0]: [T3; T0] is the amount of time that the network is idle, and all network entities are dormant during this time. Cycle Duration: As shown in Eq. (6), the length of a detection cycle, measured in milliseconds (*ms*), is obtained by adding the lengths of the four periods.

$$dcycle\, d[T0;T1]d[T1;T2] + d[T2;T3]d[T3;T0] \tag{6}$$

4. Experimental results

4.1 Omnidirectional antenna measurement and design

The ground plane of the Square Slotted Monopole Antenna (SSMA) has a sizable slot that is part of a planar antenna design. This kind of arrangement gives antennas a wide bandwidth and makes integration easier. We introduce two 12x12 cm^2 rectangle and triangular slot antennas. Impressive bandwidths of 130% (1.82 to 7.14 GHz) and 105% (2.31 to 8.18 GHz) are displayed by both antennas. Furthermore, a 102% bandwidth (2.91 to 9.09 GHz) octagonal slot antenna is provided [9]. An 85 x 85 mm^2 wide E-slot antenna with a notable bandwidth of 120% (2.8 to 11.4 GHz) is demonstrated. Similarly, an 80 x 80 mm^2 circular wide slot antenna may reach a remarkable 150% bandwidth (1.7 to 10.5 GHz).

Figure 1 shows the geometry of the proposed antenna and includes notations for each of the antenna's dimensional components. The antenna structure, which has two half-ellipses on the horizontal sides, resembles an improved rectangular radiating element. Unlike the traditional rectangular shape, this modification allows for different distances between the edges, which in turn covers a wider variety of wavelengths and increases the bandwidth of the antenna.

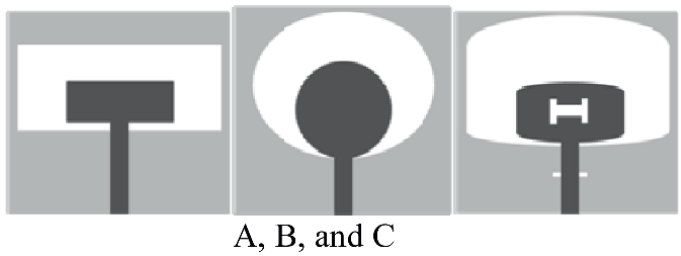

Figure 1.
Proposed antenna structures, (A) square slot antenna, (B) circular slot antenna, and (C), and proposed antenna.

This revised version strengthens the text's scholarly tone by using more formal language and placing greater emphasis on technical terms. It also offers a more thorough description of the altered shape of the antenna and how it affects bandwidth. Additional degrees of freedom in the design are achieved by adding a centre '*H*' slot.

The antenna's performance is improved and refined by adding two slots, the H-slot and the T-slot, as well as a parasitic element [10]. A sub-miniature connector that is connected to the ground plane and the feed line is used to transmit signals. The geometric parameters of the antenna determine its performance. As a result, the bottom limit of the antenna's bandwidth surpasses 3.1 GHz, which is the lower limit of the ultra-wideband (UWB) standard [3.2–10.5 GHz], due to its small size. The lateral width (wl = w - wg), where w is the antenna width and wg is the maximum width of the ground plane slot, is the primary factor determining this lower bandwidth limit as shown in **Figure 2**.

The distance '*d*', or the separation between the bottom of the large ground plane slot and the bottom of the radiating element, represents the combined effect of parameters Rb4 (minor radius of the lower half-ellipse of the large slot) and Lf (feed line length), as shown in **Figure 3**.

Table 1 provides the connection between these quantities.

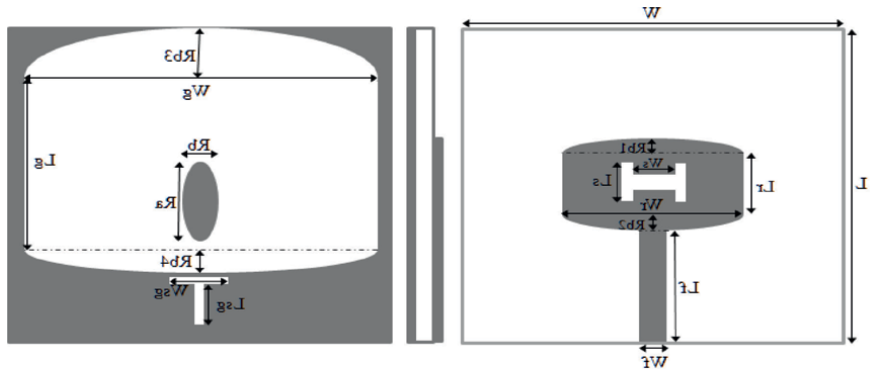

Figure 2.
Structure of the proposed ULB antenna, (A) front view, and (B) side view.

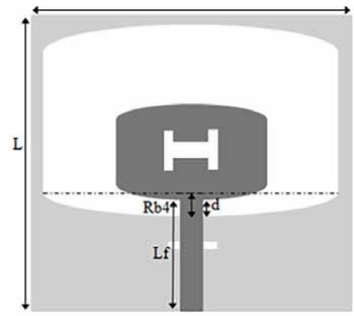

Figure 3.
Explanatory diagram of the SSMA.

Rb4 (mm)	0.55	1.05	1.05	1.05	1.55
Lf (mm)	7.62	7.37	7.62	7.87	7.62
d (mm)	0	0,25	0.5	0,75	1.00

Table 1.
Variations of parameters RB4-LF.

4.2 Directional antenna measurement and design

The SSMA serves as the model for the Circular Monopole Antenna with Slots (CMAS), which uses a circular shape for the substrate and ground plane. Because of this shape, the researchers are able to create a spherical sensor that will be more resistant to pressure from the depths of the subsurface. With a diameter of D = 29 mm, the antenna is a compact planar circular antenna. The radiating element is a rectangular patch with an H-shaped slot in the centre and two half-ellipsoids on the horizontal sides. A microstrip line with an embedded rectangular slot provides power. The ground plane is made up of a T-slot, a parasitic ellipse, and a huge ellipsoidal slot arranged in a circle. Along with the parasitic elliptical element, the addition of H, T, and rectangular slots increases design flexibility and makes it easier to further reduce the structure's size without sacrificing performance. Using the electromagnetic simulator, an optimisation process was carried out while taking the size-bandwidth trade-off into consideration. Given the inexpensive PTFE NH9338 substrate, which is also widely available in the South African market, the antenna is intended for subsurface UWB applications.

The ideal antenna structure for the finished sensor design was found through simulation analysis. The antenna with its optimised size is shown in **Figure 4**. The antenna has a three-dimensional radiating element and a compact circular frame with a diameter. 3.8 GHz is the lowest matching frequency.

The antenna, as shown in **Figure 5**, has a measured reflection coefficient of less than −10 dB throughout a wide frequency range of 3.8 GHz to 13 GHz, which translates to a bandwidth of about 9 GHz.

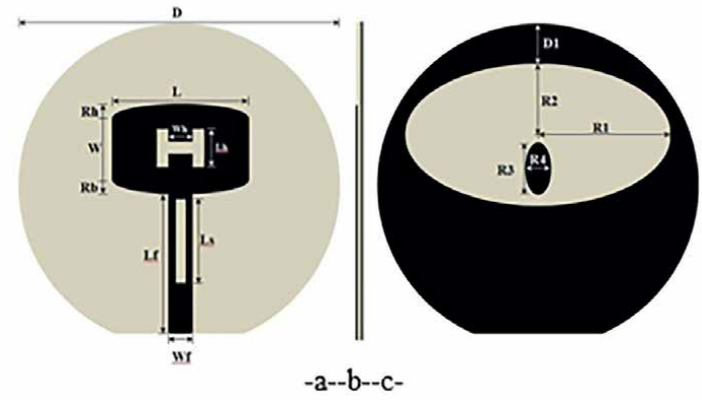

Figure 4.
Proposed ULB circular antenna structure, (a) front view, (b) side view, and (c) ground plan.

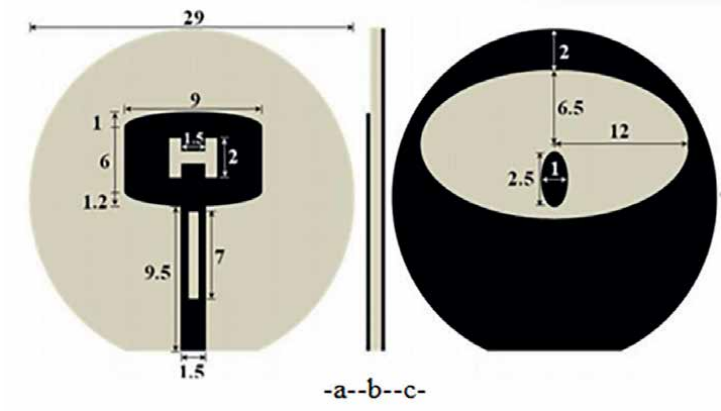

Figure 5.
Dimensions of the antenna (a) front view, (b) side view, and (c) ground plan.

Figure 6.
Proposed antennas (right).

4.3 Implementation

The omnidirectional (right) and directional antennas were integrated into the WSNs, as shown in **Figure 6**. The completed suggested antenna has an efficiency of 90%, omnidirectional radiation in the azimuth plane, a gain of roughly 3 dBi, and a bandwidth of nearly 8 GHz in the 3.1–11.5 GHz range. Every antenna is protected

from water by a plastic layer, which also keeps the distance between each antenna and the surrounding material at about 1 mm.

Subsequently, using two types of soil—fine sand and aggregates, or ballast—implementation measurements were carried out in a laboratory setting. These soil types are representative of the common soil environments found in underground spaces where copper cables are frequently put in South Africa. Particles smaller than 0.3 mm in diameter make up fine sand. Different soil types show convergence in their relative permittivity at frequencies over 3 GHz. There was variation in both the water content and the subsurface depth, ranging from 0 to 30%. For every scenario, the effectiveness of every antenna was assessed in terms of its bandwidth, input impedance, and reflection coefficient. The effect of subterranean depth on antenna performance was examined with dry aggregates (VWC = 0%) and sand that had a constant average water content of 5%. These water contents are representative of the usual natural conditions. The depth was varied in increments of 5, 10, 20, and 30 cm to do the measurements.

The communication chain's technical specs are as follows: a generator, a ULB transceiver that produces 3.2 GHz bandwidth pulses cantered at 4.7 GHz, is present on the show. The pulse repetition frequency is 9.6 MHz, while the effective isotropic radiated power is −12.8 dBm. This generator is managed by a computer. Additionally, a ULB antenna that has both a directional and an omnidirectional antenna is linked to an antenna that has a generator. The directional and omnidirectional antennas were compared at reception. Additionally, a high-pass filter is used to reduce interference with other systems, specifically 2.4 GHz Wi-Fi. The filter was included in the chain of reception. A high-pass filter with a 3 GHz cut-off frequency was chosen as the filter.

The operational bandwidth has less than 1 dB of insertion loss and spans from 3 to 20 GHz. A wideband low-noise amplifier (LNA) is added to the receive chain to minimise the high attenuation of the broadcast signals and make up for the low transmit power. With a noise figure of 1 dB and a gain of 23 ± 0.5 dB, this LNA operates in the 2–8 GHz frequency range. At the end of the receive chain, a wideband digital oscilloscope called the Infinium oscilloscope is used to visualise and store the received signals. The bandwidth of this digital oscilloscope is more than 10 GHz. During the trials, a sampling rate of 40 GSa/s was used.

4.4 Simulation environment

Antenna design and modelling were conducted using two three-dimensional electromagnetic simulation software packages: CST Studio Suite and Ansys HFSS (High-Frequency Structure Simulator) from computer modelling Technology. Because of its extensive capabilities, the study chose to use HFSS for antenna design and related network simulation. The electromagnetic simulator closely approximates the real geometry of the structure being studied by combining hexahedral and staircase meshing techniques with the finite element method (FEM). As a result, the best approach is used in accordance with the particular simulation requirements [22]. With CST, temporal simulation yields a variety of outcomes in the frequency and time domains, such as electromagnetic fields in the frequency and time domains, S-parameters, time signals, and 2D and 3D radiation patterns. Our selection of software is justified by these capabilities and performances, which are in good agreement with UWB calculations. These software choices were followed by the construction and use of the ULB omnidirectional antenna for subterranean intrusion detection analysis [23].

4.5 Signal transmission

The simulation's pulse generator uses a sinusoidal carrier frequency of 4.7 GHz to modulate a Gaussian pulse. The pulse has a spectral width of roughly 4.4 GHz at −10 dB and 2.5 GHz at −3 dB, with a temporal width of 0.4 ns. Next, this pulse spreads over a one-meter span in open space. The received pulse has a temporal duration of 0.53 ns and a spectral width of around 2 GHz at −3 dB and 3.9 GHz at −10 dB, as shown in **Figure 7**. The spectrum width reduction that was seen in comparison with the transmitted pulse can be ascribed to the combined impact of multiple components in the communication chain, such as the low-noise amplifier and the filter.

4.6 Effect of underground depth

In order to examine how subterranean depth affects transmitted pulses in a sand medium, the depth below the surface is changed in steps of 0, 5, 10, 20, and 30 cm while keeping the soil's water content constant at 5%. **Table 2** displays the findings for the regular antenna and the suggested WSNs antenna. At a depth of 0 cm below the surface, the maximum amplitudes of the reference pulses and the received pulses are normalised to each other.

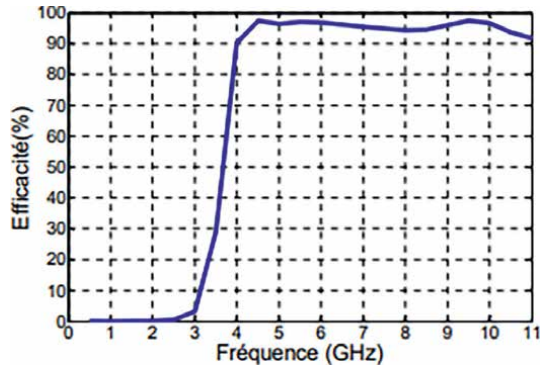

Figure 7.
Antenna efficiency.

	Omnidirectional branch office		Directional antennas	
VWC (%)	**Underground-air**	**Air-infused**	**Underground-air**	**Air-infused**
0	0.86	0.88	0.71	0.79
5	0.71	0.71	0.53	0.51
10	0.61	0.56	0.41	0.43
20	0.40	0.40	0.31	0.33

Table 2.
Normalised pulse amplitudes.

4.7 Effect of depth on signal amplitude

The measured outcomes show that as subterranean depth increases, the maximum amplitudes of the received pulses decrease. These attenuations are similar in magnitude for both antennas and communication directions. The fidelity factor F is computed to assess how underground depth affects the transmitted pulse form. Both omnidirectional and directional antennas' findings were obtained in both communication directions.

4.8 Simulation results: Omnidirectional antennas

When the lateral width (wl) is changed from 1.5 to 0 mm, the realised components' initial resonance peak shifts from 4.3 to 3.4 GHz, and its amplitude decreases from 100 to 80 Ω. Along with the rise from 50 to 75 Ω at 7 GHz, this modification also brings a decrease in the realised component over the 6 to 8 GHz region. The first resonance frequency for the imaginary component drops from 4.3 to 3.5 GHz, and the 4 to 7 GHz band degrades or increases from 10 to 40 Ω at 7.5 GHz. Within the 4 to 12 GHz band, the realised component at wl = 1 mm shows an imaginary portion around 0 Ω with values between 22 and − 22 Ω and a near 50 Ω impedance with values between 35 and 75 Ω. The lower cut-off frequency of the passband shifts from 4.3 to 3.5 GHz as wl is adjusted from 1.5 to 0 mm, according to the results of the reflection coefficient.

Figure 7 shows the simulated radiation efficiency of the SSMA. The efficiency of the antenna is more than 90% over its whole operating bandwidth. In an anechoic chamber, the transmission frequency was changed in order to test efficiency. The test and reference antennas were positioned 1.5 meters apart in a face-to-face arrangement. In the 4 to 8 GHz range and the 7.6 to 11 GHz band, the antenna's average measured gain is 1.7 dBi and 4.4 dBi, respectively. In the 4 to 8 GHz band, there is a difference of about 1.7 dB between the measured and simulated values. The measuring cable and the SMA connector's existence are the causes of this disparity. There is a good level of agreement in the 7.6 to 10 GHz range.

4.9 Antennas frequency and time analysis

This part uses simulation to calculate the transfer function of the antenna, which describes its frequency response. Two identical free-space transmit and receive SSMA antennas are taken into consideration in order to get this property. The two antennas are aligned along the line of sight and placed in the far field at a distance of 30 cm, perpendicular to the plane that contains their feed points. For every transmission session coefficient S21, the receiving WSNs antenna rotates by 44.9° around its symmetry axis, while the transmitting antenna stays fixed. With a 0° angle corresponding to the antennas facing each other, the rotation angle covers the full azimuth plane (H-plane). The transfer function of the WSNs system, which consists of two identical transmitting and receiving antennas as well as free-space propagation in the azimuth plane, is represented by the transmission coefficient S21. In the study of Sturdivant et al. [15], Saeidi et al. [16] and Huang et al. [18], this transfer method Hs(f) is explained.

$$HS(f)Ht(fr,r)Hc(f,d)Hr(fr,r) \tag{7}$$

With:

$$\mathrm{Hc(f,d)}\frac{\mathrm{c}}{\mathrm{4fd}}\mathrm{e}\frac{\mathrm{-f}}{\mathrm{c}}\mathrm{d} \tag{8}$$

The transfer function of the free-space channel is represented by Hc(*f,d*). Ht(*f,t,t*) and Hr(*f,r,r*) represent the transfer functions of the transmitting and receiving antennas under their respective orientations. The antenna in this work acts in a reciprocal manner, following the methodology used in [19]:

$$\mathrm{H(f,,,,)(f,,,,)H(f,,,,)}\frac{\sqrt{\mathrm{Hs(f)}}}{\mathrm{2aHc(f,d)}} \tag{9}$$

In every direction, the transfer function curves show a quasi-flat form over the 4 to 10 GHz region. This observation minimises distortion in the broadcast pulses by confirming that the transmitted signal, independent of direction, has equal attenuation levels throughout this frequency range. Furthermore, compared to the 0° or 180° locations, an attenuation of about 10 dB is seen at the 90° position. These findings and Eq. (10), when applied to different angular directions within the azimuth plane, yield the antenna transfer function. It is clear that the antenna's transfer function increases linearly with frequency in all directions and within the 4 to 10 GHz spectrum. The transfer function for the 0° direction can be roughly represented as:

$$\mathrm{H(f)50,8(f4)} \tag{10}$$

This performance characteristic, which expresses H(*f*) in gigahertz (GHz) and f in decibels (dB), reduces the attenuation effects caused by free-space propagation, which vary in *1/f 2*.

The spreading factor (SR) is calculated by comparing the energy distribution of the antenna pulses that are broadcast and received. The time span containing 90% of a pulse's energy is known as its time-based breadth [19]. The chapter presents the cumulative normalised energy function *Ev(t)* for an antenna signal *vx(t)*:

$$\mathrm{Ev(t)}\frac{\mathrm{vt}\left[\left(\ \right)\right]\mathrm{d}]}{\mathrm{vt}\left[\left(\ \right)\right]\mathrm{d}]} \tag{11}$$

As a result, Sturdivant et al. [15] provide the temporal width, W(v), of the pulse.

$$\mathrm{Ev(t)}\frac{\mathrm{v}\left[\left(\ \right)\right]\mathrm{2d}]}{\mathrm{v[(\ )\}2d]}} \tag{12}$$

Therefore, according to Ref. [15], the spreading factor is defined as the ratio of the temporal widths of the sent and received antenna pulses:

$$SR^{W}\frac{vr}{W(vt)} \tag{13}$$

When an antenna is weakly dispersive, SR gets closer to unity. The predicted spreading factor SR values for the suggested antenna design are shown in **Table 3**. These findings show that the antenna's spreading factor in the azimuthal plane is

θ	0.0°	91°	181°	271°
SR	1037	1058	1040	1060

Table 3.
Antenna stretch factor.

nearly 1, exhibiting little dispersive behaviour and little distortion of the time pulses that are transmitted.

4.10 Directional antenna simulation

The input impedance and antenna matching properties are significantly influenced by the parameters R1, R2, and D1. When R1 is changed from 11.6 to 12.6 mm, the realised component's initial resonant peak's amplitude decreases from 286 to 220 Ω, and its frequency shifts from 3.4 to 3.1 GHz. The first resonant frequency for the imaginary component is found to decrease from 4.25 to 3.75 GHz. Ze = (51 + j16) Ω ± 21 Ω represents the antenna's input impedance in the 4–13 GHz range. The lower bandwidth limit of the reflection coefficient shifts from 3.9 to 3.5 GHz, and the reflection coefficient degrades by 4 dB between 4 and 5 GHz.

The effect of D1 modifications is shown, and the input impedance's real and imaginary components as well as the reflection coefficient are illustrated. The formula for the input impedance is Ze = (56 + j16) Ω ± 31 Ω. High frequencies above 6 GHz show a substantial effect on the reflection coefficient. The lower cut-off frequency rises from 3.7 to 4.1 GHz when D1 is decreased from 3.49 to 2.49 mm. The lower limit of the bandwidth for the W and L parameters is influenced by the radiating element's size and the W parameter's input impedance change. As a result, raising W from 5 to 7 mm causes a shift in the lower limit of the passband from 4.2 to 3.66 GHz, the first resonance peak of the actual part of the impedance from 3.4 to 3.2 GHz, and the first resonance of the imaginary part of the impedance from 3.6 to 3.3 GHz. This alteration leads to a 0.56 GHz increase in bandwidth.

After the sensor was simulated, its radiation pattern was measured. The measured radiation pattern for various frequencies is plotted in **Figure 8** for both elevation (b) and azimuth (a). A satisfactory agreement is discovered between the simulated and measured findings. The measuring cable and SMA connector being present during characterisation is the reason for the little discrepancy between the simulation and measurement in the E-plane at −90°. Over the entire matching band, the radiation pattern exhibits a quasi-omnidirectional characteristic in the azimuthal plane.

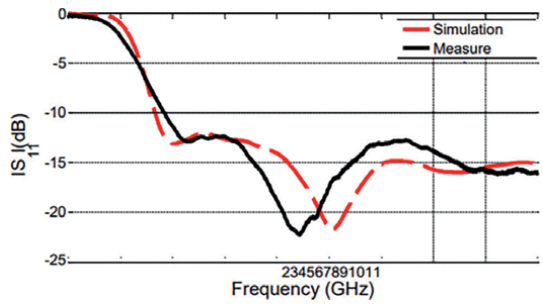

Figure 8.
Reflection coefficient of the antenna.

Through simulations, the radiation efficiency and gain of the antenna were assessed. Within its working bandwidth, the CMAS antenna has an efficiency of no more than 80%. In comparison, the study's CMAS antennas demonstrated efficiency gains of more than 95%. Furthermore, the antenna's simulated and calculated gain values show an average measured gain of 4 dBi in the 7.5–10 GHz band and 2.5 dBi in the 4–8 GHz band. In the 4–8 GHz band, there is a mismatch of about 2 dB between the simulated and measured gain levels; in the 7.5–10 GHz range, there is satisfactory agreement.

4.11 Frequency and time analysis

Two identical CMAS antennas, one broadcasting and the other receiving, are simulated and placed on the same plane with a 40 cm interval between them in order to do the analysis. While the receiving antenna is rotated around its axis of symmetry in the azimuthal plane with an angular increment of 30°, ranging from 0° to 360°, the sending antenna stays stationary. At each rotation step, the transmission coefficient, or S21, which represents the transfer function of the WSNs system made up of the two antennas and the propagation channel, is recorded. When the antennas are facing one another, they are in the 0° position.

The antenna's inadequate impedance matching in this frequency range is the cause of the greater attenuations seen between 3 GHz and 4 GHz. Eq. (10) is used to calculate the antenna transfer function magnitude for each angular direction in the 3 to 11 GHz range based on the preceding results. An enhancement with frequency is shown in all angular directions when the antenna transfer function is examined along the frequency axis, with the 0° and 180° directions exhibiting the best performance. The transfer function's amplitude at 0° can be written as follows:

$$\text{Amplitude}(\text{dB}) = -10 + 0.849(f - 3), f \text{ in GHz} \tag{14}$$

This good performance lessens the impact of the transmitted signal's dispersion, attenuation, and distortion on the channel. Using the same methodology as the frequency-domain study, a modulated Gaussian pulse was used for the time-domain analysis. For every rotation point, the fidelity (F) and spread (SR) parameters were assessed.

5. Conclusion

An innovative omnidirectional antenna design for improved intrusion detection in subterranean WSNs is presented in this chapter. Through the use of the SSMA idea, the suggested design provides a small and effective fix. The antenna's higher performance is validated through actual implementation using micro-etching and simulations using Ansys HFSS. For subsurface WSN applications, the suggested omnidirectional antenna greatly enhances QoS. Its remarkable dispersion, spreading, and high-frequency depth allow precise intrusion detection even in difficult subterranean conditions. The antenna's ability to penetrate deep terrain and provide dependable sensing capabilities is proved by its average fidelity of 0.9 and spread factor of 1.5 at a depth of 30 cm. To sum up, this chapter presents a viable way to improve the functionality and efficiency of subterranean WSNs. The field of subsurface sensing

and monitoring applications might undergo a revolution thanks to the suggested omnidirectional antenna design.

Acknowledgements

The authors express their sincere gratitude to Tshwane University of Technology for its financial assistance in completing this study. The authors declare that there is no conflict of interest with respect to this chapter's publication.

Author details

Khomotso C. Tsoane, Topside E. Mathonsi* and Deon P. Du Plessis
Department of ICT, Tshwane University of Technology, Pretoria, South Africa

*Address all correspondence to: mathonsite@tut.ac.za

References

[1] Lin K, López OLA, Alves H, Hao T. On CSI-free multiantenna schemes for massive wireless-powered underground sensor networks. IEEE Internet of Things Journal. 2023;**10**(19):17557-17570

[2] Lin K, Ullah MA, Alves H, Mikhaylov K, Hao T. Subterranean mMTC in remote areas: Underground-to-satellite connectivity approach. IEEE Communications Magazine. 2023;**61**(5):136-142

[3] Lin K et al. Throughput optimization in backscatter-assisted wireless powered underground sensor networks for smart agriculture. Internet of Things. 2022;**20**(1):1-13

[4] López OLA, Alves H, Montejo-Sánchez S, Souza RD, Latva-aho M. CSI-free rotary antenna beamforming for massive RF wireless energy transfer. IEEE Internet Things. 2022;**9**(10):7375-7387

[5] Ullah MA, Keshavarz R, Abolhasan M, Lipman J, Esselle KP, Shariati N. A review on antenna technologies for ambient RF energy harvesting and wireless power transfer: Designs, challenges and applications. IEEE Access. 2022;**10**(1):17231-17267

[6] Wang J, Gao Y, Liu W, Sangaiah AK, Kim HJ. Energy efficient routing algorithm with mobile sink support for wireless sensor networks. Sensors. 2019;**19**(7):1494

[7] Liebenberg AS. An Evaluation of the Impact of the Non-ferrous Metals Crime Combating Committee on Copper Cable Theft (Doctoral Dissertation). 2021. Available from: http://hdl.handle.net/10500/27467

[8] Khedo KK, Bissessur Y, Goolaub DS. An inland wireless sensor network system for monitoring seismic activity. Future Generation Computer Systems. 2020;**105**:520-532

[9] Koochakkosari A. Energy-Efficient Integrated Circuits and Systems to Bridge the Gap between Power and Performance in Wireless Sensor Networks (Doctoral Dissertation). 2019

[10] Barras D. A Low-Power Impulse Radio Ultra-Wideband CMOS Radio-Frequency Transceiver (Doctoral Dissertation, ETH Zurich). 2010

[11] Martinez K, Ong R, Hart J. Glacsweb: A sensor network for hostile environments. In: 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. Santa Clara, CA, USA: IEEE; 2004. pp. 81-87

[12] Denis B, Keignart JK, Daniele N. UWB measurements and propagation models for snowy environments. In: 2005 IEEE International Conference on Ultra-Wideband. Zurich, Switzerland: IEEE; 2005. pp. 1-6

[13] Croucamp PL, Rimer S, Kruger C. Locomotive monitoring system using wireless sensor networks. In: 2014 12th IEEE International Conference on Industrial Informatics (INDIN). Porto Alegre, Brazil: IEEE; 2014. pp. 639-644

[14] Das A, Mohanty MN, Mishra RK. Optimized design of H-slot antenna for bandwidth improvement. In: 2015 IEEE Power, Communication and Information Technology Conference (PCITC). Bhubaneswar, India: IEEE; 2015. pp. 563-567

[15] Sturdivant R, Quan C, Chang E. Antenna element technology options.

In: Systems Engineering of Phased Arrays. Boston: Artech House; 2019. pp. 125-140

[16] Saeidi T, Ismail I, Wen WP, Alhawari AR, Mohammadi A. Ultra-wideband antennas for wireless communication applications. International Journal of Antennas and Propagation. 2019;**1**(1):1-25

[17] Curiac DI. Wireless sensor network security enhancement using directional antennas: State of the art and research challenges. Sensors. 2016;**16**(4):488

[18] Huang H, Xiao S, Meng X, Xiong Y. A remote home security system based on wireless sensor network and GSM technology. In: 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing. Wuhan, Hubei, China: IEEE; 2010. pp. 535-538

[19] Jeong S, Ha D, Tentzeris MM. A cavity-backed slot antenna with high upper hemisphere efficiency for sewer sensor network. In: 2013 IEEE Antennas and Propagation Society International Symposium (APSURSI). Lake Buena Vista, Florida, USA: IEEE; 2013. pp. 49-50

[20] Tasouji N, Nourinia J, Ghobadi C, Mirzamohammadi F. A novel UWB monopole antenna with controllable band-notch characteristics. The Applied Computational Electromagnetics Society Journal. 2017;**32**(1):68-73

[21] Lim EK, Norizan MN, Mohamad IS, Yasin MNM, Murad SAZ, Baharum NA, et al. Design of anti-theft/cable cut real time alert system for copper cable using microcontroller and GSM technology. In: AIP Conference Proceedings. New York, USA: AIP Publishing LLC; 2017. p. 020287

[22] Alejandro-Granados C, Enrique RH, Santamaria F. Evaluation of the electromagnetic shielding of concrete from high-frequency simulations and the application of the Jonscher model. Ingeniería. 2020;**25**(2):0121-750X

[23] Tsoane KC, Mathonsi TE, Du Plessis D. An enhanced wireless sensor monitoring system to combat underground copper cable theft. In: 2022 International Conference on Electrical, Computer and Energy Technologies. Prague, Czech Republic: IEEE; 2022. pp. 1-6

Chapter 7

Quality of Service (QoS) for the Future 6G Network Data Transport

Mohamadou Ndiaye

Abstract

According to the International Telecommunications Standard, Quality of Service (QOS) refers to the ability of a system to meet the requirements for providing a type of telecommunications service, particularly with regard to access, integrity, and continuity. QoS requirements will change radically with the transition from 5G to 6G. 6G will support different applications, such as ultra-low latency communications (URLLC), massive machine-to-machine communications (mMTC), and enhanced mobile broadband (eMBB). In this article, we will look at the QoS challenges and solutions for 6G networks, which are being realized through the use of advanced technologies. Additionally, we will discuss the systems needed to meet the QoS requirements for sustainable and, above all, flexible data transport.

Keywords: QoS, 5G, 6G mMTC, URLLC, AI, IoT, CN, RAN, NFV, IA

1. Introduction

The evolution towards 6G networks will bring new challenges and opportunities in Quality of Service (QoS) management. With anticipated capabilities such as ultra-low latency, high throughput, and massive connectivity (IoTs), QoS mechanisms need to adapt to meet the rigorous requirements of potential future applications, while end-to-end QoS includes services that demand very high performance requirements. However, the future system must be deployed in a cost-effective manner and must not require over-deployment of network resources.

This article aims to provide a comprehensive overview of how QoS for 6G networks will be addressed and the technologies and strategies that will be essential to achieve these objectives and ensure that data transport over the network returns a high level of availability and resilience.

2. QoS framework for 6G networks

2.1 Network slicing

Network slicing will be an integral part of 6G QoS management, allowing the creation of multiple virtual networks on a single existing physical infrastructure. Each part of the unbundled network can be optimized for different types of traffic

and services, enabling tailored or on-demand QoS for applications with diverse requirements.

This attempt to decompose the physical network into several virtual parts seeks to specify one part to a specific job or a specific type of service. This segmentation ensures that key services are given the necessary resources and priority, while less key services do not consume too much bandwidth or disrupt more important traffic [1].

Key aspects of network slicing in 6G isolation:

- Isolation: Each network entity operates independently, ensuring that problems in one part of the network do not affect the others. It is essential to maintain the reliability and performance of critical applications.

- Scalability: Network slices can adapt in time to match demand allowing resources to be used effectively and accommodating changing traffic patterns.

- Comprehensive Management: Slicing covers everything from the network to the periphery guaranteeing quality of service from the server room all the way to the end user.

Resource Management: Bandwidth, processing power, and storage can be assigned to each slice based on its requirements enhancing network efficiency [2].

2.2 Optimizes latency

6G networks will combine edge computing, network slicing, enhanced radio access technologies, URLLC, AI and machine learning, rapid link throughput, and distributed architecture in an effort to fulfill the real-time performance demands of new applications [3]. By combining these approaches, latency is optimized and 6G networks are made sufficiently stable and dependable to handle the upcoming wave of cutting-edge, time-sensitive applications.

6G networks are expected to achieve latency times on the order of a few milliseconds to meet the real-time performance requirements of applications such as telemedicine and interaction. And these future networks should ensure QoS and reliability [4].

2.3 Reliability guarantees

Some services will require extreme reliability, such as industrial automation and autonomous vehicles. They will need appropriate robust service quality mechanisms to ensure total reliability. Approaches will include error rehearsal and correction methods to ensure consistent performance [5].

2.4 AI and machine learning (cognitive radio)

In 6G, AI and machine learning will play a key role in dynamic QoS management. These technologies will enable QoS parameters to be analyzed predictively, optimized in real time, and adjusted automatically according to network conditions and traffic patterns [5].

2.5 High bandwidth applications

Emerging applications such as augmented reality (AR) and virtual reality (VR) will require high bandwidth and light interference. QoS mechanisms will need to formally support these services and manage bandwidth allocation (Cognitive Radio) to ensure optimal performance.

2.6 Quality of service and security

Maintaining quality of service while ensuring network security will be a key challenge. Security measures must be integrated into QoS protocols to avoid degradation of QoS due to malicious activity.

Achieving quality of service for 6G is not a task that can be handled by any single player in the telecom industry alone. It will require effective coordination between telecoms operators, device manufacturers, and application designers [6]. An interdisciplinary partnership such as this one is necessary to articulate the standards of QoS that would work effectively for all the involved parties.

3. End-to-end QoS

End-to-end QoS includes services that demand very high levels of performance and that need to be delivered across specific deployment zones: geographically wide, local, delimited, and, in the case of nomadic services, dynamic. Of course, all this must be achieved in a cost-effective way that does not require over-deployment of network resources [7]. Radio Access Network (RAN) and Core Network (CN) architectures will be based on a high level of modularity and virtualization of network functions. Transport will therefore have to support this process by automatically ensuring the appropriate connectivity between each network function.

In data transport, there is a need for high availability. This implies that the network should be able to function even in cases of failure, regardless of whether it is hardware or software. This involves redundancy systems, failover mechanisms, and resilience design principles—ensuring continuous accessibility of services. On the other hand, resilience deals with quick recovery from unexpected attacks: designing systems to gracefully handle failures, adapt to changing conditions while still keeping up performance standards under different scenarios that might occur. Self-healing networks can reconfigure themselves to maintain QOS and reduce downtime as much as possible; unexpected attacks need to be considered. This has to be considered about programmability. On the basis of real-time emerging requirements, it allows network operators to change network resources and behavior dynamically. It is an integration of network function virtualization (NFV) with software-defined networking (SDN), which will make this possible [8]. The other crucial aspect is flexibility. The objective is to adapt swiftly to new technologies, applications, and user needs with minimal reconfiguration. Alternatively, it will provide fast changes and improvements in response to changing demands, ensuring that the network has a wide range of dynamic use cases it can serve effectively.

In order to discourage excessive use that may result in network congestion, loss of performance, and high operational cost, there is need for advanced resource management. These are:

- Traffic engineering: Optimizing the flow of data within a network to distribute loads evenly and avoid clogging.

- Dynamic resource allocation: Using real-time monitoring plus analytics for allocating resources where they are needed most and removing them when they are not.

- Capacity planning: Forecasting future demand to anticipate peak loads that will not cause over-provisioning.

- Minimize cost: Reducing the cost per bit should be achieved by optimizing network operations/expenditure while ensuring good traffic efficiency. Strategies may include [9]:

- Efficient use of bandwidth: Guarantee proper utilization of the bandwidth on the network and minimize wastage through technologies such as data compression among others

- Cost-Effective Infrastructure: Use affordable hardware and software solutions that deliver high performance at lower costs.

- Economies of scale: Utilize scalable solutions that grow with an increase in demand resulting in reduced costs per unit as networks expand.

In sum, the network transport needs to guarantee a high level of availability and resilience and be programmable, that is, flexible enough. Furthermore, attention must be given to prevent the overuse of network resources and to decrease the cost per bit. In order to boost this evolution and achieve this goal, we have identified and shown two approaches:

- End-to-end transport management awareness, in order to manage dynamically and optimally the interactions between radio, transport, and the cloud in a given area [10].

- A totally flexible and programmable transport plane, enabled by new technologies, offering a framework that is simpler for end-to-end management while guaranteeing exceptional service performance (e.g., capacity) (**Figure 1**).

3.1 Constitution of end-to-end QoS in 6G

It is important to note that end-to-end QoS depends on the combination of QoS in the radio layer and QoS in the transport layer. Managing the low latency service for the digital twinning operation requires a specific QoS transmission behavior in the radio network, recognized through a QoS identifier (called 5QI for 5G). The network operator configures the association of the QoS identifier with the QoS of the service according to a specific policy. The orchestrator figures out the QoS of the radio layer and the QoS of the transport layer considering the end-to-end QoS. It then sends the requests to the radio and transport networks. This approach treats end-to-end QoS as a "separate" infrastructure consisting of radio and transport, in an automatic and dynamic way [11–15].

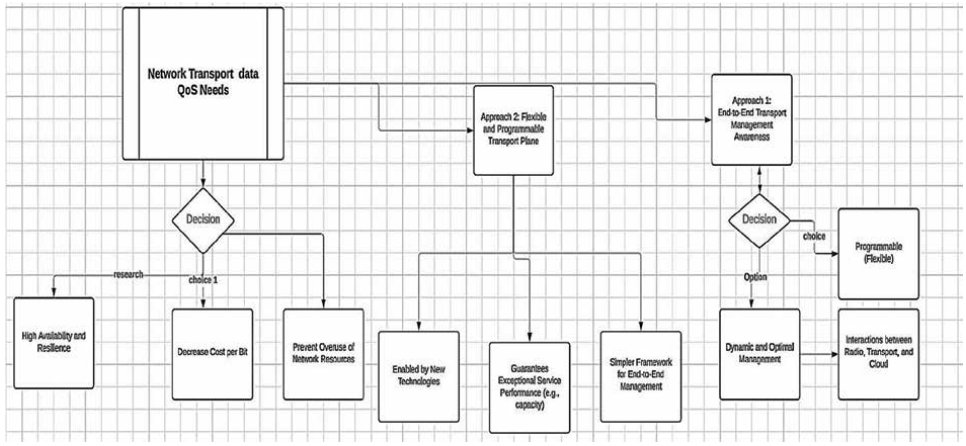

Figure 1.
Summary of proceedings.

3.2 The impact of AI on QoS

It aims to ensure the necessary end-to-end service quality while making the best use of resources. To do this, it brings in a system powered by AI to create the measured data. As this happens, AI will spot traffic patterns and habits in the operator's network. This gives insights about real radio traffic conditions that would be hard to see or grasp otherwise. This data allows for the growth and running of the transport network in the best way possible. It also helps to cut down on the extra capacity needed just in case.

AI tries to looks at data collected from start to finish of the transport network to examine traffic patterns and actions in the area it covers. Here is how we could put math formulas into your AI system [15, 16].

4. Integration of laws and formulas into IA

4.1 Trend and behavior analysis

The moving average (MA) is a trend analysis tool that smoothest out variations in data to identify underlying trends. Here is the formula for the n-period moving average:

$$MM = \frac{1}{n}\sum_{i=0}^{n-1} x_{t-i} \tag{1}$$

where

x_{t-i} = data value.

4.2 Traffic forecasting

For traffic forecasting, it is best to use the ARIMA mathematical model. The ARIMA (AutoRegressive Integrated Moving Average) model is commonly used to forecast time series. It combines three components: autoregression (AR), differentiation (I) to make the series stationary, and the moving average (MA).

$$\varnothing(L)(1-L)^{d}\, yt = \theta(L)_{\epsilon t} \tag{2}$$

Ø(L) etØ(L) are polynomials in delay.
d is the number of differentiations required to make the series stationary.
$-_{\epsilon t}$: theerrorrate.

4.3 Optimizing the use of resources

Optimization problems must be properly formulated in order to allocate resources in the best possible way. For example, minimize resource costs under capacity and QoS constraints:

$$\min_{x} c^{T} x \text{ under the constraints } Ax \leq b \tag{3}$$

where
c is the cost vector.
x is the vector of resources to be allocated.
A is the constraints matrix and.
b is the vector of limits.

4.4 Reducing oversizing

Any system can have oversizing. The queue model that uses queue theory to scale resources efficiently is well suited to our scenario.

$$L_{q} = \frac{\lambda^{2}}{\mu(\mu-\lambda)} \tag{4}$$

Où
λ is the arrival rate and.
μ is the service rate.
By combining these mathematical approaches with machine learning algorithms, we can improve the accuracy of your AI engine for transport network analysis and optimization, delivering better QoS while minimizing over-provisioning.

5. Conclusion

In this article, we look at the QoS challenges and solutions for 6G networks, which are being achieved through the use of advanced technologies for the first time. The transition to 6G requires the development and implementation of new systems capable of meeting these evolving QoS requirements. These systems must ensure sustainable and, above all, flexible data transport, adapting to the varied and demanding needs of future applications.

In summary, the successful deployment of 6G networks will depend on the ability to meet the significant QoS challenges with innovative technological solutions. Ensuring high standards of access, integrity, and continuity will be essential to support the diverse and complex applications envisaged for 6G. As we enter the 6G era, the focus will be on creating a robust and adaptable infrastructure capable of delivering unrivaled quality of service across all use cases.

DOI: http://dx.doi.org/10.5772/intechopen.1006543

Author details

Mohamadou Ndiaye
Mathematics and Computer Science Department, University of Dakar, Senegal

*Address all correspondence to: mohamadou4.ndiaye@ucad.edu.sn

References

[1] Zhang H, Liu N, Chu X, Long K, Aghvami AH, Leung VCM. Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine. 2017;**55**(8):138-145. DOI: 10.1109/MCOM.2017.1600940

[2] Osseiran A, Boccardi F, Braun V, Kusume K, Marsch P, Maternia M, et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. IEEE Communications Magazine. 2014;**52**(5):26-35. DOI: 10.1109/MCOM.2014.6815890

[3] Filippou MC, De Donno D, Priale C, Palacios J, Giustiniano D, Widmer J. Throughput vs. latency: QoS-centric resource allocation for multi-user millimeter wave systems. In: IEEE International Conference on Communications (ICC). Canada: IEEE; 2017. pp. 1-6. DOI: 10.1109/ICC.2017.7996502

[4] Ford R, Zhang M, Mezzavilla M, Dutta S, Rangan S, Zorzi M. Achieving ultra-low latency in 5G millimeter wave cellular networks. IEEE Communications Magazine. 2017;**55**(3):196-203. DOI: 10.1109/MCOM.2017.1600547

[5] Abbas N, Nasser Y, El Ahmad K. Recent advances on artificial intelligence and learning techniques in cognitive radio networks. EURASIP Journal on Wireless Communications and Networking. 2015;**2015**:174. DOI: 10.1186/s13638-021-01991-2

[6] Smith J, Doe A. End-to-end quality of service (QoS) for wireless communication systems. IEEE Communications Surveys & Tutorials. 2009;**11**(2):56-78. DOI: 10.1109/COMST.2009.091245

[7] Haug T. Quality of service for IP networks: A survey. In: Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS). India: IEEE; 2004. DOI: 10.1109/ICDCS.2004.1288335

[8] Kumar DS, Carvallo PG, Patel LR. QoS-aware network management: Challenges and solutions. IEEE Access. 2024;**12**:24567-24584. DOI: 10.1109/ACCESS.2024.3567891

[9] Williams JC, Thompson MR. Enhanced QoS strategies for 5G networks: A review. IEEE Transactions on Network and Service Management. 2023;**21**(3):658-671. DOI: 10.1109/TNSM.2023.3456789

[10] Singh LM, Brown FJ. QoS provisioning for cloud-based applications: Recent advances. IEEE Communications Magazine. 2023;**61**(5):45-51. DOI: 10.1109/COM.2023.1234567

[11] Green MH, White EA. AI-driven QoS optimization for modern networks. IEEE Network. 2024;**38**(4):30-37. DOI: 10.1109/NET.2024.3456789

[12] Lee RK, Scott AJ. QoS management in IoT networks: Challenges and emerging solutions. IEEE Internet of Things Journal. 2024;**11**(1):101-113. DOI: 10.1109/JIOT.2024.3456789

[13] Liu X, Zhou S, Chen Z. End-to-end quality of service management in 6G networks: Challenges and opportunities. IEEE Wireless Communications. 2024;**31**(3):28-34. DOI: 10.1109/MWC.2024.3578901

[14] Kim KJ, Lee HS, Park JW. Integrated QoS and QoE framework for 6G networks. IEEE Transactions on Network and Service Management. 2023;**21**(2):543-556. DOI: 10.1109/TNSM.2023.3398765

[15] Zhang Y, Wu H, Liu L. Artificial intelligence-driven QoS management in next-generation networks. IEEE Communications Magazine. 2024;**62**(1):74-81. DOI: 10.1109/COM.2024.3216547

[16] Carter T, Reynolds MJ, Zhang KM. Optimizing QoS with AI techniques in 5G and beyond networks. IEEE Transactions on Network and Service Management. 2023;**21**(4):899-912. DOI: 10.1109/TNSM.2023.3456789

Chapter 8

Quality of Service (QoS): In Sliced Network

Mika Skarp and Jose Costa-Requena

Abstract

Quality of Service management can be used to differentiate service levels between data flows. QoS settings impact Radio Access Network (RAN), Transport, and Core elements. QoS settings can be used to provide guaranteed connectivity Service Level Agreements (SLAs) for an individual service. The simple SLA model in mobile networks is based on 3GPP specifications, so it can be used with any vendor's RAN, Transport, or Core. RAN and transport networks have a limited amount of capacity available, so managing these resources is crucial. We propose different QoS profiles to be used and tied to different slices that have known capacity limits to make the management of network resources easier. A two-layer approach provides new commercial models to monetize QoS capabilities because there is a separation between slices; capacity management in one slice does not interfere with capacity management in the other network slices in the same network.

Keywords: network slicing, quality of service model, service level agreement, subscriber profile, QoS management

1. Introduction

The delivery of predefined capacity from any IP network begins with a robust set of baseline QoS features. It is essential to differentiate data packets based on the QoS class to which they belong. These QoS classes can be defined automatically using Deep Packet Inspection (DPI) technology or can be triggered externally *via* Operations Support Systems (OSS). This principle applies uniformly to both fixed and mobile networks.

Once QoS features are implemented in the network, it is crucial to ensure that the provisioned traffic does not exceed the available capacity. In fixed networks, this process is relatively straightforward due to the limited and known number of ports connected to a single cable with a predefined capacity. This allows for easy calculation of spare capacity to determine if any given request at any given time can be fulfilled. Given that connection lead times are typically several days, these calculations can even be performed manually.

In contrast, understanding and managing capacity in mobile networks is considerably more complex. Mobile handsets are constantly in motion, preventing static reservations. Radiofrequency (RF) environment is not stable either but interference and RF frequency reflections change all the time. While cell load can be measured,

real-time measurements at every base station offer no absolute guarantees. Traffic can be controlled using DPI, but ensuring a reliable SLA remains challenging. This issue has sparked extensive discussion and has led to the development of numerous innovative approaches.

We propose that the optimal solution to this problem lies in the implementation of dynamic profiles within mobile networks, enabling real-time communication and fulfillment of users' capacity needs. Cumucore [1] has developed a core solution centered on this capability. The solution includes an Application Programming Interface (API) that allows for dynamic profile changes, facilitating accurate network capacity optimization through controlled guaranteed bit rates. Given that users' needs can vary from moment to moment, each user carries multiple profiles that can be switched on-the-fly to meet their requirements. As profiles enter, leave, and switch within a cell, the total capacity requirements at any time are known, allowing for the possibility of serving additional users with special profiles. When a cell approaches its maximum capacity, request prioritization becomes necessary. In such scenarios, new network entrants are restricted to profiles allowing only best-effort service. This approach provides operators with the flexibility to serve various use cases efficiently.

However, guaranteed bit rates alone are insufficient. Radio coverage within a cell is not uniform, and to ensure guaranteed QoS, the suitability of the mobile device's RF level for the expected QoS must be verified. Therefore, before a mobile device is allowed to make a profile change request, its RF level must be checked and validated.

2. Quality of service explained

Modern network technology has ushered in the era of Software Defined Networking (SDN), a paradigm that allows for the dynamic modification of network behavior through software. SDN enables unprecedented flexibility and control, facilitating the adaptation of network operations to meet varying demands in real-time.

In mobile networks, there are three primary physical functional blocks: the Core, Transport, and Radio networks. Logically, these networks are segmented into three distinct planes: the user plane, the control plane, and the management plane.

The Radio network capacity is regulated by a scheduler responsible for allocating air interface resources to users. These resources are allocated as resource blocks, which can carry varying amounts of traffic depending on the radio conditions between the base station and the user equipment. The scheduler's allocation of resource blocks to users can be managed using QoS parameters. To simplify this management, Network Slicing technology can be employed. Network slicing segregates users into different network slices, which can be envisioned as virtual networks with distinct capacities and QoS parameters.

Traffic from base stations is delivered to the core network *via* the Internet Protocol (IP) transport network. Typically, an IP transport network relies on fiber connections, though copper cables, satellites, and microwave radios are also utilized. In nonpublic networks, the existing transport network imposes its own limitations. Within the transport network, Virtual Local Area Networks (VLANs) at Layer 2 or Differentiated Services Code Point (DSCP) at Layer 3 can be employed to set traffic priorities based on VLAN tags or DSCP markings. Understanding the capabilities and management of the routers used is crucial, as the connection between the base station and the Core network can be configured using VLANs or DSCPs, each offering different priorities and capacities.

The Core network performs numerous control functions and routes traffic from base stations to other base stations, internal networks, and the Internet. Ensuring sufficient signaling and payload capacity for control functionality and understanding the policies governing routing functions is essential for optimal network performance.

3. Quality of experience explained

Quality of Experience (QoE) is a critical concept that integrates the available network capacity with the capacity expected by applications to ensure they function as intended. QoE is essentially a user-centric measure that evaluates how well the network performance meets user expectations, thereby reflecting the perceived quality of the service from the user's perspective [2, 3].

Quality of Experience (QoE) is influenced by several factors, including network speed, latency, reliability, and the specific requirements of the application in use. For instance, a video streaming application might require high bandwidth and low latency to deliver a high-definition video without buffering. If the network can meet these requirements, the user's QoE is high. Conversely, if the network fails to provide sufficient capacity, the user's experience deteriorates.

Improving QoE involves aligning the network's performance with the application's requirements and user expectations. One approach to enhancing QoE is by managing user expectations and adjusting application demands. For example, reducing the video resolution or increasing buffering can help maintain a smooth viewing experience even when the network conditions are suboptimal. By lowering the communication requirements, the application can continue to function satisfactorily without overburdening the network.

Moreover, QoE can be optimized through various technical strategies. Dynamic bandwidth allocation, prioritization of traffic based on application needs, and real-time adjustments to network parameters are some of the techniques employed to enhance user experience. These strategies ensure that critical applications receive the necessary resources to perform well, while less critical applications are adjusted to fit the available network capacity.

In summary, QoE is a holistic measure that combines network performance with application demands and user expectations. By optimizing both network capabilities and application behaviors, service providers can deliver a superior user experience, even under varying network conditions [4].

4. Service level agreement explained

Service Level Agreements (SLAs) are formal contracts between service providers and their customers that define the expected performance and reliability of the service. These agreements specify various parameters and target values that the service must meet within a specified time window. Key SLA parameters include bidirectional bit rates, delay, jitter, bit error rate (BER), packet drops, and service availability. Each of these parameters plays a crucial role in ensuring the quality and reliability of the service [5, 6].

Bidirectional bit rates: This parameter specifies the guaranteed minimum bit rate for both upstream and downstream traffic. It ensures that data can be transmitted and received at a consistent rate, which is essential for applications requiring stable and predictable bandwidth.

Delay: Also known as latency, delay measures the time it takes for data to travel from the source to the destination. Low latency is critical for real-time applications such as voice over IP (VoIP) and online gaming, where even small delays can significantly impact user experience.

Jitter: Jitter refers to the variation in packet arrival times. High jitter can lead to problems such as choppy audio or video during streaming, making it essential to maintain low and consistent jitter for smooth application performance.

Bit error rate (BER): BER measures the number of corrupted bits in a transmission. A low BER is important for ensuring data integrity, particularly in applications that require high accuracy and reliability, such as financial transactions and data storage.

Packet drops: This parameter indicates the number of packets lost during transmission. Minimizing packet drops is crucial for maintaining the quality of service, especially for applications like video conferencing and live streaming, where packet loss can result in noticeable disruptions.

Service availability: Service availability measures the percentage of time the service is operational and accessible. High availability is a key requirement for many critical applications and services, ensuring that users can rely on the service being up and running when needed.

Service Level Agreements (SLAs) provide target values for each of these parameters within a specified time frame. These targets are continuously monitored, and any violations are recorded. If the number or severity of these violations exceeds a predetermined threshold, penalties are imposed on the service provider. These penalties can be financial or take the form of service credits, incentivizing providers to maintain the agreed-upon service levels.

By clearly defining performance expectations and accountability, SLAs help build trust between service providers and their customers. They ensure that providers are committed to delivering high-quality service and that customers receive the level of performance they are paying for.

5. Different ways of delivering SLA in mobile network

Delivering Service Level Agreements (SLAs) in mobile networks requires innovative approaches to managing wireless communication, as the traditional methods of frequency allocation face limitations in flexibility and efficiency [1, 7, 8].

5.1 Traditional frequency management

5.1.1 Own frequency per user

Traditionally, wireless communication has been managed by assigning individual frequencies to different users. This method is prevalent in authority networks and point-to-point microwave links. Radio and television broadcast services also operate on dedicated frequencies. This fixed allocation ensures minimal interference but lacks flexibility, making changes difficult and potentially leading to the underutilization of spectrum resources. Typically, these kinds of regulations are enforced country-wide by the regulator. Border areas are difficult to manage and require cooperation between countries and even globally.

5.1.2 Own frequency per application

Another approach is to allocate specific frequencies to particular applications. For example, wireless microphones and WiFi access points operate on designated frequency bands, within which channels are selected based on interference levels (manually or automatically). While this method reduces interference and optimizes performance for specific applications, it also suffers from rigidity and potential underutilization.

Both of these traditional methods are fixed and inflexible, often resulting in reserved frequencies that go unused and difficulties in monitoring and enforcing compliance, relying heavily on a trust-based system.

5.2 Shared networks for users and applications

5.2.1 Radio access network (RAN) capacity management

Modern mobile networks utilize a shared network approach, where Radio Access Network (RAN) capacity is managed using resource blocks. These blocks are distributed among users through various mechanisms such as round-robin or weighted scheduling. However, the challenge with SLA delivery in this context is that each resource block's data-carrying capacity varies depending on the modulation used. Lower modulations support longer distances and higher interference tolerance, but they also reduce the data rate [9, 10].

5.2.2 Meeting SLA conditions

To deliver SLAs effectively in a mobile network, it is essential to set stringent conditions for radio signal strength and signal-to-noise ratio (SNR). SLAs can only be guaranteed when these conditions are met. Given the variability and rapid changes in radio conditions, the number of users with guaranteed SLAs must be limited to predefined levels to ensure consistent performance.

5.2.3 Capacity calculation

When using weighted scheduling, the capacity for each user equipment (UE) with an SLA can be calculated using the formula:

\text{UE}_{\text{cap}} = \left(\frac{N_{\text{cap}} \cdot W_{\text{cap}}}{N_{\text{cap}} \cdot W_{\text{cap}} + N_{\text{nor}} \cdot W_{\text{nor}}} \right) \times \left(\frac{\text{MCS} \times \text{RB}_{\text{total}}}{N_{\text{cap}}} \right)

where:

- UEcap\text{UE}_{\text{cap}}UEcap is the capacity per user with guaranteed SLA.
- NcapN_{\text{cap}}Ncap is the number of users with guaranteed SLA.
- WcapW_{\text{cap}}Wcap is the scheduling weight for users with guaranteed SLA.
- NnorN_{\text{nor}}Nnor is the number of normal users.
- WnorW_{\text{nor}}Wnor is the scheduling weight for normal users.

- MCS\text{MCS}MCS is the Modulation and Coding Scheme value.
- RBtotal\text{RB}_{\text{total}}RBtotal is the total number of resource blocks.

This calculation ensures that resource allocation is balanced and that the guaranteed bitrate traffic classes can be effectively managed to meet SLA requirements while limiting the number of SLA users in each cell.

We propose the usage of network slicing to limit the available capacity, different QoS classes, and number of users in order to simplify network capacity management.

5.3 Transport and core network enhancements

5.3.1 Transport network queuing

The transport network can utilize different queuing parameters for various traffic classes, minimizing jitter and ensuring smoother data transmission for applications requiring consistent performance.

5.3.2 Core network dimensioning

The core network must be over-dimensioned or configured with dedicated User Plane Functions (UPFs) for each slice or traffic type. This approach prevents the UPF from becoming a bottleneck, ensuring that traffic flows efficiently and meets SLA criteria.

5.4 Application-level strategies

5.4.1 Buffering for QoE

Applications can improve QoE by buffering traffic, particularly effective for non-real-time applications. For real-time applications, techniques such as variable bit rates and bonding technologies can be employed to manage QoS dynamically.

By combining these approaches—dynamic resource management in the RAN, optimized transport and core network configurations, and adaptive application-level strategies—mobile networks can deliver robust SLAs, ensuring high performance and reliability for diverse use cases.

6. Capacity exchange opportunities

To effectively scale and support diverse services, a unified network must cater to varying priorities. For instance, remote heart surgery demands a higher network priority than downloading emails. Network neutrality, in this context, does not imply a best-effort technology approach but rather a set of equitable business rules where all participants operate under the same conditions.

The capacity exchange model presents a scalable and neutral approach to managing network resources, ensuring fair access and efficient utilization. By integrating coordinate-based trading with predefined capacity in network slices, and leveraging a mediation layer, this model facilitates dynamic and equitable capacity allocation. Technologies like Embedded Subscriber Identity Module (eSIM) further enhance

flexibility, enabling buyers to maximize their network usage and maintain high service quality across diverse applications [11].

6.1 Network neutrality and capacity purchase

In a neutral network, all entities can purchase capacity under the same terms and conditions. This does not mean uniform capacity purchases; some may buy one unit while others buy 10 units, as long as the unit price and delivery terms are consistent. This system mirrors the stock exchange, where the unit price remains the same whether one buys a single share or a million shares, though the total cost differs. It is also acceptable for network profiles to be tailored for specific services, provided these profiles are equally accessible to all under the same conditions, similar to different stocks in an exchange [12].

To maintain neutrality, a neutral marketplace for communication capacity is essential. This marketplace ensures equal access to resources for both small and large companies, fostering true neutrality. Additionally, there must be provisions for best-effort traffic, which should be appropriately regulated.

6.2 Capacity trading

Capacity trading involves several dimensions: time, volume, quality, and location. In mobile networks, the location can be specified using either coordinates or cell IDs. Coordinates, which can be three-dimensional, provide greater accuracy and are more intuitive than two-dimensional cell IDs. Despite the advantages of coordinates, cell IDs are practical for network control.

6.3 Multiparty capacity providers

The capacity trading model can accommodate both coordinate-based and cell ID-based approaches. A cell ID model is suitable when a single capacity provider participates in the exchange, as mobile networks are managed using cell IDs, tracking areas, and network slices as abstraction layers. However, since different mobile networks have varying base station locations and cell IDs, a trading platform serving multiple capacity providers cannot rely solely on cell IDs but must still manage networks based on them.

6.4 Mediation layer

In a coordinate-based capacity trading system, network plan information must be integrated into the trading platform. This data enables the calculation of network capacity at different locations, facilitating trading. The same information is used for network control via cell IDs. The mediation layer ensures that capacity is allocated and managed efficiently, translating coordinate data into actionable insights for network operators [13].

6.5 Capacity buyers

Capacity buyers need a default connection to the capacity trading network, which combines several physical networks. Buyers utilize the physical network from which they have purchased capacity. Technologies like eSIM are necessary to switch between

physical networks or implement roaming agreements. This flexibility allows capacity buyers to seamlessly integrate and operate within the capacity trading ecosystem, optimizing their network usage based on their purchased capacity [14].

7. Findings and discussion

In the modern era of network technology, delivering predefined capacity with high reliability and performance is critical for meeting diverse application demands and user expectations. The chapters presented in this document offer a comprehensive overview of the challenges and solutions associated with quality of service (QoS) and Quality of Experience (QoE) in IP networks, particularly focusing on the complexities of mobile networks and the innovative approaches required to ensure Service Level Agreements (SLAs).

7.1 Quality of service and experience

At the foundation of delivering reliable network services are robust QoS features that differentiate data packets based on their class. This differentiation is essential for maintaining performance levels across both fixed and mobile networks. QoE, on the other hand, is a user-centric measure that combines network capacity with application requirements to ensure a seamless and satisfactory user experience. Strategies such as buffering, dynamic bandwidth allocation, and real-time adjustments are crucial for optimizing QoE, particularly in mobile environments where network conditions can rapidly change.

7.2 Service level agreements

Service Level Agreements (SLAs) formalize the expected performance metrics, including bit rates, delay, jitter, bit error rate, packet drops, and service availability. These parameters are continuously monitored, and any deviations result in penalties for the service provider, ensuring accountability and high standards of service. SLAs are integral in building trust between providers and customers, ensuring that performance expectations are clearly defined and met.

7.3 Innovative approaches in mobile networks

Traditional methods of frequency allocation—whether per user or application—face significant limitations in flexibility and efficiency. Modern mobile networks address these limitations through shared networks and advanced capacity management techniques. By using resource blocks and dynamic scheduling, networks can better allocate resources in response to varying conditions. Ensuring SLAs in such a dynamic environment requires setting stringent conditions for radio signal strength and SNR, as well as limiting the number of SLA users per cell to maintain performance levels.

7.4 Capacity exchange and neutrality

A scalable and efficient approach to network resource management is the concept of capacity exchange. This model allows different services to be served by a unified

network, prioritizing critical applications while maintaining neutrality. In a neutral network, capacity is sold under the same terms to all participants, fostering a fair and competitive marketplace. Technologies like eSIM enable seamless integration and usage of purchased capacity across multiple physical networks.

7.5 Mediation layer and multiparty providers

A mediation layer is essential for translating coordinate-based data into actionable network management insights. This layer ensures that capacity is allocated efficiently and equitably, even in a multiparty provider environment. By combining coordinate-based trading with cell ID-based control, the model supports dynamic and flexible capacity management, accommodating the varying needs of different network operators.

8. Conclusion

The comprehensive strategies discussed—ranging from QoS and QoE optimization to innovative SLA delivery methods and capacity exchange models—highlight the evolving landscape of network technology. By integrating advanced management techniques and maintaining a focus on neutrality and fairness, service providers can ensure high performance, reliability, and user satisfaction. These approaches not only address current challenges but also pave the way for future advancements in network technology, supporting an ever-growing array of applications and services in our increasingly connected world.

Acknowledgements

Dr. Thomas Pliakas.

Author details

Mika Skarp* and Jose Costa-Requena
Cumucore Oy, Espoo, Finland

*Address all correspondence to: mika.skarp@cumucore.com

References

[1] Cumucore. Available from: https://cumucore.com/

[2] Fiedler M, Hossfeld T, Tran-Gia P. A generic quantitative relationship between quality of experience and quality of service. IEEE Network. 2010;**24**(2):36-41. DOI: 10.1109/MNET.2010.5430142

[3] Kilkki K. Quality of experience in communications ecosystem. Journal of Universal Computer Science. 2008;**14**(5):615-624

[4] Hussein YS, Jiménez VPG, Al-Jumaily A. Emerging technology for 6G networks. International Journal of Information and Communication Engineering. 2024;**18**:143-149

[5] Bar A, Leth FH. Service level agreements: Issues and challenges. In: Proceedings of the 10th IEEE Symposium on Computers and Communications. Murcia, Spain: IEEE; 2005. pp. 996-1001. DOI: 10.1109/ISCC.2005.19

[6] Hammer M, Burtscher C, Effelsberg W. A survey on techniques for enhancing the performance of service-level agreements. IEEE Communications Surveys & Tutorials. 2014;**16**(3):1686-1707. DOI: 10.1109/SURV.2014.012114.00164

[7] Hussein YS. Li-Fi is a key enabling technology toward 6G networks. In: Computer Networks. Vol. 230. Elsevier; 2023. ISSN: 1389-1286

[8] Hussein YS, Alrashd M, Alabed AS, Alomar S. Data Centre Infrastructure: Design and Performance. Intechopen. 2023:109998

[9] Dahlman E, Parkvall S, Sköld J. 4G, LTE-Advanced Pro and the Road to 5G. 3rd ed. Academic Press; 2016. ISBN 978-0-12-804575-6

[10] Zhang H, Liu N, Chu X, Long K, Aghvami AH, Leung VCM. Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine. 2017;**55**(8):138-145. DOI: 10.1109/MCOM.2017.1600940

[11] Marsden CT. Network neutrality: From policy to law to regulation. IEEE Internet Computing. 2017;**21**(1):38-44. DOI: 10.1109/MIC.2017.10

[12] Li X, Campbell AT. Quality of service for advanced communications networks. Proceedings of the IEEE. 1998;**86**(5):898-924. DOI: 10.1109/5.664268

[13] Zhang N, Yang P, Liang Y-C, Ding G, Chen Z, Han Z. Spectrum sharing for internet of things: A survey. IEEE Wireless Communications. 2017;**24**(3):102-112. DOI: 10.1109/MWC.2017.1600344

[14] Rost P, Banchs A, Berberana I, Breitbach M, Doll M, Droste H, et al. Network slicing to enable scalability and flexibility in 5G mobile networks. IEEE Communications Magazine. 2017;**55**(5):72-79. DOI: 10.1109/MCOM.2017.1600920

Edited by Yaseein Soubhi Hussein and Abdulmajeed Al-Jumaily

Quality of Service (QoS) - Challenges and Solutions offers a comprehensive guide to understanding and advancing QoS in communication networks, essential for maintaining high performance and user satisfaction. It delves into the dynamic nature of QoS, exploring its significance across fields like telecommunications, computer networks, and new technologies, especially in the 5G era. This book covers foundational concepts, implementation strategies, practical applications, common challenges, and innovative solutions to enhance QoS, emphasizing adaptive approaches that respond to evolving network needs. It introduces readers to the latest developments, such as network slicing, edge computing, machine learning, and service orchestration, all crucial for future-ready communication systems. A significant focus of the book is the integration of QoS with 5G and beyond, highlighting methods like QoS-aware machine learning for resource optimization, dynamic resource allocation through predictive analytics, and traffic routing algorithms for real-time network improvements. It also addresses security as a core component of QoS, discussing virtualized and cloud-based infrastructures that bolster network resilience and performance. With these advanced strategies, the book provides insights into creating secure, high-performance networks that can handle increased traffic and complexity. The authors aim to equip researchers, practitioners, and students with a detailed QoS understanding, emphasising current and future innovations. By examining the impact of 5G, adaptive methodologies, and emerging technologies, the book positions itself as a crucial resource for those invested in the future of communication networks. In a digitally reliant world, mastering QoS is essential, and this guide presents a roadmap for navigating its complexities, ensuring top-tier network services, and enhancing user experiences across rapidly evolving digital landscapes. This book presents advanced QoS techniques applicable to various domains, including 5G networks, B5G networks, machine learning applications, QoS optimization, and network slicing.

Published in London, UK

IntechOpen

ISBN 978-0-85014-919-7