A Study on Round-Trip Time Performance of Tactile Internet using Multilevel Cloud Structure

Vaibhav Harihar Fanibhare

Supervisor: Prof. Nurul I. Sarkar
Prof. Adnan Al-Anbuky

School of Engineering, Computer & Mathematical Sciences
Auckland University of Technology

A thesis submitted to Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy

April 2025

I would like to dedicate this thesis to the divine energies of Lord Ganesha and Lord Hanuman, whose blessings have illuminated my PhD journey. May their strength and guidance continue to inspire me on my journey ahead.

Abstract

The tactile Internet (TI) is poised to be evolutionary to the Internet of Things (IoT), enabling real-time immersive interactions by transmitting a sense of touch and actuation over remote distance with audio or visual haptic feedback. A few of the stringent requirements of TI are achieving a round-trip time (RTT) of 1 ms, ultra-high reliability and availability. Addressing these challenges, the emerging TI can bring a plethora of applications, such as Industry 5.0 (industry automation), teleoperation frameworks (self-driven vehicles), augmented and virtual reality (AR/VR) related applications and Healthcare 5.0 (telesurgery, telediagnosis and telerehabilitation). Whilst these applications have been promising, TI is still in the nascent stage of development. The potential enablers of TI, such as the fifth-generation (5G) framework, software-defined networking (SDN), edge/fog computing (EC/FC), network slicing (NS), and an appropriate multiple access (MA) scheme, can meet these TI applications' requirements and their challenges. The ultra-low latency and high data rates of 5G provide a foundation for real-time communication. SDN enables dynamic traffic management and efficient resource allocation, reducing network congestion. EC/FC brings computation closer to end-users, minimising transmission delays. NS allows for isolating TI traffic, ensuring dedicated resources for critical applications. Finally, an MA, specifically the Non-orthogonal Multiple Access (NOMA) scheme, enhances spectral efficiency, supporting multiple users with minimal latency overhead. Collectively, these technologies address core requirements, mitigate round-trip time and enhance the responsiveness of TI systems.

This PhD research aims to study the RTT performance in TI communication infrastructure, employing a multilevel cloud structure with FC and SDN. Moreover, NS and an appropriate MA scheme are incorporated to meet the stringent demands or requirements of TI. Firstly, a comprehensive survey on TI is presented, focussing on design architecture, crucial application areas, current issues and challenges with potential enabling technologies. To emphasise the novelty of the survey, the brainstorming mind map is presented, covering all the discussed topics. In addition, an extensive literature review of the mentioned potential enablers of TI is explored.

Secondly, a novel fog-based traffic flow framework is proposed, employing FC and SDN to address the TI challenges of having an RTT of 1 ms and ultra-high reliability for establishing haptic communications. An effective traffic flow algorithm is developed to route the traffic in a multilevel cloud structure efficiently. The combined SDN and FC approaches with the developed algorithm provide an effective solution to reduce extra processing and

waiting times at each level of the cloud-based structure. Hence, the RTT is also reduced when traffic flows from master to slave sections and vice versa. The performance metrics such as throughput, RTT, energy consumption, and reliability are evaluated using the iFogSim simulator. The simulation results for the proposed system outperform those of the existing edge, cloud, and cellular networks.

Thirdly, a novel NS mechanism based on TI communication infrastructure is proposed, leveraging SDN with Open vSwitch (OVS) to resolve the issue of provisioning and controlling the network slices on demand. Hence, the network slicing algorithms are developed for three different scenarios: topology slicing, service slicing, and emergency slicing, to have tailored and customised slices according to network requirements. This communication infrastructure is sliced under diverse traffic such as UDP, TCP and other traffic packets using pre-designed slice configurations. The system performance is evaluated with metrics such as throughput and RTT by Vagrant environment and Mininet. The simulation results align with pre-designed slices' configurations, i.e., the preset allocated spectrum and RTT constraint values, and show the efficacy of the proposed NS algorithms, thus validating these algorithms with existing algorithms.

Fourthly, a novel downlink power domain Single-Input Single-Output (SISO) NOMA communication scenario for TI is proposed, employing multiple sensors and actuators collectively treated as users and a base station. An analytical system model is mathematically derived under the NOMA scheme, incorporating signal-to-interference and noise ratio (SINR), sum rate, and fair power allocation (PA) coefficients. Considering two-user and three-user scenarios, the system performance is analysed and evaluated bit error rate (BER) and sum rate between NOMA and orthogonal multiple access (OMA) by varying path loss exponent and fixed PA coefficients using MATLAB. Moreover, the outage probability and achievable sum rate are analysed by varying fixed and fair PA coefficients in the NOMA scheme. For higher SNR, the achievable sum rate for the proposed NOMA system outperforms OMA, thus maximising spectral efficiency, minimising latency and promoting dynamic PA and user fairness. Finally, 4×4 Multiple-Input Multiple-Output (MIMO) NOMA performance is analysed by incorporating zero forcing-based beamforming and a round-robin scheduling process with SISO-NOMA concerning achievable sum rate and latency.

Lastly, the proposed models in this research have been validated against state-of-the-art models, proving performance superiority, resource efficiency and scalability.

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements) nor used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Vaibhav Harihar Fanibhare
April 2025

Co-authorship Contribution

STUDENT AND SUPERVISOR APPROVALS

By signing you are confirming that the co-author contributions stated in the table(s) below are accurate.

Student Name	Vaibhav H. Fanibhare	**Signature**		**Date**	23/08/2024
Supervisor Name	Prof. Nurul I. Sarkar	**Signature**		**Date**	23/08/2024

Chapter Number:	**Chapter 2**
Manuscript Title:	**A survey of the tactile Internet: Design issues and challenges, applications, and future directions**
Publication Status:	**Published**
Reference if published:	**V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A survey of the tactile Internet: Design issues and challenges, applications, and future directions," *MDPI Electronics*, vol. 10, no. 17, p. 2171, 2021.**

AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
V. Fanibhare (V.F.) **N. I. Sarkar (N.I.S.)**	Conceptualisation, V.F.; Methodology, V.F. and N.I.S.; Discussion, V.F., N.I.S. and A.A.-A.; Implementation, V.F.; Validation, V.F. and N.I.S.; Empirical evaluation, V.F. and N.I.S.; Writing—original draft preparation, V.F.; Writing—review and editing, V.F., N.I.S. and A.A.-A.; Project supervision, N.I.S. and A.A.-A
A. Al-Anbuky (A.A.-A.)	(V.F. [90%], N.I.S. [7%] and A.A.-A. [3%])

Chapter Number:	**Chapter 3**
Manuscript Title:	**A cloud-based traffic flow framework for tactile Internet using SDN and fog computing**
Publication Status:	**Published**
Reference if published:	**V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “A cloud-based traffic flow framework for tactile Internet using SDN and fog computing,” in *29th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–6, IEEE, 2019.**
AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
V. Fanibhare (V.F.)	Conceptualisation, V.F.; Methodology, V.F. and N.I.S.; Discussion, V.F., N.I.S. and A.A.-A.; Implementation, V.F.; Validation, V.F. and N.I.S.; Empirical evaluation, V.F. and N.I.S.; Writing—original draft preparation, V.F.; Writing—review and editing, V.F., N.I.S. and A.A.-A.; Project supervision, N.I.S. and A.A.-A
N. I. Sarkar (N.I.S.)	
A. Al-Anbuky (A.A.-A.)	(V.F. [85%], N.I.S. [10%] and A.A.-A. [5%])

Chapter Number:	**Chapter 3**
Manuscript Title:	**Toward a fog-based traffic flow framework for tactile Internet**
Publication Status:	**Published**
Reference if published:	**V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “Toward a fog-based traffic flow framework for tactile Internet,” *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 10 718–10 731, 2021.**
AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
V. Fanibhare (V.F.)	Conceptualisation, V.F.; Methodology, V.F. and N.I.S.; Discussion, V.F., N.I.S. and A.A.-A.; Implementation, V.F.; Validation, V.F. and N.I.S.; Empirical evaluation, V.F. and N.I.S.; Writing—original draft preparation, V.F.; Writing—review and editing, V.F., N.I.S. and A.A.-A.; Project supervision, N.I.S. and A.A.-A
N. I. Sarkar (N.I.S.)	
A. Al-Anbuky (A.A.-A.)	(V.F. [90%], N.I.S. [7%] and A.A.-A. [3%])

Chapter Number:	**Chapter 4**
Manuscript Title:	**TINetS3: SDN-driven network slicing enabling scenario-based applications in tactile Internet**
Publication Status:	**Published**
Reference if published:	**V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “TINetS3: SDN-driven network slicing enabling scenario-based applications in tactile Internet,” *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 4639–4654, 2024.**
AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
V. Fanibhare (V.F.)	Conceptualisation, V.F.; Methodology, V.F. and N.I.S.; Discussion, V.F., N.I.S. and A.A.-A.; Implementation, V.F.; Validation, V.F. and N.I.S.; Empirical evaluation, V.F. and N.I.S.; Writing—original draft preparation, V.F.; Writing—review and editing, V.F., N.I.S. and A.A.-A.; Project supervision, N.I.S. and A.A.-A
N. I. Sarkar (N.I.S.)	
A. Al-Anbuky (A.A.-A.)	(V.F. [85%], N.I.S. [8%] and A.A.-A. [7%])

Chapter Number:	**Chapter 5**
Manuscript Title:	**A study of downlink power-domain non-orthogonal multiple access performance in tactile Internet employing sensors and actuators**
Publication Status:	**Published**
Reference if published:	**V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A study of downlink power-domain non-orthogonal multiple access performance in tactile Internet employing sensors and actuators," *MDPI Sensors*, vol. 24, no. 22, p. 7220, 2024.**
AUTHOR SURNAME: (order as per manuscript)	**CONTRIBUTION** (May copy from the guidelines above)
V. Fanibhare (V.F.)	Conceptualisation, V.F.; Methodology, V.F. and N.I.S.; Discussion, V.F., N.I.S. and A.A.-A.; Implementation, V.F.; Validation, V.F. and N.I.S.; Empirical evaluation, V.F. and N.I.S.; Writing—original draft preparation, V.F.; Writing—review and editing, V.F., N.I.S. and A.A.-A.; Project supervision, N.I.S. and A.A.-A
N. I. Sarkar (N.I.S.)	
A. Al-Anbuky (A.A.-A.)	(V.F. [85%], N.I.S. [10%] and A.A.-A. [5%])

Acknowledgements

Since I started my PhD journey with AUT University, I have come across countless individuals who supported and encouraged me. I want to extend my immense gratitude to each one of them.

First and foremost, I am deeply indebted to my primary supervisor, Prof. Nurul I. Sarkar, for your continuous and invaluable guidance and unwavering support throughout my PhD journey. Your profound insights and critical suggestions have helped me shape and produce a top-notch PhD work that has resulted in excellent publications. I am heartily grateful for your supervision, patience, and support in navigating my professional and personal challenges, thus sculpturing my research and personal growth.

Secondly, I am equally indebted to my secondary supervisor, Prof. Adnan Al-Anbuky, for your analytical perspective and constructive discussions in refining my research. You have enriched my research by allowing me to delve deeper and consider new angles of critical thinking, ensuring the robustness of my findings. Thus, your commitment has motivated me to aspire to high research standards.

Thirdly, I would like to sincerely thank AUT University for accepting me as a PhD student and fostering a conducive platform for research and the necessary resources to undertake this PhD research. Specifically, I am grateful to the School of Engineering, Computer and Mathematical Sciences, and their staff for coordinating and assisting me with the academic processes.

In addition, I would like to acknowledge my parents, Mr. Harihar and Mrs. Devki, and my sister, Ms. Shweta, for their unconditional love and support since I started thinking of pursuing PhD studies. Their sacrifices and belief in my abilities have been the bedrock of my perseverance, thus fuelling my determination and commitment. I am eternally blessed for being a constant source of strength throughout this journey.

Moreover, I would like to express a huge thanks to my beloved wife, Mrs. Shital, for being through the crests and troughs of our life. Your wholehearted love, patience, support and sacrifices have kept me physically and mentally sane. Thank you for bearing my frustrations during the most challenging times of this journey.

Last but not least, I am also obliged to have wonderful friends, Shubham, Rohit, Pranav, Sumeet, Sahil, Nikita, Pinakin and Mrunal, for sharing their laughter, encouragement, love and being a source of immense comfort. Thank you for being a part of my life and this PhD journey.

Copyright

Chapter 1

Introduction

"Tactile Internet" (TI) broadly refers to a communication network that delivers real-time control, touch, and sensing or actuation with or without audio or visual haptic feedback in the physical and virtual world [1]. The TI is an emerging research area that is still in the nascent stage and facing many issues and challenges. Ensuring the ultra-low end-to-end delay or round-trip time (RTT) and ultra-high reliability required by TI are challenging. Specifically, the main problem that needs to be resolved and to realise the full potential of TI is encapsulated by the '1 ms challenge'. That means the RTT in the TI system should be '1 ms' or below. Otherwise, the TI system would introduce motion sickness (lag) in latency-based TI applications.

TI envisages the smooth integration of physical and virtual applications or services, enabling the end-users to communicate with the other end without noticeable delay as if they were happening in real time. Moreover, it paves the way for potential applications or services such as healthcare 4.0, Industry 4.0, education and entertainment sectors. For instance, imagine a world where telerobotic surgery is performed by a surgeon at one location while operating on a patient at a remote location. Moreover, a student experiencing an immersive virtual reality will perform a virtual lab instead of investing in buying goods or being physically present.

In addition, to achieve the stringent demands of TI, such as ultra-low latency ($\leq$1 ms), high reliability (99.999% availability), and massive connectivity to support mission-critical applications such as remote surgery, autonomous systems, and industrial automation, the communication infrastructure requires some key technological enablers including Fog Computing (FC), Software-Defined Networking (SDN), Network Slicing (NS), and an appropriate Multiple Access (MA) scheme. While other technologies, such as network coding and caching techniques, offer certain advantages, they were deemed less suitable for the real-time, low-latency requirements of TI. The rationale behind selecting these technologies over others is detailed below.

SDN enables flexible and intelligent network management by decoupling the control plane from the data plane. This separation allows centralised network control, optimising routing and resource allocation dynamically. SDN provides a more adaptive approach to traffic management by enabling real-time adjustments to network parameters based on demand, thereby ensuring ultra-reliable low-latency communication (URLLC). The programmability of SDN further enhances resource efficiency by prioritising mission-critical data flows while minimising unnecessary network overhead. Its interoperability with other technologies, such as FC, NS, and Non-orthogonal Multiple Access (NOMA), ensures that TI applications receive the required service guarantees across the communication infrastructure.

FC plays a crucial role in reducing latency and enhancing processing efficiency by bringing computational resources closer to end-users. Instead of relying on centralised cloud computing, which introduces delays due to distant data centres, FC distributes processing power to edge nodes, reducing transmission delays and supporting real-time decision-making. This capability is significant for applications such as haptic feedback in TI, where even minimal delay can significantly impact user experience. By handling data at the edge, FC also reduces network congestion and mitigates the risk of bottlenecks in high-traffic environments. Additionally, the decentralised nature of FC makes it well-suited for mobility-driven applications, as it ensures seamless service continuity across different edge nodes. On the other hand, cloud computing (CC) is unsuitable for TI due to its reliance on centralised data centres, which can result in significant delays and hinder real-time interactions. The long round-trip time to access cloud resources makes it less effective for latency-sensitive TI applications, where low latency and quick response times are critical.

NS enables the customisation of network resources to meet the specific needs of different applications. Through virtualisation, multiple network slices can be created, each optimised for particular latency, bandwidth, and reliability requirements. This capability ensures that TI applications such as remote surgery, industrial automation, and immersive haptic communication operate within dedicated network environments free from interference by other services. The isolation of network slices prevents congestion and performance degradation while allocating resources on a per-slice basis guarantees quality of service (QoS). Enabling end-to-end network slicing ensures that TI applications receive predefined performance guarantees necessary for real-time operations.

MA techniques play a crucial role in managing connectivity in TI, where a massive number of devices must communicate with minimal latency and interference. Traditional multiple access methods, such as Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), introduce excessive latency and fail to utilise spectrum resources efficiently. To address these limitations, NOMA was selected as the preferred MA scheme for TI. Unlike conventional orthogonal access techniques, NOMA allows multiple users to share the same time and frequency resources by distinguishing them through power or code domains. This approach enhances spectral efficiency and significantly improves network capacity. Additionally, by overlapping transmissions, NOMA reduces scheduling delays and optimises spectrum utilisation, making it ideal for latency-sensitive TI applications.

While network coding and caching techniques are widely used in various networking scenarios, they do not fully align with the requirements of TI. Network coding techniques such as Linear Network Coding (LNC), Random Linear Network Coding (RLNC), XOR-based coding, and Physical Layer Network Coding improve throughput and reliability in data transmission. However, these techniques introduce additional processing overhead due to the computational complexity of encoding and decoding operations. This increased burden on intermediate nodes results in added processing delays, making network coding unsuitable for ultra-low latency applications. Furthermore, network coding is primarily designed to optimise network throughput in high-bandwidth, packet-switched environments rather than minimising latency, which is a critical requirement for TI.

Similarly, caching techniques, including cache placement strategies and cache replacement policies, enhance data retrieval by storing frequently accessed content closer to users. However, caching is optimised for static content and is not well-suited for applications that require real-time, bidirectional communication. While caching reduces retrieval time for pre-stored data, it cannot replace the need for instant processing and real-time updates, which are essential for TI applications such as remote control systems, haptic feedback mechanisms, and live sensor data processing.

The enabling technologies selected for this research, viz. SDN, FC, NS, and NOMA were chosen because they best satisfy the three core requirements of TI. Ultra-low latency, which is essential for real-time applications, is achieved through the combined use of FC, SDN, and NS, as they collectively reduce processing delays, optimise routing, and allocate resources dynamically. High reliability, defined as 99.999% availability, is ensured by NS and SDN, which provide dedicated network resources and intelligent traffic management to maintain service continuity. Massive connectivity, a fundamental requirement for large-scale TI deployments, is managed efficiently through NOMA, which maximises spectrum utilisation while minimising access delays.

By integrating these technologies, this research establishes a robust framework for deploying TI applications, ensuring they meet the performance requirements for real-time, mission-critical scenarios. The selection of these technologies reflects their ability to deliver the necessary balance of minimised latency, improved reliability, and scalability, all of which are fundamental to the success of TI in future communication infrastructure.

1.1 Research background

The background of this PhD research is provided and discussed for TI and its potential enabling technologies, such as FC, SDN, NS and NOMA, in the following subsections.

1.1.1 Tactile Internet

Tactile Internet (TI) is a new concept in Internet technologies where human beings interact with each other, thus allowing us to transmit the sense of touch and actuation over a long distance in real-time.

Mobile Internet usage has tremendously increased among people of all ages globally. Everyone can be connected wirelessly without any difficulties in location and time. The tremendous research and improvements in networking and communication have led to a remarkable development in internet usage, thus creating a space for companies to produce and build up the next-generation gadgets and enhance the client experience. Depending on the Internet, communication is comprised of people communicating with single or multiple devices. There is a requirement for a communication interface where humans will interact with other humans by touch and actuation. Subsequently, this requirement has been in great demand; however, it has not been accomplished on a bigger scale.

Subsequently, mobile Internet can be used as a medium of communication between two or more people without being physically present. The mobile Internet has been an entirely dependable and comfortable mode of communication with constant upgrades through research and advancements to improve the end-user experience by minimising latency and improving data speeds. The most recent upgrade in communication technologies is 5G technologies [6], which overpassed the 4G technology.

Furthermore, the Internet of Things (IoT) has emerged as another dimension where physical and/or virtual devices are interconnected over the Internet and cloud, enabling them to collate, process and compute the data. The applications related to IoT are investigated more as they offer umpteen chances to ease the way of life.

However, TI exists at the conceptual stage, and much research is still needed to revolutionise it. TI illustrates the next evolutionary step: integrating a sense of touch and actuation with real-time haptic feedback in Figure 1.1. This figure shows the evolution of TI from mobile Internet and IoT.

To have a communication medium for transferring sense or actuation, touch and controlling information in real-time, the TI must possess competencies such as high reliability,

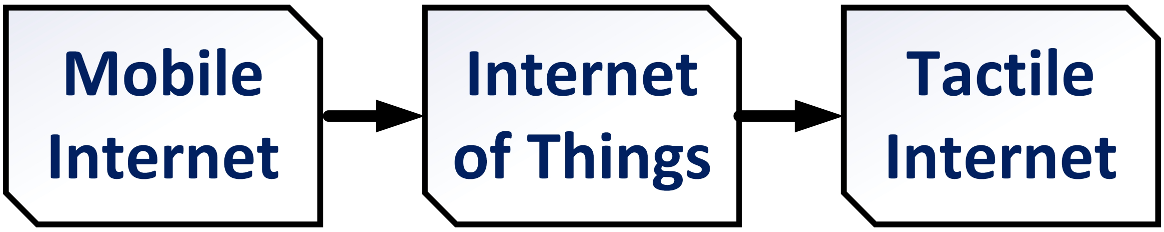


Fig. 1.1 The evolution of TI from cellular mobile Internet [1].

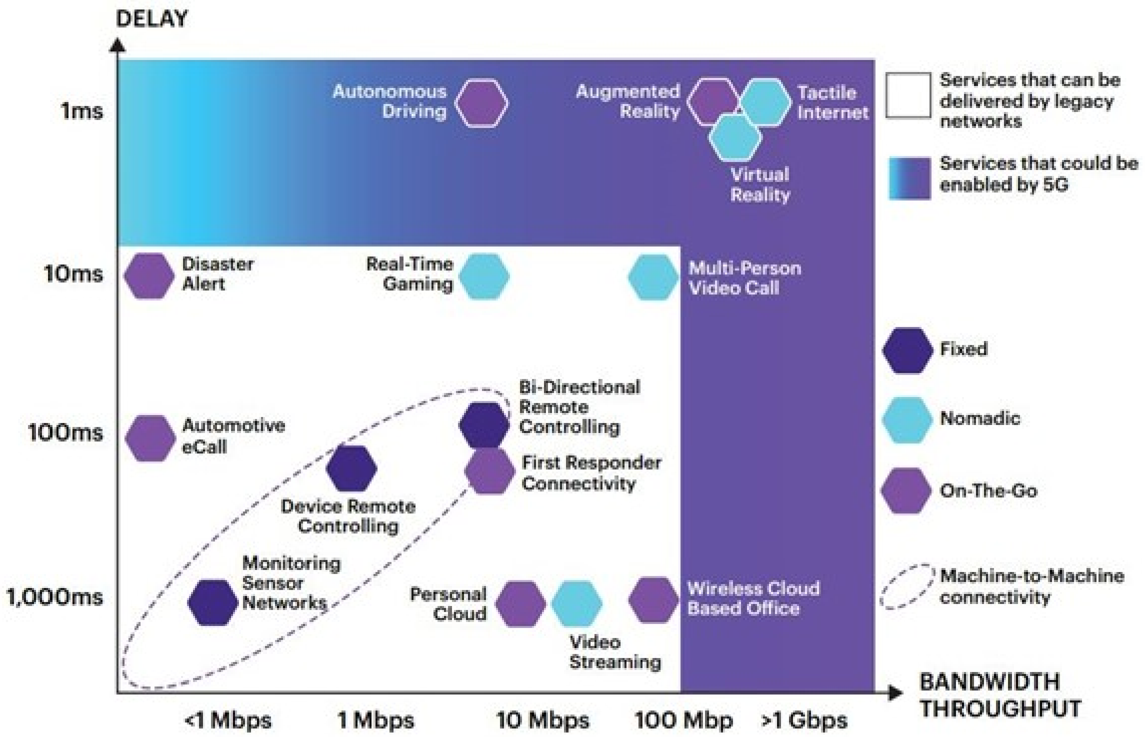


Fig. 1.2 Potential innovative applications enabled by TI [2].

responsivity and cognitive behaviour. The high accessibility, ultra-fast response times, and carrier-grade reliability of the TI will add another feature to human-machine communication by making continuous intelligent frameworks [1]. The TI is imagined to empower phenomenal applications that will change almost every fragment of the world. Figure 1.2 shows the graphs between the delay incurred in TI and the bandwidth/throughput required to carry out the functionality of TI.

1.1.2 Software-defined networking

Software-defined networking (SDN) is a relatively new concept in networking that facilitates a dynamic, flexible and cost-effective system structure by the physical partition of the data

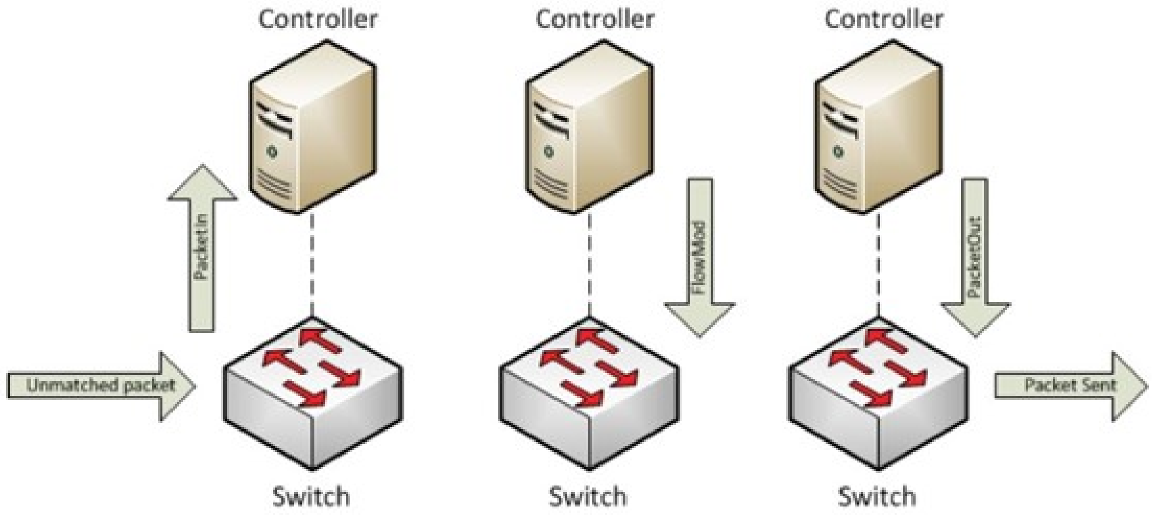


Fig. 1.3 A switch: 1) asks a controller if it knows how to handle a packet. 2) gets a rule installed by the controller if not, and 3) uses the standard to forward the packet towards the destination [3].

forwarding plane and control plane. The suitable choice to deal with traffic [7][8][9] is handled by the control plane, whereas the job of forwarding the traffic in reacting to the control plane is taken care of by the data plane. SDN grants permission to the system operator to design, configure, control and deal with the system through programming software such as Application programming interfaces (APIs). SDN is mainly utilised to disentangle the system hardware and increase network adaptability [10].

Moreover, SDN is recognised as bringing a change in how the Internet is curbed, and this is the one we need to leap forward towards TI. Referring to Figure 1.3, in SDN, the switches only execute data plane functionality, which is for packet forwarding, and depend on a controller infrastructure that facilitates the control plane functionality, that is, the intelligence like processing the forwarding rules and dumping them in the switches. In contrast, conventional networking equipment includes both features in one physical device. The decoupling of the data and control plane within SDN ensures edge-cutting decisions on the stream of per packets and facilities guaranteed quality of service (QoS), compared to the best-effort service, for example, in terms of latency.

The main potential of SDN lies in enabling modular network configuration through the use of network abstractions, just as operating systems for PCs rely on abstractions over multiple levels to manage their hardware resources.

Consequently, the controller's properties empower the framework by increasing the system efficiency regarding bit rate and end-to-end latency. Since the fundamental thing of the TI framework is to meet 1 ms RTT, i.e. ultra-low latency, the cognitive control ought to be conveyed at the central network to accomplish the desired end-to-end latency. Subsequently, the interest in SDN [11] has risen to shape an envisioned architectural design. Hence, SDN facilitates an architectural system wherein the control plane and data plane are decoupled, thus empowering the software coding of the network through its controllers.

1.1.3 Fog computing

The origin of fog computing was first proposed by Flavio Bonomi, vice president of network device manufacturing organisation Cisco, in 2011 [12][13]. Fog computing can be explained as the higher virtualised platform where an enormous number of heterogeneous, pervasive and decentralised devices connect and correspond to other devices to provide computation, storage, process tasks and facilitate networking services between devices and conventional cloud centres, but not precisely at the edge of the networks, without the interference of the third parties. The capabilities of cloud computing and services are extended to fog computing, where routers, gateways, switches and other network devices can also be a part of the network and have all the properties of fog computing. It is fundamentally utilised for automating types of equipment since fog computing has a background in IoT.

Moreover, fog computing has a decentralised computing structure dependent on fog computing nodes (FCNs). It can be fixed between the end-user device end and the cloud at any architecture location. For FCNs that utilise devices like machine-to-machine (M2M) gateways and wireless routers, fog computing is used as a computing layer between the end device layer and the cloud layer. These FCNs are utilised to process and store information from end-user devices before sending it to the cloud.

Consequently, TI is a developing idea that spotlights supporting high-fidelity, ultra-responsive, and generally accessible machine-to-human communications. To minimise the transmission delay and reduce Internet conditions, fog computing has been posted as a significant part of the TI.

The capabilities and competencies of fog computing that are beneficial for TI are as follows:

- It has location awareness and can be located at the network's edge.
- It incurs low latency, which supports large-scale complex sensor networks and storage capabilities in a wide-area distributed network.

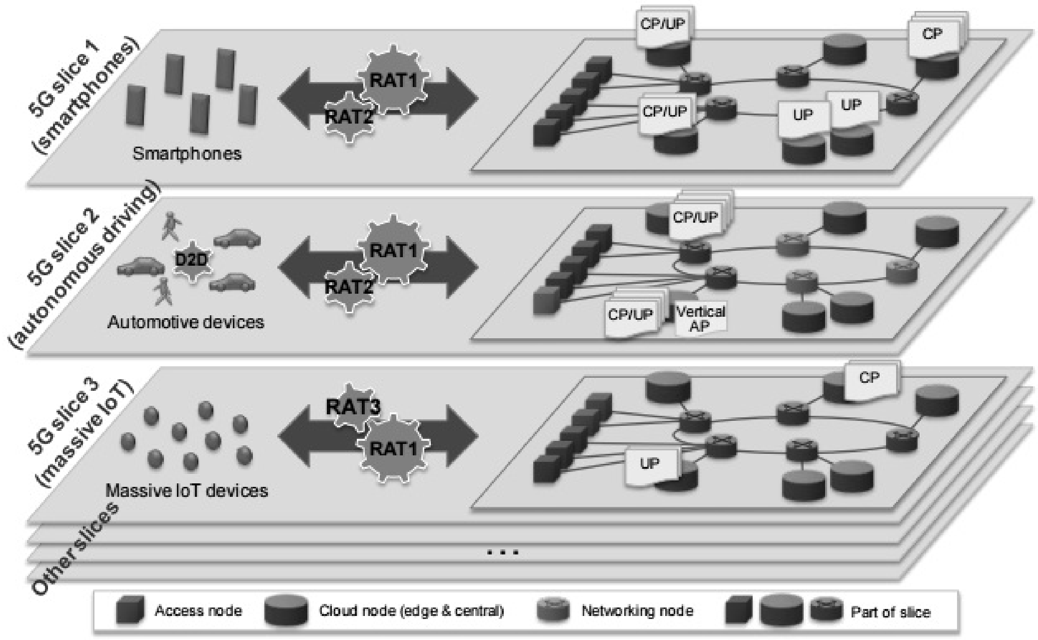


Fig. 1.4 Illustration of network slicing for 5G network [4].

- It is widely geographically distributed.
- It can support the mobility and heterogeneity of the fog nodes.
- It empowers wireless access and real-time communication with end-users and the cloud.
- Any fault in the network can be analysed online and is integrated with backend cloud computing.

1.1.4 Network slicing

Network slicing (NS) is one of the topics most discussed when structuring 5G systems. It is planned to empower operators to slice a single physical network into numerous virtual networks designed according to explicit services and application objectives.

Even though several research initiatives have been carried out on the 5G framework, there is no consistent concurrence on its architectural design. However, the general consensus is that the 5G framework needs to be planned in an adaptable way with the end goal that one network, based on typical physical infrastructure, is productively shared among various vertical applications. Here, the concept of network function virtualisation (NFV) is taken into consideration, where the network is sliced according to the requirements of the particular application.

The capacity to make various intelligently isolated network segments over a solitary physical network was one of the fundamental drivers of the industry development towards SDN long before the term ‘slicing’ appeared in the literature. The enthusiasm around 5G has restored the idea of NS to the point of making it a trending term. NS can be viewed as another complex type of virtual private network (VPN) innovation.

The idea of 5G network slicing was first depicted in a white paper on 5G, created by the next generation of mobile networks (NGMN) partnership [14]. Figure 1.4 explains this idea by demonstrating diverse network slices tending to various sorts of applications with multiple degrees of functionality, execution and dependability.

Figure 1.4 illustrates an example of network slicing with three networks. The first slice is proposed to help with the regular traffic created by smartphones. A broad scope of network function (NF) was enacted, and higher limits on virtual connections were started to help the assortment of broadband services through which smartphones can access their applications. The second slice is committed to self-driven vehicles. Self-sufficient driving requires an exceptionally low end-to-end delay and unwavering high quality to trade traffic information and warning messages with the network and different vehicles. The third slice is proposed to help ‘massive IoT’ supported applications, where smart metering equipment irregularly creates a minimal volume of information and is not expected to move. In contrast to the two past cases, the arrangement of NFs to be incorporated into the third slice is somewhat constrained. According to the publication of the white paper, NGMN facilitated increasingly in-depth definitions and terminologies to comprehend various parts of the slice [15].

As illustrated in Figure 1.5, abstracted from [15], the three layers can be recognised:

- The service instance layer signifies the end-user administration or undertaking administrations to be bolstered by a slice.
- The network slice instance layer, which has all network slice instances, handles the mentioned end-user administrative work.
- The resource layer signifies all resources (virtual or physical) from the hidden network architecture accessible to the network slice instance layer.

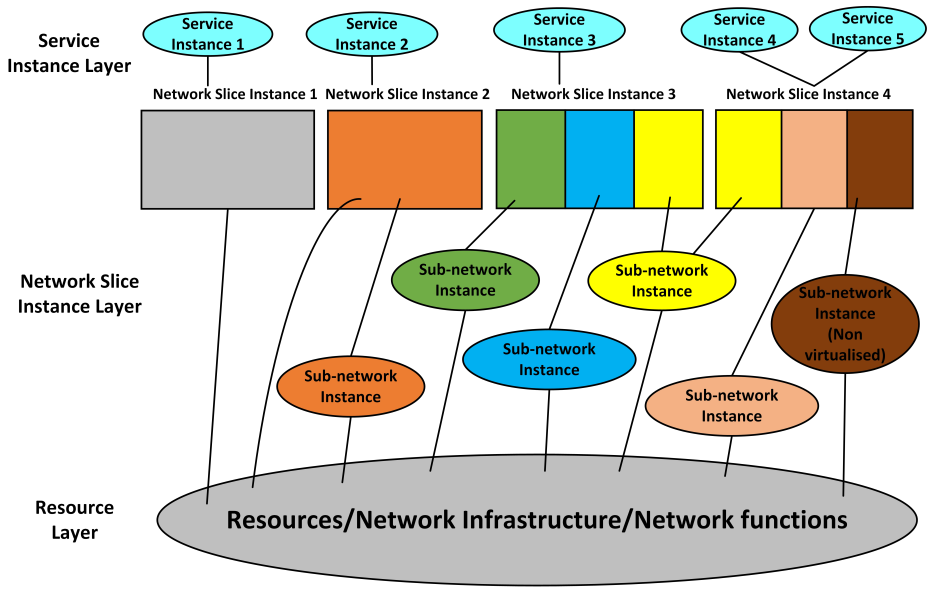


Fig. 1.5 Schematic of layers of network slicing [4].

1.1.5 Multiple access technique

The mobile Internet has become a source of connecting various global networks for information and resource sharing. In such a scenario, IoT associates smart gadgets like smartphones with home equipment using different communication standards. Such services demand reliable, low latency, and end-to-end data security, influencing the system's spectral efficiency. Nonetheless, multiple access (MA) strategies are utilised for productive spectrum utilisation, which permits the sharing of accessible bandwidth among an enormous number of clients. These strategies upgrade the system's limit by facilitating a superior quality of services (QoS) to the end clients.

Using the client-specific spreading arrangement, numerous clients are served by code division multiple access (CDMA) from the same resource block. The clients of CDMA cannot recover the original information because of the loss of orthogonality code by more flickering. In the orthogonal multiple access (OMA) strategy, the bandwidth loss happens because clients use the orthogonal resource block. Conventional MA strategies cannot give services to different clients simultaneously, so these are not reasonable for the 5G network [16][17]. It can provide services to various clients from one orthogonal resource block by shrewdly promoting clients' particular channel conditions.

Nevertheless, it is fit to give clients services according to their QoS necessity by distributing dynamic power levels. In non-orthogonal multiple access (NOMA), a superposition coding is executed at the transmitter side. On the receiver side, successive interference can-

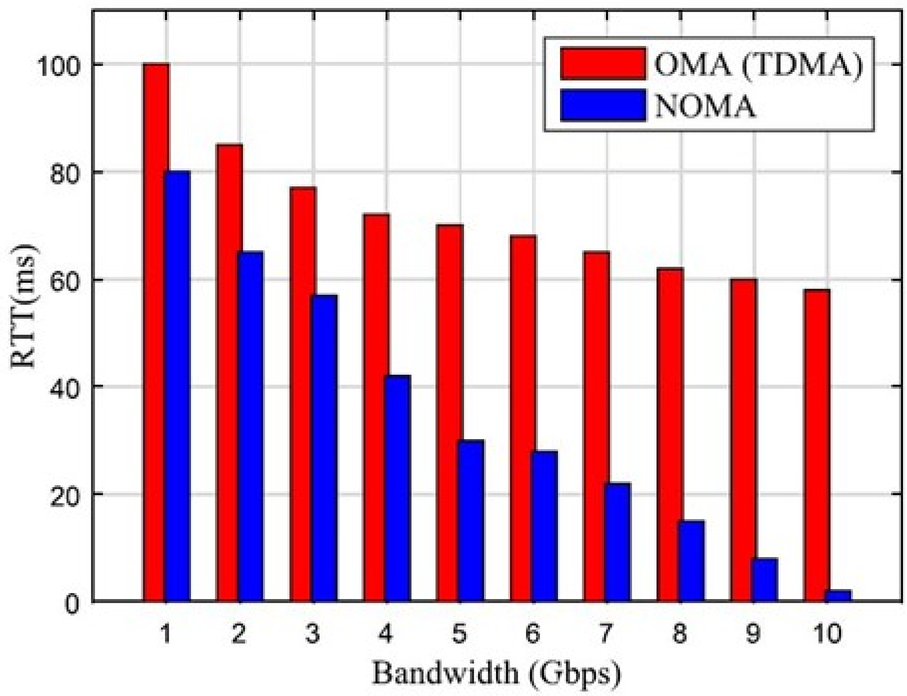


Fig. 1.6 Comparison between OMA and NOMA [5].

cellation (SIC) is performed to diminish the shared interference by utilising non-orthogonal resources. NOMA is categorised into two types: code-domain NOMA and power-domain NOMA. Figure 1.6 shows the RTT and bandwidth analysis for NOMA and OMA, where OMA facilitates low latency for lower bandwidth, but if bandwidth increases, then NOMA is a better option.

1.2 Rationale and significance

TI is imagined to make a change in perspective from content-oriented communications to steer/control-based communications by transmitting real-time haptic information, such as touch, actuation, friction, vibration, and motion over the Internet along with the existing audio-visual data traffic. This developing TI innovation, additionally considered as the following next phase of IoT (as shown in Figure 1.1, is required to make various open doors for innovative markets in a wide assortment of applications, including teleoperation frameworks (automation industry), augmented reality/virtual reality (AR/VR) related applications, and e-health services (telesurgery, telediagnosis and telerehabilitation). Nevertheless, the realisation of TI over wireless medium in the era of 5G technology and next-generation networks makes different non-traditional communication challenges and stringent demands as far as RTT

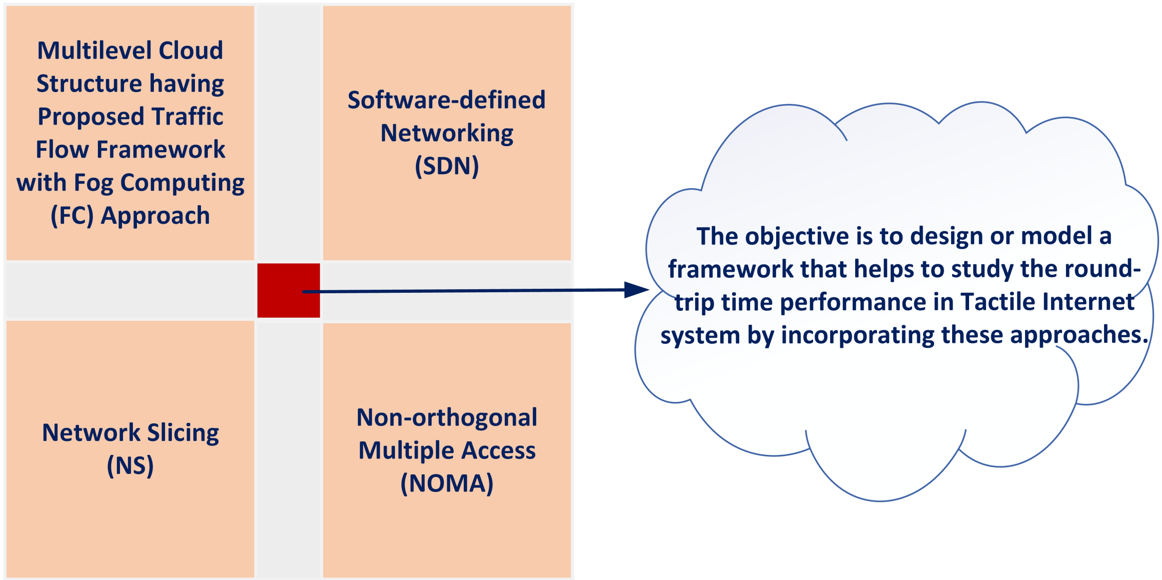


Fig. 1.7 The research gap that is unaddressed in TI.

of 1 ms, ultra-low latency, ultra-high dependability, high data-rate connectivity, dynamic resource allocation, multiple access (MA) techniques, and quality-latency rate trade-off are concerned.

Therefore, this research aims to study the RTT performance in the TI communication infrastructure, enabling the system to function smoothly with cyber or motion sickness (lag). There are many challenging solutions to reduce the delay over the network. Hence, the potential enablers of TI, such as FC, SDN, NS and NOMA techniques, can be employed to address the challenges incurred in TI, thus viewed as promising paradigms for aiding the TI applications for 5G and next-generation networks.

1.3 Research gap identification

In order to meet the stringent demands or requirements of TI, the incorporation of the proposed multilevel cloud structure with traffic flow framework, Fog Computing (FC), Software-defined Networking (SDN), Network Slicing (NS) and NOMA techniques will resolve the issues related to the low-latency based TI services/applications. Figure 1.7 depicts the unaddressed area (intersected area), which will fulfil the needs and improve the system performance of TI.

1.4 Research questions

The main research question is formulated below to achieve the ultimate aim mentioned in Section 1.2.

"*What framework can be developed to reduce the round-trip time (RTT) in the tactile Internet system?*"

To address our main research question, the following sub-questions are framed to accomplish our TI research.

RQ 1 What framework can be proposed to manage the traffic flow efficiently by adopting a multilevel cloud structure in the tactile Internet?

RQ 2 What strategies can be employed to improve the RTT of tactile Internet incorporating SDN and NS approaches?

RQ 3 What multiple access technique, such as NOMA, can be developed to support low-latency tactile Internet applications/services?

1.5 Research methodology adopted

Design science research methodology (DSRM) [18] is a systematic, robust and flexible methodology that facilitates a structured approach for creating and evaluating artefacts to resolve practical issues and challenges, thus building a research bridge between theory and practice. We have modified DSRM to accomplish our work in TI. In addition to modified DSRM, analytical modelling and system simulation methodologies are adopted in this research to comprehend the network's behaviour, optimise the network's performance, and provide valuable insights into potential bottlenecks to address research questions.

The graphical representation of the adopted DSRM is illustrated in Figure 1.8 to explain its flow and the iterative processes/activities that were conducted and contributed through Manuscripts 1, 2, 3, 4 and 5 individually in this PhD research. This adopted method consolidates standards, practices, and strategies to determine the research and meet the following four objectives. Firstly, the research starts consistently with the prior extensive literature/knowledge base and analysis. Secondly, it characterises a minimum procedure model to conduct the research. Thirdly, it facilitates and emphasises the wholistic modelling of research to present, demonstrate, implement and evaluate, supported by a knowledge base

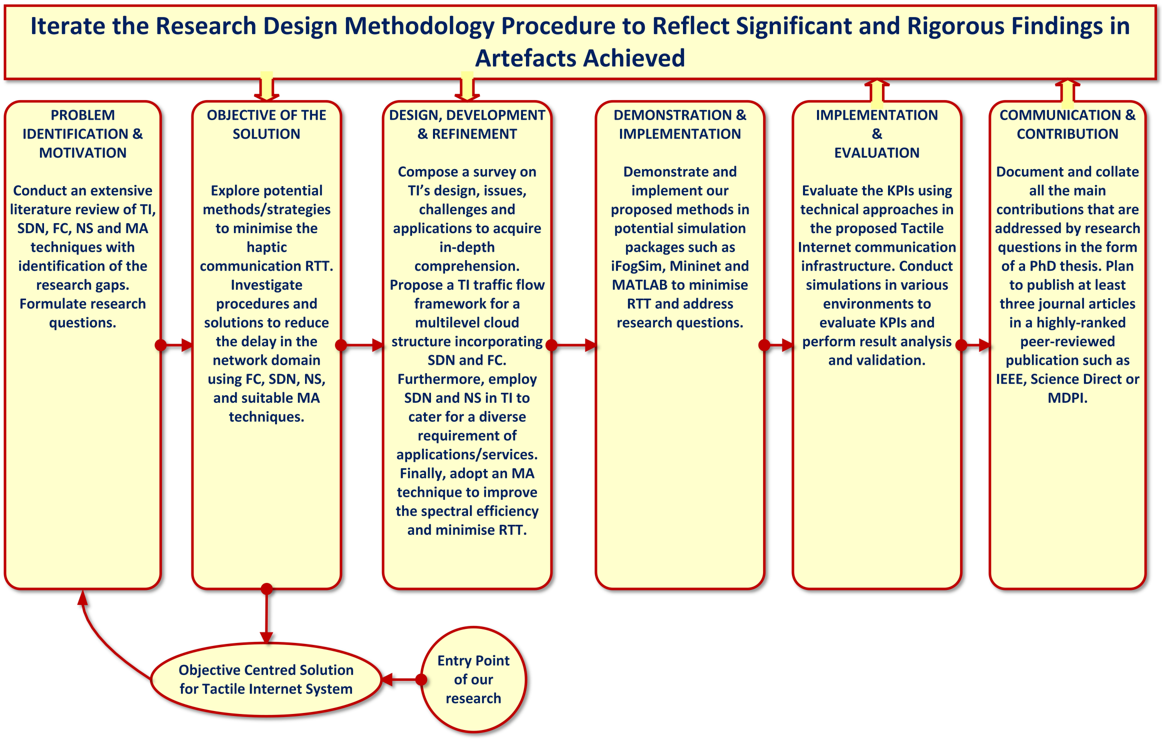


Fig. 1.8 The adopted research methodology.

and design model. Finally, the research is summarised and communicated through scholarly articles (Manuscripts 1, 2, 3, 4 and 5).

In this research, the analytical modelling method employs mathematical equations to build models that represent and predict the network's behaviour and performance, thus providing a theoretical framework to derive performance metrics and evaluate the impact of performance metrics under different traffic loads. This modelling methodology includes network topology models, which represent the design layout of the network and performance metrics, which evaluate network performance. Moreover, the system simulation method includes creating a software-based network model to critically assess its behaviour under different traffic conditions, thus offering a flexible strategy to test multiple complex scenarios that might be challenging for real-time deployment. This methodology includes adopting various system simulators and emulators and validating the network model to mimic the functioning of real-world networks.

Furthermore, the modified DSRM methodology comprises six phases: problem identification and motivation, solution objective, design-development-refinement, demonstration and implementation, implementation and evaluation, communication and contribution. Each phase comprises the list of tasks attempted to accomplish particular objectives. The entry point of the proposed research work is problem-centred initiation.

The various six phases of the adopted DSRM are discussed below. The analytical modelling and system simulation methodologies are also incorporated in Phases 3, 4 and 5.

- Phase 1: Problem identification & motivation.

  In this phase, we have conducted an extensive literature review of TI, SDN, FC, NS, and MA techniques that are needed to maintain haptic communication between master and slave environments. We have also explored existing methodologies and techniques to minimise RTT, which is crucial to avoiding motion sickness (lag). Hence, this exploration has helped to comprehend the research status, challenges, and future directions. Besides, we have also identified the research loopholes (gaps) and formulated research questions (RQ 1, RQ 2 and RQ 3) that need to be addressed in the TI communication infrastructure. The TI-based applications have motivated us to research the TI system and investigate and propose a few methods/strategies to reduce RTT to help avoid a system lag.

- Phase 2: Objective of the solution.

  We have conducted an in-depth literature review to assess the existing research related to TI and how it possibly meets the challenge of 1ms RTT. We have also explored potential methods/strategies to minimise the haptic communication RTT.

  One of the methods to minimise the haptic communication delay is to reduce the delay in the network domain by pushing all the computing parts to the cloud (FC) to avoid processing and storage limitations. Other techniques can also reduce delays in the network domain, such as incorporating the SDN, NS, and multiple access techniques.

- Phase 3: Design, development & refinement.

  First and foremost, we have composed an extensive survey on TI's design, issues, challenges and applications to acquire in-depth comprehension. The mind map is created to brainstorm ideas on the survey's design, applications and challenges. A comparative analysis is done using existing contributions in the fields of IoT and TI. Moreover, we have designed and analytically modelled a TI communication framework to meet the demands of the different applications of TI and proposed an efficient traffic flow framework by adopting the multilevel structure of cloud units through a credible simulation environment. This framework incorporates SDN and FC approaches, avoiding unnecessary processing and waiting times at each cloud unit. It depends upon the priorities and requirements of the TI, thus resulting in less network congestion, better traffic/task/request flow and RTT minimisation.

On similar grounds, SDN and NS have been employed in the TI communication framework to cater to a diverse requirement of applications/services. Efficient network slicing algorithms are proposed and tested to meet the fluctuating demands of the end-user with real-world traffic scenarios. Finally, we have adopted an MA technique (NOMA) to improve the spectral efficiency and minimise RTT. Our proposed system model is represented mathematically and tested with a downlink power-domain communication scenario, signal-to-interference ratio analysis, sum rate analysis and fair power allocation analysis.

- Phase 4: Demonstration & implementation.

  After developing the TI framework or model with proposed methods/techniques/ algorithms, we have demonstrated and implemented our proposed methods in potential simulation packages such as iFogSim, Mininet and MATLAB to minimise RTT and address our research questions (RQ 1, RQ 2 and RQ 3), thus meeting the objectives stated in Phase 2. For each contribution of our research, we have identified and recorded various key performance indicators (KPIs) influencing the proposed TI system.

- Phase 5: Implementation & evaluation.

  In this phase, we have conducted simulations in various environments to evaluate KPIs, perform result analysis, and validate our proposed system model. Upon performance evaluation of the recorded simulation results for each research contribution, the KPIs have shown the efficacy and effectiveness of our proposed approaches. Thus, it provides insight into the potential areas for future directions.

- Phase 6: Communication & contribution.

  The communication section is a crucial part of the research. We must keep in touch with researchers in academia, labs and industry. Our research is summarised and communicated in journal/conference publications to authenticate and validate our work globally.

  This phase will also include documenting and collating all main contributions that are addressed by research questions (RQ 1, RQ 2 and RQ 3) in the form of a PhD thesis. During PhD research studies, we plan and emphasise publishing at least

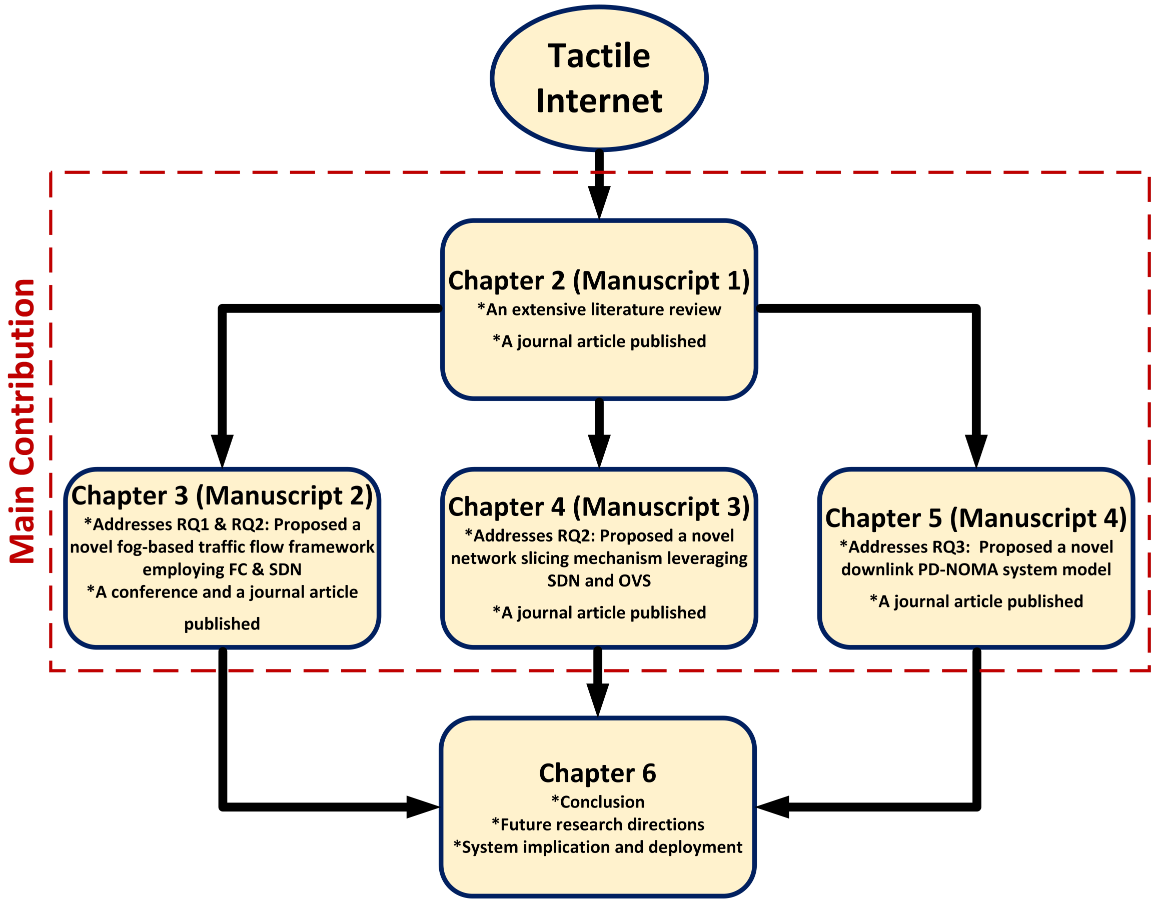


Fig. 1.9 The schematic representation of this thesis structure.

four scholarly articles, including conferences and journals, in highly ranked peer-reviewed publications such as IEEE, ScienceDirect, and MDPI. The research artefacts resulting from this study included a designed framework or mathematical model and evaluated various KPIs. These artefacts provide a valid and effective measure for use in development practice at the organisational and project level in assessing and evaluating the effectiveness and performance of the system.

1.6 Structure of the thesis

The schematic representation of this PhD thesis structure is shown in Figure 1.9. This thesis follows the manuscript-based thesis format and is organised as follows. Chapter 2 provides a comprehensive survey of TI, mainly focussing on design aspects, potential applications, related issues and challenges, thus identifying research gaps. Chapters 3, 4 and 5 present the manuscripts specifically addressing RQ 1, RQ 2 and RQ 3. These chapters comprehensively discuss the proposed system model, evaluate the network model and simulation results, and summarise the work. Chapter 6 provides a conclusion summarising the contributions, research findings, future research direction, and system implication and deployment. Finally, Appendix A lists the abbreviations used in this thesis and their explanation.

Chapter 2

A Survey of the Tactile Internet: Design Issues and Challenges, Applications, and Future Directions (Manuscript 1)

2.1 Manuscript 1 prelude

The tactile Internet (TI) is an emerging area of research involving 5G and beyond (B5G) communications to enable real-time interaction of haptic data over the Internet between tactile ends, with audio-visual data as feedback. This emerging TI technology is viewed as the next evolutionary step for the Internet of Things (IoT) and is expected to bring about a massive change in Healthcare 4.0, Industry 4.0 and autonomous vehicles to resolve complicated issues in modern society. This vision of TI makes a dream into a reality.

Manuscript 1, entitled "A Survey of the Tactile Internet: Design Issues and Challenges, Applications, and Future Directions[1]" [19], aims to provide a comprehensive survey of TI, focussing on design architecture, key application areas, potential enabling technologies, current issues, and challenges to realise it. To illustrate the novelty of our work, we present a brainstorming mind map of all the topics discussed in this manuscript. We emphasise the design aspects of the TI and discuss the three main sections of the TI, i.e., master, network, and slave sections, with a focus on the proposed application-centric design architecture. With the help of the proposed illustrative diagrams of use cases, we discuss and tabulate the possible applications of the TI with a 5G framework and its requirements. Then, we extensively address the currently identified issues and challenges with promising potential enablers of the TI.

Moreover, a comprehensive review focusing on related articles on enabling technologies is explored, including Fifth Generation (5G), Software-Defined Networking (SDN), Network Slicing (NS), Network Function Virtualisation (NFV), Cloud/Edge/Fog Computing (CC/EC/FC), Multiple Access (MA), Network Coding (Net. Cod.) and Machine Learn-

[1]V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A survey of the tactile Internet: Design issues and challenges, applications, and future directions," *MDPI Electronics*, vol. 10, no. 17, p. 2171, 2021.

ing/Artificial Intelligence (ML/AI). Finally, we conclude the survey with several research issues that are open for further investigation. Thus, the survey provides insights into the TI that can help network researchers and engineers contribute further towards developing the next-generation Internet.

In addition, this manuscript mainly covers the systematic literature review of the TI along with the description of design, application and related challenges/issues.

2.2 Introduction

The tactile Internet (TI) is an innovation that facilitates interaction between human beings (possibly over a distance) with visual presence and haptic feedback [1]. In June 2015, the Technical Activities Board Future Directions Committee launched the most current activity for the Institute of Electrical and Electronics Engineers (IEEE) Digital Senses Initiative (DSI). The DSI is devoted to propelling advancements that catch and recreate human activities such as sense, touch, hearing, taste, and sight from the external world. These human activities are received and reacted to by machines and humans differently.

In a meeting held in 2013, a future financial "golden age" of technological association was considered during the 2020s, when robots might progressively carry out a significant amount of regular day-to-day work [20]. For example, the alluring chance of monitoring is imagined by sitting at one site and reacting remotely through the Internet at another site far from the original one. However, the conception of the Internet as TI was broadly announced by Fettweis in mid-2014 [21, 22], where he mentioned that the TI has the potential for a plethora of use cases that influence our lives and the world economy.

One of the requirements of real-time Cyber-Physical Systems (CPSs) is to have a round-trip time (RTT) of below 1 ms. A significant example is the CPS model for the smart grid and the requirement of rapid reaction time in situations with power-network failures. Recent cellular and wireless local area networks (WLANs) fail to achieve this objective by a considerable order of magnitude. By achieving an RTT of 1 ms, the user can explore the new era of the TI by changing the experience of current mobile broadband applications. Along with voice and data communications, current fourth-generation (4G) systems empower continuous access to massive data for applications, such as machine-to-machine (M2M) or machine-type communication (MTC). The next major step is to associate multiple machines together and control them from a far distance. Therefore, for control communication, this will create another approach to guide and control factors of our environment and its conditions [21]. With the help of carrier-grade robustness and accessibility, the demand for an RTT of 1 ms will empower the TI for guiding and controlling real and virtual objects [22].

However, the TI comes with a warning that it should be able to tell the difference between humans and machines. This should be applicable where there is a large requirement for machines and less interference from humans. Machines should complement humans instead of acting as a substitute for them [23].

Ultimately, features such as delivering up to 1000 times more capacity, maximum data rates at 10 Gbps, and interconnections of at least 100 billion end-user devices are expected to be realised for the 5G communication system, compared with 4G. The primary purpose of 5G systems is to construct and handle novel machine-driven applications that are not tended to by cellular networks. Compared to the past four generations, the vision of 5G will be considerably flexible in integrating with cellular and wireless-fidelity (Wi-Fi) technologies and their standards. Furthermore, the crucial part of the envisioned 5G communication systems is transforming the distributed core infrastructure into a device-based infrastructure and employing Artificial Intelligence (AI) at the end-user device location or on user equipment (UE) with assisted mobility and device-to-device (D2D) communication.

2.2.1 Vision of the TI

For a wide range of application areas such as Industry 4.0, smart e-learning and education, and Healthcare 4.0, TI envisions the interaction of human-to-machine (H2M) and relies upon it to add another paradigm to see its influence on society. The envisioned TI desires to meet the stringent design requirements by empowering a suitable communication framework. First, it needs to give an extremely low round-trip time (delay) of 1 ms or below and possible ultra-high reliability to avoid motion sickness. It likewise needs to guarantee data security and high availability of a framework without compromising the ultra-low latency requirement. These crucial design goals of TI can be achieved by placing tactile devices close to each other, which is often based on distributed or decentralised architecture dependent on recent technological advancements such as Edge/Fog Computing, Software-Defined Networking (SDN), and Network Coding, Network Function Virtualisation (NFV), and suitable multiple-access (MA) techniques. Moreover, adaptable techniques at all protocol layers are expected to minimise the RTT from sensors to actuators.

Significantly, the TI will establish vital prerequisites for next-generation access networks as far as latency, availability, and reliability are concerned. Wired access networks are ultimately fulfilling these prerequisites already, and wireless access networks are not yet

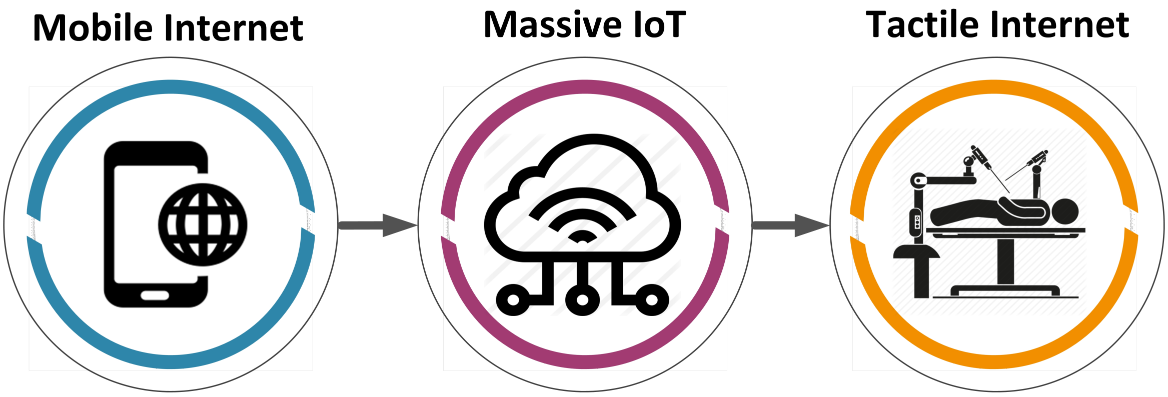


Fig. 2.1 The evolution of TI from mobile Internet.

proposed to fulfil these prerequisites. As per the International Telecommunication Union—Telecommunication Standardisation Sector (ITU-T) Technology Watch Report on the TI [1], research in these areas will need to speed up substantially, introducing new thoughts and designs to promote access network characteristic redundancy and a decent variety to resolve RTT issues and reliability constraints of TI applications.

As TI is going to serve critical applications of society, it should have competencies such as ultra-reliability [24], very low latencies, and be able to connect to numerous devices for communicating with each other simultaneously and autonomously. Interacting with ongoing and conventional wired Internet, mobile Internet and the IoT should also be feasible, thus forming a network of entirely new possibilities and opportunities.

2.2.2 Evolution of the TI

With regard to TI, Figure 2.1 depicts the progressive evolution of TI, as indicated by an ongoing ITU-T Technology Watch Report [1]. According to the definition given by the ITU-T Technology Watch Report, the Internet that comprises ultra-low latency, ultra-high availability, ultra-high reliability, ultra-responsive, fast reaction times, scalability and security is referred to as TI. On the other hand, the interconnection of physical devices with the Internet with requirements such as power efficiency, surveillance and data security, reliability, good energy management, relatively low throughput and effective low latency is referred to as the Internet of Things (IoT). In contrast, the Internet, which transmits voice and data, video-streaming content, web browsing, and telephony, is referred to as mobile Internet.

Mobile communication contribution is crucial to the modern telecommunication infrastructure. In the 1990s, various versatile computers such as laptops, pagers, cell phones and personal digital assistant (PDA) phones were announced. Generations of mobile communication, i.e., 2G, 3G, 4G and 5G and their wireless communication services, such as Universal Mobile Telecommunications Service (UMTS), General Packet Radio Services (GPRS), and Enhanced Data Rates for GSM (Global System For Mobile) Evolution (EDGE), have profited businesses as well as delivered comfort to the end user. These services have effectively linked tonnes of end users to physical devices. The 5G framework permits human-to-human (H2H) communication to trade multimedia content such as audio and video information. The primary centre of attention for the mobile Internet is facilitating the integration between end-user devices and machines, thus implied as the IoT.

After the evolution of the mobile Internet, the IoT is a concept that involves the communication of two or more devices that integrate the Internet and the cloud. Applications related to IoT have been investigated more as they offer considerable chances to ease life. They depend on H2M and D2D communication [25]. However, the IoT has a few demerits, such as low data rate, decent latency, surveillance and security, low compatibility, and privacy. Hence, the shortcomings of the IoT could be overcome by TI, which is the progressive evolution of the IoT that manages networks in a real-time environment.

The TI currently exists at the conceptual stage, and a lot of research is still needed to revolutionise it. The idea of the TI is becoming broader, where a human can interact with another human, allowing us to transmit a sense of touch and actuation over a long distance in real-time with or without audio/visual feedback. These requirements in TI permit tactile end devices to sense and actuate in real-time transmissions over distance/remote communications. It will reform all aspects of society, culture, and use cases, such as those related to e-healthcare, industry, and much more.

Thus, information-delivery networks can be fundamentally transformed into skillset-delivery networks. Low-power massive connections, ultra-responsive, ultra-reliable and ultra-fast reaction times of the TI will enumerate another H2M interconnection paradigm by empowering haptic features. Conversely, 5G and beyond should be able to manage the extraordinary increment of mobile data traffic and the immense proportion of information from intelligent devices that will energise IoT applications. Each evolution step states some features that are required for an accessible mode of communication.

2.2.3 Recent advances in the TI

Recently, because of the emergence of the TI, a remarkable amount of research has been substantially carried out. These studies are generally inspired by several applications such as self-driving vehicles, augmented reality (AR), virtual reality (VR), Industry 4.0, Healthcare

4.0, immersive virtual reality (IVR), teleoperation systems, telementoring, remote driving, and tactile robots. The existing system models and architectures are still theoretically and technically inadequate for realising the arising TI applications. Due to the non-traditional requirements and demands for TI in future B5G networks, factors such as Fog/Edge clouds, sensors and actuators at the master and slave sides, RTT, reliability, system models, and architecture are taken into consideration.

In addition, the speed of light is 3×10^8 m/s, i.e., light travels 300 km in 1 ms. To meet the requirements of the TI's RTT, the control server and tactile ends can be at a maximum of 150 km. Thus, the limiting parameter, i.e., the speed of light, also affects the speed of human interaction with humans/machines.

Furthermore, considering the stringent demands and requirements of TI involving ultra-low latency and ultra-high reliability features such as data integrity, security, encryption, and network availability need to be established with fewer trade-offs between these. However, the existing centralised infrastructure is not adequate to meet these features. Therefore, more distributed infrastructure based on Fog/Edge computing should be adequately examined to bring TI applications nearer to end users [26].

Fog computing and Edge computing are emerging paradigms designed to address the limitations of cloud computing by bringing computational resources closer to end devices. Fog computing extends cloud services to the network edge, providing distributed computing, storage, and networking capabilities to enhance efficiency and reduce latency. It enables real-time data processing by leveraging intermediate nodes, such as gateways and routers, before transmitting data to the cloud. A closely related concept, Edge computing processes data directly at the source or near-end devices, minimising dependency on centralised cloud infrastructure. This paradigm is crucial for ultra-low latency applications such as autonomous systems, industrial automation, and the TI. While Fog computing enhances distributed processing, security, and scalability, additional network hops may pose challenges for time-critical TI applications. Both approaches optimise network bandwidth usage, enhance security, and improve scalability. By decentralising computation, Fog and Edge computing play a significant role in enabling next-generation technologies, including 5G, SDN and TI applications.

Additionally, it is crucial to refresh next-generation wireless access networks by exploring novel resource allocation in network slicing, feedback mechanisms of the sensors/actuators, interference management and multiple-access techniques to satisfy the potential demands and requirements of TI applications [1].

Moreover, considerable research has been carried out on haptic devices such as tactile gloves, etc., and kinematic devices such as capacitive-based, magnetic-based, piezoelectric and image sensors, and tactile support engines with AI compatibility for information integration and decision boxes [27].

The demand for interdisciplinary research is crucial in planning and designing intelligent tactile sensors considering fabrication techniques/technologies, the material of sensors/devices, intelligent signal processing and machine-learning algorithms to analyse the behaviour of the sensing device and process complex data coming to the sensors. Several factors are necessary to realise the TI, such as security, reliability, availability, and latency in wireless transmission of tactile sensing data, thus giving the experience of remote real touch feeling and control. A few standardised works are forecasted to collaborate between the TI's use cases and improve current wireless sensor networks.

Considering ongoing standardisation, a new IEEE standard family has been defined for the TI, i.e., IEEE P1918.X [28, 29]. IEEE P1918.X describes the architecture technology and assumptions in the TI, whereas IEEE P1918.X.1, IEEE P1918.X.2 and IEEE P1918.X.3 are focussed on codecs, AI and MAC for the TI, respectively. In addition, a considerable amount of work has been done to develop a working group for low-latency Industrial IoT (IIoT), such as intelligent transportation systems, Industry 4.0, and Health 4.0 [30].

2.2.4 Research motivation

One of the reasons for carrying out the research is TI's emerging research area, where there is a need to improve the round-trip time (RTT) and reliability of the system. This improvement will ensure and enable the smooth functioning of the system without cyber or motion sickness (lag). The crucial challenge to achieve a tolerable RTT of 1 ms is essential to facilitate the services and applications related to the TI. Still, there are many challenging solutions to reduce the RTT over the network. Therefore, we have been motivated to adopt TI technology to incorporate technological advancements such as SDN, NFV, Network Coding, physical MAC-layer protocols and cloud networking technologies, which promise to serve the requirements of the TI.

2.2.5 Main contribution

Although quite a few survey papers have talked about the overview of the TI, they are not mainly focussed on the design architecture, applications, and current issues. Thus, they do not provide illustrative diagrams for the reader to comprehend the gist of the research altogether. Here, we briefly highlight the following main contributions of this survey.

- We emphasise the TI design aspects with the proposed application-centric design architecture, where the main sections of the architecture, i.e., master, network, and slave sections, with their functions, are explored.
- We identify the key application areas and provide discussions with proposed illustrative diagrams of use cases. Moreover, the multiple application scenarios with required TI competencies and their performance metrics are tabulated.
- We provide an in-depth discussion of the identified current issues and challenges with potential enablers of the TI. In addition, some possible future research directions are provided to give researchers insight.
- We present a comprehensive review focussing on the related articles on enabling technologies such as 5G/B5G, SDN, NFV, NS, CC/EC/FC, MA, Network Coding and ML/AI to realise the TI. Here, the contribution of the related articles is summarised according to enabling technologies.

2.2.6 Structure of the manuscript

Figure 2.2 shows a graphical overview of all the topics covered for design, applications, and challenges related to the TI in the form of a mind-map. The remainder of this manuscript is structured as follows. The design aspects of the TI are discussed in Section 2.3. The applications of the TI are elaborated in Section 2.4. The current issues and challenges in realising the TI are discussed in Section 2.5. Section 2.6 thoroughly reviews related articles, and a summary of research challenges is presented in Section 2.7. Finally, Section 2.8 concludes the manuscript along with future research directions.

2.3 Design aspects of the TI

TI has some features that are not apparent on the traditional Internet. The traditional Internet facilitates audio and visual communication, whereas the TI facilitates transmitting a sense of touch and actuation along with haptic feedback.

TI depends on haptic and non-haptic control to communicate between end-to-end operations. Figure 2.3 is the pictorial representation of the architectural design of the TI, which consists of three sections, i.e., master, network, and slave sections. With haptic command and haptic feedback signals, the master and slave sections communicate via the network section. There is a two-way exchange of information between the master and slave sections.

2.3.1 Master section

The master section is located at the front end of the TI system. It comprises a human operator with a tactile human–system interface (HSI). The input a human gives is then converted into tactile input by HSI. HSI is a master machine or robot which acts as a haptic device. A user is permitted to touch, feel, and manipulate in the real world. Through the command signals, the operation of the slave is controlled by the master section. It ought to be noticed that numerous operators can cooperatively curb the activities of the slave section in most of the applications. The audio and visual feedback features are essential for haptic and non-haptic control of the slave section. Thus, it is crucial to increase perceptual performance as the human brain coordinates typically diverse with tangible approaches [31].

There are some haptic devices on the market that are accessible to merchants, such as Phantom Premium [32] and Geomagic [33]. These are meant for structuring a linkage-based framework comprising a robotic arm linked to a stylus. The robotic arm follows the movement of the stylus and is fit for applying force on its tip. To genuinely understand the vision of the TI, further advancements in haptic devices are required, especially in expanding the Degree of Freedom (DoF) to fulfil the needs of future applications. These haptic devices need to be interfaced with the network to communicate with the core network.

2.3.2 Slave section

The slave section comprises a teleoperator (slave robot) and is legitimately constrained by the master section through numerous coding techniques and haptic input command (instruction) signals. The teleoperator performs multiple actions as instructed by the master section in the remote environment. In addition, the slave section does not have previous information about the surroundings. Subsequently, a global control loop is formed when communication is initiated between the master and slave sections through feedback and command signals. Here, the Da Vinci telesurgical system [34] example involves robotic arms being manipulated by a surgeon sitting at one console (master section) with video monitoring and a joystick controller, which is approved by the United States Food and Drug Administration (FDA).

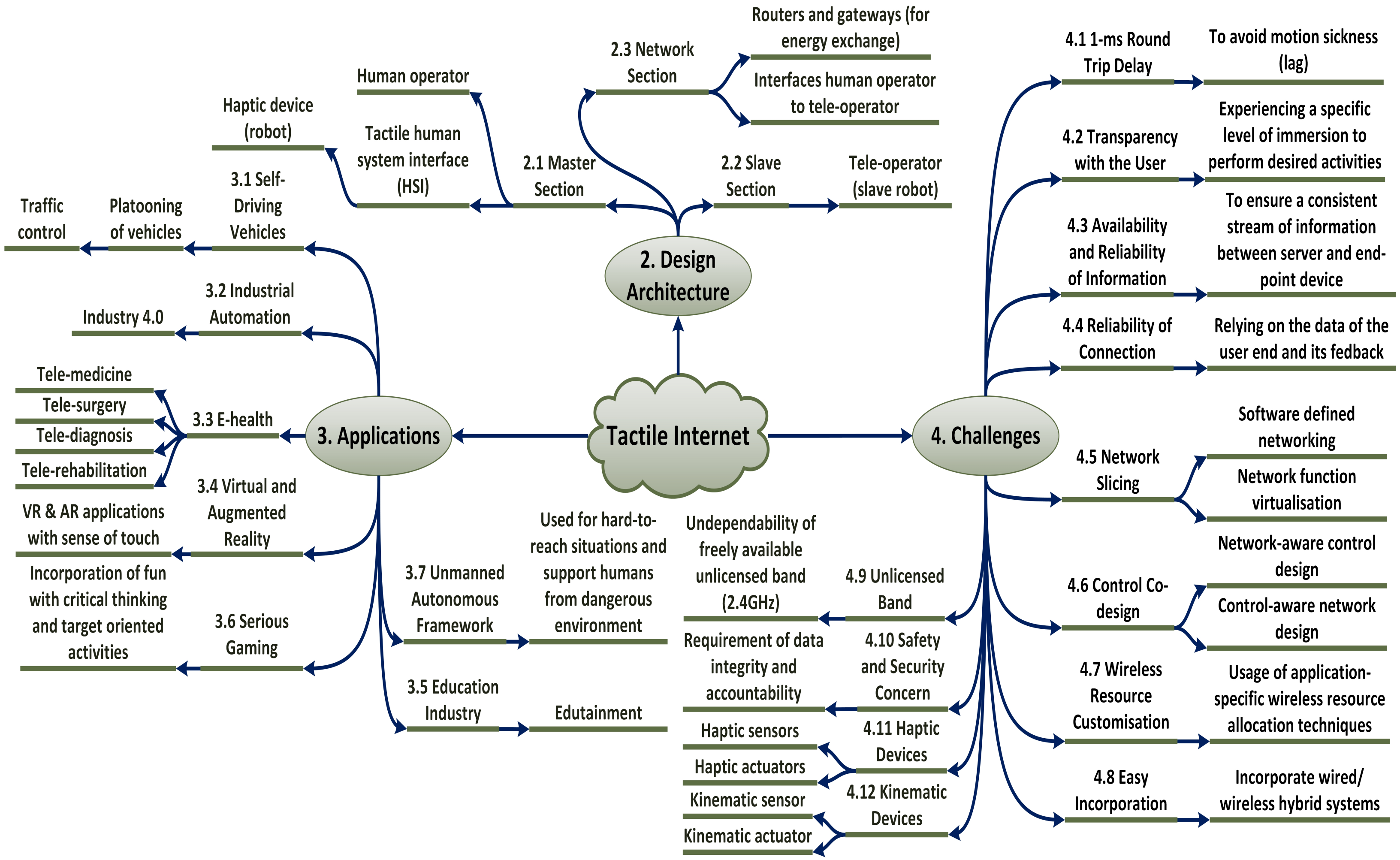


Fig. 2.2 A mind-map of the topics related to the TI discussed in this research.

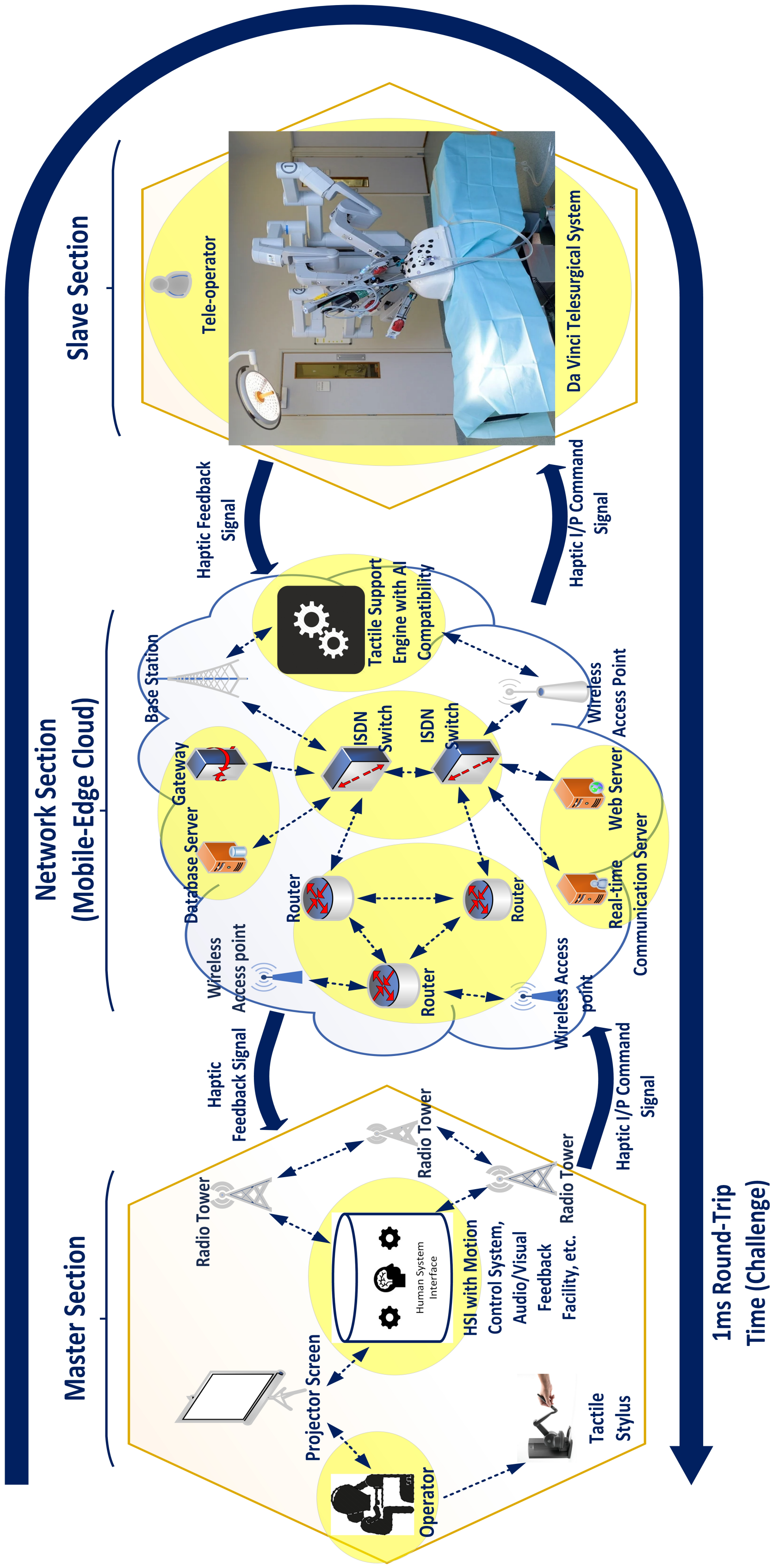


Fig. 2.3 The design architecture for the end-to-end da Vinci telesurgical system.

2.3.3 Network section

This section is typically situated in the middle of the master and slave sections and acts as a medium for two-way communication, thus interfacing the operator with the teleoperator (remote) environment.

It comprises routers and gateways, which facilitate a medium for two-way communication. The haptic input command signals are sent from the master section. Thus, the input signals pass through routers, switches, gateways, base stations, access points, database servers and tactile support engines equipped with AI and reach the slave section. AI is also facilitated by a tactile support engine and plays an essential job in processing latency-critical applications.

For real-time haptic communication, the network should have ultra-reliable and ultra-responsive connectivity, which reduces RTTs. To understand the requirement and vision of the TI, the foundation of 5G connectivity is required in the architecture of the TI system, which is composed of eNBs nodes (radio access network) and a core network.

Even though many kinds of research have been carried out on the 5G framework, there is no consistent concurrence on its architectural design. However, the consensus implies that the 5G framework must be planned in an adaptable environment. The end goal of having one network based on typical physical infrastructure is productively shared among various vertical applications. Here, the concept of NFV is taken into consideration, where the network is sliced according to the particular application's requirements.

Subsequently, interest in NFV [35] and SDN [11] has risen to shape an envisioned architectural design. NFV facilitates the detachment of network functions from the hardware system; hence, network slicing programming can be done and deployed in any standard distributed computing framework. On the contrary, SDN facilitates an architectural system wherein the control plane and data plane are decoupled, thus empowering the network's software coding through its controllers.

2.4 Applications of the TI

In various scenarios, communication will be enhanced through the TI, leading to increasingly reasonable social interaction. A portion of the imagined 5G-enabled tactile applications requires competencies such as RTT (1 ms or below), ultra-high availability and ultra-high reliability, whose reliability should be significantly higher orders than in the current radio access network. It is also seen that existing wireless or cellular networks do not yield anything to accomplish an ultra-low latency of 1 ms and ultra-high reliability of 99.999%,

Table 2.1 Possible applications of the TI with 5G framework and their requirements.

Application/Scenario	**TI Competencies**	**Performance Requirements**
• Self-driving Vehicles • Remote Driving • Industrial Automation • Virtual and Augmented Reality (AR and VR) • Unmanned Ariel Vehicle (UAVs) • Smart Grids • E-learning • Serious Gaming	• Ultra-high Reliability • Low Latency	• RTT ($>$1 ms) • Reliability (99.999%) • Date Rate ($<$250 Mbps)
• eHealth (Telesurgery) • Human-to-Machine Interaction • Immersive Virtual Reality (IVR) • Telementoring	• Ultra-low Latency	• RTT ($\approx$2 ms)
• Cloud-based Telemedicine • Industry 4.0 or Industrial Internet	• Ultra-high Reliability • Ultra-high Availability	• RTT ($<$1 ms) • Reliability ($\leq$ 99.999%) • Availability ($\approx$ 100%)
• Artificial Satellite Communication for Emergency	• Ultra-high Availability	• Coverage and Excellent Service • Frequency (1 GHz to 50 GHz)

which are essential for TI applications. In this way, grasping all the possible emerging tactile applications at the early stage of 5G development is troublesome. It could play a crucial role in eradicating the socio-economic boundaries of society [1]. Once the challenges are fulfilled, numerous applications can be explored and achieved with 5G mobile communications. Hence, the taxonomy of possible applications of TI with the 5G framework and their requirements are presented in Table 2.1. However, TI can support various applications as some of the ideas are shared below.

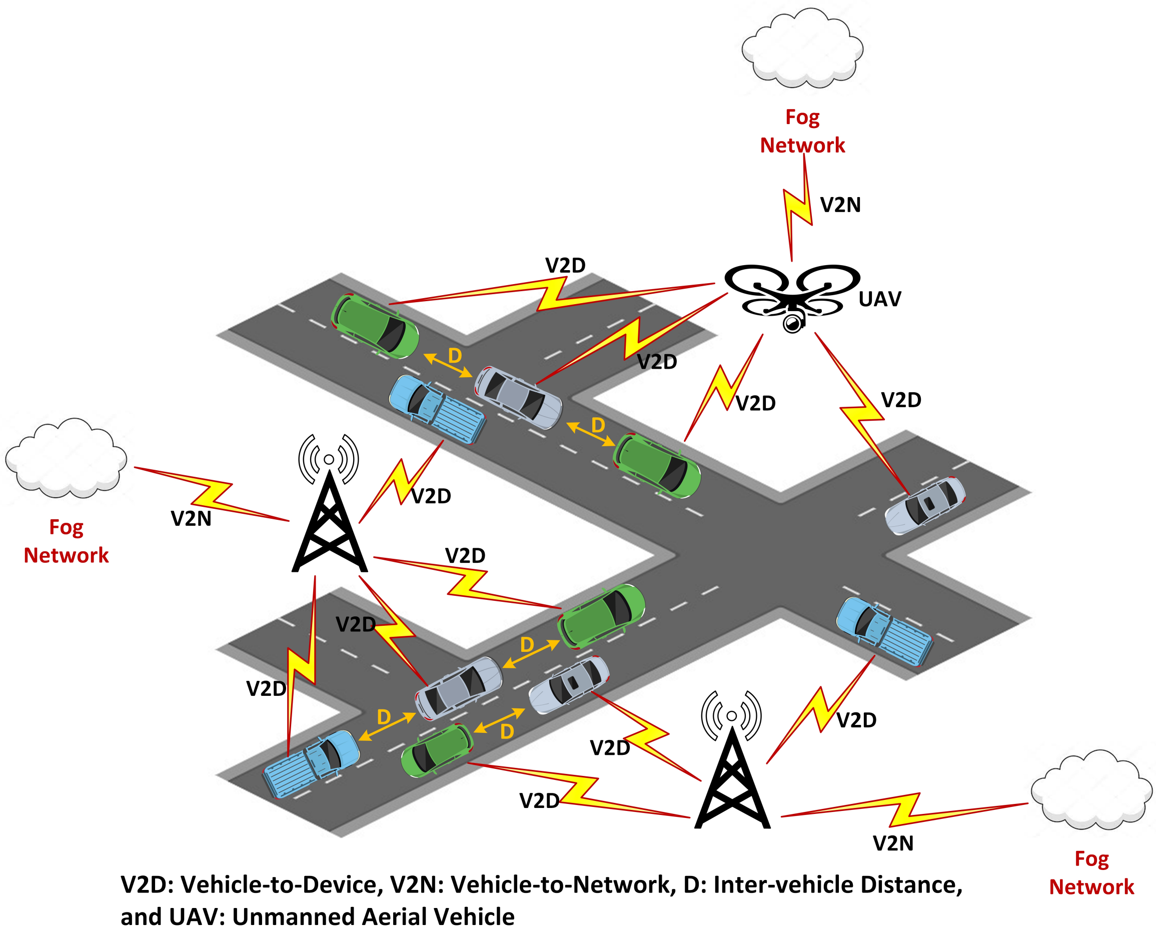


Fig. 2.4 The platooning of self-driven vehicles maintaining a fixed distance 'D' between them.

2.4.1 Self-driving vehicles

Self-driven vehicles are broadly classified as designed and constructed, along with advancements in technological fields. Most of them can be partially driven or fully self-driven. The highlights of self-driven vehicles are that they will deliver a new driving experience. By incorporating the unique features of autonomous vehicles, the consumer will take some time to have hands-on and manage those features. The highlights of such vehicles can be electronic braking systems with emergency braking, automatic parking, and versatile cruise control, where a driver can put his/her vehicle on autopilot mode. Most of these highlighted features are present in selected applications of the controlled environment. The futuristic thought of a fully self-driven vehicle is no longer just a thought. Different organisations are already working on the driving features that enable a vehicle to drive itself.

The benefits of a self-driven vehicle are nominally reduced accident rates, self-parking, and less time consumption. Other highlights of the autonomous vehicle are having access to entertainment and digital media and more relaxation time, which will significantly impact the financial budget. Nonetheless, implementation of all the features of the autonomous vehicle comes at the expense of satisfying a few challenges, which further require enormous improvements in (1) communication technology, (2) infrastructure for the sensor domain, and (3) communication between the vehicles to fetch significant information [36].

To encourage completely self-driving competencies, the development of the Intelligent Transport System (ITS) [37] has been enabled because of technological advancement in 5G wireless communication to connected and communicating vehicles. Hence, wireless communication technologies will significantly impact Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication.

Considering enhanced security, there would be a manageable reduction in traffic jams and road accidents by adopting fully automated driving. In today's world, the time required to avoid any collision for the safety of the vehicle is 10 ms [38]. To serve the purpose of a bidirectional data exchange for automated vehicles, the RTT of communication should be in the order of milliseconds [39]. This can be acknowledged by the TI and its RTT (latency) of 1 ms. Figure 2.4 illustrates the platooning of self-driven vehicles by maintaining a fixed distance 'D' between them. All vehicles establish a connection to the nearest base station, thus reaching a fog network to execute the complex computation of the traffic.

2.4.2 Industrial automation

For the TI, a steadily developing application field is in automation for the industry. It not only boosts the overall production of infrastructure but also increases efficiency. One of the requirements of automation in industry is having RTT underneath 1 ms per sensor for industrial robots, where the sensitivity of control circuits matters the most when the controlling gadgets are moving quickly. Executing this entire base concept would require high-speed connectivity of millimetre-wave technology.

Due to the global requirements for intelligent robots, smart devices, sensors, etc., traditional industries are transforming into digital ones. There were been previously three generations in the Industrial Revolution, which resulted in profitable products in the industrial market [40].

A real-time empowered cyber-physical system creates a paradigm shift in the coming fourth industrial revolution (Industry 4.0). It comes with prime modifications in assembly, de-

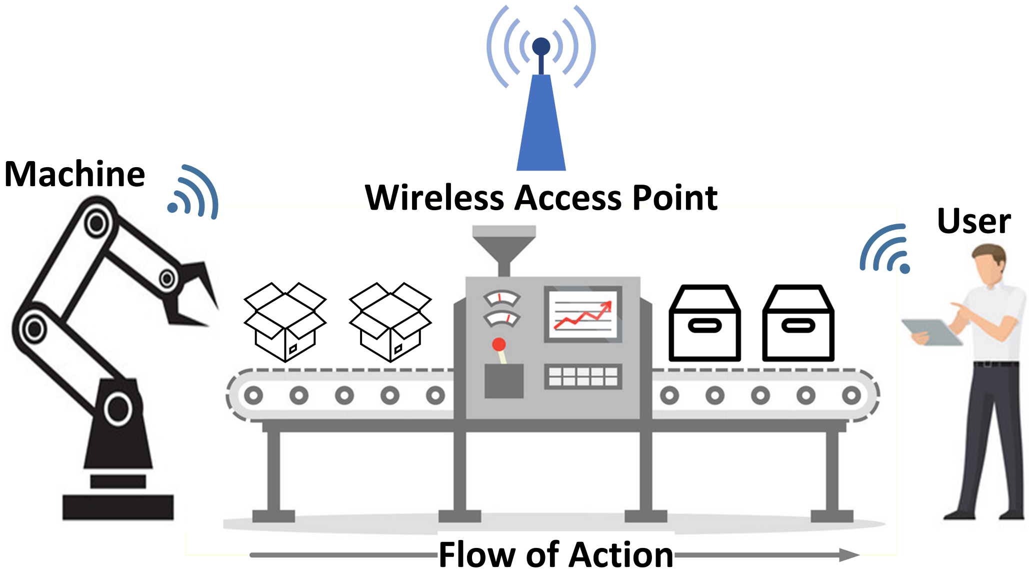


Fig. 2.5 An automated production assembly line where the user is giving instructions wirelessly on a tablet in industry.

sign, material usage and inventory networking, thus prompting adaptable and self-composed smart factories [41]. With increased features and adaptability, Industry 4.0 will empower individuals to individualise any item and transform it according to market needs. Figure 2.5 illustrates an automated production assembly line where the machine and user communicate through a wireless access point. The user is equipped with a tablet that controls the machines present on the industry premises.

2.4.3 eHealth

eHealth is a significant concern for human health and safety. A large portion of the TI sector is networked wirelessly to propel and provide proper services to humans. An efficient way could be found where information-based healthcare services can make superior use of information to upgrade patient security. Furthermore, competencies such as efficiency, robustness and reliability of the healthcare system's services in teleservices (rehabilitation, surgery, and diagnosis), TI infrastructure and IEEE wireless standards need improvement [42]. It is observed that any delay in a non-real-time system can prompt a mistake in patient behaviour in the case of remotely assisted (robotic) telepresence surgery [43].

To overcome distant geographic places and increase access to medical services, telemedicine uses the TI infrastructure with wireless communication. A time could come when medical experts could be present (virtually) anywhere and without paying attention to doctors' location

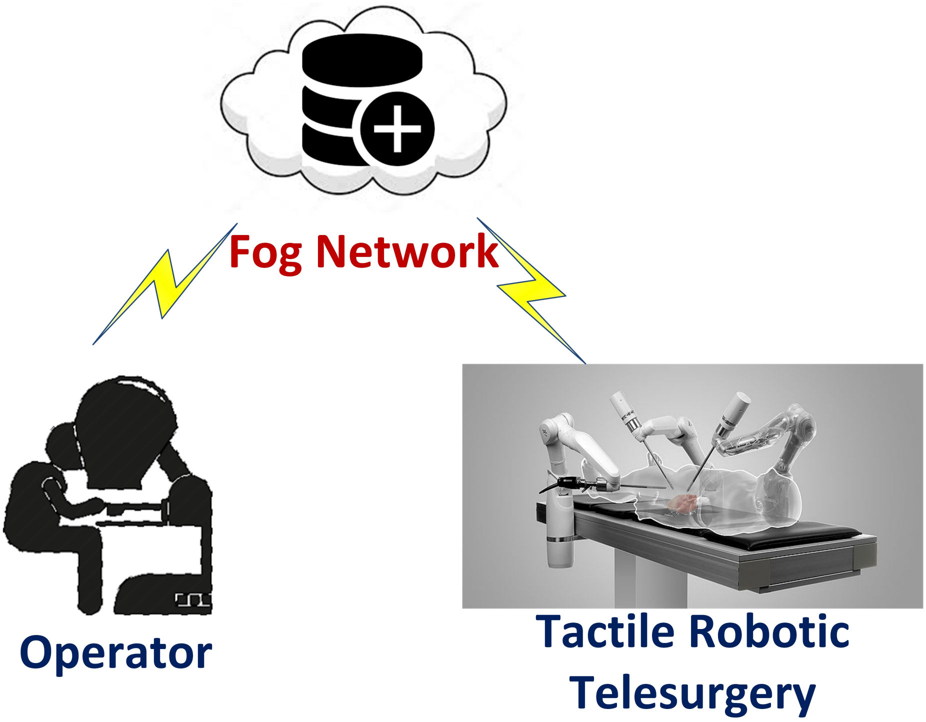


Fig. 2.6 Robotic telesurgery with a master as ‘operator’ and slave as ‘patient’.

using advanced tele-diagnostic tools. Hence and so forth, without a doctor physically present at the patient’s location, a telerobot can treat the patient operated by a doctor (remotely). Additionally, audio and visual information will be provided, along with haptic feedback. This will improve the possible future application of TI in health care, incorporating telediagnosis, telesurgery and telerehabilitation.

It is fundamental to see that the capabilities of machines are growing. Therefore, the TI should aid humans rather than find a replacement for them. This emergence of the TI should facilitate new options that enable humans by providing a hike in their career based on yields. Figure 2.6 illustrates robotic telesurgery with a master (operator) and slave (patient), which are physically present in remote locations. The patient’s image is extracted from MicroSurge – Institute of Robotics and Mechatronics [44]. The operator’s location will be equipped with a high-quality camera and controlling devices with haptic feedback. Therefore, the user (doctor) would feel physically present in the patient’s location.

2.4.4 Virtual and augmented reality

The availability of the TI can significantly aid by virtual reality (VR) and augmented reality (AR) applications. VR is a common, haptic virtual condition in which a few clients are

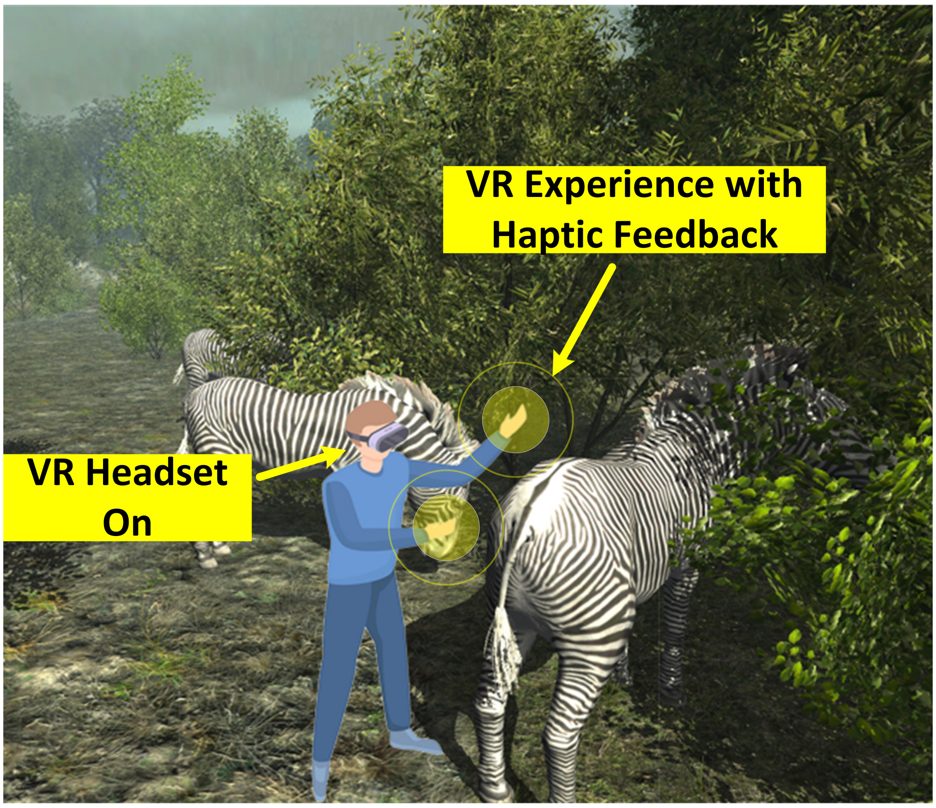


Fig. 2.7 The user experiencing virtual reality of touching the wild animal.

physically coupled through software tools to mutually/cooperatively execute an operation by virtually seeing physical objects, not only by audio and video but also with a sense of touch.

The actual real view of this world is envisioned by incorporating computer-produced objects in AR. The significant objective of future AR applications is the representation of dynamic substance and up-to-date data. For high-dependency conversations happening between devices, haptic feedback in VR is essential. Exceptionally, depending on the sense of touch, many applications rely on object size and precision. This must be acknowledged if the RTT between users and VR is a couple of milliseconds.

There is a need to improve numerous assistance systems requiring the augmentation of extra information into the client's field of view, e.g., maintenance and education. The objects in AR can be dragged from static to dynamic with the help of the TI. This offers a clear picture of real-time objects so that dangerous events such as car accidents can be distinguished and avoided.

Based on the 5G user experience along with AR and VR, there are six types of expected experiences: behavioural, emotional, social, cognitive, sensory, and basic. Each experience impacts the constituent elements of AR and VR related to the TI. The constituent elements

Table 2.2 Mapping of types of AR and VR experience with the constituent components that influence the 5G User Experience (UX).

Types of Experience	Influential Constituent Elements of AR and VR	References
Behavioural Experience	Interactivity, Context-aware, Portability and Wearability, Information Embedding	
Emotional Experience	Visual Enhancement (Input), Information Embedding and Context-aware	
Social Experience	Context-aware, Information Embedding and Visual Enhancement (Input)	
Cognitive Experience	Information Embedding, Content Creation and Visual Enhancement (Input)	[45–49]
Sensory Experience	Portability and Wearability, Geolocation and Visual Enhancement (Input and control)	
Basic Experience	User and Technical Guidance, Context-aware, Information Embedding, Visual Enhancement and Interactivity	

of AR and VR include Visual Enhancement (Input and Control), Context-aware (Initiative), Environment and Familiarity, Information Embedding, Content Creation, Interactivity, Geolocation, Portability and Wearability, and User and Technical Guidance. Table 2.2 will provide the constituent elements of the AR and VR experiences. Figure 2.7 shows that the user is experiencing the virtual reality of touching a wild animal with haptic feedback. The education industry will also benefit if VR-based study-related applications are incorporated into a learner's curriculum. The learner would feel the animals, creating a high sense of immersive tactile learning.

2.4.5 Education

The TI will empower enhanced learning dependent on the haptic overlay of the educator and learner. By consolidating TI with VR and AR, comprehension-based learning will boost learners' interests and surpass traditional learning methods. Subsequently, the learner will pursue an opportunity to adapt generally through exploring and discovering the content available. A plus point will be that there will be less interference from the educator and no educator pressure on the learner to study. With the incorporation of VR-based study-related applications in the learner's curriculum, the learner would be fully immersed in tactile learning by touching virtual things provided with haptic feedback, as illustrated in Figure 2.7.

2.4.6 Serious gaming

Serious gaming is an activity intended for a reason other than entertainment. For example, these activities incorporate fun with critical thinking challenges and target-oriented activities. In other terms, serious games also influence various application fields that are imperative to the general public, including health, education, and training.

What restricts its advancement is the RTT of the communication frameworks, as the recognisable RTTs legitimately impact the apparent authenticity of applications. Thus, the task could be made conceivable, e.g., designing a game application for a learner to simulate the dental care of a patient. The intended purpose will be to promote knowledge about oral care and how to interact with haptic devices.

2.4.7 Other TI applications

Another TI application includes the unmanned autonomous framework. For instance, industrial manufacturing products lined up for assembly line-based production require mobile robots to help deliver the assembly parts on demand. This process requires a highly reliable tactile communication framework to carry out the operation. These frameworks are now extensively used in hard-to-reach situations and provide support to humans in dangerous environments. For example, a drone with a high-precision remote control and with minimised RTT operated by a human can be significantly used as a TI application.

2.5 Current issues/challenges of realising TI

The two most challenging objectives for accomplishing communication features are to achieve ultra-low latency (1 ms or below) and ultra-high reliability (99.999%, i.e., a one-in-one-million chance of failure). There are numerous challenges related to the requirements for TI communication. These requirements focus on Quality of Experience (QoE) [50] to improve user experience. The following is a list of requirements needed for the TI. In addition, Table 2.3 provides a discussion on open challenges with descriptions and potential enablers/solutions.

Table 2.3 Open challenges with descriptions and potential enablers/solutions.

Challenges	Description	Potential Enabler and Solution with References
Ultra-low Latency (1 ms RTT)	To avoid noticeable motion sickness (lag).	• Emerging Passive Optical Local Area (POL) technology. [51] • Employing short frame transmission structure. [52] • Needing physical layer and control (MAC) layer techniques. • Relating to core network: SDN-based methods, NFV-based methods, Fog/Edge-based designs and Directory-based architectures. [53–60]
Transparency with the User	Action performed in real time must be the same in the remote scenario to have desired feedback to the end user. Experience a specific level of immersion to perform desired activities.	• Faultless simulation needed when dealing with physical interaction of virtual scenario emulating the remote scenario. [61] • To overcome "1 ms challenge" • Improved multiplexing model for video and audio together. • Enhanced perceptual performance in haptic telepresence system that improves transparency requirements.
Ultra-high Availability and Reliability of Information /Connection	To ensure a consistent stream of information between server and end user. Relying on the data of the end user and its feedback.	• Making use of concurrent connections with several links. [62] • Using multi-path communication to avoid a single point of failure. • Relies on dynamics of the Channel State Information (CSI). • Need for re-exploring the control layer, transport layer protocols and session layer. [63, 64]

Table 2.3 *Cont.*

Challenges	**Description**	**Potential Enabler and Solution with References**
Network Slicing	Trade-off between the three generic services of 5G, viz. eMBB, mMTC and URLLC.	• Non-orthogonal allocation resources among the heterogeneous services through slicing of radio access network (RAN). [65] • Dynamic network slicing for facilitating network in demand functionalities. [66, 67] • Integration of network-aware and control-aware for maximising Quality of Experience (QoE). • Need for dynamic control switching methodologies for TI telepresence applications. [68]
Wireless Resource Customisation	Usage of application-specific wireless resources allocation techniques.	• Need for joint uplink and downlink symmetric resource allocation with constant rate. [69] • Radio resource slicing, i.e., Flexible resource allocation techniques with network slicing [67] and adaptive flow management for traffic.
Easy Incorporation	Incorporation of wired and wireless system, i.e., hybrid systems.	• Incorporation of haptic communications in the 5G TI system. • Incorporation of Haptic codecs involving the kinaesthetic and tactile data.
Unlicensed Band	Undependability of freely available unlicensed band.	• To determine the trade-off between the RTT (delay) and energy efficiency in unlicensed band. • A traffic balancing scheme and scheduling needed. [70] • Collaborating licensed and unlicensed access keys in 5G unlicensed. [71]
Safety and Security Concern	Regarding the data integrity and accountability.	• Imbibe the trade-off for the required security level with the increment in RTT (delay). So distinct levels of security would be opted across broad range of tactile applications. [21]

Table 2.3 *Cont.*

Challenges	Description	Potential Enabler and Solution with References
Haptic Devices	Dealing with Haptic Sensor and Haptic Actuators.	• Optimising the spatial resolution, sensitivity, scan time and placements while selecting haptic sensors. • Designing a lightweight haptic display that gives an excellent feeling to user as similar as it would feel in the real-world context.
Kinematic Devices	Dealing with Kinematic Sensors and Kinematic Actuators	• With regard with post-processing time of reading data from sensors, algorithms are needed to be devised which will dynamically select the sensor and read the data. • Computational work is needed to push to the cloud for those with high orders of post-processing time. • Choosing physical actuators and driver circuits according to the use cases that help to minimise the RTT.

2.5.1 Transparency with the user

Reliable and susceptible communication is required when executing a task at a remote distance. Therefore, the user needs to experience a specific level of immersion to deliver the desired activities when communicating remotely. These desired activities must be user-friendly in real time and be effectively executed in the remote environment to create the appropriate acknowledgement to the user. Without this non-cognitive transparency, the efficiency and authenticity of this communication loop cannot be guaranteed [72][73][74]. When a measurable round-trip time delay appears, complete transparency becomes nearly impossible. Additionally, there is a need for correct simulation of physical interaction when the user tries to mirror the virtual environment with the remote environment.

Another challenge is the transparency of the system being affected by lossy haptic data compression and communication delay. Thus, the perceptual dead-band-based kinaesthetic data compression approach could be used to improve the quality of learning for haptic teleoperation in a TI environment [75].

2.5.2 Round-trip time of 1 ms

To have an option to guarantee a transparent experience, the most challenging factor is to achieve a round-trip time (RTT) of 1 ms [76, 21]. The motion sickness delay that occurs while performing any activity in one place, where another person perceives it in another, is called lag [77]. For haptic communication, various examinations demonstrate that lag will be observable if RTT is greater than 1 ms [78, 79]. If this lag is noticeable, it influences transparency.

One of the crucial factors affecting the smooth flow of any haptic communication in a network is RTT. However, RTT is significantly affected by queuing and processing delays at the intermediate nodes and packet transmission times. The speed of light is among the most constraining elements for RTT. Suppose the propagation time taken by the packet to reach a destination is more than 10 ms. Therefore, considering the 1 ms challenge, the propagation time taken by the packet should be less than 1 ms. With a 1 ms time delay constraint, it is workable for short-range communication.

With the availability of 4G mobile cellular networks, the RTT for short-range communication incurs a 15 ms delay, which cannot accomplish the necessity of 1 ms for a tactile response [78]. Furthermore, due to the evolution of 5G cellular networks, the RTT will incur a < 5 ms delay, which supports the necessity of haptic communication [79].

For haptic data communication, the packet size required should be less than 1500 bytes [74]. Taking all parameters, such as transmission and processing delays, into consideration, the RTT can be further minimised for smaller packet-size data [76]. This is a step that we take closer to achieving haptic data communication with the challenge of 1 ms.

In addition, the essential point of interest is executing the TI concept, which supports infrastructure or architecture with RTT below 1 ms. 5G technology is still being explored and is not available to consumers. To satisfy the conditions of 1 ms, the control server and end-point device should not be placed at distant places as the end-user information needs to be executed immediately, and acknowledgement must be sent to the origin. Therefore, to avoid any delays, the control servers should be profoundly reliable and process all the end-user information.

Subsequently, when encountering the design concern for 1 ms RTT, one-way physical-layer transmission is empowered. Thus, the one-way transmission must have a minimum packet size and around 100μs duration. To accomplish this necessity, every packet cannot surpass a 33μs packet duration [80]. The explanation for this is the need for additional structural delay that demands encoding, detection, and decoding at the transmitter and receiver ends. This restricts the packet size to less than 0.33 times the desired delay. Using the current LTE cell framework, the modulation used is not a feasible proposition, as the length of one OFDM symbol is nearly 70 μs long. A total round trip of the cellular physical layer is essential for the TI, possibly with 5G cellular networks.

2.5.3 Availability and reliability of information

It is expected that control servers must process and execute the end-user information frequently without any communication breakdown. Additionally, the communication path opted should ensure a reliable stream of information between the servers and end users. To ensure that there is no corruptive information, the processed data ought to be exemplified to guarantee that any threats or vulnerabilities will not have any access to the information [22].

2.5.4 Reliability of connection

Communication reliability relies on the data received from the end user and feedback to the remote place. If the data received is incorrect, then the same is carried forward to the remote place. The TCP and UDP protocols have their drawbacks, which are not reliable for this type of communication. Sadly, this protocol algorithm leads to undesired propagation delay, and it is not acceptable to have a communication based on real-time information [81]. Therefore, there is a need to design a highly reliable infrastructure that will support audio and video data, with a packet loss probability of a maximum of 0.001% [22]. Therefore, modifications are required for existing protocols [81], or a new protocol should be created that goes with the requirements of 5G [26].

2.5.5 Network slicing

A network slicing technique is essential in accomplishing the framework for supporting the prerequisites of various applications. This technique arose with 5G mobile systems, but other application systems can also adopt this technique. Slice-specific customisation and performance isolation of resources could be done to effectively empower various applications and guarantee the effective usage of resources. Using these techniques, it is necessary to slice the network with core network resources and radio access.

Regardless of a lot of research on network slicing, various issues related to slicing require further analysis, including productive virtualisation of the network framework, applying services to network slice, the disintegration of network function, end-to-end network orches-

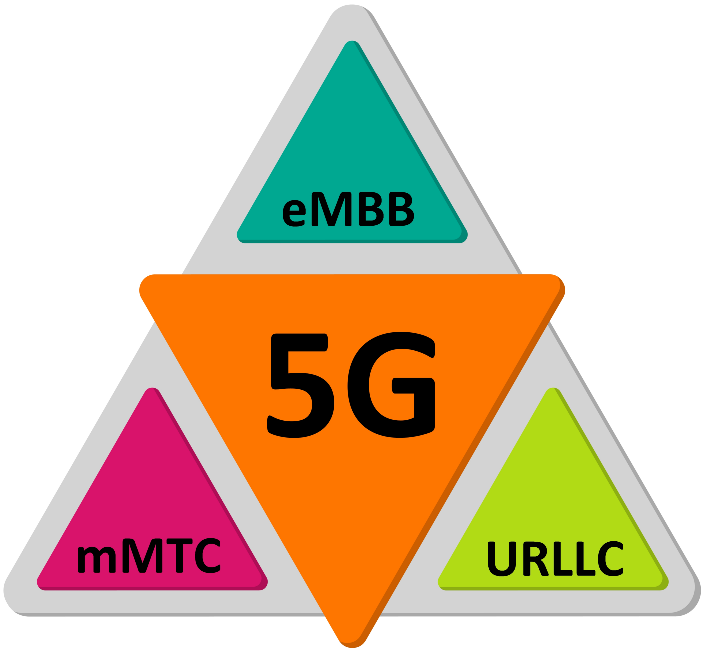


Fig. 2.8 The three generic services offered by 5G.

tration, and management mechanism [82]. Without depending on the tedious architecture and protocol-level alternations [83], radio/remote resource slicing must be accomplished dynamically, which is an essential part of any network slicing resolution.

In addition, mission-critical IoT (mc-IoT) services have different requirements in terms of latency, availability, reliability, network resources, etc. The work done in [84, 85] does not considered the network infrastructure that caters to mc-IoT's strict requirements. To address the above challenges, a flexible network infrastructure based on NFV was designed [86]. Depending upon the need for different mc-IoT applications, the services that guarantee slicing are deployed, and the network resources are distributed based on the shared physical infrastructure.

According to the 3GPP releases, three generic services under the 5G – New Radio (NR) framework address new verticals and scenario deployments, as shown in Figure 2.8. The generic services are:

- Enhanced Mobile Broadband (eMBB): Mainly focuses on faster speeds for applications demanding higher rates.
- Massive Machine-Type Communication (mMTC): Mainly focuses on sensing and monitoring systems.

Table 2.4 The three generic services under 5G framework and their attributes.

Attributes/Generic Services	**eMBB**	**mMTC**	**URLLC**
RTT	Moderately Low	Moderately Low	Extremely Low
Throughput	Moderately High/High	Moderately Low	Moderately Low/Moderate/ Moderately High
Availability	Moderate	Moderate	Extremely High
Frequency of Data Transfers	Moderately High	Moderately Low	Moderately High

- Ultra-reliable Low-Latency Communications (URLLC): Mainly focuses on providing very low latency for mission-critical systems.

In network slicing, the 5G framework is relied upon to oblige most devices that will interface with the network with a different Quality of Service (QoS) and Quality of Experience (QoE). It incorporates round-trip time (RTT), throughput, nature of information shared between the end devices and application servers, and frequency of data transfer [87, 88]. The three generic services and their attributes are mentioned in Table 2.4.

2.5.6 Control co-design

Two reciprocal standards are used to acknowledge controlling applications over the wireless system, i.e., network-aware control design and control-aware network design. Previously, the objective was to improvise network layer protocols to upgrade real-time execution. This modification holds as evidence that we can advance the word towards the TI. Subsequently, the design and configuration of control feedback codes match the shortcomings of the communication medium. For the teleoperation framework, the structure of adaptive controllers [89] is a classic case of this modification. Thus, this modification needs to be closely associated with networks to mutually improve control-aware network design and network-aware control design.

To consolidate the dynamic nature of B5G networks, RTT minimisation will always be a concern while having communication between master and slave sections. The rising timing requests are put on the control loop, thus causing a new challenge to reduce the overall RTT of the system. Hence, new approaches are needed to solve these RTT issues. However, the decoupling of the control loops can be done to make the two sections independent of each other by employing parameter weights on each side. Hence, Hence, incorporating machine learning or prediction models would potentially resolve the RTT issues in B5G networks [90].

2.5.7 Wireless resource customisation

To customise wireless resources, network slicing techniques can be used as an open door. Wireless resource customisation alludes to the usage of streamlined application-specific wireless resource allocation techniques within the assigned slice. Either human-related or machine-related communications mostly depend on the traditional wireless resource allocation strategies. Such strategies may not be helpful for a TI application, but these traditional strategies mainly concentrate on configuring of the uplink and downlink of tactile communication. Nevertheless, combined uplink or downlink configuration is alluring for an application considering the control loop [91]. Except for a few ongoing types of research [92, 93, 69], there are still loopholes (unexplored areas) in wireless resource allocation for TI applications.

A few works have been done to allocate resources to emerging technologies such as SDN, NFV, MA techniques, etc. The work in [94] focuses on understanding the traffic requirements of the TI environment, wherein strategies are discussed to control and allocate the bandwidth of the network. The standard FibreWireless (FiWi) infrastructure is merged with SDN to mobilise the bandwidth allocation. The SDN controller programmatically allocated the required resources based on the current global traffic status. Hence, the required QoS was also achieved under different TI environments.

2.5.8 Easy incorporation

Regardless of the accessibility of high-end specifications of the wireless protocols, some hybrid systems probably exist, ranging from small to medium geographical areas. The goal of hybrid systems is to easily incorporate wired and wireless systems without affecting performance and efficiency. This kind of incorporation is accomplished through particular Internet-enabled devices. The problem of consistent incorporation has been researched in a few past investigations [95], coinciding with the introduction of wireless for modern systems. Nevertheless, a concise methodology is required to research this problem, as prior research may not be straightforwardly good with continuous tasks regarding Time-Sensitive Networking (TSN) and 5G - New Radio (5G-NR).

A teleoperation system was proposed based on the 5G-NR to examine and study the performance of haptics over Internet protocols [96]. The feasibility of the proposed system ensured reliable communication and did not violate the requirements of QoS.

2.5.9 Unlicensed band and more

A significant number of wireless technologies work in the unlicensed band, i.e., 2.4 GHz to 5 GHz. There is a crowded situation regarding wireless technology's reliability within these freely available unlicensed bands. Because of this reliability, an examination was initiated to check for the substitute-free spectrum for TI applications. With the availability of enormous bandwidth, millimetre waves (mm-wave) are especially appealing for TI [97]. Although mm-wave technology comes with a set of limitations and challenges for shorter-range communications, these issues also need to be taken care of [98].

Several finger augmentation devices are wired and aim to recover lost human senses wherein the user has lost the ability to feel objects. Because of the wired system, it limits the free mobility of hand gestures and creates a feeling of being tied to some objects. These devices have an electronic surface and should be worn on the tip of the fingers. In [99, 100], the radio-frequency finger-augmented device (wireless) is considered to be a promising enabler for TI. . This wireless system, comprising a reader worn on the wrist and a sensing tag worn on the fingertip, makes use of an ultra-high frequency band ranging from 860 MHz to 960 MHz.

Additionally, there is a chance of impedance mismatch between the transmitting antenna and the electronic circuit, thus affecting the communication link. Self-tuning microchips address this issue. However, the concern remains regarding the distance between a transmitting antenna and an electronic circuit. If the distance is large between them, the tag will be disconnected and disrupt the normal functioning of the radio-frequency finger-augmented device.

2.5.10 Safety and security concerns

Safety and security issues have been significant concerns for a wide range of information regarding the TI and its applications. Data integrity, reliability, and accountability must be catered for if we rely on the TI to facilitate daily applications and avoid any harm if it does not work appropriately. Depending upon the application and various types of information, various dimensions of security are required. If we consider the latest security advancements available today, haptic communication can take place over the network. However, there has been a significant increase in RTT when using the current security implementation.

Moreover, the security requirement exclusively constrains access to information and restricts the RTT. It entirely depends on the application and whether priority is given to security or RTT. For instance, in a military organisation, information requires high privacy. On the contrary, some VR applications require minimal latency and lower requirements for data security. Subsequently, it can be concluded that there is always a trade-off for security and latency requirements depending on the type of application.

Furthermore, to provide network coverage to catastrophically affected areas following events such as earthquakes, tsunamis, etc., potentially emerging technologies such as 5G, blockchain and drones can be adopted in a TI environment. Therefore, drones can be mounted with small cellular base stations to provide cellular services in affected areas. The main challenge is to have data security in such flying cellular base stations. To tackle the concerns about data security, the blockchain-based framework [101] was used to provide a security layer in the communication between (1) drone to drone and (2) user to the drone.

2.5.11 Challenges related to haptic devices

There are also challenges related to the sensors and actuators at the operator and teleoperator ends. Related to haptic devices, haptic sensors and actuators are used for communication. These sensors are usually mounted at the teleoperator end, which is used to sense the tactile data. It also senses the activity going on at the teleoperator end and relays it back to the operator/user end as feedback. The haptic actuators handle the feedback process, often known as haptic feedback devices.

Haptic sensors

Haptic sensors are useful for detecting pressure for a delicate body, body control, and for the worst-case control scenario of <10 kPa, 10 kPa to 100 kPa and 325 kPa, recognisable to humans [102, 103]. There are different methods to measure pressure. The most famous methods are resistive and capacitive. In the resistive method, the material used is pressure-sensitive and is used to create a sensor [104]. In the capacity method [102], the capacitance difference is noted using a dielectric material between two conducting plates. Then, the difference in capacitance is converted and digitised to measure pressure.

The dielectric material used for capacitive haptic sensors is the elastomer polydimethylsiloxane (PDMS), famously called silicone. PDMS features include (i) excellent dielectric constant, (ii) low Young's modulus, and (iii) very flexible placement of sensors. However, there are some different non-idealities in PDMS pressure strain sensors, thus making the structure complicated. The non-idealities include creep unwinding, hysteresis, and so on. Thus, the need to design the capacitive weight sensor is taken into account by considering all non-idealities of PDMS.

Another vital parameter to consider when building the capacitive sensor is sensitivity. Sensitivity comes into play whenever there is a change in pressure. When the sensor's sensitivity is high, the sensing process, conversion time and the sensor's estimation scope are reduced because of saturation effects. Therefore, the trade-off is that if the blend of the high and low-sensitive sensors is considered, the fabrication process becomes more tedious and thus increases its fabrication cost. In addition, PDMS can also be modelled to adjust the sensor's sensitivity range. Nevertheless, the model will also add complexity and cost to the fabrication.

When it comes to designing the haptic framework, a commonly neglected factor in the literature is the placement of sensors. There are two ways in which haptic sensors can be fabricated: (1) single-element sensor and (2) multiple-element sensor arrays [105–109].

However, single-element and array sensors are commonly expensive. Therefore, these sensors must be placed in the detection zone to increase the coverage radius using the correct placement techniques. The increase in array size is directly proportional to the scanning rate of the number of sensors. Thus, the overall consumption of power and the time required for sensing are incremented. There will always be a trade-off between spatial resolution and the coverage area. Subsequently, the chances of collaborative design of the single-element and array sensors are more likely achievable. However, these designs could make the placement of sensors tedious.

There is an advancement issue to investigate [108] about the conceivable arrangement to have various sensing circuit-handling sensors in different areas of the array sensor. Hence, finishing up the construction of a haptic sensor for applications of the TI is not a straightforward process. There are many challenges that need to be resolved, including scanning time, placement of the sensors, spatial resolution and sensitivity.

Haptic actuators

There are two types of haptic feedback present in the TI system [110, 111]: (1) tactile touch and (2) kinaesthetic touch. Tactile touch is touch perceptible by human skin, for example, pressure, force, friction, vibration, texture, and skin temperature. In contrast, kinaesthetic touch is the feeling of a neighbouring body part's general position and pressure related to muscle tension.

Throughout the years, there has been a tremendous increase in the research on haptic feedback, where it is challenging to construct or design a haptic feedback framework. This challenge expects a sense of touch to be delivered to the user as they feel it in reality. However, it should be noted that haptic feedback does not fundamentally convey the realistic sense of touch, and a limitation is that it does not imply actual touch.

In [112], the authors performed an experiment where users tried identifying texture using tactile gloves and a custom-built haptic device. The experiments showed that tactile gloves reasonably recognised the texture. Depending on the desired use, haptic feedback can be conveyed in numerous forms. Experimentation has been done on numerous haptic displays with user and haptic feedback. The results vary from the non-contact tactile display, e.g., displays using the radiation effect on ultrasound [113], to feeling vibrations on fingertips using vibrotactile gloves [114–117]. However, there is a limitation with non-contact displays where the display area is barely movable. Hence, this limitation gives rise to a scalability problem where the end user is restricted from moving around freely in a given area.

There is another dimension for research regarding the dependency on gloves-based displays. This display type is non-grounded, meaning it can freely move around. It is not tied to any heavy equipment to limit the free movement of the display. However, this display lacks kinaesthetic feedback.

In remote operating systems, instead of the virtual system and because of limited kinaesthetic force feedback, the user will virtually experience holding, thumping, squeezing, or moving objects. This happens because of excessive force at the remote system. This might result in a limitation or complete prevention of the movement. There is another dimension for perceiving the object's weight [118].

The recent implementation of glove-based haptic displays includes kinaesthetic feedback often restricted by the region of working space. This happens due to the glove being linked to the user's arm or the entire system being grounded, e.g., in [118]. In addition, ungrounded trails are carried out with a glove-based haptic display where the system algorithm makes gloves too heavy for further testing. Thus, this process is more time-consuming.

Based on the examination, as mentioned earlier, there is no fixed or total solution for kinaesthetic feedback yet. A trial demonstration is mentioned in [116], where tests have recognised friction, weight, pressure, smoothness, and hardness with the help of gloves with only vibrotactile feedback. Nonetheless, it was quite different to make identifications in one place compared to achieving the same with a remote object. Because of these adverse effects, we require sensible recognition to create real kinaesthetic feedback.

The design of a real kinaesthetic display will have current challenges, including tactile and kinaesthetic feedback, where displays are not attached to any ground and can move around freely. Another idea is to have a lightweight system at the same time. For some applications, such as telesurgery, a haptic display would not be helpful. In such cases, a replica of the remote and the real-life scenario could be developed so that realistic feedback mechanisms will solve the problem.

2.5.12 Challenges related to kinematic devices

There are different types of sensors available for capturing or tracking the movements at the operator's end. They catch or track the movements of the operator and reproduce the same at the teleoperator end. Based on these devices, there are also kinematic sensors and kinematic actuators. Kinematic sensors capture the movements at the operator end, while kinematic actuators reproduce the same at the teleoperator end.

Kinematic sensors

There are different types of sensors available for capturing or tracking the movements at the operator end. We will discuss a few of the approaches that are being used to devise kinematic sensors. Simple mechanical frameworks are adopted to capture the moment and are generally equipped with some connections and electrochemical transducers, such as a potentiometer, to follow the motion. There are also other methods for tracking the movements of the body, such as [119–124]: (a) Magnetic-based strategy is used wherein magnetic sensors are deployed at different parts of the body and as the body moves, the sensor tracks the movement of the body and records its position; (b) In radio-frequency identification (RFID)-based strategies, the radio-frequency tags and sensors are used to gather the movements of the body; (c) Image sensors are also strategised to catch an image at stipulated period, and then it is processed to evaluate the position/motion of the body; (d) Capacitive-based techniques are also used to capture the body movement with the deployment of capacitive sensor arrays, placed in proximity to the body. Here, a parallel plate of the capacitance concept is used between the body and sensors, noting a change in capacitance as the body moves and many more. Many recent techniques have been developed to sense and track the movement of the body at the operator's end.

Despite possessing such techniques, there are always some limitations regarding the recreation of the movements at the teleoperator end. Moreover, many are helpful in catching just motion [123, 125, 126]. This is not helpful in haptic frameworks where the operator movement is to be steadfastly followed and replicated. Even though these strategies can be used to track movements, such attempts will give coarse information focusing on the movement being detected.

Although these strategies execute admirably in a few haptic cases, they fail when the user handles the duplication of the remote activity. There will also be a limitation for image sensors, which demands the operator focus on the sensors and restricts the user's working zone. This imaging strategy also experiences the blockage of the blood vessel, resulting in the loss of free movement, gripping, etc.

Notwithstanding a couple of exceptional cases, there are a few strategies such as capacitive, magnetic and RFID-based strategies, which request the operator to connect sensors as wearable gadgets. Wearable gadgets such as tactile gloves can be developed with necessary sensors to track or capture the operator's motion using such strategies. In addition, a similar glove can likewise consolidate haptic actuators to facilitate haptic feedback. On the contrary, it can be awkward to wear glove-based gadgets in certain use cases.

Another significant consideration is the output time or scanning time. This additionally incorporates the time it takes to scan and track the operator's kinematic movement. If multiple sensors are present in the TI system, each individual sensor will have its scan time constraint, and processing it could be a challenge to achieve a 1 ms TI constraint. Because of the limitations of sensors that support low-baud rate embedded protocols to transfer the sensor information, the master section seeks sensor information, each in turn. This seeking time can add up to the general output time of the kinematic strategy. A solution here is for the device calculation to powerfully pick the sensor of intrigue and dynamically read the information instead of scanning the sensor one at a time or in a stipulated sequence. Therefore, more research is required to solve this problem.

As discussed above regarding processing time or scanning time, still kinematic strategies adopt refined kinematic algorithms to capture movement. These algorithms contribute to more processing time. Therefore, conceivable arrangements can be adopted by computing all the information in the cloud.

Kinematic actuators

Kinematic actuators are used at the teleoperator end. Here, a robot will be placed at the teleoperator end to copy the action from the operator. However, there could be a variation in the Degrees of Freedom (DoF) of movement between an operator and a robot. In such instances, the reverse version of kinematic equations is applied. These equations calculate the link angle of the robot to guarantee that the final stance of the robot intently resembles that of the operator.

The reverse versions of kinematic equations are responsible for solving basic kinematic conditions for robot links. If the number of robot links increases, these kinematic equations fail and become tedious to unravel. Therefore, numerical solvers are computationally concentrated, which can increment the activation time and the RTT. Subsequently, one solution could be to transfer all the computational work to the cloud and save computational time.

2.6 Review of related articles

To formulate this survey manuscript, relevant TI-related articles contributing to 5G/B5G, SDN, NFV, Network Slicing (NS), Cloud/Edge/Fog Computing (CC/EC/FC), Multiple Access (MA), Network Coding (Net. Cod.) and Machine Learning/Artificial Intelligence (ML/AI) are reviewed and investigated to realise TI. These articles are selected from several refereed journals, conference proceedings and white papers obtained from credible and well-known databases such as IEEE Xplore, ScienceDirect, Springer, and ACM. We carried out an extensive analysis and synthesis of research findings to make a meaningful conclusion. Table 2.5 provides a review of the related articles on the TI for enabling technologies.

2.6.1 The potential of the TI

In a characteristic advancement to various Internet encapsulations, thoughts of the TI [22] are rising in which ultra-responsivity and ultra-reliability will empower the remotely operated real-time control and physical haptic communications. The TI will include a new area of measurement of human and machine interaction through constructing interactive and intelligent frameworks. The TI will facilitate a genuine change in outlook from content-delivery to skillset-delivery networks and, subsequently, change every part of society. As of now, the Institute of Electrical and Electronics Engineers (IEEE) and European Telecommunications Standards Institute (ETSI) [127] standards are also incorporating standardised activities for the TI.

Furthermore, at the wireless edge, 5G mobile communication systems are essential to support the TI. A research project has been carried out at the Mobile and Wireless Communications Enablers for the Twenty-twenty Information Society (METIS) [79] along with the telecommunication industry Next-Generation Mobile Networks (NGMN) alliance [14] regarding the early evaluation of 5G scenarios. Dealing with the TI, there has been an increase in new requirements and challenges for designing a 5G network [128].

As of late, thoughts of the TI [22] have risen, which is imagined to give a change in perspective by empowering wireless networks for ongoing steering and control communications. The working group with a standard as IEEE P1918.1 [29] refers to the TI as a metadata set of networks, mainly for accessing the data wirelessly, receiving, manipulating, or managing real and virtual data, thus processing in perceived real-time. Abd El-Latif et al. [129] explored deep learning, specifically long short-term memory (LSTM) networks, for predicting delays in M2M networks. It highlighted the integration of SDN and NFV to enhance network efficiency. It also discussed NS, EC/FC, and network coding to optimise data transmission. ML/AI techniques were applied to improve the performance and reliability of TI and other networks, addressing challenges in delay prediction and network management. On the other hand, Chaudhari [130] explored the potential of 6G technology to enable the TI, emphasising its stringent requirements, such as ultra-low latency, high reliability, and high data rates. It discussed the integration of SDN, NFV and EC to meet these demands. It highlighted the role of AI in predicting user actions and optimising network performance. Key applications like remote surgery, autonomous vehicles, and virtual reality were examined, focusing on their specific requirements and the design considerations necessary for 6G networks.

Moreover, Schulz et al. [131] proposed the concept of negative latency in TI to enable global Metaverse immersion. It addressed the challenge of ultra-low latency communication below 1 ms, which is crucial for haptic interactions. By predicting user actions with high confidence, they demonstrated that several hundred ms negative latencies are achievable. The proof-of-concept involved smart sensor gloves for grasping tasks. This approach, leveraging advanced technologies in TI, aimed to support seamless, immersive, and interactive experiences in the Metaverse, enhancing human-machine cooperation on a global scale. Patil et al. [132] developed a simulation and assessment system for TI, focusing on control mechanisms and predictive AI. It evaluated proportional-derivative (PD) and model predictive control (MPC) controllers for reference signal tracking, with PD showing superior performance. They explored a Smith predictor and an AutoRegressive with eXogenous inputs (ARX) neural network for predictive modelling, enhancing prediction within a 200 ms horizon. Despite processing limitations introducing delays, the system demonstrated effective control performance. This work laid a foundation for future TI, robotics, teleoperation, and AI research, accommodating new technologies and control approaches.

Table 2.5 Review of related articles on TI for enabling technologies (5G/B5G, SDN, NFV/NS, CC/EC/FC, MA, Net. Cod. and ML/AI).

Authors	**Year**	**5G/ B5G**	**SDN**	**NFV/ NS**	**CC/EC /FC**	**MA**	**Net. Cod.**	**ML/ AI**
Chaudhari [130]	2025	✓	✓	✓	✓	×	×	✓
Sun et al. [133]	2025	✓	✓	✓	✓	×	×	✓
Dubey et al. [134]	2025	✓	✓	✓	✓	×	×	✓
Hornik et al. [135]	2024	×	×	×	✓	×	×	✓
Arbaoui et al. [136]	2024	×	✓	✓	×	×	×	✓
Okello et al. [137]	2024	✓	✓	✓	×	×	×	✓
Otieno et al. [138]	2024	✓	✓	✓	×	×	×	✓
Silva et al. [139]	2024	✓	✓	✓	✓	×	×	✓
Hamdi et al. [140]	2024	✓	✓	✓	✓	×	×	✓
Goiz et al. [141]	2024	✓	×	✓	×	×	×	✓
Liem et al. [142]	2024	×	✓	×	×	×	×	✓
Alnajar and Barnawi [143]	2024	✓	✓	✓	✓	×	×	✓
Patil et al. [132]	2024	×	×	×	×	×	×	✓
Shaik et al. [144]	2024	✓	✓	✓	×	✓	×	✓
Sthankiya et al. [145]	2024	✓	✓	✓	×	✓	×	✓
Belmekki and Alouini [146]	2024	✓	×	×	×	✓	×	×
Ahmed et al. [147]	2024	✓	✓	✓	✓	✓	×	✓
Clerckx et al. [148]	2024	✓	✓	✓	✓	✓	×	✓
Luo et al. [149]	2024	×	×	×	×	✓	×	×
Mayarakaca and Lee [150]	2024	✓	×	×	×	✓	×	✓
Abd El-Latif et al. [129]	2023	×	✓	✓	✓	×	×	✓
Cheng et al. [49]	2020	✓	×	×	✓	×	×	×
Mekikis et al. [151]	2020	✓	✓	✓	×	×	×	×
Sharma et al. [152]	2020	✓	✓	✓	✓	✓	✓	×
Gokhale et al. [153]	2020	✓	✓	×	×	×	✓	×
Na et al. [154]	2020	✓	×	×	×	×	×	×
Zhani and ElBakoury [155]	2020	✓	✓	✓	✓	×	×	×
Meshram and Patil [156]	2020	✓	✓	✓	✓	×	×	×
X. Wei et al. [157]	2019	×	×	×	✓	×	×	×
I. Budhiraja et al. [158]	2019	✓	×	×	×	✓	×	×
Vora et al. [159]	2019	✓	×	×	✓	×	×	×
Ge et al. [86]	2019	✓	×	✓	×	×	×	×
Aggarwal and Kumar [160]	2019	✓	✓	✓	✓	×	×	×
Fanibhare et al. [161]	2019	✓	✓	×	✓	×	×	×
Maier and Ebrahimzadeh [162]	2019	✓	×	×	✓	×	✓	×
Arshad et al. [163]	2019	✓	×	×	✓	×	✓	×
Jinke Ren et al. [164]	2019	×	×	×	✓	×	×	×
Kim et al. [24]	2018	✓	×	×	×	✓	✓	×
Grasso and Schembra [67]	2018	✓	×	×	×	×	✓	×
Alextian et.al [165]	2018	✓	✓	×	×	×	×	×
J. Cabrera et al. [166]	2018	✓	✓	✓	✓	×	✓	×
Dmitry et al. [167]	2018	×	✓	✓	×	×	×	×
Y. Xiao et al. [168]	2018	×	×	×	✓	×	×	×
Ateya e al. [169]	2018	✓	✓	✓	✓	×	×	×
S. Troia et al. [170]	2018	×	✓	✓	×	×	×	×
C. Grasso et al. [67]	2018	✓	×	×	×	×	×	×
Li et al. [128]	2018	✓	×	×	×	✓	×	×
M. Gharbaoui et al. [171]	2018	✓	✓	✓	×	×	×	×

Table 2.5 *Contd.*

Authors	**Year**	**5G/ B5G**	**SDN**	**NFV/ NS**	**CC/EC /FC**	**MA**	**Net. Cod.**	**ML/ AI**
Popovski et al. [65]	2018	✓	×	×	×	✓	✓	×
Chatras et al. [4]	2017	✓	×	✓	×	×	×	×
Yi-Wei Ma et al. [172]	2017	×	✓	✓	✓	×	×	×
T. Theodorouan et al. [173]	2017	×	✓	×	×	×	×	×
K. Wang et al. [174]	2017	×	✓	×	×	×	×	×
Ateya et al. [175]	2017	✓	×	×	✓	×	×	×
Feng et al. [176]	2017	✓	×	×	×	✓	×	×
Pilz et al. [177]	2016	✓	✓	×	×	✓	×	×
Athmiya et al. [178]	2016	×	✓	×	×	×	×	×
Intharawijitr et al. [179]	2016	✓	×	×	✓	×	×	×
Simsek et al. [180]	2016	✓	✓	✓	✓	✓	✓	×
Maier et al. [26]	2016	✓	✓	✓	✓	✓	✓	×
Tong et al. [181]	2016	✓	×	×	✓	×	×	×
N. Truong et al. [182]	2015	×	✓	×	✓	×	✓	×
P. Iovanna et al. [183]	2015	✓	✓	×	×	×	×	×
D. Szabo et al. [184]	2015	✓	✓	×	×	×	✓	×
F. Bonomi et al. [12]	2011	×	×	×	✓	×	×	×
Cuervo et al. [185]	2010	×	×	×	✓	×	×	×

2.6.2 SDN-based design

While the widespread adoption and application of SDN are relatively recent, it facilitates a dynamic, flexible, and cost-effective system structure by physically partitioning the data forwarding and control planes in networking. The control plane handles a suitable choice to deal with traffic, whereas the data plane takes care of forwarding the traffic according to the control plane. SDN grants permission to the system operator to design, configure, control and deal with the system through programming software such as application programming interfaces (APIs). SDN is mainly used to disentangle the system hardware and increase network adaptability.

Gharbaoui et al. [171] presented a service-chaining orchestration framework comprising dynamic virtual function (VF) selection and intent-based traffic-steering control competencies to upgrade Management and Orchestration (MANO) framework parts with latency-minimised and self-adaptive service-chaining highlights over geographically distributed SDN-based cloud data-centres (DCs) interconnected through a SDN Wide-Area Network (WAN).

Furthermore, Yi-Wei Ma et al. proposed a software-defined infrastructure (SDI) [172] with NFV, which is then deployed in an Industry 4.0 network environment. This infrastructure uses the technological concepts of SDN and NFV, which enhance the overall activities of the network environment, the rate of information transmission, and the QoS provided.

On similar lines, Sebastian Troia et al. [170] presented their work on the SDN/NFV Orchestrator, a new service orchestrator called SENATUS. It facilitates a service orchestration for network segments deploying Openstack as infrastructure manager and Open Network Operating System (ONOS) as SDN controller. Moreover, Theodorouan et al. [173] proposed an SDN solution for wireless sensor networks (WSNs) which (i) uses intelligent centralised control mechanisms to adjust the protocol operations dynamically, (ii) underpins versatility to difficult prerequisites of the WSNs, (iii) keeps an adapted architecture, and (iv) shows enhanced network management and operation regarding execution and resource use. In addition, Christian Grasso et al. [67] proposed an architecture for unmanned aerial vehicle (UAV)-based video surveillance with TI constraints, where the drone-based Markov chain model is examined.

Various methodologies talk about using SDN at the central network of the cellular framework. In addition to working towards the TI, the need to code the system is incorporated with SDN to minimise the RTT in the 5G framework. A software router within the network is encrypted and acts as a virtual network function (VNF). They are mainly worried about coding and SDN but have not considered mobile-edge computing (MEC). This framework is predominantly presented for 5G and IoT [158].

A 5G-based SDN infrastructure is presented in [174] with the massive deployment of small cells. Therefore, having an idea of small cells incurs difficulties of continuous handover and delay during the handover process. Here, the authors have facilitated a framework to defeat these difficulties by incorporating an SDN controller in the central system. By deploying an SDN controller, the framework is ready to anticipate that the end user or a client is moving and subsequently can deal with the handover process in less time by characterising three sorts of program interfaces. Since the fundamental thing of the TI framework is to meet 1 ms round-trip time, i.e., ultra-low latency, cognitive control ought to be conveyed at the central network to accomplish the desired RTT.

Taking SDN architecture into consideration for ultra-reliable and ultra-low-latency communication, Alextian et al. [165] have introduced Residue-Defined Networking Architecture (RDNA) as a unique strategy for enabling critical competencies of micro-datacentre networks. This architecture is built on the concept of SDN to meet complex and stringent demands from next-generation networks.

In addition, the work in [166] by Juan Cabrera et al. featured the significance of SDN and NFV for 5G communication along with TI applications. It describes that SDN and NFV complement each other and facilitate empowering technologies to have facilities such as computing, storage, and networking functions. Furthermore, Dmitry et al. [167] proposed an effective algorithm and software for virtual slice formation in SDN dependent on the information of base network and links connectivity information.

In addition, Liem et al. [142] proposed an AI-driven resilience mechanism for Next-Generation Ethernet Passive Optical Networks (NG-EPONs) to enhance TI reliability. It integrated the ML/AI and SDN-Enabled Broadband Access (SEBA) platform to detect and localise fibre faults, using Radio Frequency over Glass (RFoG) for backup communication. The AI model, validated through 5-fold cross-validation, achieved an average accuracy of 81.49%. The system ensured uninterrupted, high-quality network service by automatically establishing backup links during faults. Simulation results confirmed the mechanism's efficacy in managing delay, throughput, packet drop rate, and bandwidth waste, meeting stringent QoS requirements for critical TI applications.

2.6.3 FC-based design

FC was originally proposed by Flavio Bonomi, vice president of a network device manufacturing organisation called Cisco, in 2011 [12]. FC can be explained as a highly virtualised platform where an enormous number of heterogeneous, pervasive, and decentralised devices connect and correspond to other devices. These devices provide computation, storage, and networking services between devices and conventional cloud centres, but not precisely at the edge of the networks, without the interference of third parties.

The capabilities of cloud computing and services are extended to FC, where routers, gateways, switches, etc., can also be a part of the network with all properties of FC. It was fundamentally used for the automation of types of equipment since FC has a background in IoT.

In [168], Xiao et al. focussed their research on an energy-efficient design of a fog network that helps to have low-delay-supported TI applications. Here, they have explored the performance parameters, such as the service reaction time of the end clients and the power use efficiency of fog nodes. Additionally, a novel cooperative FC concept is presented with fog nodes with various processing and energy resources to coordinate with one another. This led to fulfilling the objective of balancing the workload or the request executed by multiple fog nodes.

The idea of edge-assisted cloud computing and its connections is related to the developing area of FC. Such frameworks use low-power and efficient embedded computers to facilitate nearby processing close to end users or the cloud. The Fog of Things (FoT), analogous to Fog

Computing (FC), alludes to the amalgamation of various fog nodes that could be connected and communicated with each other with IoT.

The FC has a decentralised computing structure dependent on FC nodes (FCNs) and can be fixed at any architecture location between the end-user device end and the cloud. With FCNs that use M2M gateways and wireless router devices, FC acts as a computing layer between the end device and the cloud layers. These FCNs are used to process and store information from end-user devices before sending it to the cloud.

Nonetheless, the hypothetical establishments for advancing distributed FC systems to satisfy TI needs are still deficient. Specifically, computationally escalated services needing low delay usually consume more energy from fog nodes. Simultaneously, many TI applications include lightweight equipment such as drones, robots, and vehicles with constrained power supplies.

The International Telecommunication Union (ITU) and Next-Generation Mobile Networks (NGMN) alliance recognise FC as one of the crucial factors for the TI to accomplish ultra-low service delay for clients [26, 14, 180]. Rather than cloud data centres (CDCs), fog nodes can be deployed near the end clients so that the task transmission delay can be substantially diminished. However, because of each fog node's limited computational or processing ability, carrying out a considerable number of tasks to fog nodes will bring about a high processing delay. Therefore, previous works have discussed the most efficient resource provisioning of fog nodes to decrease the processing delay. For instance, a virtual machine (VM) synthesis technique is presented to enable each end client to rapidly arrange the resources of neighbouring fog nodes and make the required VM images achieve the desired TI application.

Furthermore, Cuervo et al. [185] presented a framework that permits fine-grained energy-aware offloading of client mobile codes to the system. Therefore, the topic of tactical cloudlets is explored for resource provisioning, and the provisioning component is introduced for the system to help with processing, offloading, and data staging at the tactical edge.

As far as the architectural aspects of FC are concerned, Bonomi et al. [12] exhibited one of the primary tasks of FC surveying the appropriateness of FC for the IoT. They mention the necessities of developing the application regarding real-time user communication, awareness of the location and the requirement for geo-distributed end nodes, and FC attending to these problems.

The authors have contributed more knowledge into the appropriateness of FC for IoT applications with a couple of use cases, including smart wind farms and intelligent traffic light systems, in the accompanying paper [186]. Furthermore, V. Fanibhare et al. [161]

proposed a multi-level cloud structure where the traffic flows over the cloud units efficiently to reach the slave section from the master section in the TI system to reduce unnecessary waiting times.

Along similar lines, a multitier infrastructure [181] is introduced in which edge cloud servers are differentiated into various levels as per their separations from the edge. If the task on the edge cloud server of a given level surpasses its computational limit, the additional outstanding task is sent to a higher-level service. One of the FC areas is that many fog nodes are distributed and deployed over vast geographical territory. Nonetheless, this can bring about a noteworthy increment in energy consumption.

Moreover, Truong et al. [182] influenced the distributed approach of fog nodes to enhance awareness of the location in Vehicular Ad hoc Networks (VANETs). They tested their framework for the application case of lane-changing assistance. Additionally, they used the fog layer's services to execute and find an optimal parking spot, where the aggregated data were gathered from different fog devices, demonstrating the concept of the distributed idea of fog devices.

In [187], an adaptive energy scheme was developed for fog nodes to work at various transmit powers with variable data rates. Jalali et al. [188] correlated the energy consumption of applications using centralised (CDCs) with applications employing nano data centres under the FC infrastructure. Xuejiang Wei et al. [157] proposed a new fog-based sensor cloud architecture for IoT. They facilitated the fundamental ways for physical sensor virtualisation, dynamic provisioning of virtual sensor groups and service instances for key issues. Furthermore, Jinke Ren et al. [164] explored the joint communication and computation resource allocation to minimise the weighted-sum delay (RTT) of all devices in a cloud-edge collaboration system.

Moreover, Hornik et al. [135] explored the integration of FC-enabled AI to enhance smart marketing management. It examined how FC and EC improved data processing by reducing latency and decentralising computational tasks, making them valuable for IoT applications. ML and AI were utilised to analyse consumer behaviour, optimise decision-making, and enhance marketing strategies. It demonstrated how AI-driven fog computing could improve network efficiency, responsiveness, and intelligence in next-generation IoT networks and applications by leveraging distributed intelligence. Alnajar and Barnawi [143] proposed TactiFlex, a flexible Tactile Internet of Things (TIoT) architecture in 6G networks, integrating federated learning (FL), blockchain, SDN, NFV, and MEC. It emphasised in-content awareness for precise resource allocation based on user intent. The architecture dynamically scaled resources, optimised bandwidth, and ensured low-latency, high-reliability communication.

FL enabled collaborative model training without compromising data privacy. They evaluated lightweight DL methods as local models in FL and analysed various FL algorithms. These advancements aimed to revolutionise TIoT applications, enhancing intelligent, context-aware, and efficient communication in 6G networks.

In summary, the advantages of FC compared to traditional cloud computing are featured in [189, 190]. In [191], the programming framework is designed to deliver resources to numerous IoT end-user devices by considering mobile FC. Therefore, at the primary level within FC, the examination of resource allocation has been accomplished using the FC framework.

Finally, FC has its capabilities and competencies as follows. It has location awareness and can be located at the edges of a network. It incurs low latency, which supports large-scale complex sensor networks and storage capabilities in a wide-area distributed network. Network faults can be analysed online and integrated with the back-end cloud system.

2.6.4 NS-based design

NS was initially proposed as a key enabler for 5G networks, allowing the creation of multiple virtual networks on a shared physical infrastructure. NS can be explained as a highly virtualised platform where an enormous number of heterogeneous, pervasive, and decentralised devices connect and correspond to other devices. These virtual networks, or slices, provide customised computation, storage, and networking services tailored to specific applications and user requirements without the interference of third parties.

The capabilities of traditional network infrastructures are extended to NS, where routers, gateways, switches, and other network elements can also be part of the network with all properties of NS. It was fundamentally used to optimise resource utilisation and provide dedicated resources for different types of services, such as enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communication (mMTC).

NS enables network resource automation and dynamic management, ensuring that each slice meets its associated services' specific performance, security, and reliability requirements. This approach leverages technologies such as SDN and NFV to create and manage these virtual networks efficiently, providing a flexible and scalable solution for the diverse needs of modern communication systems.

Arbaoui et al. [136] reviewed the implementation of NS in the Internet of Vehicles (IoV) networks, emphasising its role in fulfilling diverse V2X communication needs. It highlighted how SDN and NFV enhanced network slicing to provide ultra-low latency, high reliability, and mMTC. It categorised research into management and orchestration, NS mechanisms, and challenges, showcasing ML and blockchain integration for improved performance, security, and resource optimisation in IoV networks. Okello et al. [137] investigated the 5G core network performance enhancement using network slicing and deep reinforcement learning (DRL). It demonstrated how SDN and NFV facilitated dynamic and efficient resource allocation. By modelling the problem as a Markov Decision Process (MDP), this work used DRL to optimise User Plane Functions (UPFs), improving QoS and throughput. The results showed that sliced networks with DRL achieved better performance, reduced delays, and enhanced load balancing compared to unsliced models.

Moreover, Otieno et al. [138] systematically reviewed deploying and managing intelligent end-to-end network slicing in 5G and B5G networks. It addressed challenges in integrating ML for network slice lifecycle management, emphasising the roles of SDN and NFV. It highlighted the lack of comprehensive datasets and practical implementations, proposing solutions using open-source tools and Commercial-Off-The-Shelf (COTS) devices. It also explored the potential of unsupervised ML techniques in designing network slice templates and managing network resources dynamically. Sun et al. [133] surveyed the role of Explainable AI (XAI) in 6G communications, focusing on network slicing, SDN, NFV, and ML/AI. It highlighted how XAI enhanced transparency and reliability in resource management, fault diagnosis, and real-time decision-making. It emphasised the integration of XAI with FC, EC, and IoT to optimise network performance and meet the stringent requirements of TI applications. Real-world use cases demonstrated XAI's potential in improving resource allocation and ensuring ethical AI deployment in complex, dynamic networks.

Subsequently, Dubey et al. [134] reviewed AI-based resource management for 5G network slicing, focusing on SDN, NFV, and EC. It discussed how ML/AI techniques were employed to optimise traffic classification, admission control, resource allocation, and scheduling in dynamic network environments. It highlighted the deployment of network slicing in various industrial use cases, including smart transportation, healthcare, and Industry 4.0, emphasising the need for efficient resource management to meet diverse QoS requirements. It also identified research gaps and future directions for AI-driven network slicing. Silva et al. [139] proposed an ML-based inter-slice load balancing control for proactive offloading of virtual services in 5G networks. It utilised SDN and NFV to manage network slicing, ensuring QoS by predicting and mitigating Quality of Network-Slice Service (QoNSS) degradation. Artificial Neural Networks (ANN), k-Nearest Neighbors (kNN), and Support Vector Machines (SVM) models are evaluated for predicting throughput and packet loss, demonstrating superior performance over rule-based methods. The approach enhanced resource utilisation and service quality in dynamic network environments, highlighting its potential for B5G, TI, IoT, FC, and EC applications.

In addition, Hamdi et al. [140] reviewed the integration of NS and ML techniques for the Internet of Vehicles (IoV) in 5G and B5G networks. It discussed the role of SDN, NFV, FC, and EC in enabling dynamic and efficient resource allocation. It highlighted the benefits of network slicing in addressing the diverse QoS requirements of IoV applications. Additionally, it explored the use of ML/AI for optimising network performance and enhancing security in IoT and TI environments. Finally, it provided insights into future research directions for NS-IoV integration. Goiz et al. [141] used ML techniques to analyse IP traffic classification for network slicing in 5G networks. It focused on identifying key attributes influencing IP traffic classification and proposed a Smart Network Slicing Manager (SNSM) for dynamic resource allocation. The simulation results demonstrated that the Random Forest model, trained with five statistical features, achieved high accuracy and performance. The SNSM utilised ML models to classify and group network flows, predicting resource demands for each network slice enhancing network management and efficiency in 5G environments.

2.6.5 MA-based design

MA techniques have been fundamental in enabling efficient communication in wireless networks by allowing multiple users to share the same resource block. MA can be defined as a highly optimised method where numerous heterogeneous, pervasive, and decentralised devices connect and communicate while minimising interference. These techniques provide access to transmission, reception, and network resources among users and base stations, ensuring efficient spectrum utilisation. The evolution of MA has extended beyond traditional methods to advanced techniques such as Orthogonal Multiple Access (OMA) and Non-Orthogonal Multiple Access (NOMA), which enhance spectral efficiency and support massive connectivity.

With the emergence of the TI, MA techniques have become even more crucial in enabling ultra-reliable low-latency communication (URLLC) and haptic feedback applications. The integration of MA with NS, EC, and AI ensures seamless connectivity for mission-critical applications such as remote surgery, industrial automation, and immersive virtual reality. Originally designed for cellular and IoT applications, MA has now evolved to meet the TI's stringent latency and reliability requirements, making it a key enabler for future intelligent networks.

Taking MA with ML/AL strategies into account, Shaik et al. [144] surveyed AI/ML-aided strategies for capacity maximisation in URLLC within 5G/6G systems. It focused on resource allocation, MA techniques, and beamforming with massive MIMO. They highlighted the role of ML/AI in optimising URLLC performance, addressing challenges like latency and reliability. Techniques such as NOMA, SDN, NFV, and NS were explored. This work emphasised the importance of dynamic resource management and the integration of ML/AI to enhance system efficiency and meet stringent URLLC requirements in TI and other IoT applications. Sthankiya et al. [145] surveyed AI-driven energy optimisation strategies in terrestrial Next Generation Radio Access Networks (NG-RANs), focusing on 5G and B5G networks. It reviewed power consumption models and categorised energy-saving techniques by time, frequency, power, and spatial domains. It highlighted the role of ML/AI in optimising energy efficiency and addressing challenges like latency and computational complexity. Techniques such as SDN, NFV, network slicing, and NOMA were explored. This work emphasised the need for consistent reporting of AI energy costs to balance energy savings and operational costs in TI and other IoT applications.

Furthermore, Belmekki and Alouini [146] investigated the integration of NOMA with nonterrestrial networks (NTN) for next-generation multiple access (NGMA) in 6G systems. It explored how NOMA, combined with NTN, addressed connectivity challenges by enhancing spectral efficiency and enabling massive connectivity. This work also examined the interplay of NOMA with emerging technologies like millimetre-wave (mmWave), terahertz (THz) frequencies, reconfigurable intelligent surfaces (RIS), and integrated sensing and communication (ISAC) systems, highlighting their potential to meet 6G requirements and improve NTN functionality. Ahmed et al. [147] explored the potential of NOMA as a key MA technique for NGWNs, particularly in the context of 5G and B5G. It highlighted NOMA's ability to enhance spectral efficiency, connectivity, and energy efficiency. It reviewed various NOMA variants and their integration with advanced technologies such as SDN, NFV, FC, EC, and ML/AI. It also addressed the challenges and future research directions for implementing NOMA in IoT and TI networks, emphasising its role in achieving massive connectivity and low-latency communications.

Moreover, Clerckx et al. [148] provided a comprehensive overview of MA techniques for 6G, emphasising their role in optimising wireless systems. It covered the evolution from 5G to beyond 5G (B5G), highlighting the integration of SDN, NFV, NS, FC, EC and ML/AI. This work focused on NOMA and other MA schemes, discussing their applications in the IoT and TI and their potential to enhance network performance and multi-functionality. Luo et al. [149] proposed a distributed control strategy for semi-grant-free (SGF) NOMA to address the excessive control signalling in IoT uplink transmission. By leveraging distributed control, the SGF protocol provided services for multiple grant-free (GF) devices while ensuring reliable communications for grant-based (GB) devices with minimal control signalling. This study introduced the service outage probability (SOP) concept and derived its exact and approximate expressions. Simulation results validated the theoretical findings and demonstrated the protocol's advantages over existing SGF protocols.

Finally, Mayarakaca and Lee [150] surveyed the application of NOMA in Unmanned Aerial Vehicle (UAV) networks, emphasising the integration of ML techniques to enhance UAV communication in 5G and B5G networks. It explored how NOMA improves spectral efficiency and supports diverse UEs in UAV-enabled networks. This study also highlighted the role of ML in optimising resource allocation, interference management, and trajectory planning for UAVs. Various ML methods, including Supervised Learning (SL), Unsupervised Learning (USL), Semi-Supervised Learning (SSL), and Reinforcement Learning (RL), were discussed for their potential to enhance NOMA performance in UAV networks.

2.7 Summary of challenges

TI applications such as self-driving vehicles, augmented reality (AR), virtual reality (VR), Industry 4.0, immersive virtual reality (IVR) and eHealth are the hot topics of research, and a remarkable amount of research is being carried out. Still, existing network infrastructures are finding it hard to cope with the stringent requirements of the TI. Moreover, the speed of light also hampers the speed of TI communication as tactile ends can be at a maximum of 150 km from each other; otherwise, there would be noticeable motion sickness, also referred to as lag.

As TI applications mentioned earlier, the demand for core competencies such as ultra-low latency, ultra-high reliability, ultra-high availability, ultra-responsive and ultra-fast reaction times shows that current cellular infrastructure cannot support these competencies. It is worrisome to fulfil all the requirements of TI applications at the early stage of 5G development.

The challenges related to haptic devices such as Phantom Premium [32] and Geomagic [33] in the market have a robotic arm attached to a stylus, which restricts the free movement of the arm. Thus, it reduces the Degrees of Freedom (DoF) required to realise the TI truly. Since the 5G framework has not developed significantly, the consensus implies adapting one network and sharing the resources among various vertical applications.

Challenges are also experienced with the transparency of the user. An action executed in one location must be replicated in remote locations with a specific level of immersion. If the

RTT becomes noticeable or measurable, the transparency of the user is affected drastically. The RTT of haptic communications should be 1 ms or below. If this RTT is not achieved, it again influences the transparency of the user. There is another challenge between the control server and the user to have a reliable data stream so that any vulnerabilities cannot mess with the original data. The reliability of the connection also relies on the end-user data and its feedback.

Furthermore, there is a need for qualitative and quantitative research on network slicing, including the productive virtualisation of the network framework, applying services to the network slice, the disintegration of network function and end-to-end network MANO. Here, the concept of SDN and NFV would be helpful as potential enablers to address the challenges of TI communication. Network slicing can be used to customise wireless resources. However, human-related or machine-related communications rely on conventional wireless resource allocation strategies focusing on uplinking and downlinking communication. Therefore, more research is needed for wireless resource allocation for applications, considering the global control loop required for TI applications. Thus, incorporating hybrid systems, including wired and wireless systems, could solve the challenges of examining and studying the performance of haptics over Internet protocols.

Because of the freely available unlicensed band, there was a crowded situation surrounding the reliability of wireless technologies. To address this issue, research should be conducted to find the substitute free spectrum for TI applications. Though a mm-wave could solve the TI's problems, it comes with a set of limitations for the shorter range of communications. In addition, there is always a trade-off between security and RTT requirements. A few applications require high data privacy, and others rely on low RTT. Hence, emerging technologies, such as 5G, Blockchain, etc., can be adopted in the TI scenario to resolve security issues.

Regarding haptic devices, there are challenges related to dielectric material used in fabricating sensors, power consumption, the placement and the sensitivity range of the sensor, scanning time, spatial resolution and much more. Moreover, there are challenges surrounding the realisation of the realistic sense of touch through haptic feedback. Since haptic devices cannot move freely and are grounded, scalability would annoy users in a given area. More work is also needed to increase the DoF. Regarding kinematic devices, there are always some issues in replicating fast actions from master to slave sections. Therefore, refined kinematic equations or algorithms are required to capture the activities of the master section. On the other hand, applying reverse kinematic equations becomes more complex for unravelling and following the master's actions at the slave sections.

2.8 Summary and open research problems

A comprehensive review of the TI is provided in this manuscript. This research has explored TI design aspects with the proposed application-centric design architecture and applications. In addition, an in-depth discussion of key applications with proposed illustrative diagrams of use cases and current issues and challenges with potential enablers of the TI is provided. Finally, an extensive review of the related articles on enabling technologies such as 5G/B5G, SDN, NFV, NS, CC/EC/FC, MA, Network Coding and ML/AI is presented, keeping in mind that the importance of 5G network design, EC/FC, 1 ms RTT, compatible tactile gadgets, core software networking and artificial intelligence competencies come at the leading edge of research challenges.

The various epitomes of the Internet will be predominated by the rise of the TI, which will probably bring real-time control and tactile experiences operating from a remote distance. TI facilitates a new transformation and development in wireless communication for real-time and virtual applications. It will transform pretty much every portion of the general population so that they become familiar with TI services. It is forecasted that cutting-edge 5G network frameworks will support the TI at the wireless edge.

Three possible future research directions and open issues regarding TI are outlined below: (i) Traditional cloud-based computational techniques are adding up to the communication delay in wireless links due to the non-presence of cloud infrastructure near the end user. Therefore, the concept of Edge/Fog computing could be used to bring computational capabilities nearer the end user and thus would be proved as a promising technology to reduce the RTT compared to the traditional cloud-based infrastructure; (ii) Due to requirements of high data rates and dynamic orchestration of resources in the core network and radio access network (RAN), the concept of network softwarisation in 5G networks would be the potential technology to guarantee end-to-end reliability and meet the less than 1 ms RTT requirement. Thus, to further improve RTT, emerging latency minimisation techniques such as SDN-based and NFV-based design could be the enabling technologies for the core architecture of RAN; and (iii) In a wireless system, various protocol layers take part in the overall RTT. Some fundamental delay components, such as transmission and queuing, should be considered in TI implementation. Therefore, effective physical and MAC-layer techniques are needed to improve RTT in the wireless transmission system.

Chapter 3

Towards a Fog-Based Traffic Flow Framework for Tactile Internet (Manuscript 2)

3.1 Manuscript 2 prelude

Empowered by sufficient robotics and haptic hardware at the edges of a communication network system, tactile Internet (TI) is the upcoming transformation that will empower the control of the Internet of Things in real-time and facilitate a true paradigm shift in making a dream a reality. However, one of the most crucial challenges in realising TI is to achieve a round-trip time (RTT) of 1 ms or less. This RTT includes transmission, queuing, processing (operator's end), and acknowledgement times (controlled end). Another challenge is to achieve ultrahigh reliability for establishing haptic communications for TI.

Manuscript 2, entitled "Toward a Fog-Based Traffic Flow Framework for Tactile Internet[1]" [192], proposes a novel fog-based traffic flow framework employing software-defined networking (SDN) and fog computing (FC) to address these challenges. An efficient traffic flow algorithm is developed to effectively manage complex and critical traffic in the network and reduce extra processing and waiting times at each level of the cloud-based structure. The SDN and FC approaches provide an effective solution to route the traffic efficiently in the system. Hence, the traffic flow paths from the master to slave sections are reduced, and consequently, RTT is also reduced. The performance of the proposed system is evaluated by an iFogSim simulator in terms of throughput, RTT, energy consumption, and reliability. The simulation results obtained show that the proposed system outperforms the existing edge, cloud, and cellular networks.

Furthermore, the preliminary work[2] done in [161] and this manuscript[1] introduce how SDN and FC in TI can help improve the proposed system's performance, followed by

[1]V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "Toward a fog-based traffic flow framework for tactile Internet," *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 10 718–10 731, 2021.

[2]V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A cloud-based traffic flow framework for tactile Internet using SDN and fog computing,"in *IEEE* 29th *International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–6, 2019.

existing works, design choices and the proposed system design description with results and discussions. Specifically, the first research question (RQ 1) and partially second research question (RQ 2) are addressed in this manuscript as follows: “*RQ 1. What framework can be proposed to manage the traffic flow efficiently by adopting a multilevel cloud structure in the tactile Internet?*” and “*RQ 2. What strategies can be employed to improve the RTT of tactile Internet incorporating SDN and NS approaches?*”.

3.2 Introduction

The term “Tactile Internet” (TI) is defined as a new communication network where machine-to-human (M2H) communication takes place, which allows delivering a real-time sense of touch and actuation information with or without audio/video content over a significant distance. The TI is a development that encourages communication between human beings (preferably over a distance) with visual presence [1]. The two most challenging objectives for accomplishing the communication are to attain a round trip time (RTT) (1 ms or less) and ultra-high reliability (i.e. one-in-one-million chance of failure). These requirements will ensure smooth haptic communication between tactile ends and enable human-machine (real-time) interaction.

The fifth-generation (5G) is the next-generation communication framework with x10 or even x20 faster speeds than current 4G communication, with greater bandwidth, much faster upload and download speeds & more stable connections. As a result, the features of 5G open up new rays of opportunities with connected devices, including real-time interaction in ways that have not been possible before. Thus, the demands for TI are fulfilled by 5G, which has previously been difficult due to the number of data requirements and the need for ultralow RTT connections for real-time interactions. Once these demands are met, various practical implications can be investigated and accomplished with 5G and beyond mobile communications. However, TI can be utilised to reinforce multiple impacts in areas such as ICT, military service, the healthcare industry, e-learning and pedagogy. It could play a pivotal role in demolishing society’s financial limits [1].

The usage of mobile internet has tremendously increased in people of all ages and nationalities. Everyone can be connected wirelessly without facing any difficulties of place and time. The tremendous research and improvements in networking and communication have led to a remarkable development in internet usage, thus creating a space for companies to produce and build up the next generation of gadgets and enhance the user experience. Depending on the Internet, communication is comprised of human beings communicating

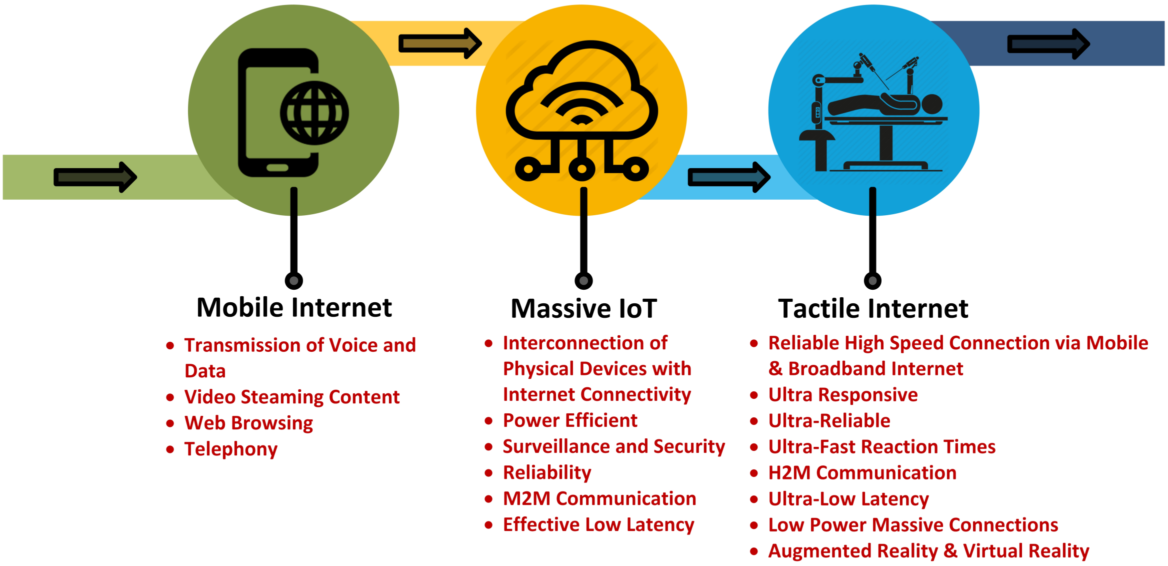


Fig. 3.1 The evolution of TI highlighting its key competencies.

with single or multiple devices. There is a requirement for a communication interface where humans interact with other humans by touch or actuation. Subsequently, this requirement has been in great demand; however, it has not been accomplished on a larger scale.

Ensuring transparency in the system, the most crucial objective is to attain the RTT of 1 ms. When the first person executes any task at one location and the second person recognises it at another location/s with some latency, it is called motion sickness or lag. Numerous analyses exemplify that motion sickness will be noticeable if RTT is more than 1 ms in haptic communication [63]. If this motion sickness is observable, it adversely affects transparency.

The factors contributing to the RTT are queuing, processing and transmission times. These are the fundamental elements that influence haptic communication in the network. However, the most compelling component of RTT that restricts smooth haptic communication is the speed of light. To reduce the RTT up to 1 ms, the transmission times should be limited to $\leq$ 1 ms and are appropriate for short-range communication. Figure 3.1 depicts the evolution of TI from mobile internet to massive Internet of Things (IoT) and eventually to TI. Each evolution step states some key competencies that are required for an accessible mode of communication.

After the evolution of mobile Internet, IoT is a concept that involves the communication of two or more devices through the integration of the Internet and cloud. The applications related to IoT are investigated more as they offer ample opportunities to ease the way of life. However, TI exists at the conceptual stage, but much research is still needed to revolutionise it.

Nevertheless, as IoT is a promising cutting-edge technology, adding tactile sensors increases human-machine (real-time) interactions in the TI paradigm. Hence, features such as ultralow latency, ultrahigh reliability, high security, and many more [193] are achieved.

Therefore, to have traffic flow in the network, RTT, reliability, and other performance metrics depend on the aggregation of nodes, such as sensors and actuators available in the network. Hence, the nodes should be located near the end-user devices and server stations, and thus, the distance between them should be minimised. Subsequently, the task can be processed quickly, and acknowledgement can be sent to the source. TI technology depends on cloud computing (mobile-edge or fog computing) for sensory and haptic controls. It will add a new paradigm to human-machine interaction by providing a low latency enough to implement real-time interactive systems [194]. Thus, by utilising the leading technologies such as Mobile Edge Computing (MEC), Fog Computing (FC), Software-Defined Networking (SDN) along with Network Function Virtualisation (NFV) and 5G communication framework, these technologies facilitate practical solutions to overcome the challenge of 1 ms.

Although the previous studies looked at embedding and implementing cloud and SDN based solutions precisely for TI systems, they are not focused on efficiently routing the user's workloads (requests) through multi-path communication. Also, the mathematical model includes the wait and processing times that further add to the overall RTT of the system. Furthermore, the existing works do not discuss and evaluate crucial performance metrics such as ultra-high reliability and energy consumption. Here, we briefly summarise the main contributions that will resolve the shortcomings of the existing works as follows.

- We propose a fog-based traffic flow framework for realising TI by employing SDN and FC approaches.
- We develop an efficient traffic flow algorithm to reduce extra processing and waiting times at each level of cloud in the system.
- We implement the proposed TI framework in the iFogSim simulator and perform an extensive simulation study to compare the system performance with the existing edge, cloud, and cellular networks.

3.2.1 Structure of the manuscript

The rest of the manuscript is structured as follows. Section 3.3 discusses motivation and provides a review of the literature on TI. The design choices and description of the proposed TI system are discussed in Section 3.4 and 3.5, respectively. The performance evaluation of the proposed system is discussed in Section 3.6. Finally, a brief conclusion in Section 3.7 ends the manuscript.

3.3 Motivation and background

3.3.1 Motivation

One of the reasons for carrying out the research is the TI's emerging research area, where there is a need to improve the RTT and reliability in the system. This improvement will ensure and enable the smooth functioning of the system with cyber or motion sickness (lag). The crucial challenges to achieving tolerable RTT (1 ms) and ultra-reliability (99.999%) [24] of communication are essential to facilitate the services and applications related to the TI. Still, there are many challenging solutions to reduce the delay over the network. Therefore, we have been motivated to adopt the TI technology by incorporating one of the technological advancements, SDN and cloud networking technologies, with the 5G framework, which promises to serve the requirements of TI.

As TI will serve critical applications in society, it will ensure connectivity to numerous devices for communicating with each other simultaneously and autonomously. Interacting with ongoing and conventional wired Internet, mobile Internet, and IoT should also be feasible, thus forming a network of entirely new possibilities and opportunities.

3.3.2 A review of literature

The potential of TI

In a characteristic advancement to various internet encapsulations, the thought of TI [22] is rising in which ultra-responsive, ultra-reliable, and perception of the connectivity of network will empower the remotely operated real-time control and interpersonal haptic communications. The TI will include a new area of measuring human and machine interactions by constructing interactive and intelligent frameworks. The TI will facilitate a genuine change in outlook from content delivery to skill-set delivery networks and, subsequently, change every part of society. As of now, the Institute of Electrical and Electronics Engineers (IEEE) and European Telecommunications Standards Institute (ETSI) [127] standards are also incorporating standardised activities for the TI.

Furthermore, 5G mobile communication systems are required at the wireless edge to support TI. The research project has been carried out at Mobile and wireless communications

Enablers for the Twenty-twenty Information Society (METIS) [79] along with the telecommunication industry Next Generation Mobile Networks (NGMN) alliance [14] regarding the early evaluation of 5G scenarios. Dealing with TI, there has been an increase in the new requirements and challenges for designing the 5G network [128].

Moreover, content delivery such as voice and video communication, instant messaging, etc. and data accumulation applications are being utilised in the era of wireless communication. Lately, the thought of TI [22] has risen, and it is imagined that it will change the perspective by empowering wireless for ongoing steering and control communications. The working group having a standard as IEEE P1918.1 [29] refers to the TI as a metadata set of network, mainly for wirelessly accessing the data, receiving, manipulating, or managing the real and virtual data, thus processing in perceived real-time. To have a communication medium for transferring sensing or actuation data, touch and controlling information in real-time, TI must possess competencies such as high reliability, responsivity and cognitive behaviour. The high accessibility, ultra-fast response times, and carrier-grade reliability of the TI will add another feature to human-machine communication by making continuous intelligent frameworks [1].

SDN-based design

While the widespread adoption and application of SDN are relatively recent, it facilitates a dynamic, flexible, and cost-effective system structure by physically partitioning the data forwarding and control planes in networking. The control plane handles the suitable choice to deal with traffic, whereas the data plane takes care of forwarding the traffic according to the control plane. SDN grants permission to the system operator to design, configure, control, and deal with the system through programming software, such as application programming interfaces (APIs) or controllers. SDN is mostly utilised to disentangle the system hardware and increase network adaptability.

Taking existing literature into account, M. Gharbaoui et al. [171] have presented a service-chaining orchestration framework comprising dynamic virtual function (VF) selection and intent-based traffic steering control competencies to upgrade Management and Orchestration (MANO) framework parts with latency-minimised and self-adaptive service chaining highlights over geographically distributed SDN-based cloud data centres (DCs) interconnected through SDN Wide Area Network (WAN). Furthermore, Yi-Wei Ma et al. have proposed a Software-defined Infrastructure (SDI) [172] with NFV, which is then deployed in an Industry 4.0 network environment. This infrastructure uses the technological concept of

SDN and NFV, which enhances the overall activities of the network environment, the rate of information transmission and the Quality of Service (QoS).

Along similar lines, Sebastian Troia et al. [170] have presented their work on the SDN/NFV Orchestrator, a new service orchestrator called SENATUS. It facilitates a service orchestration for network segments, deploying OpenStack as infrastructure manager and Open Network Operating System (ONOS) as SDN controller.

With a vast arrangement of small cells, a 5G-based SDN framework is introduced by K. Wang et al. [174]. However, there are some challenges to constant handover and delays incurred through the handover procedure. Therefore, they found a solution by providing a design that introduces an SDN controller at the core system. However, the framework is mainly concerned with latency incurred through a handover procedure and is prepared to track the movement of end users, not allocating radio access resources. Nevertheless, cognitive control should be deployed at the core system to attain the desired RTT of 1 ms in the TI framework. In addition, an unmanned aerial vehicle (UAV) based video surveillance with TI requirements is designed by Christian Grasso et al. [67], where the drone-based Markov chain model is examined.

Taking the SDN architecture into consideration for ultrareliable and ultralow latency communication, Liberato et al. [165] have introduced a residue-defined networking architecture (RDNA) as a unique strategy for enabling critical competencies of micro-DC networks. This architecture is built on the concept of SDN to meet complex and stringent demands from next-generation networks.

In addition, the work in [166] by Juan Cabrera et al. have featured the significance of SDN and NFV for 5G communication along with TI applications. Moreover, they have also described a testbed to create 5G infrastructure. It describes that SDN and NFV complement each other and facilitate empowering technologies for facilities such as computing, storage, and networking functions. Furthermore, Perepelkin et al. [167] have proposed an effective algorithm and software for virtual slice formation in SDN dependent on the information of base network and links connectivity information. This research focuses on utilising the proposed approach for addressing the issue of adaptive routing in SDN dependent on the OpenFlow protocol.

FC-based design

Considering the evolution of FC, Flavio Bonomi [12] has envisioned the key characteristics of FC with new services and applications. FC can be referred to as a highly virtualised platform used to facilitate a significant proportion of computation, storage, and networking resources, where a huge number of end devices communicate with traditional cloud DCs. However, the end devices are not exclusively located at the edge of the network. The end devices comprise routers, switches, gateways, etc., and are equipped with all FC functionalities that are expanded to traditional cloud computing and its services.

TI is a developing idea that spotlights on supporting high-fidelity, ultra-responsive and generally accessible machine-to-human (M2H) communications. Having a foundation of IoT for FC, it is primarily used to control the end devices remotely. FC has been posted as a significant part of the TI to minimise the RTT and transmission times. The FC nodes (FCNs) are based on the FC decentralised structure and can be placed between the traditional cloud and end-user devices at various positions. Before pushing it to the cloud, it processes and saves data from devices like wireless routers and machine-to-machine (M2M).

Xiao and Krunz [168] have focussed their research on an energy-efficient design of fog network that helps to have low latency supported TI applications. Here, they have explored the performance parameters such as the service reaction time of the end-users and the power utilisation efficiency of fog nodes. Also, a novel cooperative FC concept was presented with a set of fog nodes with various processing and energy resources that can coordinate with one another. This led to the fulfilment of balancing the workload or the request executed by multiple fog nodes.

Furthermore, the idea of edge-assisted cloud computing and its connections is related to the developing area of FC. Such frameworks use low-power and efficient embedded computers to facilitate nearby processing close to end-users or the cloud. FC becomes feasible by moving some of the computation and decision-making units to the edges, either close to the end-user layer or cloud backend. The Fog of Things (FoT), analogous to Fog Computing (FC), alludes to the amalgamation of various fog nodes that could be connected and communicated with each other. All the extensive processing data, long-haul storage, and investigation could happen in the cloud layer as the last destination.

FC has been acquainted with a bright arrangement to hold the extreme necessities of the TI. It supplements the cloud framework by including an enormous amount of low-cost, often decentralised equipment, usually alluded to as fog nodes. Fog nodes epitomise the numerous devices or equipment between end-users and cloud DCs, including routers, access points (APs), routers, base stations (BSs), and also light weighted gadgets such as robots, drones, and vehicles with processing and storage capabilities. The achievement of the TI will rely on the extensive deployment of fog nodes with high processing capabilities and a reliable energy supply. Nonetheless, the hypothetical establishments for advancing distributed FC systems to satisfy the needs of the TI are yet deficient. Specifically, computationally escalated services needing low delays usually request more energy consumption from fog nodes.

The International Telecommunication Union (ITU) and Next Generation Mobile Networks (NGMN) alliance have recognised FC as one of the crucial factors for the TI to accomplish ultra-low service delay for users [26, 14, 180]. Rather than CDCs, fog nodes can be deployed near the end-users so that the task transmission delay can be substantially diminished. Notwithstanding, because of each fog node's limited computational or processing ability, carrying out a considerable number of tasks to fog nodes will bring about a high processing delay. Hence, previous works have discussed the most proficient method for resource provisioning the fog nodes to decrease the processing delay. For instance, a virtual machine (VM) synthesis technique was presented to enable each end-user to rapidly arrange the resources of neighbouring fog nodes and make the required VM images to help a particular application.

Taking the architectural designs of FC into consideration, Bonomi et al. [12] have demonstrated one of the essential jobs of FC, examining the usefulness of FC for IoT. Therefore, FC managed the issues related to the prerequisites of emerging information on the site, real-time user communication, services, and its application for IoT. When FC (edge computing-based) applications are considered, the philosophy of distributed computing is connected to different developing innovations and IoT applications that exploit various merits offered by edge computing. Also, V. Fanibhare et al. [161] proposed a multi-level cloud structure where the traffic flows over the cloud units efficiently to reach the slave section from the master section to reduce unnecessary waiting times. However, the authors did not include the mathematical model and extensive performance evaluation in their research.

Along similar lines, a multi-tier infrastructure [181] is introduced in which edge cloud servers are differentiated into various levels as per their separations to the edges. If the task on the edge cloud server of a given level surpasses its computational limit, the additional outstanding task is sent to the higher-level service.

In addition, Xuejiang Wei et al. [157] have proposed a new fog-based sensor cloud architecture for IoT, in which they have facilitated the fundamental ways for physical sensor virtualisation, dynamic provisioning of virtual sensor groups and service instances for key issues. Furthermore, Jinke Ren et al. [164] have explored the joint communication and computation resource allocation to minimise the weighted-sum delay (latency) of all devices in a cloud-edge collaborated system.

FC has its capabilities and competencies as follows. It has location awareness and can be located at the edges of the network. It incurs low latency, which supports large scale complex sensor networks and storage capabilities in a wide-area distributed network. Any faults in the network can be analysed online and integrated with backend cloud computing.

Challenges and requirements related to TI

The software-defined network and fog/edge computing are a few potential enablers that can be leveraged to tackle the anticipated challenges related to computation and communication between end devices. If these potential enablers are combined, more benefits can be achieved by tackling the challenges such as RTT, high reliability, an excellent exchange rate between the end devices and fog/edge computing components in the TI system. However, the effectiveness of these enablers is limited to a maximum distance of 150 km between the end devices, as the speed of light restricts them. To address the challenge of 1 ms RTT, prediction-based algorithms (AI) [64] are required to be utilised in local and remote environments along with the current potential enablers. Seemingly, the achieved 1 ms RTT can deliver a real-time user experience, eradicate the 150 km limitation and maximise the communication distance beyond the limit. In addition, there is a further requirement to increase the network bandwidth, which improves RTT and ensures QoS user requirements through the adoption of SDN and FC.

Since an SDN controller and switches are mainly responsible for efficiently routing the traffic between fog (cloud) units and end devices, dynamic network slices can be created to cater for the requirements of multiple applications. For instance, one network slice can be used to improve the RTT criterion, and another can be used to optimise the throughput or reliability criterion depending on the tactile user's requirements. Thus, it helps conserve the network bandwidth, minimise RTT, achieve high reliability, and enhance the QoS of the system. Furthermore, fog/edge computing carries out a substantial amount of computing and storage in TI, offering significant minimisation in RTT. However, resource allocation in fog/edge computing is also another challenge to meet the stringent requirements of TI. A feasible option of offloading the computational task closer to the wireless end devices could be considered.

3.4 Design choices for TI system

In the context of haptic communication, TI mainly demands competencies such as ultra-low RTT and ultra-high reliability. To reduce the motion lag, RTT has to be $\leq$ 1 ms. This RTT arises due to processing, communication, queuing and transmission times. Therefore, the tactile action and tactile feedback should be within the limit of 1 ms. Hence, this limit enables smooth haptic communication, which is relatively less than human reaction time [1].

For smooth, haptic communication between master and slave sections, we surveyed the existing literature review. After identifying and analysing the existing work, the main technologies such as 5G, cloud computing, fog computing and SDN; will ensure effective communication between the master and slave sections. The following are the few design choices that the proposed system can adopt.

3.4.1 Cloud computing with 5G framework

This is the first design choice that the proposed system can adopt. By adopting cloud computing, all the computationally complex operations will be carried out in the cloud. Cloud computing has several benefits such as reduced IT costs, ease of implementation, scalability and flexibility for growth. However, cloud computing incurs some demerits such as network connectivity dependency, downtime in cloud services, high dependency on the internet connection, higher bandwidth requirements, more latency while transferring information to the cloud.

3.4.2 Fog computing with 5G framework

This is the second design choice that the proposed system can adopt. Fog Computing (FC) serves as merit over traditional cloud computing (CC). FC provides a decentralised computing infrastructure where the processing costs such as system response time and CPU-memory-bandwidth utilisation, storage cost, operational costs such as maintenance & configuration costs are considered. In addition, FC is an extension of CC services. It comprises numerous fog nodes which are associated with physical end devices. These nodes are located very near the end devices, so nodes offer better services than CC. Also, the fog nodes possess better processing power to execute complex & faster computational tasks and utilise the node resources of CC. Thus, it has a more secure infrastructure than CC. In order to realise FC, the fog nodes are deployed physically closer to the end devices instead of directing data from the centralised server to the cloud. However, all the computational processing will be carried out in a decentralised FC infrastructure, leading to more computational load. Depending upon the nature of the data, various levels of cloud structure might be needed to process and execute the complex data, which might require higher computing resources.

3.4.3 Fog computing with multilevel cloud structure and 5G framework

This is the third design choice that the proposed system can adopt. Here, we can opt for the multilevel cloud-based structure for TI, keeping an idea of fog nodes. The multilevel cloud-based structure comprises three levels, i.e. micro clouds, mini clouds and a central cloud. This multilevel cloud-based structure employs the working of a decentralised FC infrastructure. In addition, three micro clouds are clustered, which are referred to as a group of fog nodes, present at the first level. In contrast, mini clouds can also serve as fog nodes located at the second level of the cloud-based structure. These fog nodes get attention and are flexible to comply with the varying availability and performance demands of critical applications. As fog nodes, they can reduce reliance on centralised cloud data centres, improving speed, security, and scalability in the TI system. They can share the workload (task) and centralised storage. Even if one node fails, it provides failover (backup) or isolates services. They improve performance by moving the task from one server to another, and they support horizontal scalability, non-failure reliability, and ease of maintenance.

Based on the geographic locations, the clusters of microclouds are deployed at the first level of the cloud-based structure. The microclouds have a relatively low task computational capacity to compute, store, communicate, and process the information within a single cellular BS range. In other words, micro clouds are small, localised computing units deployed at the edge to process real-time sensory and haptic data, ensuring minimal latency. This information is received from user equipment (UE) connected to BS.

At the second level, the mini clouds are deployed at the cellular aggression point and process the information from BSs present in the proximity if the tasks are left unprocessed at the micro cloud's level. In other words, mini clouds aggregate data from multiple micro clouds, acting as regional processing hubs before forwarding critical data to the central cloud. As shown in Figure 3.2, a cluster of three micro clouds is associated with one mini cloud. Both micro and mini clouds serve as fog nodes, thus reducing latency and enhancing reliability.

Finally, the unprocessed tasks will be passed further to the central cloud level (i.e. main server) following the hierarchy of the multilevel cloud structure. The central server has a relatively higher computational capacity than the mini clouds and is followed by micro clouds. Being a central cloud at the core, it is deployed at a distant location from end-user devices. Some controllers still need to direct/route the data into the central cloud.

3.4.4 Fog computing with multilevel cloud structure and SDN with 5G framework

This is the fourth design choice that the proposed system can adopt. With the functions of central, mini, and micro clouds in a multilevel cloud structure, we can opt for a decentralised SDN controller and Vswitches to efficiently control and route the traffic (task) in the 5G framework. Such design allows efficient use of computing resources, such as servers, memory, storage, etc. The resulting system can achieve the following merits by integrating the multilevel cloud-based structure for TI along with SDN and FC.

- Minimises system load due to distributed decentralised framework. Hence, reduces the congestion and RTT issue.
- Enhances flexibility and, therefore, control over traffic flow in the system.
- Boosts up the performance and security of the system.
- Implements new services through cellular networks.
- Achieves better system efficiency because of dynamic task offloading from end users.

3.5 Description of the proposed system

To develop a system by looking at the design choices for the TI system, it is crucial to distinguish between the merits and demerits of each design choice. After analysing each design choice, the fourth design choice, i.e. the multilevel cloud-based structure along with SDN and FC, is adopted as the proposed system for TI, as illustrated in Figure 3.2. This manuscript aims to facilitate insight into design choices for the proposed system, which will ensure smooth haptic communication between two tactile ends.

In the proposed system, the SDN controller administers the flow of traffic and routes the traffic according to the proposed traffic flow algorithm, as shown in Algorithm 1, for enhanced system management and efficient service performance. Constructing the cluster of micro clouds, the excellent utilisation of cloud space is made. Most of the cloud vendors such as Microsoft Azure [195], Amazon Web Services (AWS)[196], Google Cloud Platform [197], IBM Cloud [198] and Alibaba Cloud [199], have an optimum value of around 70%-80% for cloud CPU capacity (upper threshold) while configuring/deploying the cloud and its services. Therefore, every cloud in the proposed system is limited to use up to 80%, referred to as

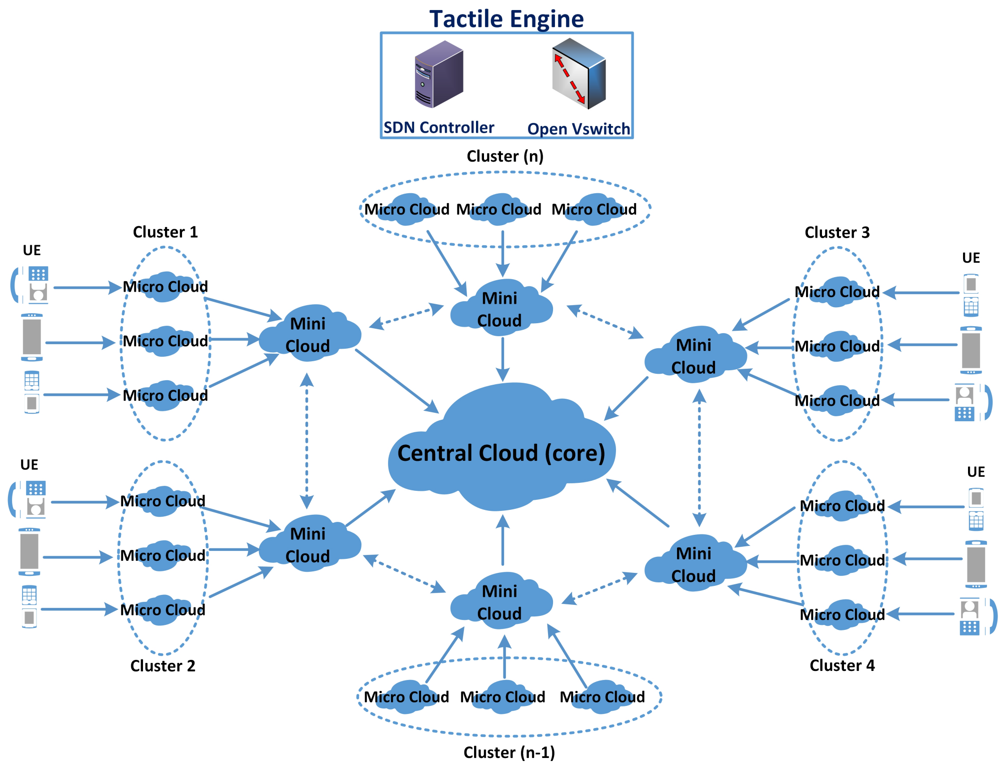


Fig. 3.2 The proposed fog-based framework for TI.

cloud capacity. Hence, the traffic routing rate and cloud capacities are managed by the SDN controller and Vswiches. Ultimately, the primary purpose is to reduce the extra processing and waiting times at each level of the cloud-based structure. Figure 3.3 presents a flowchart depicting efficient traffic flow patterns to enhance understanding and analysis of the proposed TI system.

In a multilevel cloud-based structure, the following explanation is for level-wise paths followed by traffic (task) from master to slave sections, as represented in Figure 3.4.

- **At** 1^{st} **level:** The master section allocates the task to the slave section. The task from the end-user first reaches the first level of the cloud-based structure, where a cluster of three micro clouds is deployed. Here, the incoming task checks the condition of cloud capacity (i.e. 80%). If the existing micro cloud is busy executing its previous task and operating at $< 80\%$, then the task is forwarded to the next micro cloud within the same cluster with the help of a task (load) balancer, and so on. In a situation where all the micro clouds are busy executing their previous task and operating at $< 80\%$, the task balancer forwards the incoming task, which skips the first level of micro clouds to the mini cloud, i.e. the second level. If this cloud level is busy executing its previous

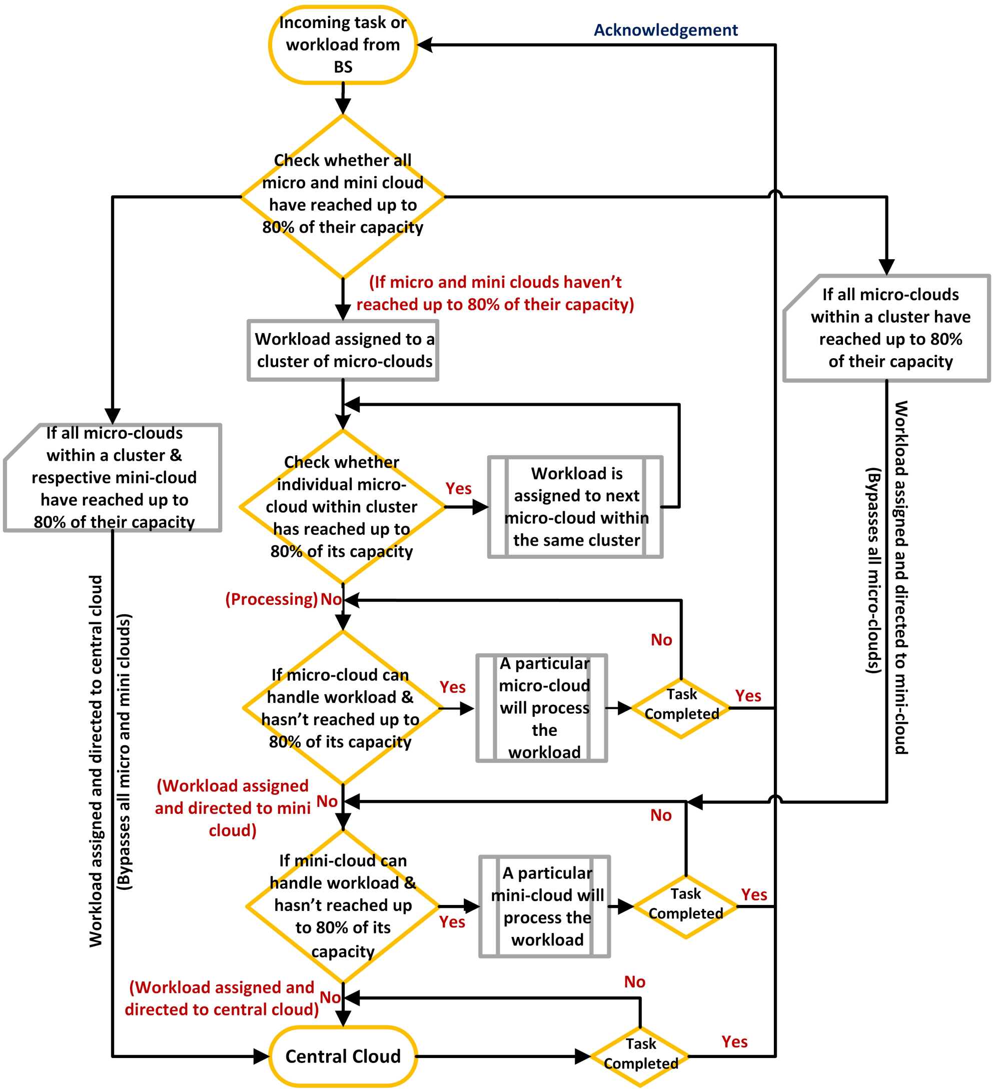


Fig. 3.3 A flow chart for efficient traffic flow patterns in the proposed TI system.

task and the cloud capacity condition is not applied, the new incoming task will have to wait and queue up at that level. Therefore, the new task will wait for its turn to get processed. Consequently, this waiting time will add delay to the processing of the task.

- **At** 2^{nd} **level:** The incoming unprocessed and skipped task from the first level reaches the mini cloud, where it again checks for the condition of cloud capacity. If the particular mini cloud is operating at $< 80\%$ cloud capacity, the task is processed, and the acknowledgement is sent back to the master section. Moreover, if the condition of cloud capacity is not satisfied, the task skips the second level of the mini cloud with the help of the task balancer and reaches the central cloud, i.e., the third level. Suppose the task is getting routed from one level to another of the multilevel cloud-based structure

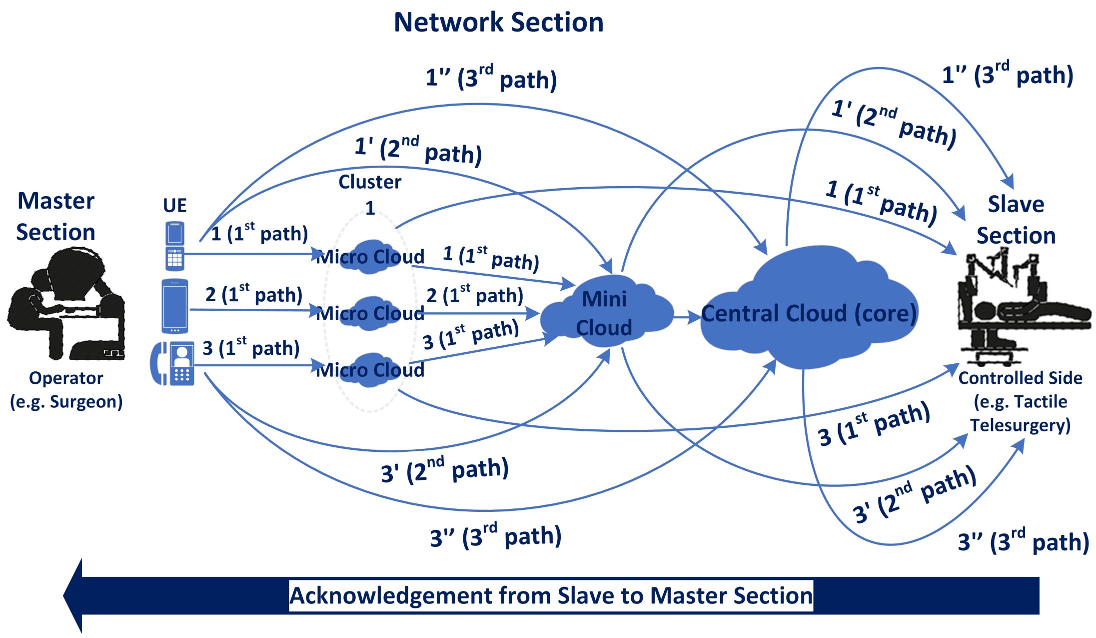


Fig. 3.4 Illustrating the traffic flow paths in multilevel cloud structure of fog-based framework.

(for instance, from micro cloud to mini cloud) depending on the condition of cloud capacity. In that case, the waiting and processing times are saved significantly at the existing level of the cloud-based structure. Also, by moving to the upper level of the cloud-based structure, the task will be processed faster as it would have higher computational capabilities. The principle objective of skipping from one level to another is to reduce RTT and improve system efficiency.

- **At** 3^{rd} **level:** Finally, the task that is left unprocessed and skipped the first and second levels of the cloud-based structure reaches the central cloud to handle complex processing. Hence, after each processing and completion of the task at any level, the slave section will be teleoperated, and acknowledgement will be sent back to the master section.

3.5.1 Mathematical model

For the proposed system for multilevel cloud-based structure in the FC environment, we develop a mathematical model for improving RTT (i.e. end-to-end delay). Let t be the incoming task or workload from the master section where the user equipment (UE) is present.

This task is a set of i independent tasks sent to the cloud as follows:

$$t = [t_1, t_2, ..., t_i] \tag{3.1}$$

Let μ_{cloud}, m_{cloud} and c_{cloud} denote the micro, mini and central clouds, respectively. Here, we are mathematically representing the model for just one section having three μ_{cloud}, one m_{cloud} and one c_{cloud}. Hence, this model can be iterated across the remaining paths for other sections.

When particular cloud processes the task (t_i), let t_c be the task completed at that level of the cloud-based structure and ACK be the acknowledgement sent back to the master section. Based on the user's (UE) demand, the rate of tasks (t_i) generated by UE coming to the μ_{cloud} varies. Therefore, according to the poison's process with a poison rate of λ_j, we assume that the tasks are randomly coming to the μ_{cloud}.

The RTT in our proposed system is comprised of a task's response time and a communication delay. The average response time is given by the sum of waiting (queue) time and the processing time for tasks at μ_{cloud} and m_{cloud}.

For the calculation of processing time ($P(t_i)$), we are adopting a multi-server queuing model *M/M/s* [200], wherein S_μ and S_m represent the number of servers in μ_{cloud} and m_{cloud}. Thus, the average processing time is given by a function of arrival rate (λ), based on the Erlangs' C formula derivated in [201, 202].

The average processing time for μ_{cloud} is given by,

$$P_\mu(t_i) = \frac{C(s_j, \frac{\lambda_j}{\mu_j})}{s_j\mu_j - \lambda_j} \tag{3.2}$$

The average processing time for m_{cloud} is given by,

$$P_m(t_i) = \frac{C(s_m, \frac{\lambda_m}{\mu_m})}{s_m\mu_m - \lambda_m} \tag{3.3}$$

where, λ_j and λ_m are the arrival rates of μ_{cloud} and m_{cloud}, respectively, whereas μ_j and μ_m are the corresponding service rates of μ_{cloud} and m_{cloud}, respectively.

Hence, if the task (t_i) is processed at μ_{cloud}, then the total RTT is given by,

$$RTT_\mu = t_\mu(wait) + P_\mu(t_i) + d_{UE \to \mu_{cloud}} \tag{3.4}$$

where, RTT_{μ}, $t_{\mu}(wait)$ and d_{UE} represent the RTT, waiting time and communication delay from UE to μ_{cloud} at μ_{cloud}.

Secondly, if the task (t_i) is processed at m_{cloud}, then the total RTT is given by,

$$RTT_m = t_m(wait) + P_{\mu}(t_i) + P_m(t_i) + d_{UE} + d_{\mu_{cloud} \to m_{cloud}} \tag{3.5}$$

where, RTT_m, $t_m(wait)$ and $d_{\mu_{cloud} \to m_{cloud}}$ represent the RTT, waiting time and communication delay from μ_{cloud} to m_{cloud} at m_{cloud}.

Thirdly, if if the task (t_i) is processed at c_{cloud}, then the total RTT is given by,

$$RTT_c = t_c(wait) + P_{\mu}(t_i) + P_m(t_i) + P_c(t_i) + d_{UE} + d_{\mu_{cloud} \to m_{cloud}} + d_{m_{cloud} \to c_{cloud}} \tag{3.6}$$

where, RTT_c, $t_c(wait)$ and $d_{m_{cloud} \to c_{cloud}}$ represent the RTT, waiting time and communication delay from m_{cloud} to c_{cloud} at c_{cloud}.

3.5.2 Proposed traffic flow algorithm

The existing traffic-flow algorithms [203, 175, 184, 169] mainly focused on the total latency of various cloud levels. It is not clear how the workloads (requests) from users can be routed to get processed in the cloud capacity threshold. Our proposed traffic-flow algorithm efficiently reduces the waiting and processing times at every cloud level, which Vswitches and an SDN controller thus manage. Hence, the multipath communication further decreases the RTT and increases the throughput and resiliency of the system.

The proposed efficient traffic-flow framework algorithm is shown in Algorithm 1, where the algorithm is iterated across all paths (p) having three μ_{cloud}, one m_{cloud} and one c_{cloud} sections of the multilevel cloud-based structure. To set the threshold of all cloud units, $C_{cap}^{80\%}$ (cloud capacity) was set to 80%. The traffic flow algorithm has been developed to reduce extra processing and waiting times at each level of the cloud-based structure in the system.

The proposed traffic flow algorithm demonstrates the flow of traffic between the master and the slave sections. The master section receives the incoming task (t_i) as input. Firstly, we have initialised the count (j) as zero. This algorithm iterates over all end-user devices (UE) to the central cloud (c_{cloud}) in the physical system simulation model and places t_i on each UE for each such path. For each UE in a path (p), the t_i that can be placed on it is identified and reaches μ_{cloud}. Firstly, the algorithm first executes the *For Loop*, where the *variable individual*$\mu_{cloud}(j)$ is given the value of ≤ 2. Secondly, the t_i is assigned to the *variable individual*$\mu_{cloud}(j)$ and processes the task $P(t_i)$ only if all *individual*$\mu_{cloud}(j)$ is

Algorithm 1: Proposed traffic flow algorithm

Input: Incoming task (t_i) from master section.
Output: Efficient flow of traffic in proposed framework that optimises the cost in FC environment.

Initialise: $j = 0, t_i$
for $p \in across\ all\ PATHS$ **do**
    for $individual\ \mu_{cloud}(j) \leq 2$ **do**
        if $all\ individual\ \mu_{cloud} \leq C_{cap}^{80\%}$ **then**
            $individual\ \mu_{cloud}(j) \leftarrow t_i$
            $P(t_i) = t_i$
            if $t_c = t_i$ **then**
                $master \leftarrow ACK$
            else
                goto $step$ 17
            end
        else
            $j + 1 \leftarrow j$
            goto $step$ 3
        end
    end
    if $respective\ m_{cloud} \leq C_{cap}^{80\%}$ **then**
        $m_{cloud} \leftarrow t_i$
        $P(t_i) = t_i$
        if $t_c = t_i$ **then**
            $master \leftarrow ACK$
        else
            goto $step$ 26
        end
    end
    if $c_{cloud} \leq C_{cap}^{80\%}$ **then**
        $c_{cloud} \leftarrow t_i$
        $P(t_i) = t_i$
        if $t_c = t_i$ **then**
            $master \leftarrow ACK$
        end
    end
end

$\leq C_{cap}^{80\%}$. If the condition is true, the algorithm checks whether t_i is completed ($t_c = t_i$), and then an ACK is sent to the master section. Otherwise, the incomplete task will be forwarded to the $respective\ m_{cloud}$.

Considering a $individual\mu_{cloud}(j)$, if $\leq C_{cap}^{80\%}$ condition is not fulfilled, then the count (j) is incremented, i.e. $j+1$ and the execution process is jumped to the initial step where the task is allocated to the next $individual\mu_{cloud}(j)$. Again, the condition of $\leq C_{cap}^{80\%}$ is checked. If true, the t_i is assigned to the current $individual\mu_{cloud}(j)$ within a cluster of three $individual\mu_{cloud}(j)$. For each subsequent time, the algorithm is executed when the loop $variable\ individual\mu_{cloud}(j)$ is incremented by 1, i.e. $j+1$ until the condition $individual\ \mu_{cloud}(j) \leq 2$ satisfies. When the $individual\mu_{cloud}(j)$ exceeds the value 2, the particular *For Loop* is no longer executed.

In a case where the 1^{st} μ_{cloud} is busy processing its previous task, the incoming task t_i is directed to the 2^{nd} μ_{cloud} within the same cluster. Moreover, t_i will not have to wait for its turn to be processed. Therefore, the Equation 3.4 will deduce to the following equation.

$$RTT_{\mu} = P_{\mu}(t_i) + d_{UE \to \mu_{cloud}} \tag{3.7}$$

In a case where all $individual\mu_{cloud}(j)$ are operating at $> C_{cap}^{80\%}$, the incoming task t_i will skip the μ_{cloud} level and reaches m_{cloud}. Hence, the waiting time, processing time and communication delay from μ_{cloud} to m_{cloud} will not be contributed to the calculation of RTT. Consequently, the Equation 3.5 will deduce to the following equation.

$$RTT_m = P_m(t_i) + d_{UE} \tag{3.8}$$

When the forwarded (incomplete) task reaches the respective m_{cloud} from μ_{cloud}, it checks for $\leq C_{cap}^{80\%}$ condition. If the condition is true, then t_i is assigned to m_{cloud}, processes $P(t_i)$, checks for the completion status ($t_c = t_i$) and an ACK is sent to the master section.

In a case where both $individual\mu_{cloud}$ and m_{cloud} are busy processing their previous task and operating at $> C_{cap}^{80\%}$, the incoming task t_i will skip μ_{cloud} and m_{cloud} levels, and reaches c_{cloud} and thus processes t_i. Similarly, the Equation 3.6 will deduce to the following equation.

$$RTT_c = P_c(t_i) + d_{UE} \tag{3.9}$$

Overall, the proposed traffic flow algorithm saves waiting times, processing times, and communication delays between each level of the fog-based framework.

3.5.3 Computational complexity

The computational complexity of the proposed traffic-flow algorithm is low in terms of CPU time/utilisation to process user's requests/tasks.

To analyse the computational complexity of the proposed algorithm (Algorithm 1), let us break the algorithm into fragments. Consider the first fragment having an '*If*' statement starting from steps 7 to 11. This fragment will always run a constant amount of time of order 1. Hence, the computational complexity for the first fragment is $O_1(1)$. For the '*If*' statement starting from steps 4 to 15, we consider these steps as the second fragment. Here, the condition is checked to determine whether all individual micro clouds are operating under 80% of cloud capacity. This second fragment will also have a constant time of order 1, and thus the computation complexity of the second fragment is $O_2(1)$.

Assuming '*For*' loop starting from steps 3 to 16 as the third fragment, the loop will always run precisely 3 times. Since '*For*' loop is running for a constant time, the computational complexity of the third fragment is $O_3(1)$. Similarly to the second fragment, the fourth and fifth fragments, starting from steps 17 to 25 and steps 26 to 32, the computational complexities are $O_4(1)$ and $O_5(1)$, respectively.

Therefore, when you have '*For*' loop and multiple '*If*' statements from steps 3 to 32, the equivalent computational complexity can be obtained by multiplying the individual computational complexities, i.e. $O_e(1*1*1*1*1) = O_e(1)$.

However, the outer '*For*' loop (from steps 2 to 33) consist of the inner '*For*' loop and two '*If*' statements. As paths p increase, you can multiply their computational complexities of '*For*' loop having paths p as $O_p(P)$ and equivalent $O_e(1)$ to achieve the overall computational complexity. Hence, the (worst case) overall computational complexity becomes $O_o(P*1) = O_o(P)$ for the complete block of code. Therefore, it is easy to implement in practical scenarios.

3.6 Performance evaluation

3.6.1 Simulation environment

We evaluate the efficacy of the proposed system using the iFogSim simulator, a Java-based high-performance open-source toolbox. This simulator is built upon the CloudSim framework, which is comprehensively used for cloud-based scenarios. We opted to use the iFogSim

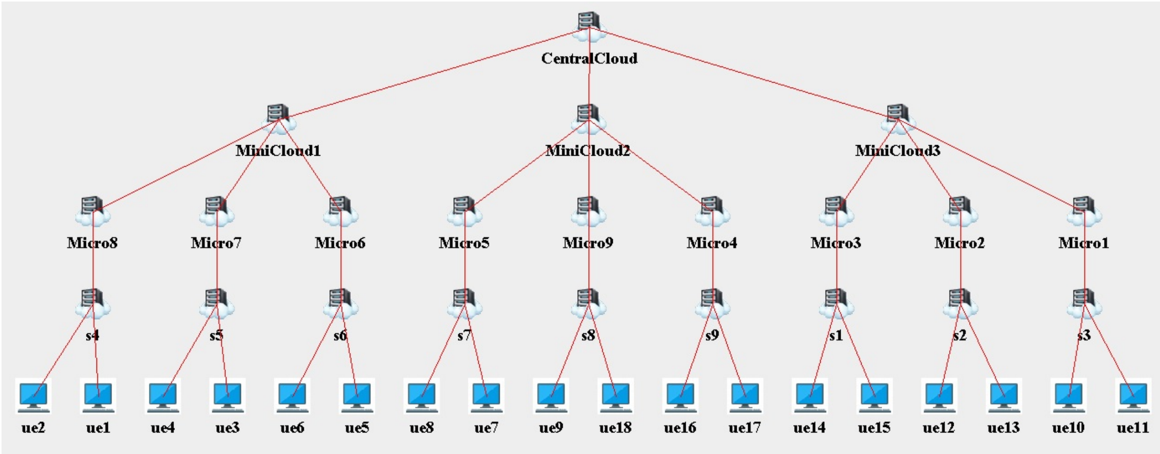


Fig. 3.5 The system simulation model.

simulator since it is feasible and available to carry out our simulation tasks of IoT networks with more TI requirements.

The simulation results from the proposed system showed steady-state behaviour with a relative statistical error of $\leq 5\%$ at the 95% confidence level. A relative error of $\leq 5\%$ will be introduced because of the placements of the fog nodes in the network. Besides, there might be a chance of not attending to the task when coming to the cloud levels. On the contrary, a 95% confidence level signifies the reliability of the estimation procedure of traffic flow in the framework but not to a specific interval. This confidence level represents that the task coming to the clouds are successfully routed according to the proposed traffic flow framework. The length of the simulation is set to 60 minutes, where an initial transient period is concise. Also, no observations are made during the first minute of the transient period and, therefore, excluded from the final simulation results.

3.6.2 Modelling the network

The simulation model of the proposed system is composed of a multilevel cloud-based structure along with SDN and Fog Computing competencies. The hierarchy of the proposed system is shown in Figure 3.5. It comprises one central cloud (core) at the top, followed by three mini clouds. Each mini cloud has a cluster of three micro clouds, so nine micro clouds are connected to nine open Vswitches and subsequently to eighteen end-user devices.

We have evaluated the performance of the proposed system and compared it with edge, cloud, and cellular networks. To simulate the cellular network, we consider a long-term evolution (LTE) cell tower for communicating between the master and slave sections. To simulate the cloud network, we consider a database server, switches, access point (AP), and a gateway with all cloud computational compatibility such as storage and accessing the data over the internet. To simulate the edge network, we use a model similar to the cloud network model with edge computing abilities that extend the cloud computing services to the edges of the network.

After a comparative analysis, we emphasise that the proposed system outperforms the edge, cloud, and cellular networks with respect to the selected performance metrics discussed in the upcoming Section 3.6.4. Table 3.1 lists the simulation parameters used in the system simulation.

Considering the small working scenario, the simulation parameters of interest are kept at optimum values to achieve the possible lowest latency and optimise the performance of the system. For instance, the cloud units are kept less to analyse the preliminary data generated from the simulation. Also, the computing resources of VMs such as bandwidth, memory and storage are chosen to meet the requirements of expected workload demands in our simulation environment. It is noted that for the slower arrival rate, the micro clouds may remain idle and increase waiting times for the arrival of the incoming task in the system. Thus, the time taken to complete the task will also increase. So, the optimised value for an arrival rate of 10 Mbps is chosen. The service rate is chosen on the higher side (3 Mbps for micro cloud and 5 Mbps for mini cloud) to lessen the completion time of the incoming task since the service rate is inversely proportional to service time. All the expected processing times are preset for an open Vswitch and an SDN controller.

3.6.3 Cost considerations

While considering the cost in the FC environment, the services or resources are allocated to the users, and accordingly, users execute their tasks. In addition, access to services or resources relies on the active fog nodes and consistently includes the cost of services or resources. Few works [204, 205] have considered the several approaches for FC costs which involve both fog nodes (equipment) and execution of the task for resource allocation, provisioning and scheduling. The cost is determined per equipment or task. It encapsulates various types of costs, such as communication, processing, cloud-network, and power costs, which are considered to evaluate the performance of the proposed system. The cost metrics used in the proposed system are as follows:

1. Communication cost: It is directly dependent on the number of requests coming to the cloud.

Table 3.1 Simulation parameters.

Parameters	Values
No. of micro cloud units	9
No. of mini cloud units	3
No. of central cloud units	1
No. of users	18
No. of VMs (micro, mini and central)	13
Open Vswitch processing time	5 μs
SDN controller processing time	0.5 μs
Communication latency inside BS	80 μs
Micro-cloud unit arrival rate	10 Mbps
Micro-cloud unit service rate	3 Mbps
Mini-cloud unit service rate	5 Mbps
Downstream transmission speed	1.5 Mbit/s
Upstream transmission speed	512 Kbit/s
Bandwidth of micro-cloud units	1000 Mbps
Bandwidth of mini-cloud units	2000 Mbps
Bandwidth of central-cloud unit	4000 Mbps
Micro-cloud RAM, Storage	512 Mb, 1 Gb
Mini-cloud RAM, Storage	2048 Mb, 2 Gb
Central-cloud RAM, Storage	4096 Mb, 4 Gb
Radio propagation	Large-scale propagation
Network area	15 Km $\times$ 15 Km
Simulation time	60 minutes

2. Processing cost: It is dependent upon the total number of desired resources, CPU utilisation and executed tasks and their failures. This cost is associated with the user's demands for either local or virtual processing of the task.

3. Cloud-Network Cost: This cost is involved when there is communication between cloud (fog) units. Here, the task is directed/forwarded to another cloud to execute a particular task and its respective delay.

4. Power Cost: This cost is related to the energy consumption of all fog nodes present in the FC environment while executing the tasks.

3.6.4 Results and discussion

To evaluate the performance of the proposed system, we consider the following performance metrics as they are appropriate in our system performance study: throughput (communication

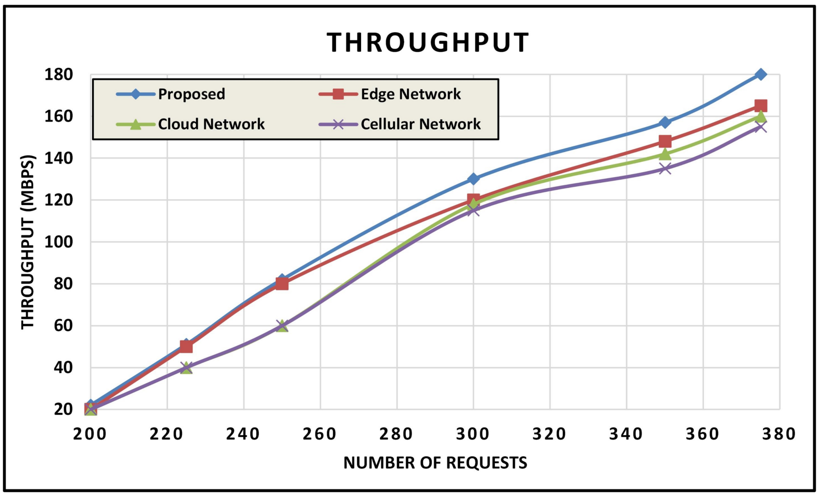


Fig. 3.6 Throughput versus number of requests.

cost), round-trip time (cloud-network cost), jitter (cloud-network cost), bandwidth utilisation (processing cost), energy consumption (power cost), cloud capacity (processing cost), and reliability (processing cost).

Throughput Performance: Throughput signifies how many packets arrive at the destination or how much data has been successfully transferred from source to destination. Figure 3.6 depicts the outcome of throughput performance with increasing number of requests. When the number of requests is set to 200, the throughput is approximately 22 Mbps for all networks. For a number of requests $N_R = 380$, the throughput of the proposed system, edge, cloud, and cellular networks are 180 Mbps, 165 Mbps, 160 Mbps, and 155 Mbps, respectively. We observe that the throughput of the proposed system is about 8%, 11%, and 14% higher than the edge, cloud, and cellular networks, respectively.

RTT Performance:RTT is the time required by the packets to travel from source to destination and its acknowledgement back from destination to source. In Figure 3.7, we plot the number of user equipment against RTT. As a number of equipment (N_E) is incremented from 3 to 18, the RTT for the proposed system started with 9 ms and ended with 25 ms, respectively. We observe that the proposed system offers 11%, 41%, and 43% lower RTT than the edge, cloud, and cellular networks, respectively.

Jitter Performance: Jitter is referred to as the variance in packet delay. In Figure 3.8, the jitter in *ms* is plotted against the number of user equipment (N_E). The jitter for the proposed system started from 0.04 ms for 3 pieces of equipment and reached 0.49 ms for 18 pieces of equipment. As N_E is varied, it is observed that the proposed system has approximately 11%, 43%, and 47% lesser jitter than the edge, cloud, and cellular networks, respectively.

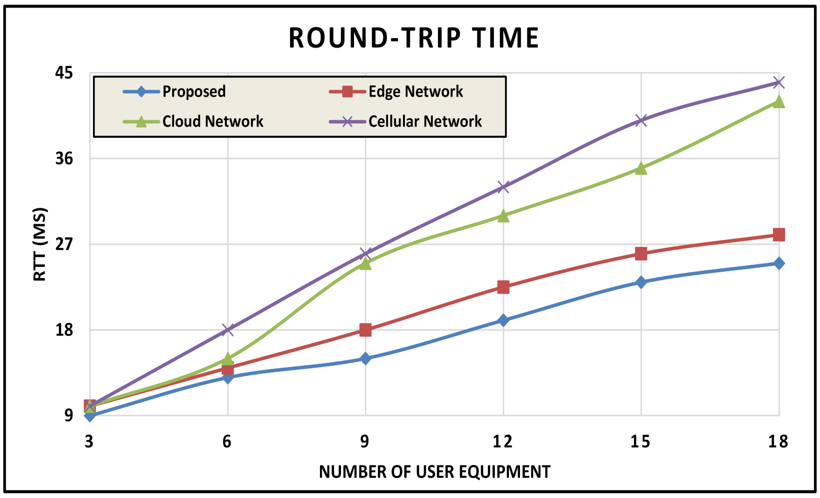


Fig. 3.7 Round-trip time versus number of user equipment.

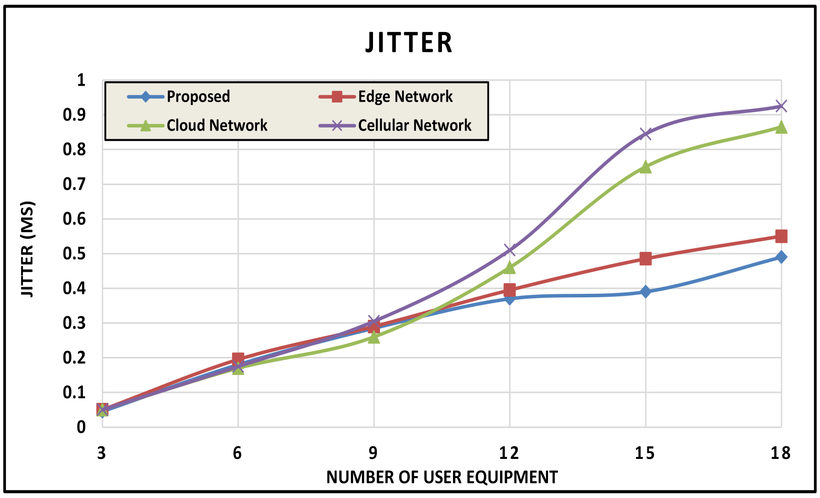


Fig. 3.8 Jitter versus number of user equipment.

Bandwidth Utilisation: It is defined as the total amount of bandwidth actually used to maximise the rate of data transmission in the system at a specific time. Here we have varied N_R from 10 to 70 and have analysed the bandwidth utilised for the networks as depicted in Figure 3.9. We have compared the utilised bandwidth with an increasing N_R. We observe that the proposed system has used around 6 Kbps for 10 requests and utilised up to 25.2 Kbps for 70 requests. The proposed system has utilised lesser bandwidth than the edge, cloud, and cellular networks, with around 39%, 62%, and 71%, respectively. We observe that the proposed system used less bandwidth than the other three existing networks.

Energy Consumption: It is the amount of energy used in the network. Figure 3.10 compares the energy consumption of the proposed system, edge, cloud, and cellular networks.

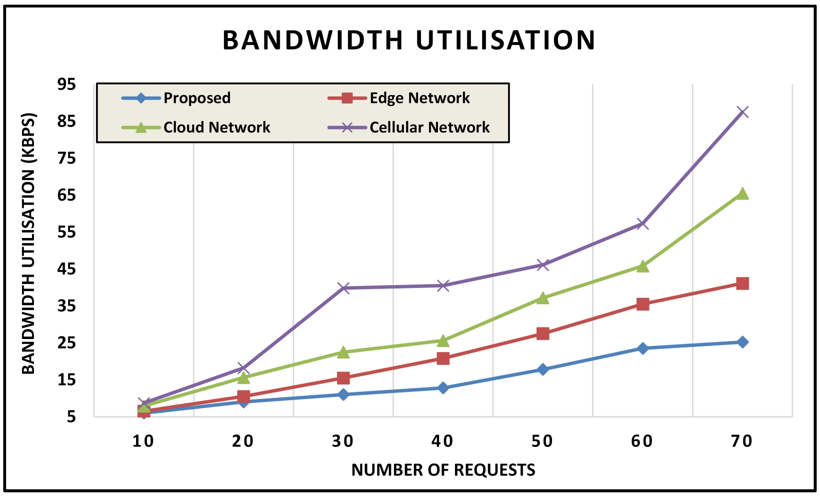


Fig. 3.9 Bandwidth utilisation versus number of requests.

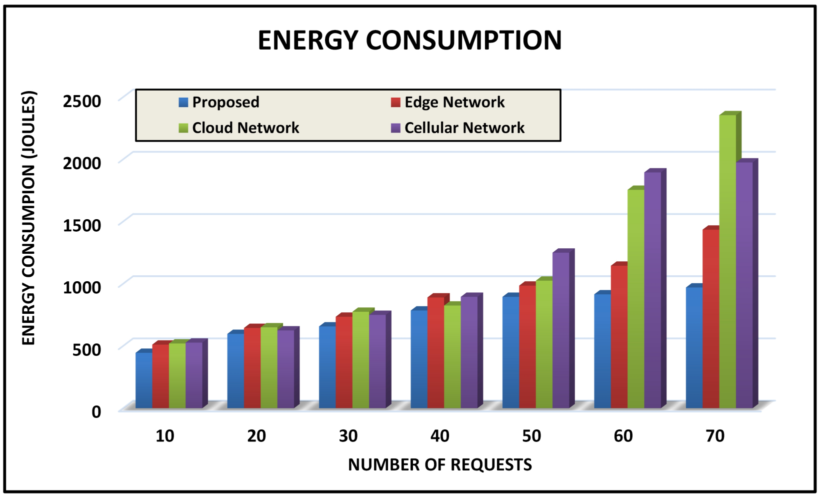


Fig. 3.10 Energy consumption versus number of requests.

We observe that the proposed system consumed less energy than the other three existing networks. Initially, the energy consumption is almost the same as 500 joules for 10 requests. However, when N_R is increased from 10 to 70, the proposed system consumed approximately 32%, 59%, and 51% lesser energy than the existing edge, cloud, and cellular networks, respectively.

Cloud Capacity: It measures and concerns how much workload the network can hold and process. The measurement of cloud capacity is needed to make excellent utilisation of cloud space in the network. It is calculated by evaluating how much percentage of particular cloud-capable networks are operating. In Figure 3.11, we plot cloud capacity (in %) against the number of user equipment. One can observe that the proposed system used less cloud

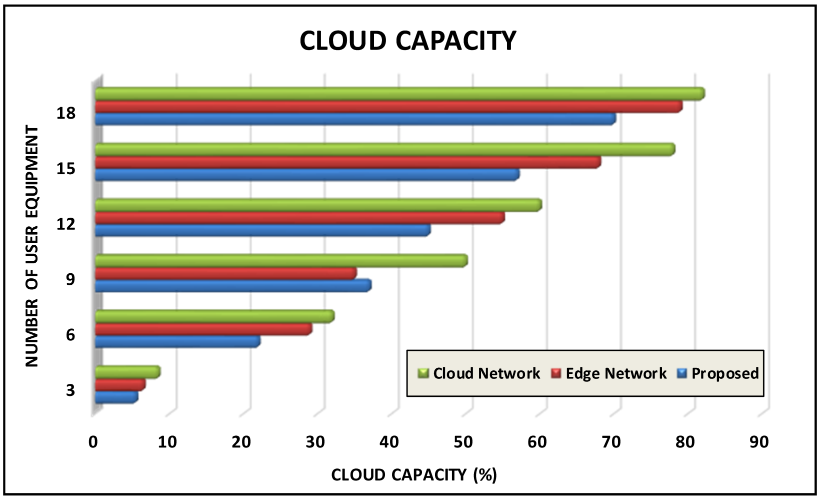


Fig. 3.11 Cloud capacity versus number of user equipment.

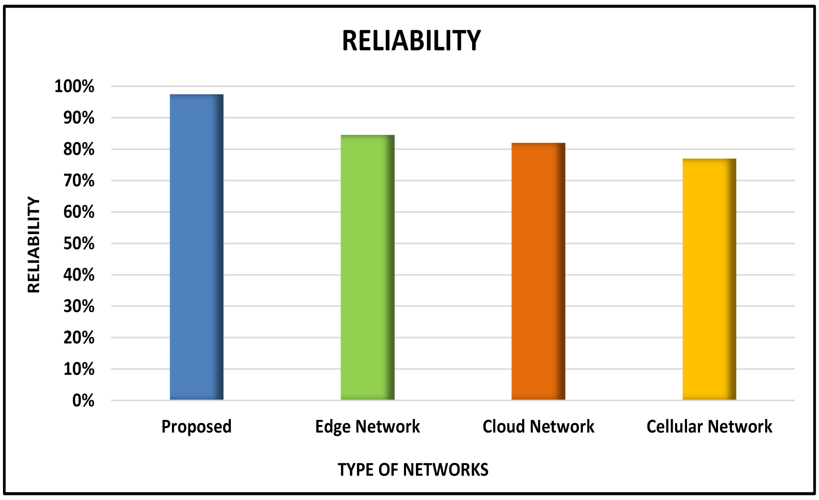


Fig. 3.12 Network reliability comparison.

capacity than the existing edge and cloud networks. For instance, for $N_E = 18$ user requests, the proposed system used about 11.4% and 15% less cloud capacity than edge and cloud networks, respectively.

Reliability Performance: Reliability refers to the accessibility of a specific level of communication service. It is measured by the frequency of failure, i.e. the time it takes to recover from failure, which refers to the network's robustness. The factors can help measure the reliability of a network such as downtime, which means the required time to recover; catastrophe effects like fire and earthquake; and failure frequency when it fails to work the way it is expected.

Mathematically, if R and F denote the overall probability of reliability and failure in micro, mini and central clouds, then R with respect to time (t) is given by,

$$R = 1 - F = e^{(-\lambda t)} \tag{3.10}$$

where, λ = overall failure rate in the system.

Figure 3.12 compares the reliability (in %) of the proposed system, edge, cloud, and cellular networks. We observe that the reliability of the proposed system, edge, cloud, and cellular networks are 98%, 84.5%, 84%, and 77%, respectively.

3.6.5 Model validation

iFogSim enables the modelling and simulation of FC to evaluate and define infrastructure, resource management and scheduling policies across fog, edge and cloud resources under different computing environment scenarios. In spite of the fact that iFogSim is one of the extraordinary simulation toolboxes, it might deliver irrational outcomes if the simulation parameters are inaccurately specified. To guarantee the smooth running of the system simulation model, the simulation log files were verified to confirm that there were no errors. Also, to ensure the proper working of the proposed system, the simulation results were compared with similar related works [206, 207].

In [206], H. Gupta et al. have proposed a simulator called iFogSim to model and simulate resource management techniques in IoT, edge and FC environments. Here, they have demonstrated two case studies of an IoT environment and the effectiveness of the iFogSim simulator. Furthermore, they have evaluated the first case study on “A Latency Sensitive Online Game” and the second case study on “Intelligent Surveillance through Distributed Camera Networks”. Moreover, the scalability of the simulation is also verified. In [207], R. Mahmud et al. have discussed several fog environment scenarios that can be simulated through iFogSim. Subsequently, the critical features of iFogSim are highlighted by facilitating the instructions to install and simulate the Fog environment. They have demonstrated “IoT-enabled Smart Healthcare” as a case study using iFogSim.

We have set the system parameters correctly while modelling the multilevel cloud-based structure. Our results demonstrate that iFogSim is a potential open-source simulator that can carry out end-to-end modelling and simulation in the context of IoT (TI) and FC.

3.7 Summary

In this manuscript, we have proposed a fog-based traffic flow framework employing SDN and FC approaches to realise TI. Furthermore, an efficient traffic flow algorithm has been developed to reduce extra processing and waiting times at each level of the cloud-based structure in the system. Finally, the proposed system is implemented in an iFogSim simulator (a Java-based high-performance open-source modelling toolbox). Also, the impacts of increasing the number of user equipment and the number of flow requests on system performance are analysed. The results obtained from the simulation have demonstrated that the proposed system offers up to 14% higher throughput, 43% lower RTT, 47% lower packet jitter, and 51% lower energy efficiency than the cellular networks. The proposed system also consumes about 71% less bandwidth to run haptic applications. Based on the reliability study, we found that the proposed system is about 21% more reliable than the existing cellular networks. These research results deliver a clear vision in realising TI that can support network researchers to impart further towards building up the next-generation (xG) Internet. Incorporating network function virtualisation and multiple access techniques in the proposed system is recommended for future research work.

Chapter 4

TINetS3: SDN-driven Network Slicing enabling Scenario-based Applications in Tactile Internet (Manuscript 3)

4.1 Manuscript 3 prelude

Network slicing (NS), a key enabler, plays a crucial role in implementing tactile Internet (TI). This TI promises scalable, customisable, isolated, and logical end-to-end networks (i.e., network slices) by offering dedicated and tailored networking solutions on demand to meet the customer's need. Research in NS for TI is essential to realise haptic communications to achieve ultra-low latency, high throughput, and high reliability for real-time immersive and interactive environments in fifth-generation (5G) and beyond 5G (B5G).

Manuscript 3, entitled "TINetS3: SDN-driven Network Slicing enabling Scenario-based Applications in Tactile Internet[1]" [208], proposes a novel network slicing mechanism based on TI communication infrastructure for three different scenarios (TINetS3) leveraging software-defined networking (SDN) with Open vSwitch (OVS) to resolve the issue of provisioning and controlling the network slices on demand. This communication infrastructure is sliced under various traffic/application loads using pre-designed slice configurations. We develop algorithms for three scenarios, i.e., topology slicing, service slicing, and emergency slicing, to have a tailored slice for the physical infrastructure according to the network requirements.

Moreover, we validate the system performance by measuring the key performance indicators (KPIs), such as throughput and round-trip time (RTT), to optimise the network resources by network emulator. These KPIs show the expected experimental results aligned with the preset allocated spectrum for slices and RTT constraint values in various scenarios. The simulation results show the efficacy of our NS algorithms in designing tailored and

[1]V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "TINetS3: SDN-driven network slicing enabling scenario-based applications in tactile Internet," *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 4639–4654, 2024.

customised slices as per the network-required KPIs. Finally, we validate our algorithms by comparing them with state-of-the-art algorithms.

Furthermore, this manuscript focuses on implementing NS under the TI communication infrastructure to improve the network KPIs with proposed slicing algorithms. SDN with OVS is also incorporated to manage the network slicing behaviour. Specifically, this manuscript addresses the second research question: "*RQ 2. What strategies can be employed to improve the RTT of tactile Internet incorporating SDN and NS approaches?*".

4.2 Introduction

The concept of the tactile Internet (TI) has revolutionised the way we see how human operators and remotely present robots can be connected and carry operational information over the Internet. The area of real-time applications has served a new level of immersive experience due to an emerging concept of TI [22]. The recent advancements in haptic and kinematic sensors and actuators [19], along with artificial intelligence (AI) in TI, have risen beyond the traditional audio and visual feedback [209][210]. Particularly, TI can be defined as a new communication framework that allows remote access, perception, reception, analysis, manipulation, control, sense and/or touch the real and virtual bodies with or without audio and visual content over a significant distance [1].

The progression of TI started with mobile Internet and moved through the massive Internet of Things (IoT), which has stringent requirements such as ultra-responsiveness, ultra-reliability, ultra-low latency, and ultra-high availability [192]. To achieve such stringent criteria, the fifth-generation (5G) communication framework that offers better speed (up to $\times 20$), latency (with a target of 1 ms or less), capacity and bandwidth support of connected devices, and engineered to support a diverse range of applications such as augmented reality (AR)/virtual reality (VR), autonomous vehicles and factory automation, when compared to fourth generation (4G) communication framework [211]. Hence, the TI stringent requirements are fulfilled by 5G, which were challenging to meet the need for high data speeds and ultra-low latency communications for real-time applications. Once these requirements are met, umpteen practical implications can be explored and accomplished through 5G and B5G. [212].

In addition, TI technologies such as software-defined networking (SDN), network function virtualisation (NFV), network coding (NC), mobile edge computing (MEC), network slicing (NS), etc., aim to facilitate a digital closed-loop communication between human-to-human (H2H) or human-to-machine (H2M) real-time communication, providing new

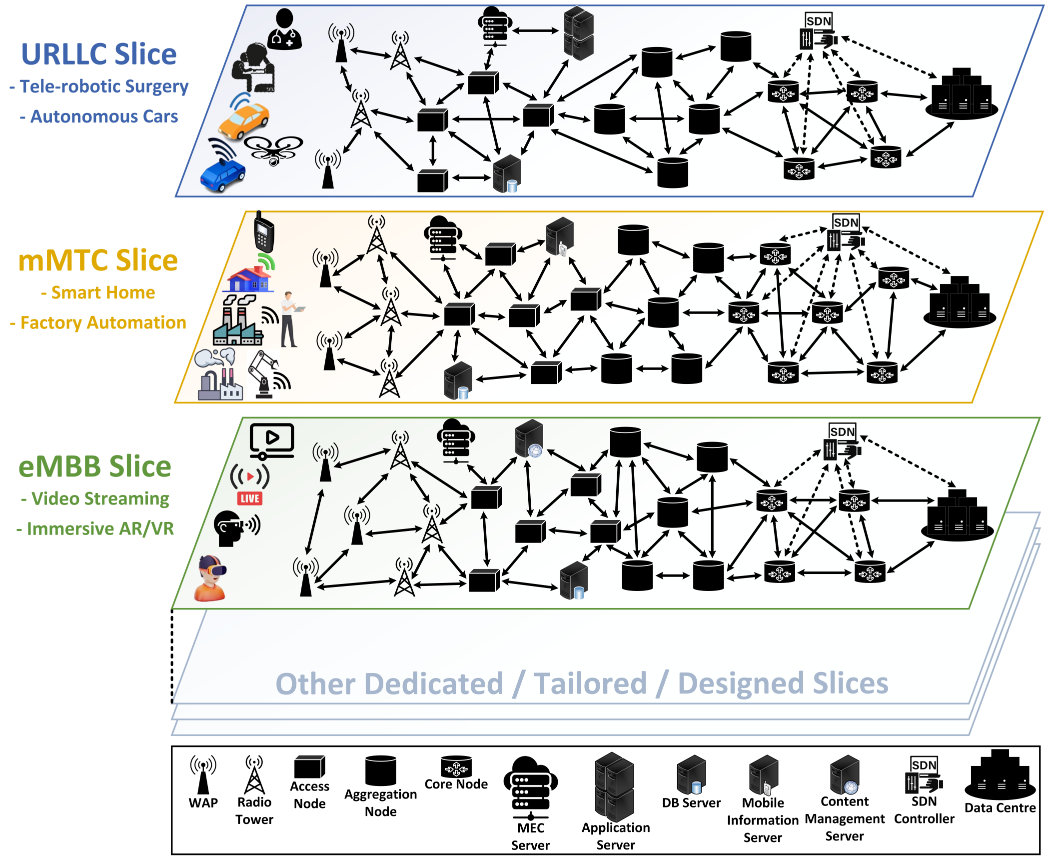


Fig. 4.1 Illustrating designed and dedicated E2E network slices.

avenues for haptic (multisensory) with VR and AR experiences. Therefore, AR, VR and mixed reality solutions can bring a plethora of practical use cases and their implications, such as e-health (tele-medicine, tele-surgery, tele-diagnosis and tele-rehabilitation), platooning of vehicles, smart homes, Industry 4.0 and many other use-cases. Thus, TI will form a new era of possibilities and opportunities for such use cases.

As in 5G networks, the 3^{rd} Generation Partnership Project (3GPP) Release 15 [211] formally mentions NS with logical and end-to-end (E2E) networks. In particular, NS is made possible through the incorporation of key enablers such as SDN, cloud computing (CC), and edge computing (EC) [213]. Moreover, according to [214], the predicted market size of NS globally is expected to grow from \$759.0 million in 2023 to $\$13,662.9$ million by the year 2030, showcasing a compound annual growth rate (CAGR) of $\$51.1\%$ over the forecast period.

NS is introduced to meet and deliver the stringent and distinctive traffic demands related to 5G generic services such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable low latency communication (URLLC) [19].

These traffic demands include customisable bandwidth, round-trip time (RTT), throughput, and high availability or reliability for specific services. This is achieved by creating multiple logical and virtualised networks that are built on the shared physical infrastructure. Each logical network (slice) is designed and tailored to dynamically serve a set of users, networks, applications, and services by network managers. Because of the aforementioned technologies, a single and shared physical infrastructure can facilitate several services according to the user's demand. Figure 4.1 depicts the designed and dedicated E2E network slices tailored to meet the stringent requirements of the services or applications. In this figure, for instance, URLLC-based use cases require high reliability and extremely low latency, whereas mMTC-based use cases can support up to 1 million devices per square kilometre and require low energy and bandwidth. However, eMBB-driven use cases focus on speed, capacity and mobility and thus require high bandwidth. Consequently, a slice can be designed and tailored to the network service that demands and enhances quality of service (QoS) and quality of experience (QoE), having low latency and high availability, and another slice having high throughput with efficient bandwidth consumption. Furthermore, it helps the network manager to efficiently reduce capital expenditure (CAPEX) and operational expenditure (OPEX), enhancing the efficacy of the delivery of network services.

On the other hand, SDN centralised architecture simplifies and intelligently manages the traffic flow of the extensive networks by decoupling the data (forwarding) plane and control plane (where the SDN controller resides). Thus, SDN provides several advantages, such as increased control over the network with great speed and agility, deeper and centralised network management, customisable network infrastructure, and programmatical configuration. As the demand for agility, flexibility and customisable services rose, network architecture started incorporating SDN technologies. SDN creates umpteen opportunities to meet the deployment requirements in 5G and beyond, where it was not possible and achievable in 4G. The market for SDN is estimated to achieve approximately USD 73M by 2027, with an expected CAGR of 28.2% between 2020 and 2027 [215].

4.2.1 Research challenges

The aforementioned use cases in TI are predominantly latency-dependent. One of the research challenges is to attain the latency required to realise tactile and haptic communication, which must have an RTT of around 1 ms [216]. The RTT varies depending on the use cases but is mostly latency-aware/sensitive. Furthermore, to realise the full potential of TI, the challenges also lie in creating and designing effective and dynamic E2E network slices, allocating and managing the resources, ensuring and guaranteeing required QoS, and isolating created network slices. The exponentially ever-increasing devices and the demand for multimedia services have led to the scarcity of radio resources and difficulty in their allocation.

Moreover, the scalability of network slices created will pose a challenge due to the increasing number of TI applications. Thus, more network slices will be needed in a given service area to offer critical services with TI's stringent requirements. So, the network slice controller should mitigate the difficulties related to managing and orchestrating created network slices. Creating and designing a customised and completely new network slice each time is tedious. However, the recursive network slices concept can be used where a new network slice is created using an existing slice in a hierarchical manner. In other words, sub-slices within a network slice can be implemented. Nevertheless, its practical implementation can be highly complex and challenging. Implementing multi-tenant network slicing can also be challenging as the resources will be shared from one network slice to another. However, dynamic resource sharing can solve this issue. While resource sharing brings advantages to infrastructure providers, it brings additional challenges, such as slice isolation, which hinders QoSs.

4.2.2 Research contributions

Although the previous studies based on 5G are well-defined with their vision and objectives, some underlying unexplored solutions based on SDN and NS would potentially need to enable effective haptic communication in TI. The research is still developing, as previous studies did not explicitly focus on specific types of slices, such as topology, service, and emergency slicings. Hence, novel slicing approaches must be proposed for the TI communication framework to design and optimise a dedicated slice to meet the stringent and diverse demands of applications or services. The contributions of this manuscript are briefly summarised below.

1. We propose novel network slicing mechanisms (topology, service and emergency slicings) based on TI communication infrastructure by employing SDN with OVS.

2. We develop an efficient topology slicing technique that focuses on the division of the proposed infrastructure's physical topology in TI communication infrastructure.

3. We design an efficient service slicing technique that focuses on designing, tailoring or customising slices to facilitate distinct services and meet the specific demands of customers.

4. We design an automated/dynamic emergency slicing method to support critical and immediate communication requirements during disaster or emergency scenarios and effectively prioritise the network resources and services without intervention. To this end, we evaluate the network's KPIs to exhibit the efficacy of our developed algorithms and draw meaningful insights for TI communication infrastructure.

4.2.3 Structure of the manuscript

The rest of the manuscript is structured as follows. Section 4.3 reviews the existing work on deploying network slices with SDN. Section 4.4 describes the proposed system of TI communication infrastructure. This section also proposes algorithms for deploying NS in three different scenarios. Section 4.5 presents simulation results. Finally, Section 4.6 provides a brief conclusion with future directions.

4.3 Motivation and related work

4.3.1 Motivation

One of the motivations to conduct research on NS is the need to create an efficient, dynamic, adaptable, flexible and optimised slice(s) to cater for the tailored network requirements. Moreover, the slice(s) can support a broad spectrum of distinguished services and applications over 5G infrastructure in an ever-increasing connected and data-driven world, specifically enhanced mobile broadband (eMBB), IoT and low to very low latency required services and applications. It seemingly addresses the difficulties or issues related to enhancing QoS and managing network resources. Thus, this opens the opportunity to increase the client base and facilitate tailored network services and applications.

4.3.2 Related work

Several research studies have been carried out and focussed on achieving an efficient, dynamic and optimal slice(s) and supporting a diverse range of application and service requirements in 5G networks.

With regards to 5G NS incorporating network softwarisation technologies such as SDN, Network Function Virtualisation (NFV), cloud computing (CC) / Edge computing (EC) / Fog Computing (FC), commercial off-the-shelf (COTS) hypervisors, and so on, the authors in

[217] provide a comprehensive survey and recent solutions. Furthermore, this survey discusses several industrial initiatives and projects adopting network softwarisation technologies to speed up the growth of 5G NS. This research compares several 5G architectural strategies concerning real-world implementation, technology integration and deployment strategies.

Moreover, Botex et al. [218] present an innovative approach for NS in 5G backhaul networks, explicitly focusing on low and very low latency communications between user equipment (UE). The authors have worked on a heuristic algorithm that outperforms Dijkstra's algorithm concerning network QoS and real-time network congestion management. Also, the SDN is incorporated in their algorithm with the Ryu controller to perform the computation and calculate the path delay incurred. Wijethilaka and Liyanage [219] provide another survey on NS to realise IoT applications in 5G. In addition, the emerging techniques, such as artificial intelligence (AI) and blockchain along with SDN, NFV and CC, are discussed to explore the possible answers to implement NS and integrate IoT applications in 5G and B5G.

In [220], the article presents the simplistic survey of 5G network slicing to a first-time reader of their interest, reviews the existing state of the art, discusses the NS architectural components and challenges, and allows identifying the umpteen open research questions. On the other hand, Li et al. [221] talk about the challenges and opportunities of NS for 5G. Moreover, the aforementioned emerging technologies will enable service-oriented 5G network infrastructure.

Another research conducted in [222] provides a comprehensive overview and analysis of NS with the necessity of slicing automation and management and orchestration (MANO) architectural components in 5G. Also, it addresses commercial NS, focusing on several recent industrial projects related to AI and machine learning (ML), leading to open research problems and possible solutions to guide network researchers.

Considering NS implementation, the authors in [223] incorporated blockchain technology into B5G to efficiently monitor, share and utilise the resources in a controlled manner. They have proposed a Blockchain-Enabled Network Slicing Model (BENS) incorporated into NS architecture to handle the spectrum resource allocation. This model delivers better energy-efficient performance and transmission success rate when compared with other decentralised and centralised learning approaches. Moreover, this article introduces dynamic spectrum sharing and discusses the 5G-CORE architecture and collaborative learning-based algorithms.

In [224], the MANO framework is presented for automating 5G E2E NS along with the integration of core network (CN) and transport network (TN) slices. This framework includes the 3GPP NS management, 5G core NS mechanisms, novel bandwidth management techniques, and recent work on cloud-native technologies. Moreover, it showcases the effectiveness of the proposed framework by evaluating the resource overhead and service throughput under bandwidth policies. Another solution in [225], an open-source 5G NS architecture, was developed and deployed automatically. The authors utilised the OpenStack, free5GC, UERANSIM and Tacker modules to execute various functions such as virtualisation, managing the 5G core network, simulating gNB (base station) and user equipment (UE), and deploying slicing environment. Hence, this proposed architecture created automatic NS with dedicated functions, the registration of dedicated slices into 5G, and the connection of UE with the corresponding slice. Thus, the experiment's outcomes depict the viability of the proposed architecture and guarantee the QoS for each slice.

Considering the 5G mobile backhaul networks for the URLLC module, Hefele and Costa-Requena [226] introduce and achieve the NS based on SDN for URLLC communication. The authors outline the design of the modules and algorithms that execute NS competencies and illustrate their practical application in an emulated mobile backhaul environment. On the contrary, in [227], monitoring the NS KPIs is demonstrated in a 5G testbed. The authors have explained and deployed an operational 5G testbed by utilising the amalgamation of open-source frameworks and tools to configure several programmable network slices based on the network requirements. They have also considered the cloud-gaming application to depict the viability of monitoring and visualising the NS KPIs. In [228], the authors have focused on and proposed an effective way of reducing service-based dedicated provisioning time for an E2E NS framework in physical and virtual environments. They have developed an intricate NS provisioning framework by applying Dijkstra's and A* strategies, and the results depict a linear scaling and reduction in provisioning time concerning nodes and slice-request volumes.

To enable different vehicle applications, the authors in [229] have presented a common framework for assigning resource slicing and associating vehicles in drone-supported vehicle networks. This framework facilitates spectrum partitioning among heterogeneous base stations (BSs) and implements the slicing for better spectrum utilisation. The simulation results show that the proposed framework and its algorithm incur better throughput than existing ones.

Taking AI into account, AI-native network slicing architecture for the sixth generation (6G) network is presented in [230] to support emerging AI services. This architecture helps to foster a synergy between AI and network slicing, thus envisioning to improve intelligent network management, referred to as AI for slicing.

Furthermore, AlQahtani [231] proposes an adaptive slice allocation approach for a 5G cloud-based radio access network (C-RAN) to dynamically allocate resources according to the evolving network environment, manage traffic delay tolerances to meet QoS requirements, and optimise network resources efficiently. It integrates slice admission control and dynamic resource allocation to meet QoS requirements and optimise the network resource efficiently. The authors in [232] work on creating network slices in a 5G virtualised core by optimising the virtual network functions (VNFs) sharing to reduce costs using the Integer Linear Programming model and propose a Heuristic Backtracking Algorithm for NS to address scalability issues and predict UE requests using a deep learning model to meet dynamic service level agreement (SLA) requirements for a user. Moreover, they propose a Heuristic Backtracking Algorithm for NS to address scalability issues and a deep learning model to predict user equipment requests to meet dynamic service level agreement (SLA) requirements for a user.

In addition, a model and algorithm are proposed in [233] on dynamic selection slicing-based offloading for vehicular tasks to mitigate network congestion and time delay using the Markov decision process and deep reinforcement learning and integrate network slices within the edge server. It adopts a directed acyclic representation for vehicular task dependencies and integrates network slices within the edge server. Along similar lines, Dangi and Lalwani [234] adopt ML models to optimise and accurately predict the optimal 5G network slices by gathering distinct datasets and applying feature selection algorithms.

A summarised related work with authors' key findings or contributions is tabulated in Table 4.1, where they have focussed on implementing and deploying NS employing SDN, NFV, cloud computing (CC)/mobile edge computing (MEC)/fog computing (FC), machine learning (ML)/artificial intelligence (AI) and their proposed algorithms to achieve efficient slicing.

Table 4.1 Summary of related work.

Reference	Technologies involved/discussed						Key findings/contributions
	NS	SDN	NFV	CC/ MEC/ FC	ML/ AI	Algor-ithm	
[218], 2023	✓	✓	×	×	×	✓	• An innovative approach for NS in 5G backhaul networks, explicitly focusing on low latency communications between user equipment (UE) with a proposed heuristic algorithm that outperforms Dijkstra's algorithm.
[229], 2023	✓	×	×	×	×	✓	• A common framework for assigning resource slicing and associating vehicles in drone-supported vehicle networks by facilitating spectrum partitioning among heterogeneous base stations (BSs).
[231], 2023	✓	✓	✓	✓	×	✓	• An adaptive slice allocation approach for 5G cloud-based radio access network (C-RAN) to dynamically allocate the resources according to the evolving network environment and manage the traffic delay tolerances to meet QoS requirements and optimise the network resource efficiently.
[233], 2023	✓	✓	×	✓	✓	✓	• A proposed model and algorithm on dynamic selection slicing-based offloading for vehicular tasks to mitigate network congestion and time delay using the Markov decision process and deep reinforcement learning and integrates network slices within the edge server.
[234], 2023	✓	×	×	×	✓	✓	• Machine learning models to optimise and accurately predict the optimal 5G network slices by gathering distinct datasets and applying feature selection algorithms.

Table 4.1 *Contd.*

Reference	Technologies involved/discussed						Key findings/contributions
	NS	SDN	NFV	CC/ MEC/ FC	ML/ AI	Algor- ithm	
[232], 2023	✓	×	×	×	✓	✓	• Creates network slice in 5G virtualised core by optimising the virtual network functions (VNFs) sharing to reduce costs using the Integer Linear Programming model, proposes a Heuristic Backtracking algorithm for NS to address scalability issues and predict UEs requests using a deep learning model to meet dynamic SLA requirements for a user.
[222], 2022	✓	×	×	×	✓	×	• A comprehensive and commercial overview and analysis of the NS with a necessity of slicing automation and management and orchestration (MANO) architectural component in 5G, and discusses several recent project objectives and industrial forms related to ML and AL.
[223], 2022	✓	✓	✓	×	✓	✓	• Proposes a Blockchain-Enabled Network Slicing Model (BENS) incorporated into NS architecture to handle spectrum resource allocation, introduces dynamic spectrum sharing and discusses 5G-CORE architecture and collaborative learning-based algorithms.
[224], 2022	✓	✓	✓	×	×	×	• MANO framework for automating 5G E2E NS along with the integration of core network (CN) and transport network (TN) slices and showcasing the effectiveness of the proposed framework.
[227], 2022	✓	×	✓	×	×	×	• Demonstrates NS KPIs monitoring in a 5G testbed, deploys an operational 5G testbed and explains cloud-gaming application as a use-case.

Table 4.1 *Contd.*

Reference	Technologies involved/discussed						Key findings/contributions
	NS	SDN	NFV	CC/ MEC/ FC	ML/ AI	Algor-ithm	
[228], 2022	✓	✓	✓	×	×	×	• An effective way to reduce service-based dedicated provisioning time for an E2E NS framework in physical and virtual environments and to develop an intricate NS provisioning framework is by applying Dijkstra's and A* strategies.
[230], 2022	✓	✓	✓	×	✓	×	• AI-native network slicing architecture for the sixth generation (6G) network to support emerging AI services and help foster a synergy between AI and NS.
[219], 2021	✓	✓	✓	✓	✓	×	• A comprehensive analysis of the contribution of NS to realise IoT applications with corresponding technical challenges that can be resolved and possible answers with a discussion on emerging techniques, such as AI, blockchain, SDN, NFV, and CC in 5G and B5G.
[225], 2021	✓	✓	✓	×	×	×	• Develops and automatically deploys an open-source 5G NS architecture utilising OpenStack, free5GC, UERANSIM and Tacker modules to execute various functions such as virtualisation, managing 5G core network, simulating gNB (base station) and UE, and deploying slicing environment.

Table 4.1 *Contd.*

Reference	**Technologies involved/discussed**						**Key findings/contributions**
	NS	**SDN**	**NFV**	**CC/ MEC/ FC**	**ML/ AI**	**Algor- ithm**	
[226], 2020	✓	✓	×	×	×	✓	• Introduces and achieves the NS based on SDN for URLLC communication in 5G mobile backhaul networks by outlining the design of the modules and algorithms that execute NS competencies.
[217], 2020	✓	✓	✓	✓	×	✓	• A comprehensive survey and recent solutions with regards to 5G NS incorporating network softwarisation technologies such as SDN, NFV, CC/EC/FC, commercial off-the-shelf (COTS) hypervisors, and so on, with discussion on several industrial initiatives and projects.
[220], 2017	✓	✓	✓	×	×	×	• Presents the simplistic survey of 5G network slicing to the first-time reader of their interest, reviews the existing state of the art, discusses the NS architectural components and challenges, and allows identifying the umpteen open research questions.
[221], 2017	✓	✓	✓	×	×	×	• Talks about the challenges and opportunities of NS for 5G, which serves as a Network as a Service (NaaS) and an enabler for service-oriented 5G network infrastructure.
Our work	✓	✓	×	✓	×	✓	• Proposes novel network slicing mechanisms for three different scenarios, i.e., topology, service, and emergency slicings, based on TI communication infrastructure by employing SDN with OVS.

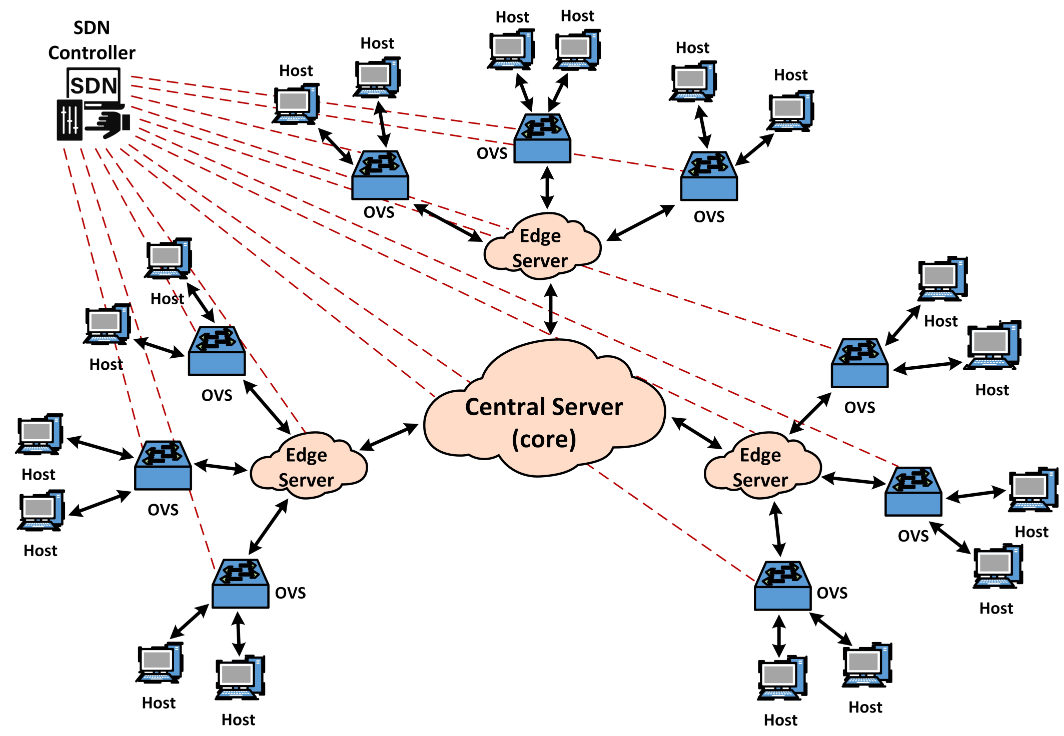


Fig. 4.2 Proposed TI communication infrastructure.

4.4 Description of the proposed system

To propose a TI communication infrastructure that meets the requirements of the services or applications existing on a shared physical network, it is crucial to design and deploy the services according to the customers' needs. By incorporating NS and SDN, the proposed infrastructure provides the following benefits.

- Flexible network management as the slices can be provisioned, E2E customised and deleted on demand.
- Slice isolation guarantees the service level agreements (SLAs) to enhance KPIs; thus, the performance of one slice does not impact the other.
- Optimised consumption of network resources by minimising the wastage of resources and maximising the network capacity.
- Network elasticity automatically permits scaling up or down to use network resources effectively.

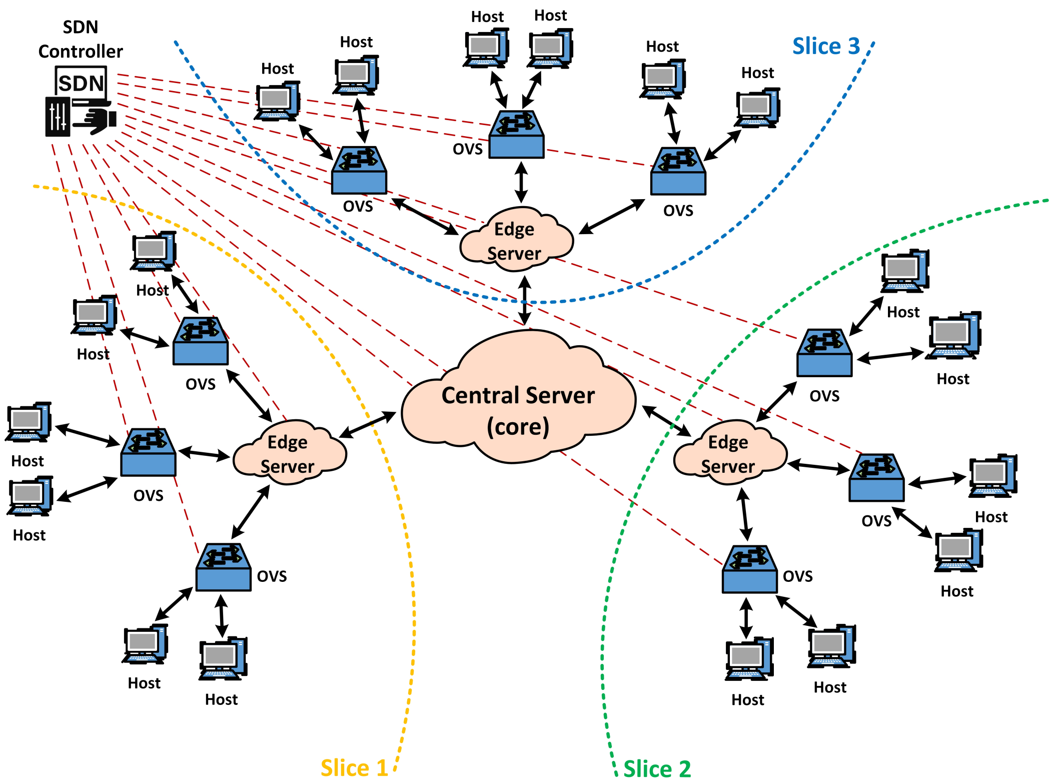


Fig. 4.3 Illustrating topology slicing scenario with three slices.

- Programmability, as the network traffic behaviour can be altered through an SDN controller, network manageability and congestion reduction.

Depending on the need for customised network services, the proposed TI communication (Figure 4.2) is envisioned in the following three scenarios.

4.4.1 Topology slicing

Topology slicing refers to the logical division of infrastructure's physical topology into slices individually having their own distinctive network structure, as depicted in Figure 4.3, to support a wide range of use cases. In simpler words, network slicing (NS) is achieved based on slicing the physical topology. This type of slicing is especially important in scenarios where various services or applications coexist on the same physical infrastructure or network but have dedicated strict networking needs.

Moreover, this slicing offers the following advantages.

Algorithm 2: Proposed topology slicing algorithm

Input: *d*: datapath, *i*: incoming port, *ovs_id*: OVS ID number.
Output: Efficient topology slicing of the TI communication infrastructure.

```
import SDN controller, OpenFlow (OF) protocol and traffic libraries.
class Topologyslicing
def topo_slicing[ovs_id][i]
def ovs_characteristics
Find and add the missing flow_entries in the OVS.
def packetOut_ovs(payload, d, i):
    sent_data = 0
    if the payload has no buffer then
        sent_data ← payload
    end
    PacketOut ← d, sent_data, i
    ovs ← PacketOut
def packetIn_ovs (payload, d, i):
    extract payload, d, i, ovs_id
    f(ovs_id, i) = topo_sclicing[ovs_id][i]
    add flow_entries to OVS directing all future
    incoming packets with the same i
    packetOut_ovs
```

- Customisable and unique network topology where the network managers can fragment and create network segments with distinct arrangements concerning network devices, such as routers, switches, interface links, and other devices.
- The network slices created through topology slicing can have their own dedicated resources, including the physical infrastructure and network KPIs.

Considering Figure 4.3, three topology slices are created where the physical infrastructure is virtually segmented to meet the specific demands of the different services or applications. The three distinct network slices can serve or handle different network traffic or services, such as data, voice and video traffic. The individual traffic has its unique network requirements. Hence, each slice with three different traffic running into it will lead to a customised virtual network environment. Subsequently, the efficient allocation of network resources and the optimised performance and management of each traffic will be achieved. Moreover, the SDN controller is used to code the programmable and flexible virtual switches, called open virtual switches (OVSs), to manage traffic. Furthermore, the working of topology slicing is explained and proposed in Algorithm 2.

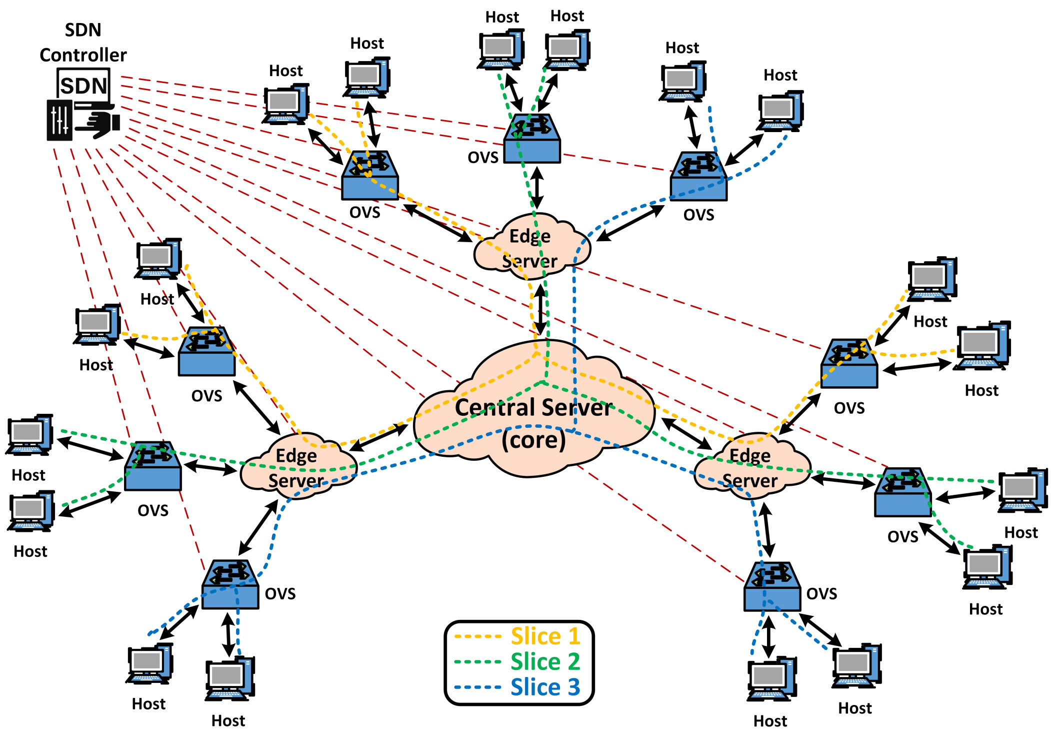


Fig. 4.4 Illustrating service slicing scenario with three slices.

$topo\_slicing[ovs\_id][i]$ is a dictionary that is defined to slice the physical topology according to the three different traffic, i.e., $traffic\_1$, $traffic\_2$ and $traffic\_3$. This dictionary maps i and directs to the o (outgoing port) with the identified ovs_id. The $ovs\_characteristics$ represent the OVS connection to the controller, set the OF protocol and extract the information such as d, $PacketIn$, $PacketOut$, FlowMod, priority and actions. The $packetOut\_ovs$ is used to send the messages to OVS. Finally, the $packetIn\_ovs$ is used when the controller gets a payload from an OVS. The function $f(ovs\_id, i)$ determines the output to o based on i and ovs_id. Overall, the controller is designed to configure OVS to direct the inbound traffic based on the slicing rules created in $f(ovs\_id, i)$.

4.4.2 Service slicing

Service slicing refers to creating multiple virtual networks or slices designed or tailored within a single physical infrastructure to facilitate dedicated services, as illustrated in Figure 4.4. Each slice is customised to meet the stringent demands of several applications or end-users. In other words, the network manager(s) can create service-based slices according to customer's demands on the same physical infrastructure. This slicing is specifically crucial in scenarios where the heterogeneous services or applications share the common physical infrastructure to acquire the desired and customisable networking resources and guarantee SLA.

In addition, this type of slicing showcases the following benefits.

- Customised service where the network manager can design and create virtual service-based slices mainly meant for the diverse needs of different applications or service categories.
- Optimised network resource utilisation and enhanced QoS for various types of applications and services in normal operating conditions.

The three service slices are formed, as shown in Figure 4.4, where each slice can be configured to meet the KPIs of the application or end-user, such as bandwidth allocation, latency constraints, QoS customisation, and other network parameters. Hence, these slices create a dedicated service environment within the same physical infrastructure. As mentioned earlier regarding different traffic, each service slice can handle dedicated traffic at a time and, thus, isolate other service slices to avoid traffic congestion. Moreover, the physical infrastructure will route different types of traffic and direct or forward them to the appropriate service slices with the help of the SDN controller. Algorithm 3 is proposed to explain how service slicing is done.

$service\_slicing[ovs\_id][mac]$ is a dictionary that is defined to create and customise the service-based slices according to the three different traffic, i.e., $traffic\_1$, $traffic\_2$ and $traffic\_3$. This dictionary maps mac addresses and directs to the o (outgoing port) with the identified ovs_id. Also, it maps the $serviceslice\_number$ to o with ovs_id. Furthermore, each traffic and traffic $destination\_port$ number is conditionally checked and compared with preset $slice\_port\_numbers$. If the condition holds true, the traffic is assigned to dedicated $serviceslice\_number$ accordingly. The function $f'(ovs\_id, serviceslice\_number)$ determines the output port to o based on ovs_id and $serviceslice\_number$.

4.4.3 Emergency slicing

Emergency slicing refers to the significantly designed network slices from a common physical infrastructure to support critical networking or infrastructure needs during unforeseen and emergency situations, as shown in Figure 4.5. These dedicated and designed slices guarantee that the crucial networking services stay effective and functional even if the network exposes to high traffic congestions and other interruptions. Therefore, the emergency slices are like contingency plans for the network to continue utilising the services or applications in case of man-made or natural disasters. These slice creations offer a business continuity and disaster recovery strategy. Some of the key characteristics of this type of slicing are as follows.

Algorithm 3: Proposed service slicing algorithm

Input: *d*: datapath, *i*: incoming port, *ovs_id*: OVS ID number, mac: MAC address of the hosts.

Output: Efficient service slicing of the TI communication infrastructure.

```
import SDN controller, OpenFlow (OF) protocol and traffic libraries.
class Serviceslicing
def service_slicing[ovs_id][mac]:
    set the slice_port_numbers for traffic_1, traffic_2 and traffic_3
    map the ovs_id with serviceslice_number and i
    set end_ovs
def ovs_characteristics
Find and add the missing flow_entries in the OVS.
def packetOut_ovs(payload, d, i):
    sent_data = 0
    if the payload has no buffer then
        sent_data ← payload
    end
    PacketOut ← d, sent_data, i
    ovs ← PacketOut
def packetIn_ovs (payload, d, i):
    extract payload, d, i, ovs_id
    set source
    set destination
if ovs_id in service_slicing then
    if destination in service_slicing[ovs_id][mac] then
        f(ovs_id,destination) = service_slicing[ovs_id][destination]
        add flow_entries to OVS directing based on
          desired f(ovs_id,i) with mapped destination
        packetOut_ovs
    else
        traffic_(1,2 or 3) and traffic_(1,2 or 3).destination_port =
          slice_port_numbers for traffic_(1,2 or 3)
        serviceslice_number = 1,2 or 3
        f'(ovs_id,serviceslice_number) =
          service_slicing[ovs_id][serviceslice_number]
        add flow_entries to OVS directing based on desired f(ovs_id,i) with mapped destination
        packetOut_ovs
    end
else
    ovs_id is not in end_ovs
    add flow_entries to OVS directing based on
      desired f(ovs_id,i) with mapped destination
    packetOut_ovs
end
```

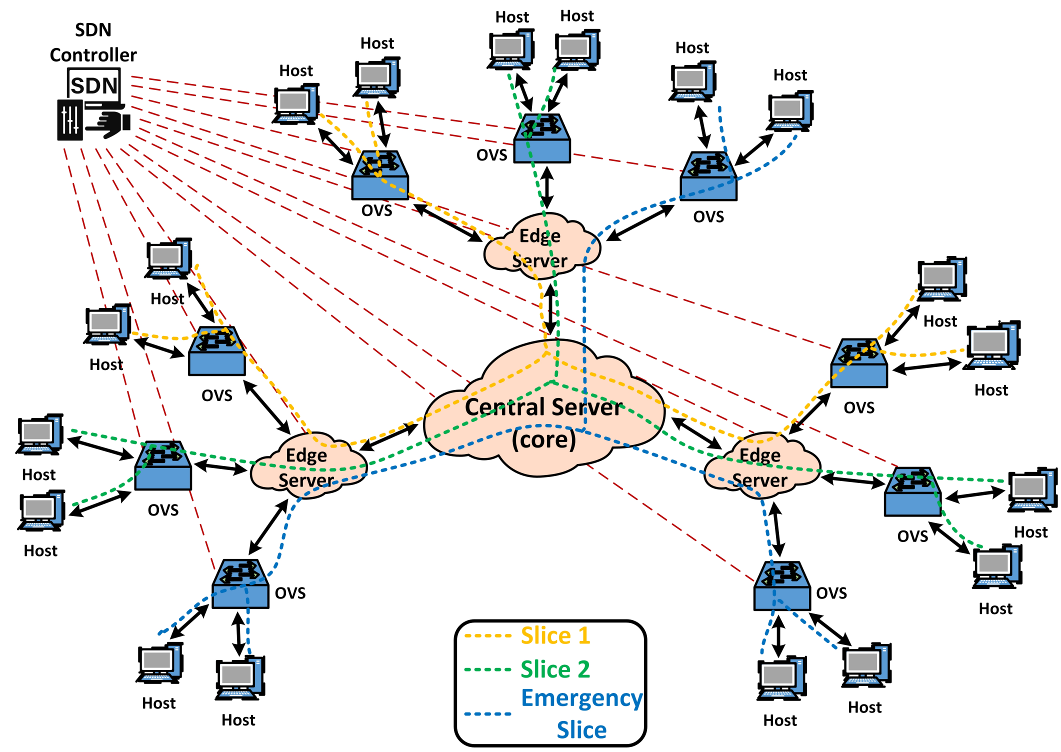


Fig. 4.5 Illustrating emergency slicing scenario with three slices (including emergency slice).

- Emergency slicing needs high-priority, dedicated and immediate access to networking resources or services for the first responders, critical workers and emergency management organisations to function optimally.
- In case of an emergency, the emergency slice created may consume a portion of the physical infrastructure to respond to critical situations and isolate emergency services from other portions of the infrastructure.
- The network abilities, such as resilience, redundancy and special QoS guarantees, can be obtained from these emergency slices, thus implying the network's failover or backup communication mechanism.

In this scenario, the emergency slice is particularly configured to segregate the critical traffic from the non-critical ones and meet the requirements accordingly. This slice prioritises the critical traffic and guarantees optimised QoS in emergencies to ensure reliable communication. During emergencies, the pre-designed emergency slice will be rapidly activated and deployed. In Figure 4.5, the two slices, i.e., Slice 1 and Slice 2, are for non-critical purposes, and the third slice, viz. emergency slice, is for critical purposes. The SDN controller

Algorithm 4: Proposed emergency slicing algorithm

Input: *d*: datapath, *i*: incoming port, *ovs_id*: OVS ID number, mac: MAC address of the hosts.

Output: Efficient emergency slicing of the TI communication infrastructure.

```
import SDN controller, OpenFlow (OF) protocol and traffic libraries.
class Emergencyslicing
def emergency_slicing[ovs_id][mac]:
    set the source mapping for the ports
    set end_ovs
def ovs_characteristics
Find and add the missing flow_entries in the OVS.
def packetOut_ovs(payload, d, i):
    sent_data = 0
    if the payload has no buffer then
        sent_data ← payload
    end
    PacketOut ← d, sent_data, i
    ovs ← PacketOut
def packetIn_ovs (payload, d, i):
    extract payload, d, i, ovs_id
    set source
    set destination
if ovs_id in emergency_slicing then
    if emergnecy == 0 then
        create a slice for trafficslice_1 and trafficslice_2 based on
          emergency_slicing[ovs_id][mac]
    else
        emergnecy == 1
        if destination in emergency_slicing[ovs_id] then
            f(ovs_id,i) = emergency_slicing[ovs_id][destination]
            add flow_entries to OVS directing based on
              desired f(ovs_id,i) with mapped destination
            packetOut_ovs
        end
        add an emergency_slice along with trafficslice_1 and trafficslice_2
          based on emergency_slicing[ovs_id][mac]
    end
end
```

will handle and route the different types of traffic through emergency and non-emergency slices. This controller will ensure that critical traffic is routed to the emergency slice while non-critical ones are routed to non-emergency slices (Slices 1 and 2). Configuring these slices assures the highest level of reliability and isolation, thus meeting desired network KPIs. Consequently, emergency slicing is achieved and configured as proposed in Algorithm 4.

$emergency\_slicing[ovs\_id][mac]$ is a dictionary defined to create and customise the traffic slices according to the three different traffic, i.e., $traffic\_1$, $traffic\_2$ and $traffic\_3$. This dictionary maps mac addresses and directs to the o (outgoing port) with the identified ovs_id. During normal situations ($emergency == 0$), $trafficslice\_1$ and $trafficslice\_2$ will be created based on the traffic rules set in $emergency\_slicing[ovs\_id][mac]$ dictionary. If an emergency has occurred ($emergency == 1$), $emergency\_slice$ along with $trafficslice\_1$ and $trafficslice\_2$ will be created, and the available resources will be allocated among themselves according to the pre-designed traffic rules in the controller.

4.5 Performance evaluation

4.5.1 Simulation environment

To evaluate the TI communication infrastructure concerning NS, Vagrant is used as a simulation environment. Vagrant [235] is an open-source tool used for building and managing virtualised development environments, i.e., virtual machines (VMs). It is easy and quick to create, configure and administer the reproducible virtualised environment to maximise the productivity and flexibility of reconfiguration for testing purposes. It uses a command line interface (CLI) to define and configure the VMs. To define the simulation environment in Vagrant, the configuration file, called ‘VagrantFile’, must specify the environment parameters such as base operating system (OS), CPU and memory allocation, and other network configurations. We used Linux-based OS in a Vagrant box as a base image and Oracle VirtualBox as a virtualisation platform provider to deploy a virtual machine. In our simulation model, six cores of CPU and 4096 Megabytes of memory were allocated to a Linux-based VM.

Moreover, the network emulator Mininet [236], an open-source tool, is used to create a virtual network topology comprising programmable virtual switches, hosts, links and an SDN controller for NS experimentation using Python3 script. The TI topology uses Open vSwitch (OVS), a Linux kernel-based software switch designed and deployed to provide network switching capabilities to manage the traffic slices based on the aforementioned NS scenarios. OVS is integrated with an SDN controller that facilitates the overall management

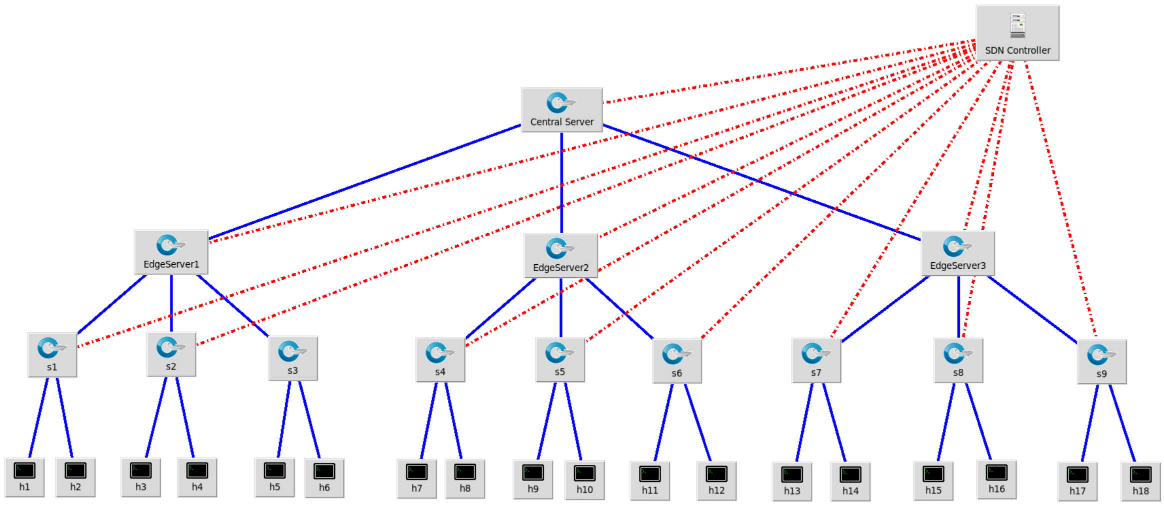


Fig. 4.6 Simulation model to enable scenario-based network slicing with SDN.

and control and defines forwarding rules in the TI communication infrastructure. Hence, this integration enables the dynamic NS configuration.

The simulation results obtained from the proposed infrastructure demonstrated stable performance validated at a 95% confidence level with an error margin of less than 5%. The 95% confidence level represents the efficient slices created for three different scenarios. The slices are tested in each scenario by generating three different traffic according to the proposed algorithms. On the other hand, a 5% error margin might arise because of inadequate resources and traffic congestion. Furthermore, the simulation is iterated for 30 trials to achieve stable, robust and optimised results in terms of throughput and round-trip time (RTT).

4.5.2 Modelling the network

The exploration of TI communication infrastructure with SDN and NS has a pivotal contribution to understanding real-world network dynamics. This subsection explains the proposed simulation model comprising one central server (core) and three edge servers. Each edge server is connected to one OVS. Subsequently, each OVS is connected to two hosts. Therefore, the TI communication infrastructure is designed with network resources, i.e., central cloud, three edge servers, three OVS and eighteen hosts, as represented in Figure 4.6. The SDN controller is also included in the infrastructure where it oversees, defines the forwarding rules and pushes to OVS. These configured forwarding rules are efficiently applied to the network traffic and managed based on the customised slicing scenarios. Table 4.2 lists the slicing configuration parameters used in the TI communication infrastructure.

Table 4.2 Slicing simulation parameters.

Parameters	Values
No. of hosts	18
No of central server	1
No. of edge server	3
No. of OVS	9
SDN controller processing time	0.5 μs
OVS processing time	5 μs
Central server's memory and storage	2048 Mb and 2 Gb
Edge server's memory and storage	1024 Mb and 1 Gb
Amount of spectrum allocated to slice 1	1000 Mbps
Amount of spectrum allocated to slice 2	100 Mbps
Amount of spectrum allocated to slice 3	10 Mbps
Amount of spectrum allocated to emergency scenario	10 Mbps
Before an emergency, amount of spectrum allocated to slice 1 and slice 2	7 Mbps and 3 Mbps, respectively
During an emergency, amount of spectrum allocated to slice 1, slice 2 and emergency slice	4 Mbps, 2 Mbps and 4 Mbps, respectively
Simulation trials	30

Considering the small complex physical topology, the simulation parameters of interest are chosen to have optimum values to achieve the robust and optimised performance of the network traffic slices. The amount of spectrum for traffic slices is set in order to meet the desired performance of configured slices. The effectiveness of configured network slices is assessed with the help of simulation parameters and proposed Algorithms 2, 3 and 4.

4.5.3 Results and discussions

To critically analyse the performance of configured slices in the proposed TI communication topology, we observe and measure the KPIs, such as throughput and RTT, for topology slicing, service slicing, and emergency slicing scenarios. In each scenario, three slices are designed and customised, and thus, the simulation is carried out using three different traffic, viz. UDP, TCP and other traffic packets. Diverse types of mentioned traffic are considered for simulation to imitate and create a realistic environment, thus allowing us to evaluate the performance of different protocols under different conditions. For example, real-time applications such as video conferencing calls may utilise UDP, whereas file transfer applications such as email conversations may rely on TCP. Specifically, based on the chosen

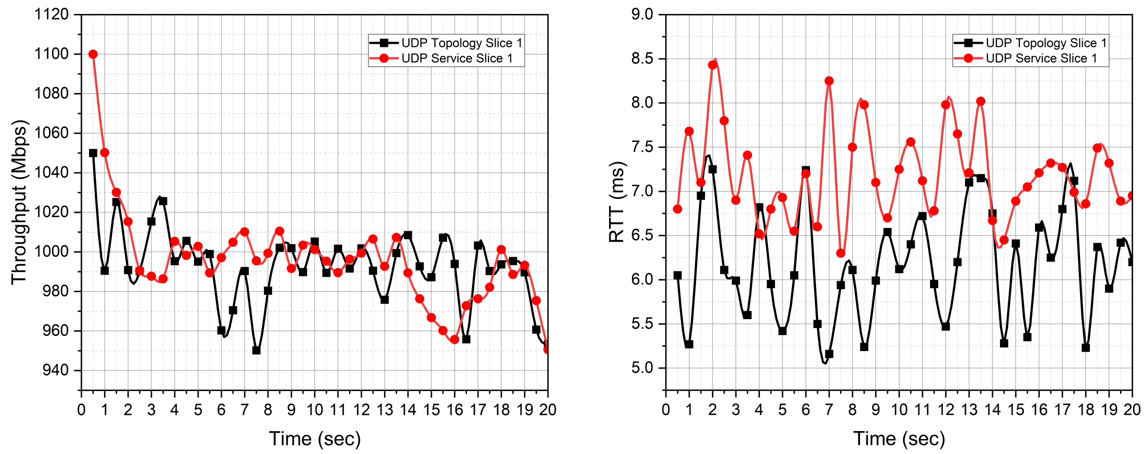
Linux-based VM (Linux distribution), TCP Cubic [237] is used as a default TCP variant in a Mininet simulation. However, UDP does not have the notion of a variant.

Furthermore, the traffic packets through the topology slicing scenario having slices 1, 2, and 3 are routed and designed with RTT constraint values of 5.5 ms, 7.5 ms and 11.5 ms, respectively. In contrast, for the service slicing scenario, slices 1, 2 and 3 are routed and designed with RTT as 6.5 ms, 8.45 ms and 13.35 ms, respectively. For an emergency slicing scenario, RTT equals 4.5 ms and 6.75 ms for slices 1 and 2 before an emergency. During an emergency, the traffic packets are routed with RTT equal to 8.35 ms, 14.25 ms and 7.55 ms, respectively. During simulation, the throughput for the mentioned traffic slices is measured using "*iperf3*". "*iperf3*" is a tool used to measure network throughput by generating UDP, TCP, and other traffic between hosts. The RTT for each traffic slice is measured using "*ping*", which includes SDN controller processing time and pre-designed links' RTT constraints. Another way to calculate RTT is the difference between sent and received timestamps for a traffic packet from source to destination, which can be aggregated to achieve average, maximum and minimum values. The mentioned traffic was generated for 20 seconds in an interval of 1 second for all three slicing scenarios. It was chosen to measure throughput and RTT for 20 seconds because the values were saturated afterwards. It reflected a reasonable time frame for observing and recording the trends in throughput and RTT. The proposed algorithms provide stable results in the first 20 seconds.

Based on our simulation results, it is observed that throughput in service slicing scenarios is recorded higher than topology slicing scenarios for the mentioned traffic because of better utilisation of available/allocated amount of spectrum to slices and reduced congestion, leading to slightly higher throughput. Conversely, RTT for service slicing scenarios has also been recorded to be higher than that for topology slicing scenarios. This is because of designed RTT constraint values (in ms) for topology slicing and service slicing scenarios.

Measurement and comparison of throughput and RTT - UDP traffic slice 1 in topology and service slicing scenarios

Considering topology and service slicing scenarios, the throughput and RTT are measured for slice 1 running UDP traffic. The amount of spectrum allocated to slice 1 in both scenarios is 1000 Mbps. It is observed that, in topology slicing, the throughput for UDP traffic slice 1 attains an average value of 992.8 Mbps with a minimum value of 950.2 Mbps and a maximum value of 1050.2 Mbps. On the other hand, in service slicing, the throughput accounts for an average of 996.1 Mbps with a minimum of 950.6 Mbps and a maximum of 1100.3 Mbps. By comparing the throughput in both scenarios with the same spectrum allocation, the average



(a) Throughput for UDP traffic slice 1.

(b) RTT for UDP traffic slice 1.

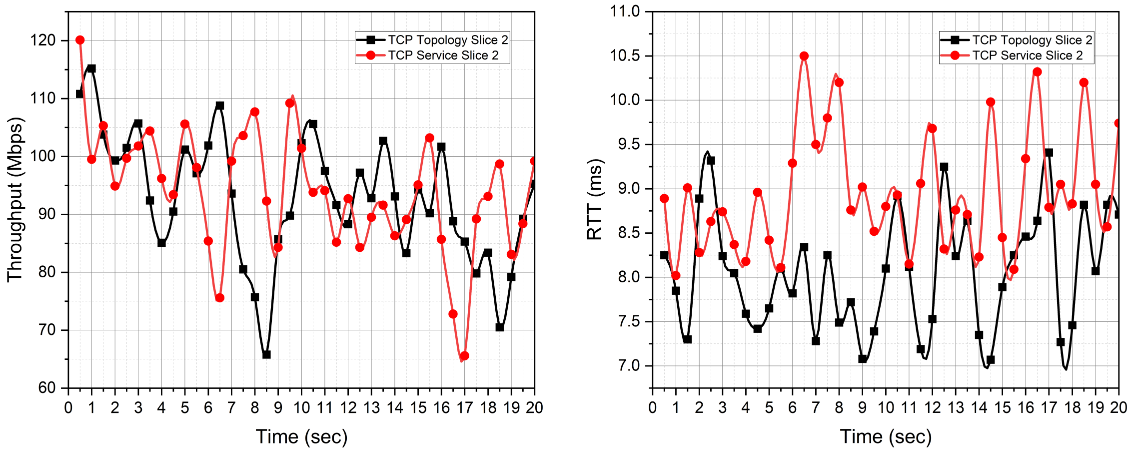
Fig. 4.7 Measurement of throughput and RTT for UDP traffic slice 1 in topology slicing and service slicing scenarios.

throughput in service slicing is approximately 3 Mbps more than topology slicing, as shown in Figure 4.7a. In other words, the average throughput in service slicing is slightly 0.33% higher than in topology slicing.

Moreover, concerning RTT in topology slicing scenario, the average RTT is measured as 6.17 ms with 5.16 ms and 7.25 ms as minimum and maximum RTT values. But, in the service slicing scenario, the average, minimum and maximum RTT is observed as 7.19 ms, 6.3 ms and 8.43 ms. It is observed that the RTT in both scenarios is almost the same, with a difference of 1.02 ms, which is 16.57% higher in the service slicing scenario than in topology slicing (see Figure 4.7b).

Measurement and comparison of throughput and RTT - TCP traffic slice 2 in topology and service slicing scenarios

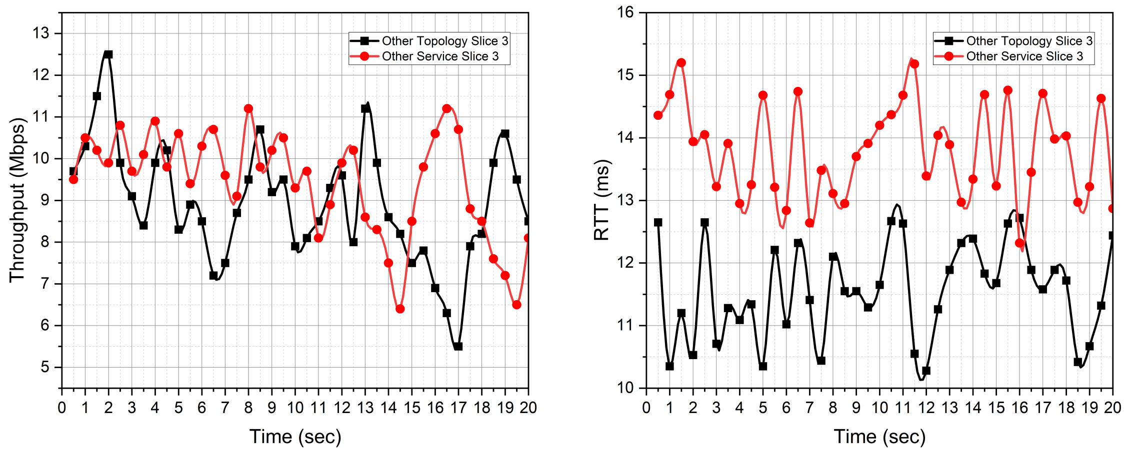
To measure the throughput and RTT, slice 2 is considered to have TCP traffic running in it. The amount of spectrum allocated to slice 2 in both scenarios is 100 Mbps. In the topology slicing scenario, the average throughput is recorded to be 92.9 Mbps and has minimum (65.8 Mbps) and maximum (115.2 Mbps) values. However, for the service slicing scenario, the average, minimum and maximum throughput are observed as 93.96 Mbps, 65.6 Mbps and 120.1 Mbps, respectively. By contrasting the throughput in both scenarios with the same spectrum allocation, as shown in Figure 4.8a, the average throughput in the service slicing scenario is slightly higher than the topology slicing scenario by 1.06 Mbps. There is a 1.14% higher throughput in service slicing than in the topology slicing scenario.



(a) Throughput for TCP traffic slice 2.

(b) RTT for TCP traffic slice 2.

Fig. 4.8 Measurement of throughput and RTT for TCP traffic slice 2 in topology slicing and service slicing scenarios.

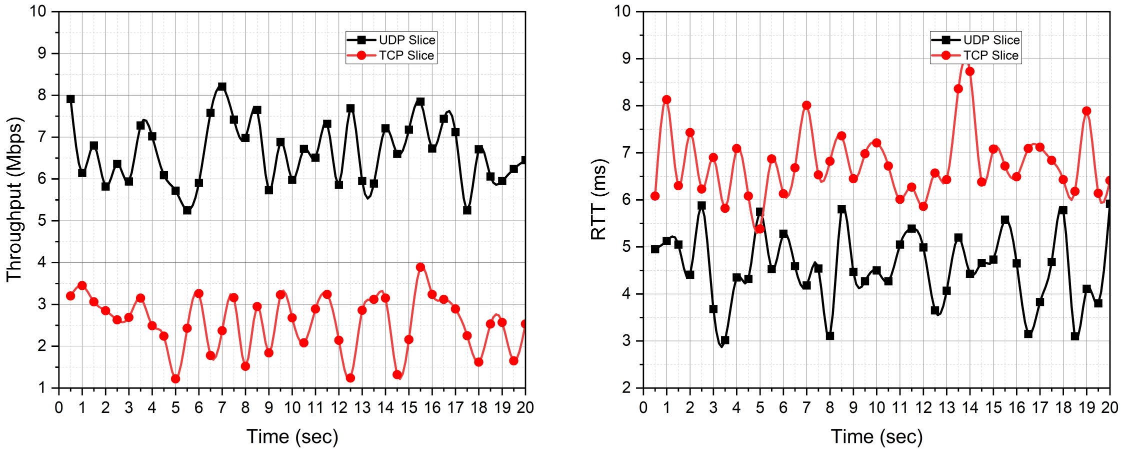


(a) Throughput for other traffic slice 3.

(b) RTT for other traffic slice 3.

Fig. 4.9 Measurement of throughput and RTT for other traffic slice 3 in topology slicing and service slicing scenarios.

For RTT in the topology slicing scenario, the average, minimum and maximum are measured as 8.06 ms, 7.07 ms and 9.41 ms, respectively. On the other hand, 8.95 ms, 8.02 ms and 10.5 ms values are observed as average, minimum and maximum RTT values in service slicing scenario. Hence, there is a slight difference in RTT between both scenarios with 0.95 ms (11% higher in the service slicing scenario compared to topology slicing), as depicted in Figure 4.8b.



(a) Throughput for UDP and TCP slices. (b) RTT for UDP and TCP slices.

Fig. 4.10 Measurement of throughput and RTT for UDP slice and TCP slice before an emergency scenario.

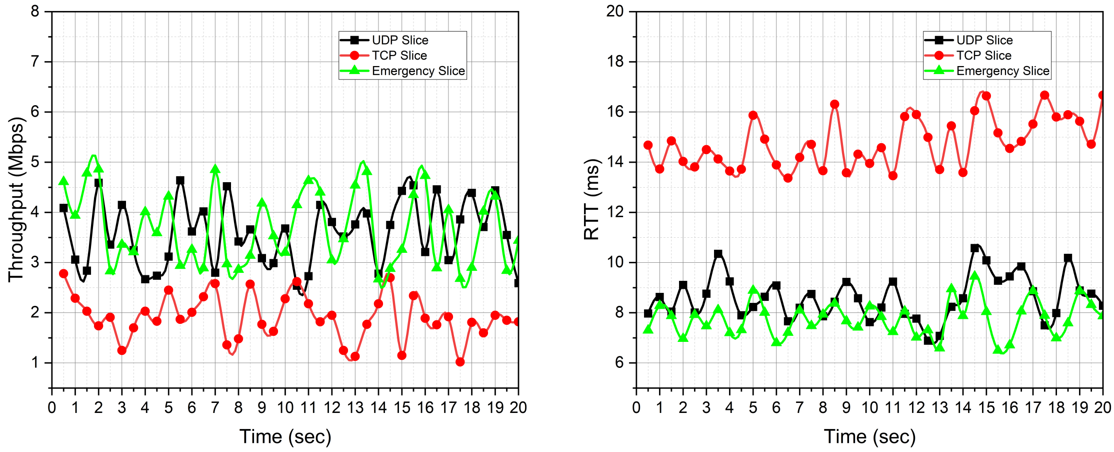
Measurement and comparison of throughput and RTT - other traffic slice 3 in topology and service slicing scenarios

The throughput and RTT are measured for slice 3, which is reserved for traffic other than UDP and TCP. In slice 3, only 10 Mbps resource spectrum is allocated in both scenarios. Considering the measurement for throughput, 8.93 Mbps, 5.5 Mbps and 12.5 Mbps are observed as average, minimum and maximum values, respectively, in topology slicing scenario. While in service slicing scenario, the average (9.43 Mbps), minimum (6.4 Mbps) and maximum (11.2 Mbps) values are observed. As depicted in Figure 4.9a, the throughput in service slicing is 1.05× greater (i.e., 5.6% higher) than topology slicing.

For RTT, the average value in topology slicing scenario is measured as 11.56 ms with minimum (10.28 ms) and maximum (12.72 ms) values, whereas in the service slicing scenario, average (13.79 ms), minimum (12.32 ms) and maximum (15.20 ms) RTT values are observed. To compare the RTT values, there is a difference of 2.23 ms in RTT values in both scenarios, where RTT in the service slicing scenario is 19.29% higher when compared to topology slicing (see Figure 4.9b).

Measurement and comparison of throughput and RTT in an emergency scenario

In emergency scenarios, the total amount of spectrum allocated is 10 Mbps. Before an emergency, only two slices (slice 1 with UDP traffic and slice 2 with TCP traffic) were designed with a spectrum allocation of 7 Mbps and 3 Mbps, respectively. During an emergency, two



(a) Throughput for UDP and TCP slices.

(b) RTT for UDP and TCP slices.

Fig. 4.11 Measurement of throughput and RTT for UDP slice and TCP slice during an emergency scenario.

slices (slice 1 with UDP traffic and slice 2 with TCP traffic) along the emergency slice are created with the spectrum allocation of 4 Mbps, 2 Mbps and 4 Mbps, respectively, with a total of 10 Mbps.

Throughput and RTT before an emergency and during an emergency are depicted in Figures 4.10a, 4.10b, 4.11a and 4.11b, respectively. The average, minimum and maximum values of throughput and RTT are tabulated in Tables 4.3 and 4.4. From these tables, it can be inferred that the network can guarantee performance when dedicated traffic packets flow through the configured slices before and during an emergency.

Table 4.3 Measurement of throughput (in Mbps) before and during an emergency.

Before emergency						**During emergency**								
Slice 1 (Mbps)			**Slice 2** (Mbps)			**Slice 1** (Mbps)			**Slice 2** (Mbps)			**Emergency slice** (Mbps)		
Avg.	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**
6.63	5.25	8.21	2.56	1.22	3.89	3.58	2.54	4.64	1.91	1.02	2.78	3.68	2.67	4.86

Table 4.4 Measurement of RTT (in ms) before and during an emergency.

Before emergency						**During emergency**								
Slice 1 (RTT constraint 4.5 ms)			**Slice 2 (RTT constraint** 6.75 ms)			**Slice 1 (RTT constraint** 8.35 ms)			**Slice 2 (RTT constraint** 14.25 ms)			**Emergency slice (RTT constraint** 7.55 ms)		
Avg.	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**	**Max.**
4.57	3.02	5.92	6.75	5.38	8.73	8.60	6.89	10.58	14.78	13.78	16.67	7.76	6.50	9.45

4.5.4 Algorithm evaluation and validation

This section evaluates and validates our proposed algorithms with existing relevant published algorithms presented below.

According to [238], the authors have proposed an algorithm based on the Markov decision process to have a dynamic network slicing in fog computing network powered harvested energy using BSs location deployed in an environment. It is managed by an orchestrator to energy and computational resources based on service demands to guarantee a specific QoS. A game theory strategy is being developed to optimise resource allocation, improving workload processing capacity. Similarly, another dynamic slicing approach in [239] has been developed for TI-based cyber-physical systems by proposing a clustering algorithm for network flow slices, mainly focusing on tactile telesurgery. These slices are managed by SDN and programmable switches to facilitate hard QoS guarantees required for TI-based cyber-physical applications, thus optimising resource utilisation.

Furthermore, in [232], integrated linear programming is formulated to create network slices in the 5G virtual core to minimise virtual network function instances and resource consumption. To address the scalability and dynamic traffic, a heuristic backtracking algorithm integrated with a deep learning model is proposed for network slicing to predict user requests, thus meeting users' service level agreement requirements. In [231], an adaptive slice allocation algorithm is proposed for 5G cloud-radio access networks, dynamically allocating resources to meet changing network demands. Therefore, through slice admission control, this algorithm maximises resource efficiency and ensures QoS. This algorithm aims to allocate network resources fairly across three 5G generic services (eMBB, mMTC and URLLC).

Moreover, in [224], the authors have integrated cloud native deployment in the proposed E2E network slicing framework. MANO framework is presented to automate E2E network slicing with the core network and transport slices. This 5G 3GPP compliant framework focuses on novel bandwidth management techniques by enabling QoS-aware network services. [240] have proposed an E2E network slicing framework for creating and deploying network slices using machine learning techniques. Using a multi-layer perceptron algorithm, the framework constructs network slices based on KPIs such as CPU capacity and bandwidth. The created network slices are categorised into 5G generic services (mMTC, eMBB, and URLLC), using Dijkstra's algorithm to maximise user access rate and resource efficiency by efficiently allocating optimal slices to the requested services.

The existing algorithms related to network slicing published in [238], [239], [232], [231], [224] and [240] are mostly equipped with deep learning and machine learning approaches and heuristic and Markov decision process models for efficient network slicing and resource allocation. Compared to existing algorithms, no published study proposes network slicing algorithms for three scenarios, i.e., topology, service, and emergency slicing, under the TI communication framework. The advantage of our proposed algorithms is the adaptability to perform and cater to diverse and tailored network slices under several traffic conditions. Though our algorithms do not encompass machine intelligence (ML/AI), the three different network slicing scenarios will be considered and expanded as future work to address the limitation of manual pre-configuration of network slicing based on users' requests/demands and enhance the adaptability and intelligence of the algorithm.

4.6 Summary

In this manuscript, we have proposed a novel network slicing (NS) mechanism based on TI communication infrastructure to address the issues of provisioning and controlling the network slices on demand. We have developed algorithms to optimise network structure (topology slicing), customise service (service slicing) and support critical communication (emergency slicing) according to the network's demand. Moreover, the SDN controller with programmable Open vSwitch (OVS) is used in these scenarios to create and design the network slices. The results obtained show that the measurement and comparison of throughput and round-trip time (RTT) for three designed slices align with the pre-designed allocated spectrum and RTT constraint values. However, in future, these algorithms and results can be further improved by considering the multiple access (MA) and ML/AL techniques. The MA techniques can positively impact NS, support many hosts, and enhance resource and spectral efficiencies. However, ML/AI techniques for NS can dynamically create and destroy customised slices based on network requirements, facilitate intelligent slice management, and detect anomalies and network faults in real time. Thus, these techniques can help implement NS in 5G and B5G. Incorporating both MA and ML/AL techniques in the proposed system is recommended for future research work.

Chapter 5

A Study of Downlink Power-Domain Non-Orthogonal Multiple Access Performance in Tactile Internet Employing Sensors and Actuators (Manuscript 4)

5.1 Manuscript 4 prelude

The Tactile Internet (TI) characterises the transformative paradigm that aims to support real-time control and haptic communication between humans and machines, heavily relying on a dense network of sensors and actuators. Non-Orthogonal Multiple Access (NOMA) is a promising enabler of TI that enhances interactions between sensors and actuators, which are collectively considered as users, and thus supports multiple users simultaneously in sharing the same Resource Block (RB), consequently offering remarkable improvements in spectral efficiency and latency.

Manuscript 4, entitled "A Study of Downlink Power-Domain Non-Orthogonal Multiple Access Performance in Tactile Internet Employing Sensors and Actuators[1]" [241], proposes a novel downlink power domain Single-Input Single-Output (SISO) NOMA communication scenario for TI by considering multiple users and a base station. The Signal-to-Interference Noise Ratio (SINR), sum rate and fair Power Allocation (PA) coefficients are mathematically derived in the SISO-NOMA system model. The simulations are performed with two-user and three-user scenarios to evaluate the system performance in terms of Bit Error Rate (BER), sum rate and latency between SISO-NOMA and traditional Orthogonal Multiple Access (OMA) schemes.

Moreover, outage probability is analysed with varying fixed Power Allocation (PA) coefficients in the SISO-NOMA scheme. In addition, we present the outage probability,

[1] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A study of downlink power-domain non-orthogonal multiple access performance in tactile Internet employing sensors and actuators," *MDPI Sensors*, vol. 24, no. 22, p. 7220, 2024.

sum rate and latency analyses for fixed and derived fair PA coefficients, thus promoting dynamic PA and user fairness by efficiently utilising the available spectrum. Finally, the performance of 4×4 Multiple-Input Multiple-Output (MIMO) NOMA incorporating zero forcing-based beamforming and a round-robin scheduling process is compared and analysed with SISO-NOMA in terms of achievable sum rate and latency.

Furthermore, this manuscript focuses on how the downlink PD-NOMA meets the stringent requirement under the TI communication infrastructure, thus enhancing and offering flexibility in allocating resources. Henceforth, this manuscript explicitly addresses the third research question: "*RQ 3. What multiple access technique, such as NOMA, can be developed to support low-latency tactile Internet applications/services?*".

5.2 Introduction

The tactile Internet (TI) in the fifth-generation communication infrastructure (5G) and beyond 5G (B5G) has been considered a potential enabler for service categories such as Enhanced Mobile Broadband (eMBB) communication, Massive Machine-Type Communication (mMTC) and Ultra-Reliable Low Latency Communication (URLLC) [19] to allow the interactions between humans, machines and Internet of Things (IoT) devices with audio and/or visual haptic feedback over a certain remote distance [22][1]. The TI communication infrastructure demands significantly high performance concerning ultra-low latency (≤ 1 ms), high bandwidth $(30\text{–}300)$ GHz, ultra-high availability (99.999%), ultra-reliable and high data rate support for augmented reality/virtual reality (AR/VR) applications.

Considering key enabling TI technologies such as software-defined networking (SDN), network slicing (NS), network function virtualisation (NFV), network coding (NC) and multiple access (MA) techniques, the TI communication infrastructure aims to improve their immersive, responsive and interactive services/applications. Thus, these technologies facilitate necessary dynamic infrastructure, optimising resource allocation and utilisation, offloading computational capabilities, and improving spectral efficiency, quality of service (QoS) and quality of experience (QoE).

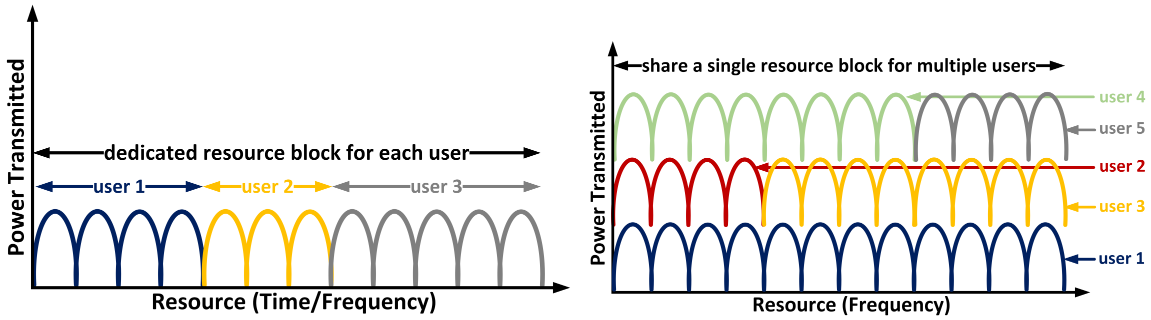
According to the Ericsson Mobility Report [242], a significant 28% surge was observed in mobile network data traffic between the 4^{th} quarter of 2022 and 2023, reaching a monthly consumption of 152 exabytes, contributing to approximately 8.5 billion mobile subscribers. In addition, the Cisco annual Internet report (2018–2023) [243] stated that by 2023, global Internet users were estimated to touch 5.3 billion, constituting 66% of the world's population. IoT devices will exceed three times the population, totalling 29.3 billion, whereas M2M (machine-to-machine) (a.k.a IoT) connections constitute half. The user devices surpassed the total connections, representing 74%, whereas the IoT applications forecasted substantial growth with a 30% compound annual growth rate (CAGR).

Despite the technological advancements in 5G communication infrastructure compared to first to fourth generations (1G - 4G) ones, the system capacity and spectral efficiency improvement fall short of the 1000$\times$ increase mentioned in the International Mobile Telecommunications - 2020 (IMT - 2020) vision [244].

To improve the system capacity and spectral efficiency, the mobile and wireless communication infrastructure have employed several MA techniques from 1G to 4G, such as 1G - Frequency Division Multiple Access (FDMA), second generation (2G) - Time Division Multiple Access (TDMA), third generation (3G) - Code Division Multiple Access (CDMA), and 4G - Orthogonal Frequency Division Multiple Access (OFDMA), or combinations of these MA techniques. These MA techniques are adopted to share the available bandwidth among multiple users and reply on the resource (time/frequency) orthogonalisation, thus assigning different resources to individual users. Moreover, the bandwidth is wasted due to users' utilisation of separate orthogonal resource blocks (RBs) using traditional MA techniques. These MA techniques are examples of orthogonal multiple access (OMA) schemes. Hence, these traditional MA techniques lack the competencies to serve multiple users having sensors and actuators simultaneously, rendering them unsuitable for a 5G communication infrastructure. In general, MA techniques can be structurally classified into OMA for 1G to 4G and non-orthogonal multiple access (NOMA) for 5G.

As mentioned, one of the high-performance enablers concerning parameters for TI communication infrastructure in 5G, the MA techniques, viz., NOMA, are complementing 5G New Radio (NR) robust, flexible and scalable framework. However, integrating NOMA can further resolve challenges and serve the demands of the proliferation of users, whether sensors or actuators, by utilising the same RB, according to 3GPP standardisation [245]. This integration facilitates high system capacity, enhances spectral efficiency, supports massive connectivity, and reduces latency in the network [246]. This technique concurrently assigns non-orthogonal resources amongst the multiple users whilst having some interference on the receiver side. In contrast, the traditional OMA techniques assign orthogonal resources solely to individual users, thus incurring challenges in achieving the required system capacity and spectral efficiency in 5G.

Figure 5.1 depicts the pictorial representation of NOMA when compared to OMA. Considering each user a sensor or an actuator, the power with respect to the resource graph is plotted to understand how the resources are allocated to the users in both cases. Figure



(a) OMA scheme (b) NOMA scheme

Fig. 5.1 Illustrating of OMA and NOMA schemes.

5.1a represents the OMA scheme - how the users are orthogonal to each other and have a dedicated/assigned RB for each user (U_1, U_2, ..., U_N). On the other hand, Figure 5.1b represents the NOMA scheme - how multiple users share a single RB, as the available bandwidth is not split across the users and assigns a single frequency channel to multiple users. Hence, the bandwidth wastage is minimised, and the system throughput is increased since the entire frequency spectrum is accessible to each user for transmission.

Consequently, it can be inferred that the NOMA environment is crucial and allows a dense deployment of sensors and actuators from users to share the assigned single RB, thus increasing the spectral efficiency and capacity compared to traditional OMA. Moreover, NOMA offers improved user fairness and QoS by having flexibility in power allocation (PA) and allowing users to be prioritised for stricter latency requirements by allocating them more power for faster data transmission. There are fewer waiting times to access the resources as users are sharing the resources concurrently. So this leads to overall reduced latency in data transmission. Finally, NOMA can effectively support heterogeneous networks with diverse user requirements and traffic characteristics. These benefits suit TI applications [19], such as enhanced mobile broadband, massive connectivity and latency-sensitive applications.

In addition, NOMA techniques can be categorised into power-domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA). In PD-NOMA, multiple users share the same RB. However, each user (whether a sensor or an actuator) is identified or allocated with varying power levels (coefficients) for transmission according to the channel conditions and distance from the base station (BS). Users with high channel gain (near BS) are allocated low power levels, whereas users with low channel gain (far from BS) are allocated high power levels. It utilises superposition coding (multiplexing) at BS and performs successive interference cancellation (SIC) at the receiver to detect and extract the users' original data signals and vice versa. In CD-NOMA, multiple users share the same RB and are assigned distinct non-orthogonal codes, such as low-density spreading (LDS), sparse code multiple access (SCMA) and multi-user shared access (MUSA) [247]. Increasing the number of users in CD-NOMA requires sophisticated algorithms for multi-user reception when non-orthogonal codes are adopted.

Furthermore, in PD-NOMA, both uplink (UL) and downlink (DL) transmissions play crucial roles in facilitating seamless message exchange between the user (whether a sensor or an actuator) and BS. Considering UL PD-NOMA, the simultaneous transmission happens from user to BS using the same RB and is distinguished by varying power levels. Conversely, for DL PD-NOMA, the simultaneous transmission happens from BS to the user, where BS assigns the proportion of power levels to multiple users based on the channel conditions and distance between BS and the user. Therefore, these PD-NOMA concepts enhance system capacity and optimise spectral efficiency, thus accommodating more sensors and actuators and reducing congestion in the TI communication infrastructure. Based on a sensor or actuator's priority, channel conditions and power requirements, it can empower low-power communication to conserve energy and extend the battery life with less frequent need to replace it.

5.2.1 Research challenges

While the PD-NOMA is broadly considered a promising multiple access technique which enhances the existing 5G-NR framework and B5G, several challenges [248][249] demand further investigation into implementing NOMA to improve capacity, spectral efficiency and latency. Considering resource allocation in NOMA, multiple users having sensors and actuators share the same RB. They are paired based on the distinct (strong and weak) channel gains, resulting in different PAs. At BS, the user signals are superimposed and transmitted, whereas at the user's location, SIC is applied to received user signals to decode and extract the original signal. However, decoding the superimposed signal using SIC considering a perfect channel state information (CSI) can be challenging. Assuming a perfect CSI can pose real-time channel estimation errors, which may degrade the system's performance [250]. Therefore, new hybrid techniques and algorithms can be developed to yield a practical channel estimation.

In addition, NOMA's system performance reduction may arise due to an imperfect SIC receiver [251]. This imperfection depends on building efficient hardware to reduce computational complexity. The efficient PA among multiple users can pose another challenge in the NOMA system. Inefficient PA may lead to interference in the received user signal, unfairness among paired users and higher outage probability, thus resulting in performance degradation. Moreover, most researchers focus on adopting a two-user pairing or sensor clustering for easy superimposition at BS and to reduce SIC complexity at the receiver's end. To cater to the growing demands of mMTC devices, multi-user pairing strategies or sensor clustering must be formed to take complete advantage of NOMA systems. Finally, due to the heterogeneous wireless network, a need for hybrid-NOMA (NOMA with other multiple access techniques) arises to address the diverse needs of devices through BS. Integrating NOMA with other multiple access techniques can be challenging, optimising the system's performance.

5.2.2 Motivation

PD-NOMA offers several significant benefits over traditional MA techniques, such as enhancing capacity, increasing spectral efficiency, improving user fairness and QoS, mitigating interference and serving multiple users with the same RB, thus addressing shortcomings of current communication infrastructure and directing future networks. Therefore, PD-NOMA flexible PA enables and prioritises eMBB, mMTC and URLLC TI applications/services with required resources for seamless operation. However, research in PD-NOMA is needed to improve system performance further and address the aforementioned challenges, such as PA algorithms, hardware impairments, receiver design with imperfect SIC and CSI, and integration of PD-NOMA with other MA techniques to cater to the diverse needs of user devices. Subsequently, it is driven by the need for an efficient, flexible, and user-centric communication infrastructure that meets the ever-growing demands of future networks.

5.2.3 Research contributions

In accordance with the findings from the literature review and to the best of our knowledge, the research still lacks in the field of PD-NOMA for TI applications. The performance analysis and evaluation of PD-NOMA in TI have not been comprehensively conducted in the past. Therefore, in this research, we have developed a novel downlink PD single-input single-output (SISO) NOMA communication scenario for TI employing multiple sensors and actuators (e.g. users), focusing on transmitted signals from a BS and processed received signals at the user's end. Furthermore, the key performance indicators (KPIs) are derived, discussed, and evaluated to compare the performance characterisation of NOMA and OMA schemes. The main contributions of this paper are highlighted below.

1. We developed an analytical system model incorporating signal-to-interference and noise ratio (SINR), sum rate, fair power allocation (PA) coefficients and latency among the available SISO-NOMA users.

2. We compared and analysed BER for SISO-NOMA and OMA schemes with varying path loss exponent and fixed PA coefficients using two- and three-user scenarios. To this end, we compared the achievable sum rate and latency trends for SISO-NOMA and OMA schemes for fixed PA coefficients.

3. The outage probability, achievable sum rate and latency trends are compared and analysed for fixed and fair PA coefficients. To this end, the performance trend of outage probability for SISO-NOMA users in varying fixed PA coefficients is also analysed.

4. Finally, the achievable sum rate and latency are compared and analysed between SISO-NOMA and 4×4 multiple-input multiple-output (MIMO) NOMA, incorporating a zero forcing-based beamforming and a round-robin scheduling process.

5.2.4 Structure of the manuscript

The rest of the manuscript is organised as follows. Section 5.3 presents the literature review relevant to the NOMA scheme used in the network. Section 5.4 describes the proposed system model for the downlink PD SISO-NOMA communication scenario for TI. This section also includes the derivation and analysis of SINR, achievable sum rate, fair PA and latency. The performance evaluation is presented in Section 5.5. The simulation results are also presented in this section. Finally, Section 5.6 concludes our work with future directions.

5.3 Related work

Several research studies have been conducted on NOMA to enhance system capacity, improve spectral efficiency and reduce latency. In [252], NOMA is adopted on 5G communication infrastructure (test-bed) with the objective of minimising latency and optimising resource allocation for autonomous vehicles (AVs). Also, some challenges for implementing and adopting NOMA for 5G-based applications are highlighted for researchers and AV manufacturers. A novel cloud-based queuing model for TI is proposed in [253] by utilising the PD-NOMA strategy. This model uses a baseband processing unit (BBU) and radio remote head (RRH) queuing delays for tactile end-users. Resource allocation is formulated to reduce the latency between tactile end-users. The transmit power is reduced by choosing a dynamic approach of fronthaul and access delays rather than a fixed approach. Finally, the energy efficiency of PD-NOMA and OFDMA is compared.

In [251], the critical features of NOMA are reviewed and highlighted merits and demerits against other OMA strategies. The features of several NOMA schemes, such as PD-NOMA, SCMA, MUSA and pattern division multiple access (PDMA), have been discussed. In addition, the KPIs such as sum rate, energy efficiency, and BER are compared between NOMA strategies. Thus, NOMA has the potential to achieve the objectives of the required KPIs. On the other hand, the authors in [254] have proposed deep learning-based grant-free NOMA to tackle and minimise the ultra-low latency and ultra-reliability requirements in massive access scenarios. They mentioned that the combination of grant-free access and NOMA can be leveraged and is a promising approach for tactile IoT. However, random interference is caused, which lowers the system's reliability. Hence, a grant-free NOMA-based neural network model and a novel multi-loss function are considered highly suited for automated applications in tactile IoT. Simulation results of the proposed model outperform the traditional grant-free NOMA strategies.

Moreover, in [255], the NOMA in mobile edge computing (MEC) enabled wireless tactile IoT scenario is investigated and optimisation algorithms for system and user performance. The proposed network model incorporates an MEC server at the access point, thus supporting computation for two sensor clusters. The system and cluster heads' performance using the successful computation probability (SCP) is assessed, mainly focusing on high signal-to-noise ratios (SNRs) for a comprehensive understanding. The simulation results of the proposed network model outperform and boost system performance compared to traditional OMA strategies. The authors in [5] explore and analyse the application-specific NOMA-based communication infrastructure for TI, allowing the non-orthogonal RB sharing among 5G generic services such as eMBB, mMTC and URLLC devices to a shared BS. A comparative analysis of NOMA and OMA is shown concerning the sum rate and number of users. Various NOMA variants are discussed to check the feasibility of future low-latency TI applications/services.

To meet the requirements of high spectral efficiency, ultra-low latency, and multi-user connectivity, the authors in [246] have considered NOMA a potential solution which allows some degree of interference on the receiver side. In contrast, the OMA technique may not meet the strict previously mentioned requirements. Here, they have focussed on providing a novel NOMA model, including uplink (UL) and downlink (DL) transmissions in MIMO and cooperative communication scenarios. Hence, the performance of NOMA and OMA are compared to analyse the system in terms of spectral efficiency, sum rate, and BER. In [256], authors have predominantly investigated the implementation of NOMA on a software-defined radio (SDR) platform. They have highlighted SDR as a flexible platform for testing and implementing 5G and B5G technologies. In addition, various SIC receivers such as Ideal SIC, Symbol-level SIC, Codeword-level SIC, and Log Likelihood Ratio (LLR) based receivers are mathematically evaluated and analysed. When NOMA is compared with OMA with simulation results, NOMA showcases its superiority over OMA in terms of KPIs.

Furthermore, a survey on the NOMA system is presented in [257], focussing on error rate analysis. The enhanced NOMA strategies, which consist of constellation diagrams, multicarrier systems, and detector designs, are discussed, along with research problems and future directions. In [258], a discussion of a deep learning technique for the NOMA systems is presented. The emphasis is given to deep learning-based NOMA systems for solving communication issues. These NOMA systems can be integrated with potential technologies such as MEC, MIMO, Intelligent Reflecting Surfaces (IRSs) and simultaneous wireless and information power transfer (SWIPT). Also, the focus has been given to KPIs such as SIC, CSI, user fairness, and other valuable parameters.

Considering the IRS, [259] has addressed the issues related to harnessing the performance of wireless networks in 1G to 5G propagation environments. The issues can be resolved in the sixth generation (6G) by employing the IRS with NOMA. The designs and challenges related to IRS-based NOMA systems are comprehensively discussed with a detailed analysis of the communication framework. In [260], a survey is conducted based on combining the benefits of NOMA and cell-free massive MIMO systems and a detailed review of how the performance can be increased. Moreover, the challenges of combining cell-free massive MIMO systems with other potential technologies are discussed.

Visible light communication is a potential solution for high-speed data communications. In [261], the authors have conducted a comprehensive review of NOMA techniques with the involvement of visible light communication systems. They also discussed the limitations and challenges of integrating NOMA with visible light communication systems and the role of machine learning and physical layer security. The authors in [262] have reviewed NOMA-enabled MEC systems in depth, focusing on the issues, challenges, and shortcomings that arise. They have claimed that integrating NOMA with MEC will bring umpteen performance characteristics to 5G and B5G communication infrastructure, such as energy efficiency, latency, throughput and massive end-user connectivity.

To overcome the issues related to poor channel quality and disconnected communication from BS to users, the work in [263] has analysed the BER and outage probability for multi-hop decode and forward relay-assisted NOMA systems. The BER and outage probability equations are also derived by considering imperfect SIC and CSI. The proposed model's simulation results depict the superiority over traditional OMA systems. The authors in [247] have focussed on the efficient strategies to integrate NOMA in 5G and 6G systems. This integration can be beneficial for UL and DL application environments under the 3^{rd} Generation Partnership Project (3GPP) standardisation.

The author in [264] has focussed on optimising PA coefficients to gain optimum proportional fairness in PD-NOMA transmission with complete or limited SIC. The numerical results illustrate the impact of complete or limited SIC on the system performance with sum rate loss due to proportional fairness. Similarly, [265] considers a hybrid automatic repeat request protocol for PD-NOMA with proportional fairness on the fading channel. It analyses the system performance for different symmetric and asymmetric scenarios concerning throughput, outage probabilities and delays with varying PA.

5.4 System model

A SISO-NOMA-based power-domain multiplexing system model is considered mathematically, consisting of a BS with users (U_1, U_2, ..., U_N). Here, we are considering a downlink SISO-NOMA communication scenario. At the transmitter side, the users' messages are multiplexed using superposition coding, whereas, at the receiver end, the received users' messages are demultiplexed (decoded) to retrieve the original message and perform successive interference cancellation (SIC) to remove interference from other user's messages.

5.4.1 Downlink PD SISO-NOMA communication scenario

Figure 5.2 represents the downlink power-domain communication scenario in TI. Let m_1, m_2, ..., m_N denote users' messages to be transmitted from the BS with transmitted power as P_T. Let the PA coefficients be α_1, α_2, ..., α_N with the corresponding channel coefficients as h_1, h_2, ..., h_N.

Let us consider U_1 to be the farthest user, followed by U_N to be the nearest user to the BS.

As superposition coding is performed at the BS, the transmitted signal (t_s) from BS is given by,

$$t_s = \sqrt{P_T}(\sqrt{\alpha_1}m_1 + \sqrt{\alpha_2}m_2 + ... + \sqrt{\alpha_N}m_N) \tag{5.1}$$

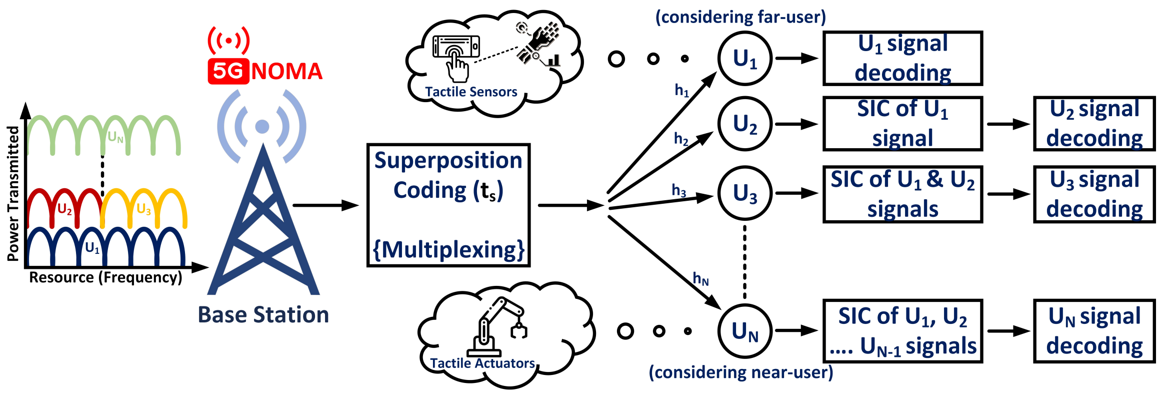


Fig. 5.2 Downlink power-domain communication scenario in TI.

In short, t_s at i^{th} user can be written as,

$$t_s = \sum_{i=1}^{N} \sqrt{P_T}\alpha_i \times m_i$$

$$t_s = \sqrt{P_T}\sum_{i=1}^{N} \alpha_i \times m_i \tag{5.2}$$

Therefore, the received signal (r_s) at i^{th} user is given by,

$$r_s = h_i t_s + n_i \tag{5.3}$$

Comparing Equations 5.2 and 5.3, r_s can be written as,

$$r_s = h_i \left[\sqrt{P_T}\sum_{i=1}^{N} \alpha_i \times m_i \right] + n_i$$

$$r_s = \sqrt{P_T} h_i \left[\sqrt{\alpha_1} m_1 + \sqrt{\alpha_2} m_2 + \ldots + \sqrt{\alpha_N} m_N \right] + n_i \tag{5.4}$$

where, n_i is additive white Gaussian noise (AWGN) with mean 0 and variance σ^2.

By considering Equation 5.4, the r_s for User 1 (U_1 = farthest user), where $i = 1$, will be,

$$r_s^{U_1} = \sqrt{P_T} h_1 \left[\underbrace{\sqrt{\alpha_1} m_1}_{Desired\ Signal} + \underbrace{\sqrt{\alpha_2} m_2 + \ldots + \sqrt{\alpha_N} m_N}_{Interference\ Signal} \right] + n_1 \tag{5.5}$$

As User 1 is allocated with the highest power coefficients (α_1), the r_s component of User 1 is the desired and dominant signal, and other components are considered undesired (interference).

Along the same lines, r_s for User 2 (U_2 = relatively near to BS), where $i = 2$, will be,

$$r_s^{U_2} = \sqrt{P_T} h_2 \left[\underbrace{\sqrt{\alpha_1} m_1}_{Undesired\ Signal} + \underbrace{\sqrt{\alpha_2} m_2}_{Desired\ Signal} + \underbrace{... + \sqrt{\alpha_N} m_N}_{Interference\ Signal} \right] + n_2 \tag{5.6}$$

As User 2 is allocated with a relatively lower power coefficient ($\alpha_2 < \alpha_1$), the r_s component of User 2 is considered as desired but not dominant. The dominant r_s signal is still the User 1 signal component, which is considered undesired (interference), along with other remaining user signal components. Hence, SIC is performed on the r_s to decode/retrieve the user's original message/signal to remove undesired (interference) signal components.

5.4.2 SINR analysis

To retrieve the original message from the received signal (r_s), the particular users' received signal must be directly decoded, considering other signal components as interference.

Therefore, to decode r_s signal for User 1 from Equation 5.5, the instantaneous SINR for User 1 is given as,

$$SINR_{U_1} = \frac{P_T \alpha_1 |h_1|^2}{P_T \alpha_2 |h_1|^2 + P_T \alpha_3 |h_1|^2 + ... + P_T \alpha_N |h_1|^2 + \sigma^2} \tag{5.7}$$

where, $|h_1|^2$ is channel gain for User 1.

Similarly, to decode r_s signal for User 2 from Equation 5.6, the r_s must decode the User 1 signal and perform SIC as User 1's received signal component is an undesired but dominant signal. Hence, after performing SIC, the resulting r_s will be,

$$r_s^{\prime U_2} = \sqrt{P_T} h_2 \left[\underbrace{\sqrt{\alpha_2} m_2}_{Desired\ Signal} + \underbrace{... + \sqrt{\alpha_N} m_N}_{Interference\ Signal} \right] + n_2 \tag{5.8}$$

Once User 1's dominant signal component is removed from Equation 5.6, resulting in Equation 5.8, the r_s signal can be directly decoded so that User 2 can have the desired and dominant signal after SIC. Hence, other remaining user signal components can be treated as interference. Hence, the instantaneous SINR for User 2 is given as,

$$SINR_{U_2} = \frac{P_T \alpha_2 |h_2|^2}{P_T \alpha_3 |h_2|^2 + ... + P_T \alpha_N |h_2|^2 + \sigma^2} \tag{5.9}$$

where, $|h_2|^2$ is channel gain for User 2.

Therefore, the SINR for i^{th} user can be expressed as,

$$SINR_{U_i} = \frac{P_T \alpha_i |h_i|^2}{P_T \alpha_{i+1} |h_i|^2 + ... + P_T \alpha_N |h_i|^2 + \sigma^2}$$

$$SINR_{U_i} = \frac{P_T \alpha_i |h_i|^2}{\sum_{j=i+1}^{N} P_T \alpha_j |h_i|^2 + \sigma^2} \tag{5.10}$$

where, $|h_i|^2$ is channel gain for i^{th} user.

5.4.3 Sum rate analysis

The achievable sum rate of i^{th} user for downlink SISO-NOMA can be computed as,

$$R_{U_i} = \log_2(1 + SINR_{U_i}) \tag{5.11}$$

From Equation 5.10, Equation 5.11 (R_{U_i}) can be written as,

$$R_{U_i} = \log_2\left(1 + \frac{P_T \alpha_i |h_i|^2}{\sum_{j=i+1}^{N} P_T \alpha_j |h_i|^2 + \sigma^2}\right) \tag{5.12}$$

Therefore, the overall achievable sum rate for SISO-NOMA downlink can be expressed as,

$$\begin{aligned} R_{overall} &= \sum_{i=1}^{N} R_{U_i} \\ &= \sum_{i=1}^{N} \log_2\left(1 + \frac{P_T \alpha_i |h_i|^2}{\sum_{j=i+1}^{N} P_T \alpha_j |h_i|^2 + \sigma^2}\right) \\ &= \sum_{i=1}^{N-1} \log_2\left(1 + \frac{P_T \alpha_i |h_i|^2}{\sum_{j=i+1}^{N} P_T \alpha_j |h_i|^2 + \sigma^2}\right) \\ &\quad + \log_2\left(1 + \frac{P_T \alpha_N |h_N|^2}{\sigma^2}\right) \end{aligned} \tag{5.13}$$

To evaluate SISO-NOMA scheme at higher SNR, the variance σ^2 tends to 0 ($\sigma^2 \to 0$). Hence, from Equation 5.13, the achievable sum rate can be approximated as,

$$R_{overall} \approx \sum_{i=1}^{N-1} \log_2\left(1+\frac{P_T\alpha_i|h_i|^2}{\sum_{j=i+1}^{N} P_T\alpha_j|h_i|^2+\sigma^2}\right)+\log_2\left(1+\frac{P_T|h_N|^2}{\sigma^2}\right)$$

$$R_{overall} \approx \log_2\left(\frac{P_T|h_N|^2}{\sigma^2}\right) \tag{5.14}$$

5.4.4 Fair PA analysis

The fair PA in SISO-NOMA downlink communication is crucial to ensure that all system users (U_1, U_2, ..., U_N) have a guaranteed QoS and maximise the sum rate. Thus, fair PA promotes user fairness, balances interference (undesired signal components), and enhances spectral efficiency.

Moreover, the PA coefficients (α_1, α_2, ..., α_N) are directly dependent on channel conditions. Without considering the channel conditions, the fixed PA coefficients are used, where the outage probabilities of the users are higher with a lower sum rate (bps/Hz).

Considering the CSI, the PA coefficients can be fairly optimised to improve the system performance. The fair PA focuses on the far user (User 1 in our case), as the far user is weak and located relatively away from BS compared to other users.

To better evaluate the SISO-NOMA scheme, let us consider two users (User 1 – far user and User 2 – near user) presented in the system. From Equation 5.12, the achievable sum rate for Users 1 and 2, respectively, can be given as,

$$R_{U_1} = \log_2\left(1+\frac{P_T\alpha_1|h_1|^2}{P_T\alpha_2|h_1|^2+\sigma^2}\right) \tag{5.15}$$

$$R_{U_2} = \log_2\left(1+\frac{P_T\alpha_2|h_2|^2}{\sigma^2}\right) \tag{5.16}$$

Hence, in this case, the PA coefficients (α_1 and α_2) are designed/derived by setting up the target rate (R_T) for the far user (User 1), which is less than or equal to the sum rate, i.e., $R_T \leq R_{U_1}$, from Equation 5.15. Once R_T for User 1 is met, the optimised PA coefficients can be derived instead of having a fixed PA.

Therefore, Equation 5.15 becomes,

$$R_T = \log_2\left(1+\frac{P_T\alpha_1|h_1|^2}{P_T\alpha_2|h_1|^2+\sigma^2}\right) \tag{5.17}$$

$$2^{R_T}-1 = \frac{P_T\alpha_1|h_1|^2}{P_T\alpha_2|h_1|^2+\sigma^2}$$

Let us consider $2^{R_T}-1=\beta$

$$\beta = \frac{P_T\alpha_1|h_1|^2}{P_T\alpha_2|h_1|^2+\sigma^2}$$

$$\beta P_T\alpha_2|h_1|^2+\beta\sigma^2 = P_T\alpha_1|h_1|^2 \tag{5.18}$$

We know that the sum of the PA coefficients is equal to 1 as we have two users (User 1 and User 2 having PA coefficients α_1 and α_2, respectively), $\alpha_1+\alpha_2=1$. Therefore, Equation 5.18 implies,

$$\alpha_1 = \frac{\beta(P_T|h_1|^2)+\sigma^2}{(1+\beta)P_T|h_1|^2};\alpha_1<1 \tag{5.19}$$

Hence, Equation 5.19 can be also written as,

$$\alpha_1 = \min\left(\frac{\beta(P_T|h_1|^2)+\sigma^2}{(1+\beta)P_T|h_1|^2},1\right)$$

Once the PA coefficient (α_1) of a far user (User 1) is calculated, the PA coefficient (α_2) of a near user (User 2) can be calculated using,

$$\alpha_2 = 1-\alpha_1 \tag{5.20}$$

In this case, if the calculation for the term $\left(\frac{\beta(P_T|h_1|^2)+\sigma^2}{(1+\beta)P_T|h_1|^2}\right)>1$ (let us consider 20), then $\alpha_1=\min\left(\frac{\beta(P_T|h_1|^2)+\sigma^2}{(1+\beta)P_T|h_1|^2},1\right)=1$, which does not satisfy the condition $\alpha_1+\alpha_2=1$.

If α_1 is designed to have 20, then User 1 will meet the target rate (R_T). But, if $\alpha_1<20$, User 1 will not meet R_T, leading to an outage condition. Moreover, α_1 cannot hold the value 20, which violates the condition $\alpha_1+\alpha_2=1$. Consequently, even if User 1's α_1 holds the value 1 (not meeting the considered condition $\alpha_1=20$), User 1 will be in an outage situation.

In contrast, if $\alpha_1 = 1$, then $\alpha_2 = 1 - \alpha_1$ will be 0. This infers that User 2 will not have any PA coefficient. Hence, User 2 will also be in an outage condition.

The solution for this outage problem of Users 1 and 2 can be resolved by setting up α_1 as 0 (no PA) and α_2 as 1 (entire PA). To keep User 1 away from the outage situation, the considered value of α_1 must be kept at 20. But, the ideal value of α_1 cannot exceed 1, i.e., for $\alpha_1 < 20$, User 1 will still be in an outage and not meet the desired target rate. Therefore, allocating the entire PA to User 2 ($\alpha_2 = 1$) can be considered to a point where allocating any PA to User 1 will not matter or affect the outage condition, i.e., User 1 will be in outage condition even if PA is done. Thus, User 2 will not be in the outage, achieving a higher sum rate (bps/Hz).

5.4.5 Beamforming with scheduling process for 4×4 MIMO use-case scenario analysis

For the downlink PD-NOMA scenario, a SISO-based NOMA has been considered and discussed in the previous sections, which is extended for the MIMO-based NOMA scenario.

In this scenario, consider a use-case having a BS with a uniform linear array (ULA) of 4 antennas with half-wavelength spacing. The BS serves 5 clusters/pairs of a total of 10 legitimate users with 2 users in each cluster. Each user is equipped with 4 receiving antennas. Assume the users in each cluster have almost the same angles with clusters equally spaced between $-60°$ to $60°$, i.e. $-60°$, $-30°$, $0°$, $30°$ and $60°$, concerning ULA.

In our multi-user 4×4 MIMO-NOMA scenarios, the beamforming technique proposed in [266][267][268], especially zero forcing-based beamforming (ZF-BF), can be used to mitigate the interference caused by multiple clusters of users, where multiple users' clusters are simultaneously served within the same RB. Such interference can lead to performance reduction and signal quality degradation. In ZF-BF, the channel matrix is formed to represent a BS and clusters of users' channel conditions and is used to calculate the precoding matrix by pseudo-inverting the channel matrix. Each column of the precoding matrix represents the beamforming vectors. So, the beamforming vectors are designed in such a way that they carefully direct/beam/steer the transmitted signal spatially to the desired/targeted users, forcing nulls to undesired/untargeted users or users' clusters, thus without causing interference between them. In this scenario, having 5 clusters, ZF-BF can serve 2 clusters (with a total of 4 users) simultaneously at a time, thus ensuring users within a cluster and inter-cluster do not interfere among themselves. Hence, ZF-BF helps improve the achievable sum rate, resource management, and signal quality.

Moreover, the scheduling process proposed in [269][270], specifically the round-robin scheduling process, can be utilised by a BS to serve clusters of users simultaneously during each time slot. After serving a cluster of users, the BS selects the next cluster of users in a round-robin sequence, thus managing equal access for each cluster to channel over time and promoting user fairness. Hence, in our case, with 4×4 MIMO-NOMA, 2 clusters with 2 users each (with a total of 4 users) can actively get channel access at a given time slot. The remaining cluster users will get their turn for the channel evenly over the following time slots. Imperfect CSI and SIC are not taken into consideration in this scenario.

It is worth noting that joint dynamic user clustering, beamforming, and a scheduling process can be mathematically modelled for MIMO-NOMA in Tactile communication infrastructure, which is beyond the paper's scope and will be included in our future work.

5.4.6 Latency analysis

In the downlink power-domain communication scenario in TI, latency can be defined as the delay incurred when the signal (ts) is transmitted from BS until it is received by users (U_1, U_2, ..., U_N). This latency includes delays such as transmission delay, propagation delay, queuing delay and processing delay.

The transmission delay (D_T) is the time required to transmit data packets over the communication channel. This delay considers the sum rate (Eq. 5.15 and Eq. 5.16) or set target rate (Eq. 5.17) of users and the amount of data to be sent (packet size). Thus, D_T is given by,

$$D_T = \frac{Packet\ size\ in\ bits}{Transmission\ rate\ in\ bps} \tag{5.21}$$

where, transmission rate = sum rate in (bps/Hz) $\times$ system bandwidth (Hz).

Moreover, the propagation delay (D_P) is the time required for the data packets to travel the physical distance between the BS and the user. This delay also relies on the transmission rate of the signal and thus is given by,

$$D_P = \frac{Distance\ between\ the\ BS\ and\ the\ user\ in\ metres}{Speed\ of\ light\ in\ m/s} \tag{5.22}$$

where, speed of light = 3×10^8 m/s.

The queuing delay (D_Q) is the time required to wait in queues at the BS before data packet transmission. This delay may arise due to network congestion or scheduled multiple users based on channel conditions. On the other hand, the processing delay (D_{Pr}) is the time required to process superposition coding (multiplexing) at the BS and SIC decoding at the receiver's side.

Hence, the overall latency ($L\_overall$) for a particular user will be given by,

$$L\_overall = D_T + D_P + D_Q + D_{Pr} \tag{5.23}$$

5.5 Performance evaluation

5.5.1 Simulation environment

To evaluate the performance of the NOMA system model, the MATLAB 2024a [271] platform is used along the 5G communication toolbox and required packages to carry out the simulations. This platform models the transmitter (BS), receiver (users) and channel condition of the NOMA system. Particularly, it can also model fading channels, such as Rayleigh fading, and simulate the path loss to mimic real-world propagation conditions.

Moreover, Monte Carlo simulation is applied to generate Rayleigh fading channels using random complex Gaussian variables, add AWGN to received signals at the receiver end, and run multiple simulation runs to achieve statistically meaningful numerical results.

5.5.2 Simulation results and discussions

To critically compare and analyse our proposed SISO-NOMA system in TI, we observe and measure KPIs, such as bit error rate (BER), achievable sum rate and outage probability. The fixed PA coefficient pairs are used to calculate the trend of the mentioned KPIs. Furthermore, to infer meaningful insights, an analysis of performance comparisons for BER and achievable sum rates is made between SISO-NOMA and OMA. In addition, the effect of fair PA on certain users is also considered whilst being compared against the fixed PA to see the outage probability and achievable sum rate trends of the users. Moreover, a comparative latency analysis is conducted for SISO-NOMA and OMA with fixed and fair PAs. Finally, the performance comparison concerning achievable sum rate and latency is analysed between 4×4 MIMO-NOMA and SISO-NOMA. The simulation results reported in this paper showed steady-state behaviour with a relative statistical error of $\leq 5\%$ at a 95% confidence level. Table 5.1 lists the simulation parameters used in the proposed NOMA-based TI system.

Table 5.1 NOMA simulation parameters.

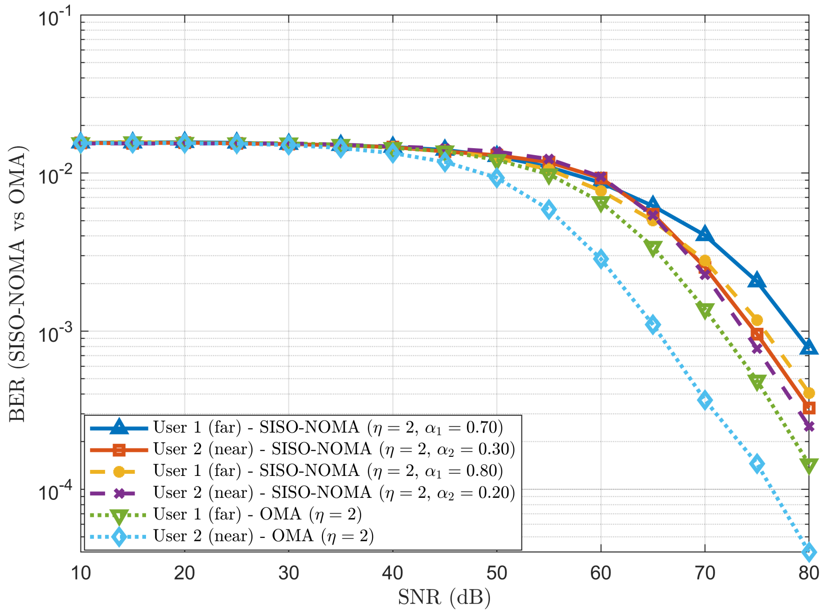
Parameter	Value
Users 1, 2 and 3 distance from BS	1000, 500 and 200 metres, respectively
For a two-user scenario, PA coefficients (α_1 & α_2)	(70% & 30%) and (80% & 20%), respectively
For a three-user scenario, PA coefficients (α_1, α_2 & α_3)	(70%, 20% & 10%) and (76%, 16% and 8%), respectively
Modulation scheme	Binary phase shift keying (BPSK)
Path loss exponent (η)	2 and 4
Channel	Rayleigh Fading
Number of OFDM subcarriers	128
Packet size	128 bytes
Noise	AWGN
System bandwidth	1 GHz
Power transmitted from BS	40 dBm

Performance comparison and analysis of bit error rate (BER) between SISO-NOMA and OMA

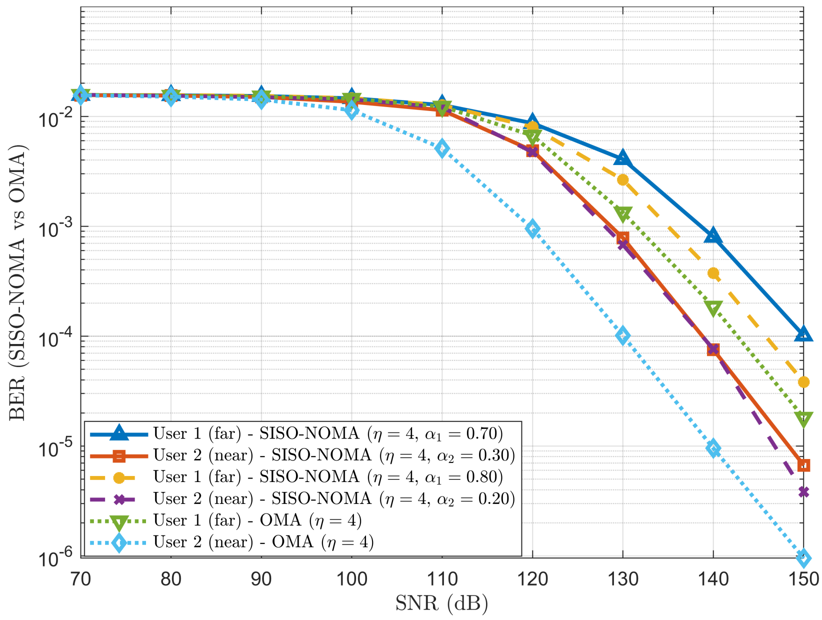
To study the effect of varying path loss exponent (η) with fixed PA coefficient values, the performance comparison of BER is observed and measured for SISO-NOMA and OMA schemes using two-user and three-user scenarios, with the following cases: *a*, *b*, *c* and *d*. The PA coefficients for Users 1, 2 and 3 are denoted as α_1, α_2 and α_3, respectively. The varying η allows system behaviour analysis under different environmental conditions. When η increases, the system represents the received signal's faster decay with distance.

Case (*a*): Considering a two-user scenario with $\eta = 2$ and fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$).

The BER performance trend is observed in Figure 5.3a against a varying signal-to-noise ratio (SNR) in decibels (dB) for SISO-NOMA and OMA schemes having a two-user scenario with $\eta = 2$ and fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$) for Users 1 and 2. User 1 (weak user) is located far away from BS, whereas User 2 (strong user) is located near BS. It is observed that User 1 with α_1 having 70% has incurred more BER than User 1 with α_1 having 80%, at an SNR = 80 dB. This performance trend is because more power has been allocated to User 1, with 80% than User 1, with 70%. In contrast, even though higher power is allocated to User 2 with 30%, it shows a higher BER than User 2, with 20%. This is mainly due to cross-user interference between User 1 with



(a) BER comparison between SISO-NOMA and OMA with η as 2.



(b) BER comparison between SISO-NOMA and OMA with η as 4.

Fig. 5.3 BER comparison between SISO-NOMA and OMA with η = 2 & 4, and fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$).

70% and User 2 with 30%. Thus, there is less interference between Users 1 with 80% and 2 with 20%. Keeping $\eta = 2$ constant, when the BER performance trend with the SISO-NOMA scheme is compared with the OMA scheme, it is seen that lower BER is noted for the user pairs (with 70% & 30%) and (with 80% & 20%). In the SISO-NOMA scheme, the extra signal processing complexity is added at the receiver end, i.e., decoding and performing SIC of received signals, thus leading to higher BER than in the OMA scheme.

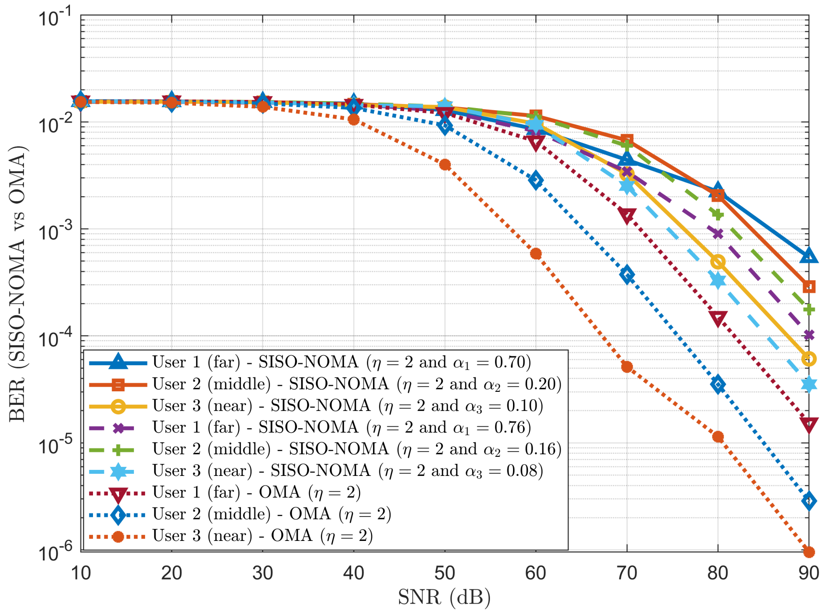
Case (*b*): Considering a two-user scenario with $\eta = 4$ and fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$).

On similar lines, the BER performance trend is observed in Figure 5.3b with the same fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$) for Users 1 and 2, but changing η from 2 to 4. For this simulation case, the effect of increasing η is noted for BER performance for Users 1 and 2. In Figure 5.3b, we observe that a slightly lower BER is encountered for User 1, with 80% than for User 1, with 70% at an SNR = 150 dB. In contrast, a similar trend is observed for Case (a), followed by User 2, which has a lower BER of 20%, than User 2, which has 30%. Moreover, in the OMA scheme, smaller BER is observed than in proposed SISO-NOMA scheme. Considering a constant $\eta = 4$, as the SNR increases, BER tends to improve as the received signal becomes more distinguishable than the noise signal. Hence, improvement in BER performance is observed when η is increased from 2 to 4, and the PA coefficient pairs are kept fixed. Consequently, a trade-off is observed in BER performance to strike a balance between η and PA coefficients to optimise the communication performance.

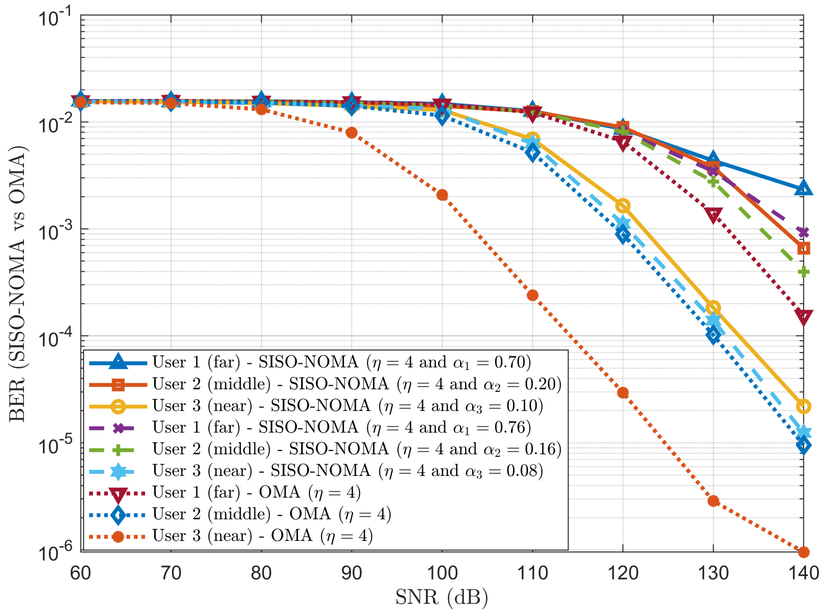
Case (*c*): Considering a three-user scenario with $\eta = 2$ and fixed PA 3-tuple coefficients as ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$) and ($\alpha_1 = 0.76$, $\alpha_2 = 0.16$ & $\alpha_3 = 0.08$).

The BER performance trend is observed in Figure 5.4a against varying signal-to-noise ratio (SNR) in decibels (dB) for SISO-NOMA and OMA schemes having a three-user scenario with $\eta = 2$ and fixed PA 3-tuple coefficients as ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$) and ($\alpha_1 = 0.76$, $\alpha_2 = 0.16$ & $\alpha_3 = 0.08$) for Users 1, 2 and 3. In this three-user scenario, User 1 is located far away from the BS, whereas User 3 is located near the BS. User 2 is located in the middle of User 1 and User 3. We observe that User 1, with 70%, has incurred higher BER than User 1, with 76%, at SNR = 90 dB. This is because less power (70%) is allocated to User 1 than to User 1 with 76%. Contrarily, lower BER is observed for User 2 with 16% than User 2 with 20%, even though higher power is allocated to User 2 with 20%. In both cases, cross-user interference is observed between the User 2 for PA coefficients are 16% and 20%, respectively. For User 3 (with 10% and 8%), almost the same BER is noted. As $\eta = 2$ is also kept constant for OMA scheme users, it can be clearly perceived that the BER performance trend is relatively less than SISO-NOMA scheme users, as OMA users do not have to undergo decoding and SIC operations for received signals.

Case (*d*): Considering a three-user scenario with $\eta = 4$ and fixed PA 3-tuple coefficients as ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$) and ($\alpha_1 = 0.76$, $\alpha_2 = 0.16$ & $\alpha_3 = 0.08$).



(a) BER comparison between SISO-NOMA and OMA with η as 2.



(b) BER comparison between SISO-NOMA and OMA with η as 4.

Fig. 5.4 BER comparison between SISO-NOMA and OMA with η = 2 & 4, and fixed PA coefficient pairs as ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$) and ($\alpha_1 = 0.76$, $\alpha_2 = 0.16$ & $\alpha_3 = 0.08$).

On the same note, Figure 5.4b shows the BER performance trend for $\eta = 4$ (instead of 2) and the same fixed PA 3-tuple coefficients as ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$) and ($\alpha_1 = 0.76$, $\alpha_2 = 0.16$ & $\alpha_3 = 0.08$). The same BER performance is observed for Case

(c) with User 1 (with 70% and 76%) at an SNR = 140 dB. However, User 2, with 20%, has shown higher BER than User 2, with 16%. User 2, with 20%, has experienced more cross-user interference with User 1 with 70%, while User, 2 with 16%, has experienced less cross-user interference with User 1 with 76%. Approximately the same BER is observed for User 3 (with 10% and 8%). When OMA scheme users are compared to SISO-NOMA users, lower BER is noted for far, middle, and near users with higher SNR values. It can be inferred that when η is increased from 2 to 4 in a three-user scenario, the degradation (increment) of BER performance is observed with $\eta = 4$ at higher SNR when compared with $\eta = 2$ due to increased interference.

Performance comparison and analysis of achievable sum rate between SISO-NOMA and OMA

Figure 5.5 shows the achievable sum rate in bps/Hz with respect to SNR values for SISO-NOMA and OMA. Considering a three-user scenario, Users 1, 2 and 3 have been allocated with fixed PA coefficients as $\alpha_1 = 0.70$ (70%), $\alpha_2 = 0.20$ (20%) and $\alpha_3 = 0.10$ (10%) and path loss exponent as $\eta = 2$. It can be observed that the achievable sum rate in the OMA performs slightly better than the SISO-NOMA at low SNR values. Due to cross-user interference and simulation transmission using the same RB, the achievable sum rate in the SISO-NOMA scheme suffers when compared to the OMA scheme at low SNR values.

However, as SNR increases, the achievable sum rate in the proposed SISO-NOMA scheme outperforms the OMA scheme.

Performance analysis of outage probability in SISO-NOMA

The performance analysis of outage probability in the SISO-NOMA scheme is carried out using a two-user scenario with fixed PA coefficient pairs as ($\alpha_1 = 0.70$ & $\alpha_2 = 0.30$) and ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$) and path loss exponent $\eta = 4$. To plot the outage probabilities, let us consider the target rates for Users 1 and 2 as 1.5 bps/Hz and 2.5 bps/Hz. These considered target rates are now compared with the achievable sum rate calculated in Eqs. 5.15 and 5.16. If the calculated achievable sum rates drop below the respective considered target rates of Users 1 and 2, the individual counters will be incremented. Hence, the outage probability is plotted as a function of power transmitted for both Users 1 and 2.

From Figure 5.6, it can be observed that User 1, with 70%, has more outage probability than User 1, with 80%. On the other hand, User 2, with 30%, has a lower outage probability than User 2, with 20%. Therefore, it can be concluded that the outage probability depends

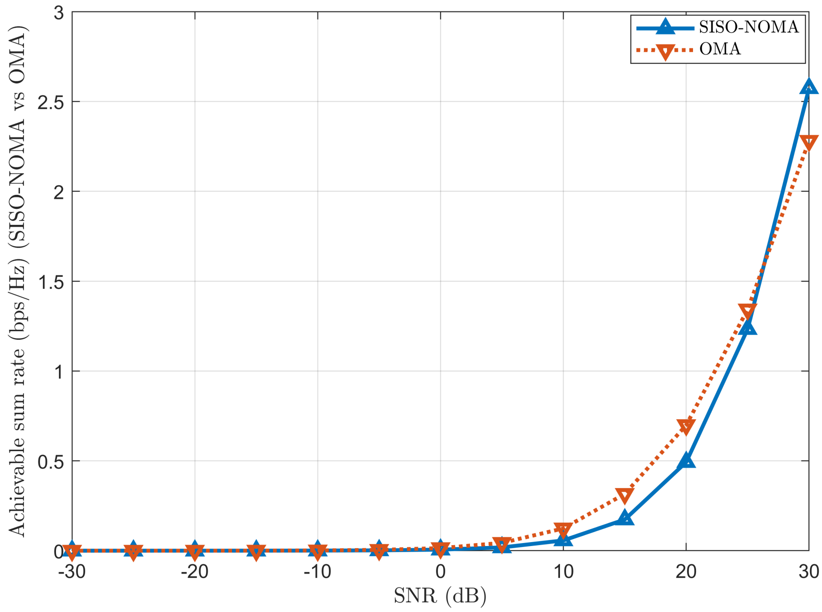


Fig. 5.5 Achievable sum rate comparison between SISO-NOMA and OMA.

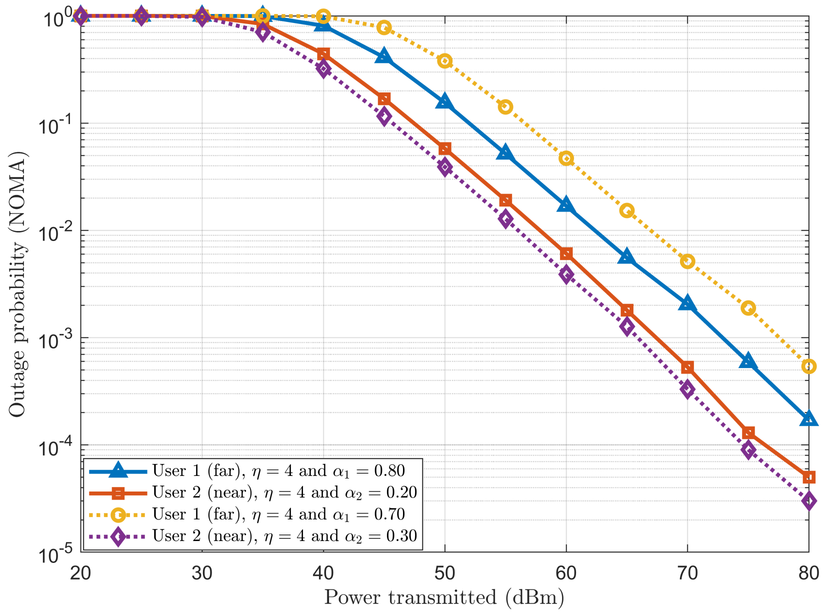
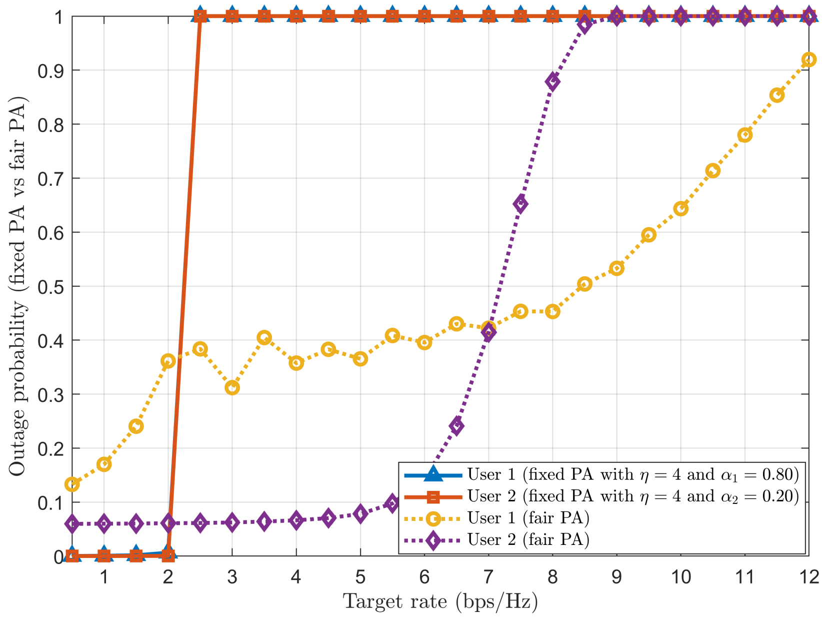
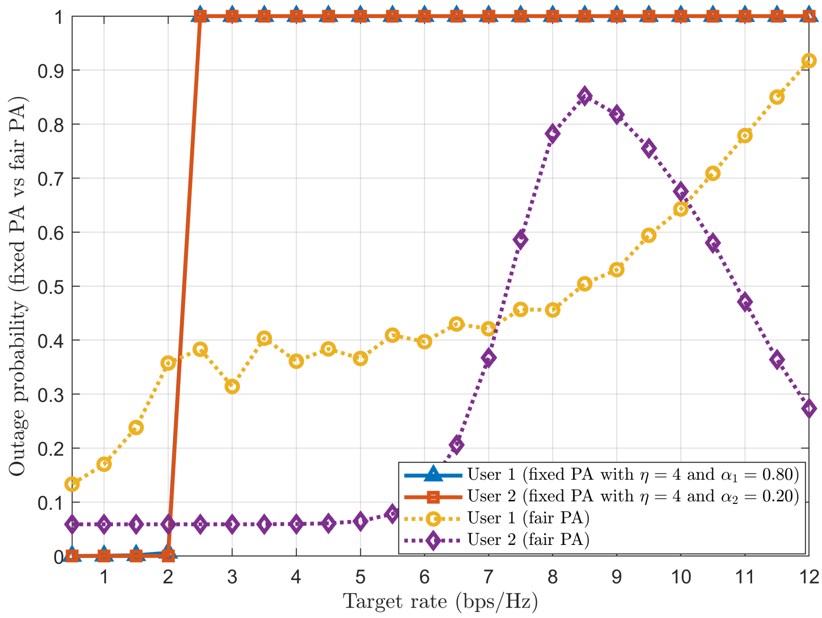


Fig. 5.6 Outage probability of SISO-NOMA scheme.

on the power percentage allocated to the particular user. Thus, in this case, User 1, with 80% and User 2, with 30%, have shown less outage probability than User 1 with 70% and User 2 with 20%, respectively.



(a) Outage probability of fair PA.



(b) Improved outage probability of fair PA.

Fig. 5.7 Outage probability of fair PA with a two-user scenario.

Performance comparison and analysis of fair PA with fixed PA

The outage probability with fair PA and fixed PA are plotted against the target rate in a two-user scenario with $\eta = 4$ (Figures 5.7a and 5.7b). For this simulation, the achievable sum rate derived in Eq. 5.15 for User 1 (far from BS) is considered the same as the target

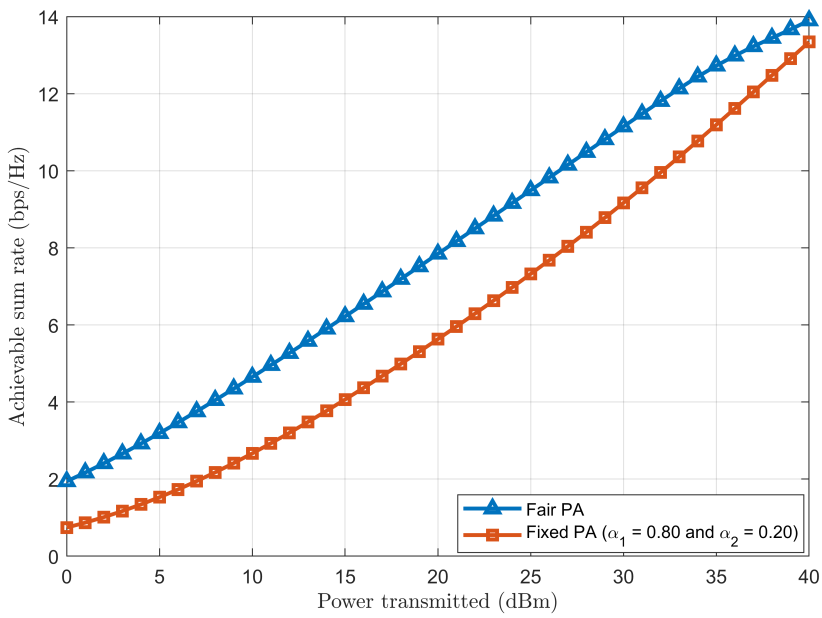


Fig. 5.8 Achievable sum rate comparison between fair and fixed PAs.

rate for User 1 (far). The PA coefficients α_1 & α_2 are calculated using derived Eqs. 5.19 and 5.20 for fair PA. The fixed PA coefficients for Users 1 and 2 are used as $\alpha_1 = 0.80$ and $\alpha_2 = 0.20$, respectively (for the transmitted power of 40 dBm).

By looking at Figure 5.7a, one can observe that Users 1 and 2 in fixed PA coefficients are in outage condition when the target rate is greater than 2 bps/Hz. This means that Users 1 and 2 in fixed PA will not experience an outage if the target rate is less than 2 bps/Hz. To infer, users in fixed PA are poorly performing and saturated to 1, as fixed PA neither considers the user's target rate requirements nor utilises instantaneous CSI conditions. Hence, it is not an optimal strategy to allocate power to users in the network. On the other hand, for a fair PA case, coefficients α_1 & α_2 are derived by considering target rate requirements dependent on the achievable sum rate of User 1 (far), i.e. CSI. As the target rate increases, User 1 (far) gradually goes into an outage condition, thus increasing the outage probability. However, User 2 (near) shows a sudden change in the outage probability trend when the target rate is between 5 bps/Hz and 8.5 bps/Hz. After 8.5 bps/Hz, User 2 also goes into an outage condition.

Figure 5.7b shows the improved outage probability for a fair PA case for the required target rate. The improved outage probability trend is achieved by forcefully setting User 1 (far) and User 2 (near) PA coefficients as $\alpha_1 = 0$ and $\alpha_2 = 1$. This setting is done when it does not affect the outage condition of User 1. When the target rate is less than 8.5 bps/Hz, derived PA coefficient α_1 is assigned to User 1 (far) without bothering about User 2 (near), thus focussing more on User 1's performance. However, when the target rate is above 8.5 bps/Hz, the PA coefficient $\alpha_2 = 1$ is assigned to User 2 (near), thus allocating the whole power to it and focussing on User 2's performance. Hence, User 2's outage probability is improved from 8.5 bps/Hz. User 1 has left no power ($\alpha_1 = 0$), which does not affect its outage condition.

Figure 5.8 shows that the achievable sum rate for having a fair PA is higher than having a fixed PA when plotted against transmitted power. Because of varying CSI conditions, the fair PA coefficients α_1 & α_2 are calculated to meet the desired target rate requirements, thus achieving an improved achievable sum rate in a fair PA case.

Performance comparison and analysis of latency

The performance comparison in terms of latency is conducted and observed with a fixed η in the following cases: *a* and *b*. In our simulation, the transmission and processing delays are the major contributors to the latency for the users, whereas queuing and propagation delays are considered negligible. For the queuing delay to be negligible, the simulation does not consider multiple users sending traffic simultaneously. Considering 2- and 3-user scenarios, the likelihood of traffic congestion is quite low without queueing the packets on the BS side, thus creating less traffic demand at the BS. Since the propagation delay depends on the distance between the BS and the user, the simulation scenario considers user distances from the BS, such as 1000 m, 500 m, and 200 m. Hence, the propagation delay will be in the order of μs, thus treating it as negligible.

Case (*a*): Considering a three-user scenario and comparing the latency trend for SISO-NOMA and OMA schemes.

The latency trend is observed in Figure 5.9 as a function of SNR for SISO-NOMA and OMA schemes having a three-user scenario with $\eta = 2$ and fixed PA coefficients ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$). The latency is plotted for each user in both cases, i.e., SISO-NOMA and OMA. This figure represents how the latency varied with varying SNR (signal quality) having three different user distances from the BS. We observe that the latency decreases with increasing SNR. At low SNR, the latencies in SISO-NOMA show higher latency than in OMA as the system consumes more time to multiplex (superposition coding) and decode (SIC) the signal. However, as the signal improves with increasing SNR and negligible noise, the rapid reduction of latencies in SISO-NOMA is observed around 5 dB to 10 dB due to appropriate PA and simultaneous access to RB. As User 1 is located farthest from the BS and most likely goes to an outage condition, more power (70%) is allocated to ensure efficient

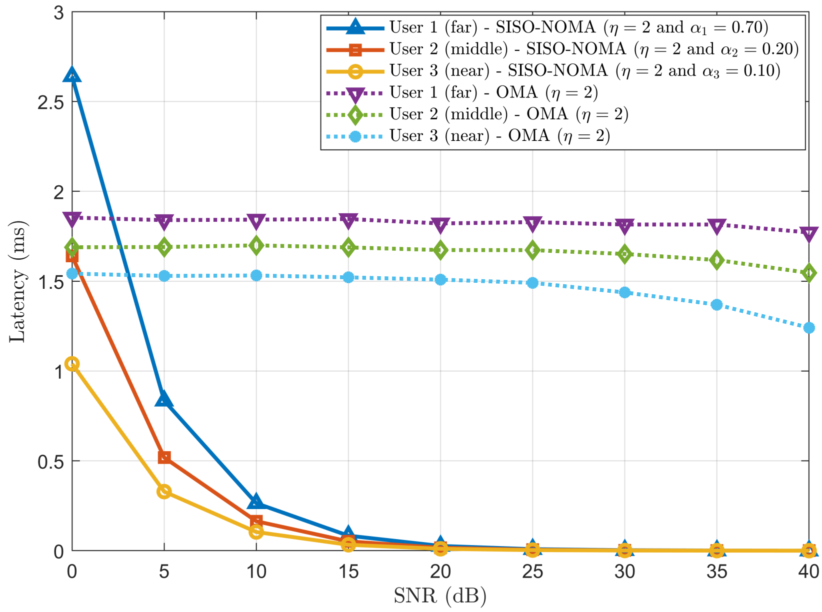


Fig. 5.9 Latency comparison between SISO-NOMA and OMA with $\eta = 2$ and fixed PA coefficient ($\alpha_1 = 0.70$, $\alpha_2 = 0.20$ & $\alpha_3 = 0.10$).

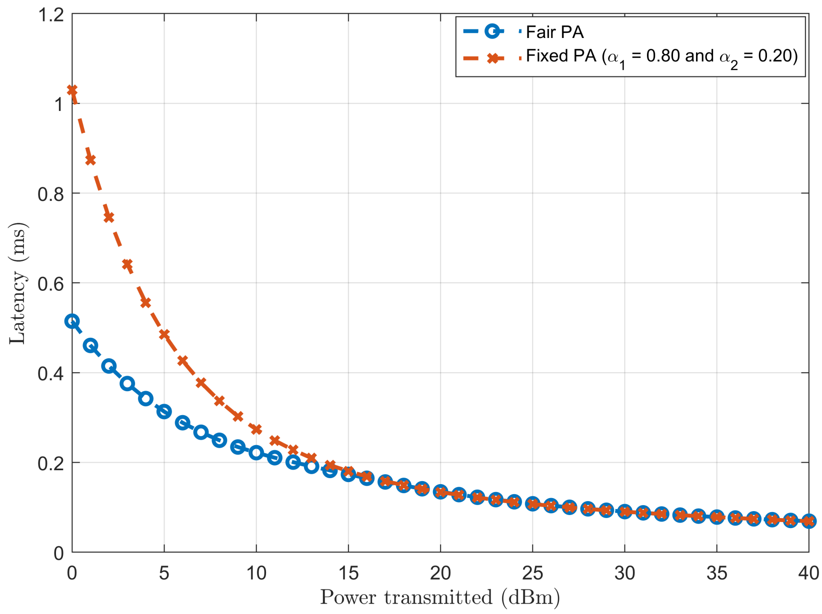


Fig. 5.10 Latency comparison between fair and fixed PAs in SISO-NOMA.

sharing of resources, thus contributing to the higher latency than User 2 (20%) and User 3 (10%) in SISO-NOMA.

The signal quality is poor for OMA at low SNR, leading to higher BER, as observed in Figure 5.4a. Hence, frequent retransmissions are required, increasing each user's latencies.

Figure 5.9 represents that at low SNR, the latencies for each user are higher. However, as SNR improves, the decreasing trend of latencies for each user is observed as SNR increases from 15 dB. Moreover, the latencies depend on users' respective distances from the BS, as a similar trend is observed for SISO-NOMA users' latency. To compare the performance in terms of latency, SISO-NOMA outperforms OMA as SNR increases.

Case (*b*): Considering a two-user scenario and comparing the latency trend for fair and fixed PAs in SISO-NOMA.

Figure 5.10 represents the latency trend for fair (derived in Eqs. 5.19 and 5.20) and fixed ($\alpha_1 = 0.80$ & $\alpha_2 = 0.20$) PAs as a function of the transmitted power in SISO-NOMA. The latency trends for fair and fixed PA cases decrease with increasing transmitted power. However, the decrease in latency for fair PA is more pronounced than in fixed PA, as fair PA dynamically adjusts the power allocated to SISO-NOMA users based on channel conditions. Such fair PA adjustments most likely provide better latency performance because of the efficient utilisation of resources. On the contrary, for fixed PA, users are allocated a fixed proportion of power independent of channel conditions, thus leading to inefficient use of the available power. In comparison, better latency performance is observed in fair PA than in fixed PA as the transmitted power increases.

Performance comparison and analysis of 4 $\times$ 4 MIMO scenario in NOMA

In our simulation, the performance comparison of 4 $\times$ 4 MIMO-NOMA with SISO-NOMA is conducted and observed in terms of the achievable sum rate and latency with a fixed $\eta = 4$ and 10 users, as previously mentioned in Section 5.4.5. The simulation scenario considers the user's distances between 1000 m and 200 m from a BS, with User 1 being farthest and User 10 being closest to BS. From User 1 to User 10, the distances from BS are 1000, 900, 800, 700, 600, 500, 400, 300, 250 and 200 metres. Five user clusters are formed with 2 users in each cluster with a near-near and far-far clustering approach. The achievable sum rate and latency comparison are analysed in the following cases: *a* and *b*.

Case *(a)*: Performance comparison and analysis of achievable sum rate between 4 $\times$ 4 MIMO-NOMA and SISO-NOMA.

The achievable sum rate comparison is observed for 4 $\times$ 4 MIMO-NOMA and SISO-NOMA in Figure 5.11 when plotted against transmitted power. This figure represents an increase in the achievable sum rate as the transmit power increases for both cases, 4 $\times$ 4 MIMO-NOMA and SISO-NOMA, thus improving SNR for higher transmitted power. It can be clearly seen that 4 $\times$ 4 MIMO-NOMA outperforms SISO-NOMA as transmitted power increases. This is because of exploiting spatial diversity and multiplexing, thus allowing

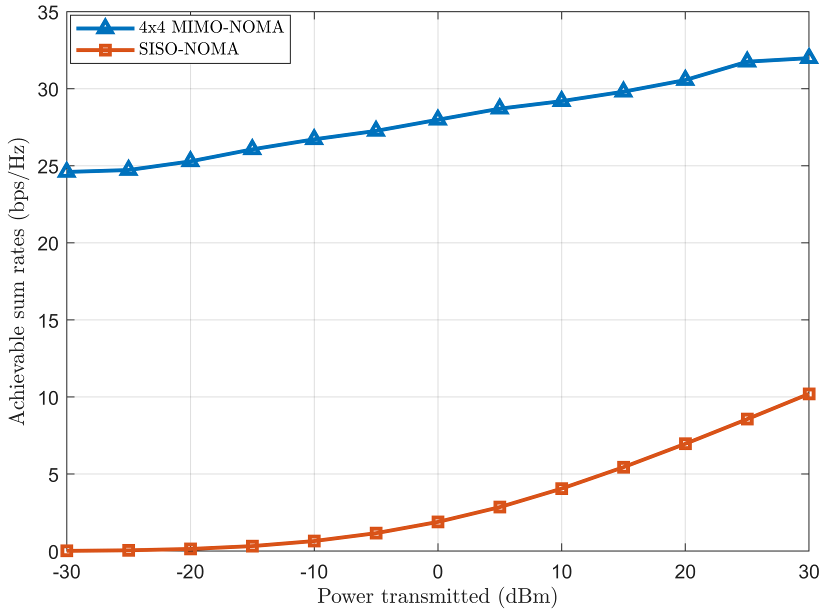


Fig. 5.11 Achievable sum rate comparison between 4×4 MIMO-NOMA and SISO-NOMA.

multiple data streams to be transmitted from the BS to multiple users simultaneously in 4×4 MIMO-NOMA. Moreover, the incorporation of ZF-BF further improves the performance of 4×4 MIMO-NOMA by reducing the spatial inter-user and inter-cluster interference. In contrast, a round-robin scheduling process promotes fair user access to the RB as it cycles through the clusters of users and allocates time slots sequentially, which does not prevent any clusters from having access to RB and does serve each cluster in each round.

Case *(b)*: Performance comparison and analysis of latency between 4×4 MIMO-NOMA and SISO-NOMA.

Figure 5.12 represents the latency plot of 4×4 MIMO-NOMA and SISO-NOMA against the increasing transmitted power. It can be observed that 4×4 MIMO-NOMA incurs less latency than SISO-NOMA. This is because of effectively nullifying interference in ZF-BF; thus, clusters of users do not have to wait for a clear transmission slot from BS and are served simultaneously. Also, it helps in interference-free streams for each cluster of users, thus reducing the need for retransmission and error correction and contributing to overall lower latency. However, the complexity of designing a beamforming matrix for many antennas or users may potentially increase the overall latency. However, with careful beamforming matrix design, reduced interference and parallel data transmissions offset latency. On the other hand, all clusters of users are served sequentially in a round-robin scheduling process, thus reducing the waiting time of clusters of users promoting user fairness and overall latency. Also, this scheduling process reduces the queuing delay as it cycles through the clusters of

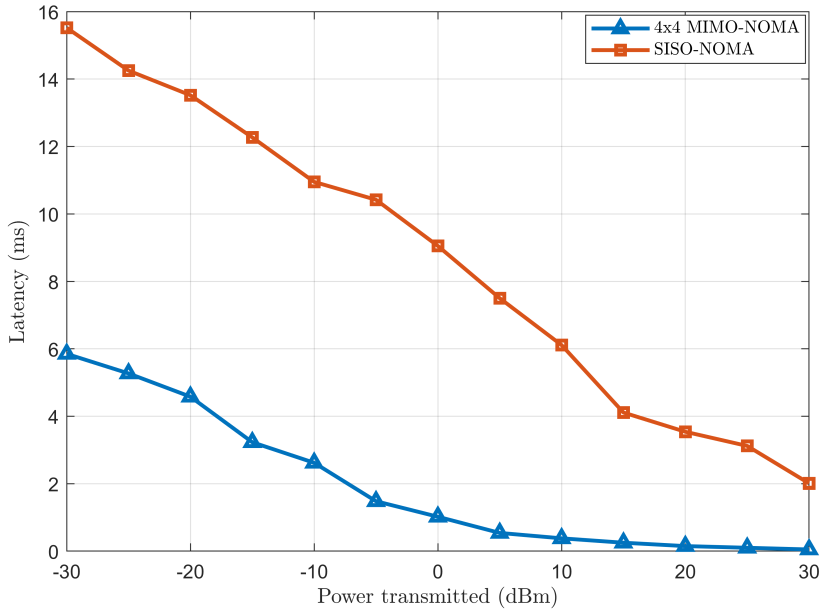


Fig. 5.12 Latency comparison between 4 × 4 MIMO-NOMA and SISO-NOMA.

users. Without any scheduling process, the clusters of users may experience delays or have to wait longer for their turn for transmission, thus making higher overall latency and lesser system responsiveness and avoiding constant allocation of RB to high-priority users or users having channel conditions.

5.5.3 Model assessment and validation

This section assesses and validates the robustness of our system model by comparing it with existing work done in incorporating NOMA in TI wireless networks.

To the best of our author's knowledge, no existing works have focussed and analysed mathematically derived SINR, achievable sum rate and fair PA coefficients altogether to meet the TI user's stringent requirements. Most of the existing work done in [246], [250], [254], [257] and [263] have focused on and shared primitive analyses on downlink PD SISO-NOMA, considering mostly two users in a scenario. The NOMA performance has not been rigorously evaluated and analysed on varying network parameters, such as PA coefficients and η, to observe users' performance effects. Moreover, the comparison between SISO-NOMA and traditional OMA concerning BER, sum rate and latency trends with varying parameters mentioned earlier with a two-user and a three-user scenario has not been studied well in the networking literature.

Nonetheless, the fair PA analysis in terms of outage probability presented in our work has focussed on deriving PA coefficients to meet the stringent target rates of users under TI communication. Therefore, the PA coefficients can be specifically calculated and tuned to meet a specific SISO-NOMA user's target rate requirement. In addition, the achievable sum rate and latency with calculated PA coefficients are compared with fixed PA coefficients, which was also missing from previous works. Along similar lines, the performance of 4×4 MIMO-NOMA incorporating zero forcing-based beamforming and a round-robin scheduling process are compared with SISO-NOMA in terms of achievable sum rate and latency, which is not exclusively mentioned in previous work.

However, this research did not focus on machine intelligence with imperfect SIC conditions, which can be considered for future research directions.

5.6 Summary

A novel downlink PD SISO-NOMA communication scenario for TI employing multiple sensors and actuators, collectively treated as users, and a base station is proposed in this manuscript. We have developed an analytical system model comprising SINR, achievable sum rate, fair power allocation (PA) coefficients and latency for SISO-NOMA users to study the system performance. The system model is validated by simulation scenarios with varying path loss exponents and fixed PA coefficients. We have evaluated and validated the analytical model by simulation. A higher BER is achieved for SISO-NOMA users than for the OMA scheme due to additional signal processing to decode and perform Successive Interference Cancellations (SICs) for received signals. In contrast, the achievable sum rate and latency trends for the proposed PD SISO-NOMA have outperformed the OMA scheme for higher SNR. The outage probability of SISO-NOMA is also analysed with varying fixed PA coefficients. Finally, to promote dynamic PA and user fairness of the proposed PD SISO-NOMA scheme, we have compared the outage probability, achievable sum rate and latency for fixed and derived fair PA coefficients. The 4×4 MIMO-NOMA incorporating zero forcing-based beamforming and a round-robin scheduling process has outperformed SISO-NOMA in terms of achievable sum rate and latency. Thus, NOMA has maximised the spectral efficiency while minimising the power consumption and latency by efficiently utilising the available spectrum/resources. Proposing a joint dynamic user clustering, beamforming and scheduling process for MIMO-NOMA and developing a deep-learning-based NOMA algorithm are suggested as future research work.

Chapter 6

Conclusion and Future Directions

This chapter presents our main contributions to the thesis, where we have synthesised the findings and insights gathered by addressing the research questions RQ 1, RQ 2 and RQ 3 of tactile Internet (TI). Also, we have discussed the potential future research directions that can be conducted for further exploration.

6.1 Conclusion

This section summarises and concludes the core contributions to the thesis, navigating through the series of interconnected Chapters 2, 3, 4, and 5, uniquely addressing our research questions.

Chapter 2 provides an extensive literature review covering all the research questions addressed in my PhD study. The complete RQ 1 and partial RQ 2 are addressed in Chapter 3, utilising a multilevel cloud structure with a proposed novel traffic flow framework and SDN. On the other hand, the complete RQ 2 is addressed in Chapter 4, proposing novel network slicing mechanisms in TI communication infrastructure. Finally, RQ 3 is addressed in Chapter 5, proposing a novel downlink PD-NOMA communication scenario and developing an analytical system model.

Starting with Chapter 2, a comprehensive survey of TI is presented, covering design aspects, key application areas, potential enabling technologies, related issues and challenges incurred in realising TI. The TI design architecture encompasses master, network and slave sections. Depending on the TI application, any change recorded through sensors in a master section is actuated in the slave section through the network section. A proposed application-centric Da Vinci telesurgical system is represented as an example of TI design architecture for better comprehension. In addition, key applications such as self-driven vehicles, industrial automation (Industry 4.0), e-health (Healthcare 4.0), AR/VR reality and serious gaming are discussed and represented with proposed scenario-based illustrative diagrams.

Moreover, these key applications are tabulated with required TI competencies and optimal performance metrics (such as RTT, data rate, reliability and availability). Envisioning these applications and their TI requirements, the related issues and challenges such as 1 ms RTT, transparency with the user, availability and reliability of information, reliability of connection, network slicing, control co-design, wireless resource customisation and easy incorporation, are also extensively encountered, tabulated and addressed. These issues and challenges are described as accompanied by potential enablers of TI. During this survey, we identified that EC/FC, SDN, NS, NFV, MA, and network coding techniques are considered potential TI enablers, along with the 5G framework. Hence, this survey also presents a comprehensive review of these enabling technologies. Finally, this survey provides in-depth insights into the TI and enables researchers to interpret the stringent demands to realise TI truly, thus developing next-generation Internet.

Chapter 3 proposes a novel fog-based traffic flow framework incorporating SDN and FC to achieve an RTT of 1 ms and ultra-high reliability for establishing haptic communication in TI. RTT comprises transmission, queuing, processing and acknowledgement times. Considering a design choice of FC with a multilevel cloud structure and SDN with a 5G framework, an efficient traffic flow algorithm is also proposed to reduce extra processing and waiting times at each level of the cloud-based structure to manage complex and critical incoming traffic in the network. The computational complexity is calculated for the proposed traffic flow algorithm, thus concluding to be easy to implement in practical scenarios.

Moreover, the combined SDN and FC approach facilitates an effective solution to route the traffic in the system, thus reducing the traffic flow paths between master and slave sections and minimising RTT. Therefore, the proposed framework is implemented in iFogSim, a Java-based high-performance open-source toolbox for modelling fog computing environments. The simulation results of the proposed framework are evaluated with performance metrics such as throughput, RTT, bandwidth utilisation, energy consumption, cloud capacity and reliability. These performance metrics are compared against edge, cloud, and cellular networks to validate the effectiveness of the proposed framework.

Subsequently, the cost consideration metrics, such as communication, processing, cloud-network, and power costs, are also discussed and aligned with the mentioned performance metrics. Furthermore, these performance metrics are analysed by varying the number of user equipment and flow requests to evaluate the system performance. It is observed that the proposed framework yields 14% higher throughput, 43% lower RTT, 47% lower packet jitter and 51% less power consumption when compared to cellular networks. In addition, the proposed framework consumes 71% less bandwidth and is 21% more reliable than existing cellular networks. Hence, these results deliver a clear vision for realising TI haptic communications.

In Chapter 4, a novel NS mechanism based on TI communication infrastructure is proposed to address the issues of provisioning and controlling the network slices on demand. Hence, the NS algorithms are developed and implemented for three different scenarios: topology slicing, service slicing, and emergency slicing scenarios, leveraging SDN with programmable OVS to tailor slices according to network requirements. To illustrate how NS works, the designed and dedicated E2E network slices tailored to meet the stringent requirements of the services or applications are represented.

Moreover, Vagrant, an open-source tool used for building and managing a virtualised development environment, is used as a simulation environment to evaluate and validate the NS-based TI communication infrastructure. A Linux-based OS (a base virtual machine image) and Oracle VirtualBox (a virtualised platform) are specified in a VagrantFile. Besides, Mininet, a network emulator, is used to create a network topology.

In addition, key performance indicators (KPIs), such as throughput and RTT, are observed to critically analyse the performance of configured network slices in the three previously mentioned scenarios. The simulation is carried out using diverse traffic, such as UDP, TCP and other traffic packets, to imitate the real-work traffic environment. The topology and service slicing scenarios have three slices (1, 2 and 3) each, having traffic packets routed and pre-designed RTT constraint values. On similar lines, before and during emergency slicing scenarios, there are two and three slices, respectively, with pre-designed RTT constraint values. "*iperf3*" is used to measure throughput for the mentioned traffic slices for each scenario, whereas RTT is measured using "*ping*", including SDN controller time and pre-designed constraint values.

Based on our simulation results, the measurement and comparison of throughput and round-trip time (RTT) for three designed slices align with the pre-designed allocated spectrum and RTT constraint values. It is observed that throughput in service slicing scenarios is higher than in topology slicing scenarios for the mentioned traffic because of better utilisation of available/allocated spectrum to slices and reduced congestion, leading to slightly higher throughput. Conversely, RTT for service slicing scenarios has also been higher than that for topology slicing scenarios. This is because of designed RTT constraint values (in ms) for topology slicing and service slicing scenarios. Finally, our developed algorithms are validated by comparing them with start-of-the art algorithms.

Lastly, in Chapter 5, we propose a novel downlink PD SISO-NOMA communication scenario for TI employing multiple sensors and actuators, collectively treated as users, and a base station. An analytical system model is mathematically derived for SINR, achievable sum rate, and fair PA coefficients for SISO-NOMA users to study system performance.

We consider two- and three-user scenarios to analyse and evaluate the simulation results by varying path loss exponent and fixed PA coefficients concerning BER and sum rate for SISO-NOMA and OMA. With fixed and fair PA coefficients, the outage probability is also measured for the SISO-NOMA system with a two-user scenario. To end this, a comparison of the achievable sum rate between fixed and fair PAs is made.

In addition, the simulation is conducted on the MATLAB 2024a platform, along with a 5G communication toolbox and required packages. This platform models the transmitter (a base station), receiver (users) and channel conditions of the proposed NOMA system scenario. Moreover, Monte Carlo simulation is applied to generate the Rayleigh fading channels using random complex Gaussian variables, add AWGN to received signals at the receiver end, and run multiple simulation runs to achieve statistically meaningful numerical results.

Furthermore, considering simulation results, analytical performance comparisons are conducted for BER and sum rate. It is observed that a higher BER is achieved for SISO-NOMA users compared to the OMA scheme due to additional signal processing to decode and perform successive interference cancellation (SIC) for received signals in a two-user and three-user scenario. However, for higher SNR, the achievable sum rate in the proposed SISO-NOMA system outperforms that obtained by OMA. The performance analysis of outage probability is carried out with varying fixed PA coefficients using a two-user scenario. The target rates for both users are considered and compared against derived achievable sum rate equations. The outage probability is observed to depend on the power percentage allocated to the particular user. Having two pairs of users with varying fixed PA coefficients, the outage probability is seen as lesser for the first user pair having the first user with higher PA than the second user pair having the first user with lower PA. The performance comparison and analysis of fixed PA with derived fair PA are conducted for outage probability and achievable sum rate using a two-user scenario. Moreover, using a 3-user scenario with a fixed PA coefficient, it is observed that the latency is less in SISO-NOMA than in OMA. However, when the latency is compared for fixed and fair PA coefficients for SISO-NOMA in a two-user scenario, SISO-NOMA with fair PA coefficients has outperformed SISO-NOMA with fixed PA coefficients.

To further analyse and compare the performance of 4×4 MIMO-NOMA with SISO-NOMA, 10 legitimate users are chosen, with two users in each cluster with a near-near and far-far clustering approach, equally spaced between 1000 m and 200 m. This scenario has a BS with a uniform linear array (ULA) of 4 antennas with half-wavelength spacing, assuming the users in each cluster have almost the same angles with clusters equally spaced between -60° to 60°. Concerning the achievable sum rate and latency, the 4×4 MIMO-NOMA incorporating zero forcing-based beamforming and a round-robin scheduling process has outperformed SISO-NOMA. It is concluded that NOMA has maximised the spectral efficiency and minimised the power consumption and latency by efficiently utilising the available spectrum.

6.2 Future research directions

This research significantly addresses the questions RQ 1, RQ 2 and RQ 3. This section identifies and outlines the potential future research directions that will extend the work done in Chapters 2, 3, 4 and 5, thus contributing to further tactile Internet (TI) literature survey and efficient approaches to achieve better network KPIs.

- **TI with sixth generation (6G).**

  Upon considering Chapter 2, incorporating EC/FC, SDN, NS, NFV and MA techniques, and 5G in TI communication infrastructure can open possible horizons and directions for real-time and immersive experiences. However, an in-depth survey is also needed to explore possible rays of application and service opportunities for TI with 6G. 6G promises to be a breakthrough beyond 5G, contributing to the following key characteristics: (i) It will further reduce the microsecond-level RTT, promoting near-instantaneous response times, and it is perfectly suitable for critical applications such as telerobotic surgery; (ii) It will utilise higher frequency bands (THz waves), thus supporting higher transmission rates and capacity for TI applications; (iii) It will incorporate AI and edge computing, making more intelligent and responsive networks to improve tactile experiences through advanced data prediction, processing, and analysis; and (iv) it will increase the security and reliability of TI communications, suitable for sensitive applications. With these improvements, 6G will unravel promising future directions for TI, such as enabling fully immersive tactile remote experiences, real-time remote holographic communication and extended reality (XR) – AR, VR and mixed, that creates virtual replicas and mirrors the physical world with a digital twins world for more immersive and interactive TI experiences.

- **Integrating machine learning/artificial intelligence (ML/AL) model on fog nodes.**

  Research done in Chapter 3 can be extended by integrating the machine learning/artificial intelligence (ML/AL) model, which can be deployed on fog nodes to perform crucial tasks such as analysing and filtering real-time data and making automated informed decisions. Therefore, the traffic (data) flowing to these fog nodes (micro and mini cloud units) can be analysed much quicker than transmitting the data to the central cloud. Due to the limited resources of fog nodes, these models must be designed and operated to be lightweight. In our case, the resources allocated to fog nodes are static. Any significant changes in the network will not be easy to handle or process the massive incoming traffic on those fog nodes, thus increasing the processing and wait times. Hence, ML models can be implemented to dynamically allocate resources such as bandwidth, storage and computational power based on fluctuating demands of incoming tasks, thus efficiently using available resources, avoiding overloading a single node and reducing RTT further.

- **Dynamic network slicing (NS) with an appropriate ML/AI model.**

  Considering Chapter 4, NS in TI communication infrastructure can further improve efficiency and capabilities by incorporating an appropriate ML/AI model. This model can analyse the present network's traffic patterns and application/service requirements, thus predicting future network demands. Using real-time data, such analysis can be used to dynamically create, customise and destroy network slices based on fluctuating demands. For instance, a dynamic slice will be created to serve a particular application needing higher throughput and lower latency, thus improving QoS and user experience. Such dynamic slices can rely on the policies set by this model, ensuring compliance with service level agreements (SLA). The resources such as compute, storage and bandwidth required for created slices can be predicted for dynamic resource allocation and decommission/destroy the slices when utilised or not in use, promoting intelligent slice management. In addition, real-time monitoring to gather slice performance metrics can aid in detecting anomalies and faults in the network with the help of supervised and unsupervised learning ML/AL models, identifying known and unknown issues. A pre-defined response can be automatically triggered to mitigate faults, such as re-configuring the slice parameters. These approaches will optimise the network's efficacy and resource allocation, thus further enhancing throughput, RTT and other performance metrics.

- **To accurately estimate dynamic and imperfect channel state information (CSI) conditions using a suitable ML/AI model.**

  In Chapter 5, a suitable ML/AL model can be adopted for a downlink PD-NOMA-based TI communication infrastructure to accurately estimate dynamic and imperfect channel state information (CSI) conditions, leading to reliable communication, dynamic/fair adjustment of power allocation/level and efficient SIC process resulting less error in signal decoding. This model can also utilise an optimal power allocation strategy to promote user fairness and achieve the desired system sum rate. Hence, the power level can be automatically adjusted based on channel conditions on the transmitter side, thus optimising resource usage, maintaining acceptable QoS, maximising spectral capacity, and minimising RTT. Moreover, managing interference between users becomes crucial due to the non-orthogonal nature of the user's signals and using the same RB. Hence, predicting effective interference among NOMA users can optimise the SIC process. To efficiently decode the superimposed signals on the receiver side, the ML/AL model can streamline the SIC process, reducing computational complexity and improving decoded signal error correction competencies. Thus, it reduces the time required to decode the user's signals, leading to more reliable communication and can further improve the sum rate and minimise RTT.

6.3 System implication and deployment

The proposed algorithms and analytical system model to study the RTT performance in TI communication infrastructure are investigated in Chapters 3, 4 and 5. The system planners play a significant role in understanding the system implication and deployment phase, which involves considering the potential impacts or consequences of the proposed solutions. The research findings gathered in these chapters will be analysed and benefitted by system planners in the following ways.

Chapter 3 proposes a fog-based traffic flow framework to meet stringent requirements such as ultra-low latency and ultra-high reliability, which have outstanding implications for the network's infrastructure and operation. Meeting such requirements requires accurate transmission, propagation, queuing and processing delay optimisation, and resilient error handling and failover procedures. Moreover, using this framework with FC and SDN, the system entails the distributed or decentralised approach in strategically deploying the fog nodes near the UE or end devices. Thus, the deployed fog nodes must be competent in handling/processing real-time network traffic and making informed decisions quickly to minimise RTT and enhance reliability through the corresponding optimised traffic flow algorithm. In addition, an SDN controller is essential to manage and route the network traffic dynamically, hence requiring optimal controller configuration. Subsequently, looking at the results standpoint, the system planners benefit from enhanced network performance, informed decision-making, optimised resource utilisation and consumption, mitigate risk by recognising potential bottlenecks through simulation and a comprehensive cost-benefit analysis, henceforth giving system planners a vision of creating a dynamic network infrastructure to support rising requirements in TI applications.

Furthermore, the proposed NS algorithms in Chapter 4 imply real-time slice provisioning where customisable, dedicated and logical E2E networks can be configured. Such configurations necessitate tailored network slices to meet diverse customer demands in TI environments and require dynamic mechanisms to manage multiple virtual networks or logical slices on a shared physical/hardware resource. SDN and OVS are mainly responsible for efficient and flexible network management and dynamic slice provisioning. So, installing a suitable SDN controller(s) and deploying a distributed OVS setup can pose a deciding factor for efficient network performance. The system planners will be able to utilise pre-designed slice configurations/templates such as topology, service and emergency slicings, which are prepared for diverse traffic or application loads, thus ensuring performance stability and optimising resource allocation across the network. Using a potential network emulator, the research findings can facilitate some insights into the system's behaviour and validate the system's KPIs before conducting full-scale testing and deployment, thus proving the effectiveness of slicing algorithms. Therefore, the proposed NS algorithms can guarantee enhanced QoS levels as each tailored slice meets the application/service requirements. Finally, the system planners can accelerate the service deployment because of the rapid slice provisioning on demand.

Lastly, in Chapter 5, a downlink PD-NOMA for TI communication is proposed, supported by multiple users and a BS. Using the same RB to accommodate multiple users in the NOMA scheme can be challenging for system planners. However, the NOMA implementation improves spectral efficiency and minimises latency, making it suitable for TI applications. Moreover, efficient (dynamic and fair) PA strategies are required to support multiple users, thus optimising system performance, reducing power utilisation, and enhancing user fairness. Developing and deploying suitable interference management strategies, such as SIC, can reduce BER, increase signal quality and ensure reliable communication. However, system planners need to ensure that existing BSs support NOMA capabilities. If not, upgrades may be required. The upgraded BSs need to be deployed with NOMA-empowered transmission and reception capacity. The end-users must be equipped with NOMA-supported modems and advanced signals processing competencies, which may hamper their cost and battery life. Constant system monitoring is required to fine-tune the specifications to cater to changing network needs. The research findings validate the mathematical model and fair PA strategy in a simulation environment before system planners consider full-scale deployment in the real world. Consequently, system planners benefit substantially from flexible network deployment, knowledge to improve KPIs and room for new service opportunities due to increased spectral capacity and minimised latency.

References

[1] G. Fettweis, H. Boche, T. Wiegand, E. Zielinski, H. Schotten, P. Merz, S. Hirche, A. Festag, W. Häffner, M. Meyer *et al.*, "The tactile Internet-ITU-T technology watch report," *Int. Telecom. Union (ITU), Geneva*, 2014, (Accessed: Apr. 15, 2025). [Online]. Available: https://www.itu.int/dms_pub/itu-t/opb/gen/T-GEN-TWATCH-2014-1-PDF-E.pdf

[2] Preetipadma, "Unlocking the power of tactile Internet via 5G network," Available: https://www.analyticsinsight.net/internet-of-things/unlocking-the-power-of-tactile-internet-via-5g-network#:~:text=In%20its%202018%20paper%2C%20titled,and%20virtual%20world%20will%20blur., Dec. 2020, (Accessed: Apr. 15, 2025).

[3] F. Kuipers, "The Internet touch," *Magazine*, vol. 21, Nov. 2016.

[4] B. Chatras, U. S. T. Kwong, and N. Bihannic, "NFV enabling network slicing for 5G," in 20^{th} *Conference on Innovations in Clouds, Internet and Networks (ICIN)*. IEEE, 2017, pp. 219–225.

[5] I. Budhiraja, S. Tyagi, S. Tanwar, N. Kumar, and J. J. Rodrigues, "Tactile Internet for smart communities in 5G: An insight for NOMA-based solutions," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3104–3112, 2019.

[6] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3098–3130, 2018.

[7] R. Kirichek, A. Vladyko, A. Paramonov, and A. Koucheryavy, "Software-defined architecture for flying ubiquitous sensor networking," in *19th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2017, Conference Proceedings, pp. 158–162.

[8] M. Hammoudeh and R. Newman, "Information extraction from sensor networks using the Watershed transform algorithm," *Information Fusion*, vol. 22, pp. 39–49, 2015.

[9] A. Abuarqoub, M. Hammoudeh, B. Adebisi, S. Jabbar, A. Bounceur, and H. Al-Bashar, "Dynamic clustering and management of mobile wireless sensor networks," *Computer Networks*, vol. 117, pp. 62–75, 2017.

[10] A. Muthanna, P. Masek, J. Hosek, R. Fujdiak, O. Hussein, A. Paramonov, and A. Koucheryavy, "Analytical evaluation of D2D connectivity potential in 5G wireless systems," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2016, pp. 395–403.

[11] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[12] F. Bonomi, “Connected vehicles, the Internet of Things and fog computing,” in *The eighth ACM international workshop on vehicular inter-networking (VANET), Las Vegas, USA*. sn, 2011, pp. 13–15.

[13] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the Internet of Things,” in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. ACM, 2012, Conference Proceedings, pp. 13–16.

[14] N. Alliance, “5G white paper,” *Next Generation Mobile Networks, White Paper*, vol. 1, 2015.

[15] A. Ngmn, “Description of network slicing concept,” *White Paper*, Jan. 2016.

[16] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, “A survey of non-orthogonal multiple access for 5G,” *IEEE communications surveys & tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[17] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, “On the performance of non-orthogonal multiple access in short-packet communications,” *IEEE Communications Letters*, vol. 22, no. 3, pp. 590–593, 2017.

[18] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[19] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “A survey of the tactile Internet: Design issues and challenges, applications, and future directions,” *Electronics*, vol. 10, no. 17, p. 2171, Sep. 2021.

[20] M. Maier, “Un âge d’or économique dont le numerique est la trame de fond,” *Webzine PlanèteINRS. ca*, 2013, (Accessed: Apr. 15, 2025). [Online]. Available: https://ieeexplore.ieee.org/document/7470948/references#references

[21] G. Fettweis and S. Alamouti, “5G: Personal mobile internet beyond what cellular did to telephony,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 140–145, 2014.

[22] G. P. Fettweis, “The tactile Internet: Applications and challenges,” *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[23] E. Brynjolfsson and A. McAfee, *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

[24] K. S. Kim, D. K. Kim, C.-B. Chae, S. Choi, Y.-C. Ko, J. Kim, Y.-G. Lim, M. Yang, S. Kim, and B. Lim, “Ultrareliable and low-latency communication techniques for tactile Internet services,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 376–393, 2018.

[25] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler, and Y. Koucheryavy, “Understanding the IoT connectivity landscape: A contemporary M2M radio technology roadmap,” *IEEE Communications Magazine*, vol. 53, no. 9, pp. 32–40, 2015.

[26] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, "The tactile Internet: Vision, recent progress, and open challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 138–145, 2016.

[27] L. Zou, C. Ge, Z. J. Wang, E. Cretu, and X. Li, "Novel tactile sensor technology and smart tactile sensing systems: A review," *Sensors*, vol. 17, no. 11, p. 2653, 2017.

[28] O. Holland, E. Steinbach, R. V. Prasad, Q. Liu, Z. Dawy, A. Aijaz, N. Pappas, K. Chandra, V. S. Rao, and S. Oteafy, "The IEEE 1918.1 "tactile Internet" standards working group and its standards," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 256–279, 2019.

[29] A. Aijaz, Z. Dawy, N. Pappas, M. Simsek, S. Oteafy, and O. Holland, "Toward a tactile Internet reference architecture: Vision and progress of the IEEE P1918.1 standard," *arXiv preprint* , 2018.

[30] T. H. Szymanski, "Securing the industrial-tactile Internet of Things with deterministic silicon photonics switches," *IEEE Access*, vol. 4, pp. 8236–8249, 2016.

[31] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, p. 429, 2002.

[32] "3D systems - Phantom premium haptic devices," Available: https://www.3dsystems.com/haptics-devices/3d-systems-phantom-premium, (Accessed: Apr. 15, 2025).

[33] "Thinglab - Geomagic haptic devices," Available: https://freedspace.com.au/thinglab/3d-design/design-tools/geomagic-phantom-premium-haptic-devices/, (Accessed: Apr. 15, 2025).

[34] "The da vinci surgical systems," Available: https://www.intuitive.com/en-us/patients/da-vinci-robotic-surgery/about-the-systems, (Accessed: Apr. 15, 2025).

[35] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualisation: Challenges and opportunities for innovations," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 90–97, 2015.

[36] C. Ainhauser, L. Bulwahn, A. Hildisch, S. Holder, O. Lazarevych, D. Mohr, T. Ochs, M. Rudorfer, O. Scheickl, and T. Schumm, "Autonomous driving needs ROS," *BMW Car IT GmbH, ROSCon, Stuttgart, Germany*, 2013.

[37] B. Singh and A. Gupta, "Recent trends in intelligent transportation systems: A review," *Journal of Transport Literature*, vol. 9, no. 2, pp. 30–34, 2015.

[38] J. D. Lee, D. V. McGehee, T. L. Brown, and M. L. Reyes, "Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator," *Human factors*, vol. 44, no. 2, pp. 314–334, 2002.

[39] S. B. Mer, "Smart vehicle-to-vehicle communication with 5G technology," *Int. J. Recent Innov. Trends Comput. Commun*, vol. 3, no. 5, pp. 3241–3244, 2015.

[40] L. Bartevyan, "Industry 4.0 - Summary report," *DLG-Expert report*, vol. 5, no. 2015, pp. 1–8, 2015.

[41] R. C. Schlaepfer, M. Koch, and P. Merkofer, “Industry 4.0 challenges and solutions for the digital transformation and use of exponential technologies,” *Deloitte, Zurique*, 2015.

[42] M. Anvari, T. Broderick, H. Stein, T. Chapman, M. Ghodoussi, D. W. Birch, C. Mckinley, P. Trudeau, S. Dutta, and C. H. Goldsmith, “The impact of latency on surgical precision and task completion during robotic-assisted remote telepresence surgery,” *Computer Aided Surgery*, vol. 10, no. 2, pp. 93–99, 2005.

[43] S. Haddadin, L. Johannsmeier, and F. D. Ledezma, “Tactile robots as a central embodiment of the tactile Internet,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 471–487, 2018.

[44] “MicroSurge – Institute of Robotics and Mechatronics,” Available: https://www.dlr.de/en/rm/research/robotic-systems/multi-arm/mirosurge, (Accessed: Apr. 15, 2025).

[45] D. Nincarean, M. B. Alia, N. D. A. Halim, and M. H. A. Rahman, “Mobile augmented reality: The potential for education,” *Procedia-social and behavioral sciences*, vol. 103, no. 0, pp. 657–664, 2013.

[46] J. M. Harley, E. G. Poitras, A. Jarrell, M. C. Duffy, and S. P. Lajoie, “Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes,” *Educational Technology Research and Development*, vol. 64, no. 3, pp. 359–388, 2016.

[47] P. J. Bank, M. A. Cidota, P. E. W. Ouwehand, and S. G. Lukosch, “Patient-tailored augmented reality games for assessing upper extremity motor impairments in parkinson’s disease and stroke,” *Journal of medical systems*, vol. 42, no. 12, p. 246, 2018.

[48] I. J. Cifuentes, B. L. Dagnino, M. C. Salisbury, M. E. Perez, C. Ortega, and D. Maldonado, “Augmented reality and dynamic infrared thermography for perforator mapping in the anterolateral thigh,” *Archives of Plastic Surgery*, vol. 45, no. 3, p. 284, 2018.

[49] Y. Cheng, “Edge caching and computing in 5G for mobile augmented reality and haptic Internet,” *Computer Communications*, 2020.

[50] F. Kuipers, R. Kooij, D. De Vleeschauwer, and K. Brunnström, “Techniques for measuring quality of experience,” in *International Conference on Wired/Wireless Internet Communications*. Springer, 2010, pp. 216–227.

[51] Y. Ruan, N. Anerousis, M. Srivatsa, J. Xiao, R. T. Christner, L. Farrolas, and J. Short, “Measuring enterprise network usage pattern & deploying passive optical LANs,” in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2015, pp. 890–893.

[52] C. She, C. Yang, and T. Q. Quek, “Cross-layer transmission design for tactile Internet,” in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.

[53] Z. Xiang, F. Gabriel, E. Urbano, G. T. Nguyen, M. Reisslein, and F. H. Fitzek, “Reducing latency in virtual machines: Enabling tactile Internet for human-machine co-working,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 5, pp. 1098–1116, 2019.

[54] R. Cziva and D. P. Pezaros, "On the latency benefits of edge NFV," in *ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*. IEEE, 2017, pp. 105–106.

[55] J. Xie, D. Guo, X. Li, Y. Shen, and X. Jiang, "Cutting long-tail latency of routing response in software-defined networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 384–396, 2018.

[56] R. Khalili, Z. Despotovic, and A. Hecker, "Flow setup latency in SDN networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 12, pp. 2631–2639, 2018.

[57] A. Asaduzzaman, K. K. Chidella, and D. Vardha, "An energy-efficient directory-based multicore architecture with wireless routers to minimize the communication latency," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 2, pp. 374–385, 2016.

[58] J. Li, X. Shen, L. Chen, D. P. Van, J. Ou, L. Wosinska, and J. Chen, "Service migration in fog computing-enabled cellular networks to support real-time vehicular communications," *IEEE Access*, vol. 7, pp. 13 704–13 714, 2019.

[59] J. Du, L. Zhao, X. Chu, F. R. Yu, J. Feng, and I. Chih-Lin, "Enabling low-latency applications in LTE-A-based mixed fog/cloud computing systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1757–1771, 2018.

[60] A. Huang, N. Nikaein, T. Stenbock, A. Ksentini, and C. Bonnet, "Low latency MEC framework for SDN-based LTE/LTE-A networks," in *IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[61] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari, "Haptic data compression and communication," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 87–96, 2010.

[62] D. Öhmann, M. Simsek, and G. P. Fettweis, "Achieving high availability in wireless networks by an optimal number of Rayleigh-fading links," in *IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2014, pp. 1402–1407.

[63] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, "Toward haptic communications over the 5G tactile Internet," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3034–3059, 2018.

[64] D. Van Den Berg, R. Glans, D. De Koning, F. A. Kuipers, J. Lugtenburg, K. Polachan, P. T. Venkata, C. Singh, B. Turkovic, and B. Van Wijk, "Challenges in haptic communications over the tactile Internet," *IEEE Access*, vol. 5, pp. 23 502–23 518, 2017.

[65] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[66] J. Park and M. Bennis, "URLLC-eMBB slicing to support vr multimodal perceptions over wireless cellular systems," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.

[67] C. Grasso and G. Schembra, "Design of a UAV-based videosurveillance system with tactile Internet constraints in a 5G ecosystem," in *4th IEEE Conference on Network Softwarisation and Workshops (NetSoft)*. IEEE, 2018, Conference Proceedings, pp. 449–455.

[68] X. Xu, Q. Liu, and E. Steinbach, "Toward QoE-driven dynamic control scheme switching for time-delayed teleoperation systems: A dedicated case study," in *IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, 2017, pp. 1–6.

[69] A. Aijaz, "Hap-SliceR: A radio resource slicing framework for 5G networks with haptic communications," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2285–2296, 2017.

[70] Y. Su, X. Lu, L. Huang, X. Du, and M. Guizani, "Tac-U: A traffic balancing scheme over licensed and unlicensed bands for tactile Internet," *Future Generation Computer Systems*, vol. 97, pp. 41–49, 2019.

[71] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, and Z. Zhang, "Enabling ultra-reliable and low-latency communications through unlicensed spectrum," *IEEE Network*, vol. 32, no. 2, pp. 70–77, 2018.

[72] I. Elhajj, N. Xi, W. K. Fung, Y. H. Liu, W. J. Li, T. Kaga, and T. Fukuda, "Haptic information in Internet-based teleoperation," *IEEE/ASME Transactions on mechatronics*, vol. 6, no. 3, pp. 295–304, 2001.

[73] E. Steinbach, S. Hirche, M. Ernst, F. Brandi, R. Chaudhari, J. Kammerl, and I. Vittorias, "Haptic communications," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 937–956, 2012.

[74] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, "Perceptual and bitrate-scalable coding of haptic surface texture signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 462–473, 2014.

[75] B. Gülecyüz, X. Xu, A. Noll, and E. Steinbach, "On the quality-of-learning for haptic teleoperation-based skill transfer over the tactile Internet," in *GLOBECOM IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[76] C. She and C. Yang, "Ensuring the quality-of-service of tactile Internet," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*. IEEE, 2016, Conference Proceedings, pp. 1–5.

[77] J. J. LaViola Jr, "A discussion of cybersickness in virtual environments," *ACM Sigchi Bulletin*, vol. 32, no. 1, pp. 47–56, 2000.

[78] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[79] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, and H. Taoka, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[80] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile Internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 82–89, 2016.

[81] S. Payandeh and J.-y. Lee, "User-centered force signal processing for internet-based telemanipulation: An overview," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014, Conference Proceedings, pp. 3978–3983.

[82] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[83] A. Aijaz, "Radio resource slicing in a radio access network," Apr. 16 2019, uS Patent 10,264,461.

[84] L. Ma, X. Wen, L. Wang, Z. Lu, and R. Knopp, "An SDN/NFV-based framework for management and deployment of service-based 5G core network," *China Communications*, vol. 15, no. 10, pp. 86–98, 2018.

[85] H. D. Chantre and N. L. da Fonseca, "Multi-objective optimization for edge device placement and reliable broadcasting in 5G NFV-based small cell networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 10, pp. 2304–2317, 2018.

[86] X. Ge, R. Zhou, and Q. Li, "5G NFV-based tactile Internet for mission-critical IoT services," *IEEE Internet of Things Journal*, 2019.

[87] "White paper 3: End to end network slicing," https://wwrf.ch/wp-content/publications/outlook/Outlook21.pdf, 2017, (Accessed: Aug. 30, 2024).

[88] E. Mohyeldin, "Minimum technical performance requirements for IMT-2020 radio interface(s)," https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Documents/S01-1_Requirements%20for%20IMT-2020_Rev.pdf, 2020, (Accessed: Aug. 30, 2024).

[89] L. Chan, F. Naghdy, and D. Stirling, "Application of adaptive controllers in teleoperation systems: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 3, pp. 337–352, 2014.

[90] P. Seeling and F. H. Fitzek, "Anticipatory networking: Negative latency for ubiquitous computing," in *IEEE* 18^{th} *Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2021, pp. 1–4.

[91] A. Aijaz, "Toward human-in-the-loop mobile networks: A radio resource allocation perspective on haptic communications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4493–4508, 2018.

[92] ——, “Towards 5G-enabled tactile Internet: Radio resource allocation for haptic communications,” in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2016, Conference Proceedings, pp. 145–150.

[93] C. She, C. Yang, and T. Q. Quek, “Cross-layer optimization for ultra-reliable and low-latency radio access networks,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 127–141, 2017.

[94] E. Ganesan, I.-S. Hwang, and A. T. Liem, “Resource allocation for tactile Internet via software-defined FiWi access network,” in *International Computer Symposium (ICS)*. IEEE, 2020, pp. 283–287.

[95] G. Cena, A. Valenzano, and S. Vitturi, “Hybrid wired/wireless networks for real-time communications,” *IEEE industrial electronics magazine*, vol. 2, no. 1, pp. 8–20, 2008.

[96] W. Monnet and T. A. Yahiya, “HoIP performance for tactile Internet over 5G networks: A teleoperation case study,” in 11^{th} *International Conference on Network of the Future (NoF)*. IEEE, 2020, pp. 48–54.

[97] M. Cheffena, “Industrial wireless communications over the millimeter wave spectrum: Opportunities and challenges,” *IEEE Communications Magazine*, vol. 54, no. 9, pp. 66–72, 2016.

[98] H. Shokri-Ghadikolaei, C. Fischione, P. Popovski, and M. Zorzi, “Design aspects of short-range millimeter-wave networks: A MAC layer perspective,” *IEEE Network*, vol. 30, no. 3, pp. 88–96, 2016.

[99] G. Bianco, S. Amendola, and G. Marrocco, “Near-field modeling of self-tuning antennas for the tactile Internet,” in *XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science*. IEEE, 2020, pp. 1–3.

[100] G. M. Bianco, C. Vivarelli, S. Amendola, and G. Marrocco, “Experimentation and calibration of near-field UHF epidermal communication for emerging tactile Internet,” in 5^{th} *International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 2020, pp. 1–4.

[101] V. Hassija, V. Saxena, and V. Chamola, “A blockchain-based framework for drone-mounted base stations in tactile Internet environment,” in *IEEE INFOCOM Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2020, pp. 261–266.

[102] S. C. Mannsfeld, B. C. Tee, R. M. Stoltenberg, C. V. H. Chen, S. Barman, B. V. Muir, A. N. Sokolov, C. Reese, and Z. Bao, “Highly sensitive flexible pressure sensors with microstructured rubber dielectric layers,” *Nature materials*, vol. 9, no. 10, p. 859, 2010.

[103] F. L. Hammond, Y. Mengüç, and R. J. Wood, “Toward a modular soft sensor-embedded glove for human hand motion and tactile pressure measurement,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, Conference Proceedings, pp. 4000–4007.

[104] G. H. Büscher, R. Kõiva, C. Schürmann, R. Haschke, and H. J. Ritter, "Flexible and stretchable fabric-based tactile sensor," *Robotics and Autonomous Systems*, vol. 63, pp. 244–252, 2015.

[105] G. Liang, D. Mei, Y. Wang, and Z. Chen, "Modeling and analysis of a flexible capacitive tactile sensor array for normal force measurement," *IEEE Sensors Journal*, vol. 14, no. 11, pp. 4095–4103, 2014.

[106] M. Mizushima, S. Takagi, H. Itano, T. Obata, T. Kasahara, S. Shoji, and J. Mizuno, "Flexible and capacitive tactile sensor sheet," in *International Conference on Electronics Packaging (ICEP)*. IEEE, 2014, Conference Proceedings, pp. 756–759.

[107] G. Liang, Y. Wang, D. Mei, K. Xi, and Z. Chen, "Flexible capacitive tactile sensor array with truncated pyramids as dielectric layer for three-axis force measurement," *Journal of Microelectromechanical Systems*, vol. 24, no. 5, pp. 1510–1519, 2015.

[108] Y. Wang, G. Liang, D. Mei, L. Zhu, and Z. Chen, "A flexible capacitive tactile sensor array with high scanning speed for distributed contact force measurements," in *IEEE 29th International Conference on Micro Electro Mechanical Systems (MEMS)*. IEEE, 2016, Conference Proceedings, pp. 854–857.

[109] D. Kwon, T.-I. Lee, M. Kim, S. Kim, T.-S. Kim, and I. Park, "Porous dielectric elastomer based ultra-sensitive capacitive pressure sensor and its application to wearable sensing device," in *Transducers-18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*. IEEE, 2015, Conference Proceedings, pp. 299–302.

[110] C. Pacchierotti, A. Tirmizi, and D. Prattichizzo, "Improving transparency in teleoperation by means of cutaneous tactile force feedback," *ACM Transactions on Applied Perception (TAP)*, vol. 11, no. 1, p. 4, 2014.

[111] C. Pacchierotti, L. Meli, F. Chinello, M. Malvezzi, and D. Prattichizzo, "Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems," *The International Journal of Robotics Research*, vol. 34, no. 14, pp. 1773–1787, 2015.

[112] J. Martinez, D. Martinez, J. P. Molina, P. Gonz, and A. Garcia, "Comparison of force and vibrotactile feedback with direct stimulation for texture recognition," in *International Conference on Cyberworlds*. IEEE, 2011, Conference Proceedings, pp. 62–68.

[113] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda, "Non-contact tactile display based on radiation pressure of airborne ultrasound," *IEEE Transactions on Haptics*, vol. 3, no. 3, pp. 155–165, 2010.

[114] J. Martínez, A. García, M. Oliver, J. P. Molina, and P. González, "Identifying virtual 3D geometric shapes with a vibrotactile glove," *IEEE computer graphics and applications*, vol. 36, no. 1, pp. 42–51, 2014.

[115] P. Weber, E. Rueckert, R. Calandra, J. Peters, and P. Beckerle, "A low-cost sensor glove with vibrotactile feedback and multiple finger joint and hand motion sensing for human-robot interaction," in *25th IEEE International Symposium on Robot and*

Human Interactive Communication (RO-MAN). IEEE, 2016, Conference Proceedings, pp. 99–104.

[116] Y. Muramatsu and M. Niitsuma, "Correspondence relationships between vibrotactile stimuli and tactile sensations determined by semantic differential," in *IEEE RO-MAN*. IEEE, 2013, Conference Proceedings, pp. 668–673.

[117] P. Galambos, "Vibrotactile feedback for haptics and telemanipulation: Survey, concept and experiment," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 41–65, 2012.

[118] K. Minamizawa, D. Prattichizzo, and S. Tachi, "Simplified design of haptic display by extending one-point kinesthetic feedback to multipoint tactile feedback," in *IEEE Haptics Symposium*. IEEE, 2010, Conference Proceedings, pp. 257–260.

[119] L. Dipietro, A. M. Sabatini, and P. Dario, "A survey of glove-based systems and their applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 4, pp. 461–482, 2008.

[120] M. Zhou and P. Ben-Tzvi, "RML glove - An exoskeleton glove mechanism with haptics feedback," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 2, pp. 641–652, 2014.

[121] T. L. Baldi, S. Scheggi, L. Meli, M. Mohammadi, and D. Prattichizzo, "GESTO: A glove for enhanced sensing and touching-based on inertial and magnetic sensors for hand tracking and cutaneous feedback," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 1066–1076, 2017.

[122] Y. Ma, Z.-H. Mao, W. Jia, C. Li, J. Yang, and M. Sun, "Magnetic hand tracking for human computer interface," *IEEE Transactions on Magnetics*, vol. 47, no. 5, pp. 970–973, 2011.

[123] A. Nelson, G. Singh, R. Robucci, C. Patel, and N. Banerjee, "Adaptive and personalised gesture recognition using textile capacitive sensor arrays," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 1, no. 2, pp. 62–75, 2015.

[124] L. Silva, R. Dantas, P. Diniz, V. Jeronimo, L. Bueno, and T. Dutra, "PHYS. IO: Wearable hand tracking device," in *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, 2016, Conference Proceedings, pp. 1–6.

[125] G. Cohn, S. Gupta, T.-J. Lee, D. Morris, J. R. Smith, M. S. Reynolds, D. S. Tan, and S. N. Patel, "An ultra-low-power human body motion sensor using static electric field sensing," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 99–102.

[126] R. Bainbridge and J. A. Paradiso, "Wireless hand gesture capture through wearable passive tag sensing," in *International Conference on Body Sensor Networks*. IEEE, 2011, Conference Proceedings, pp. 200–204.

[127] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "The 5G-enabled tactile Internet: Applications, requirements, and architecture," in *IEEE Wireless Communications and Networking Conference*. IEEE, 2016, Conference Proceedings, pp. 1–6.

[128] C. Li, C.-P. Li, K. Hosseini, S. B. Lee, J. Jiang, W. Chen, G. Horn, T. Ji, J. E. Smee, and J. Li, "5G-based systems design for tactile Internet," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 307–324, 2018.

[129] A. A. Abd El-Latif, Y. M. Lo'ai Tawalbeh, and B. B. Gupta, "Secure edge and fog computing enabled AI for IoT and smart cities," in *International Conference on Advanced Computing & Next-Generation Communication (ICACNGC)*. Springer, 2023, pp. 978–3 031 510 960.

[130] B. S. Chaudhari, "Enabling tactile Internet via 6G: Application characteristics, requirements, and design considerations top of form," *Future Internet*, vol. 17, no. 3, p. 122, 2025.

[131] J. Schulz, C. Dubslaff, P. Seeling, S.-C. Li, S. Speidel, and F. H. Fitzek, "Negative latency in the tactile Internet as enabler for global metaverse immersion," *IEEE Network*, 2024.

[132] H. Patil, H. S. Negi, P. A. Devarani, A. Barve, R. Maranan *et al.*, "Enhancing tactile Internet experiences through control mechanisms and predictive AI," in *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*. IEEE, 2024, pp. 235–239.

[133] H. Sun, Y. Liu, A. Al-Tahmeesschi, A. Nag, M. Soleimanpour-Moghadam, B. Canberk, H. Arslan, and H. Ahmadi, "Advancing 6G: Survey for explainable AI on communications and network slicing," *IEEE Open Journal of the Communications Society*, 2025.

[134] M. Dubey, A. K. Singh, and R. Mishra, "AI based resource management for 5G network slicing: History, use cases, and research directions," *Concurrency and Computation: Practice and Experience*, vol. 37, no. 2, p. e8327, 2025.

[135] J. Hornik, C. Ofir, and M. Rachamim, "Out of the fog: Fog computing-enabled AI to support smart marketing management," *Management Review Quarterly*, pp. 1–33, 2024.

[136] A. Arbaoui, S. Abdelatif, and M. Derdour, "Network slicing solutions for Internet of Vehicles (IoV) networks: A review," in *6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. IEEE, 2024, pp. 1–7.

[137] F. O. Okello, V. Oduol, C. Maina, and A. Apiyo, "Improvement of 5G core network performance using network slicing and deep reinforcement learning," *International Journal of Electrical and Electronics Research*, vol. 12, no. 2, pp. 493–502, 2024.

[138] H. O. Otieno, B. Malila, and J. Mwangama, "Deployment and management of intelligent end-to-end network slicing in 5G and beyond 5G networks - A Systematic Review," *IEEE Access*, vol. 12, pp. 190 411–190 433, 2024.

[139] F. S. D. Silva, S. N. Silva, L. M. Da Silva, A. Bessa, S. Ferino, P. Paiva, M. Medeiros, L. Silva, J. Neto, K. Costa *et al.*, "ML-based inter-slice load balancing control for proactive offloading of virtual services," *Computer Networks*, vol. 246, p. 110422, 2024.

[140] W. Hamdi, C. Ksouri, H. Bulut, and M. Mosbah, "Network slicing based learning techniques for IoV in 5G and beyond networks," *IEEE Communications Surveys & Tutorials*, 2024.

[141] G. Goiz, C. M. de Farias, and F. C. Delicato, "Traffic classification for network slicing using machine learning techniques," in *IEEE Conference on Pervasive and Intelligent Computing (PICom)*. IEEE, 2024, pp. 32–39.

[142] A. T. Liem, I.-S. Hwang, R. Kharga, and C.-H. Teng, "Enhancing tactile Internet reliability: AI-driven resilience in NG-EPON networks," in *Photonics*, vol. 11, no. 10. MDPI, 2024, p. 903.

[143] O. Alnajar and A. Barnawi, "TactiFlex: A federated learning-enhanced in-content aware resource allocation flexible architecture for tactile IoT in 6G networks," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108934, 2024.

[144] R. B. Shaik, P. Nagaradjane, I. Ioannou, V. Sittakul, V. Vasiliou, and A. Pitsillides, "AI/ML-aided capacity maximization strategies for URLLC in 5G/6G wireless systems: A survey," *Computer Networks*, vol. 249, p. 110506, 2024.

[145] K. Sthankiya, N. Saeed, G. McSorley, M. Jaber, and R. G. Clegg, "A survey on AI-driven energy optimisation in terrestrial next generation radio access networks," *IEEE Access*, 2024.

[146] B. E. Y. Belmekki and M.-S. Alouini, "NOMA as the next-generation multiple access in nonterrestrial networks," *Proceedings of the IEEE*, 2024.

[147] A. Ahmed, W. Xingfu, A. Hawbani, W. Yuan, H. Tabassum, Y. Liu, M. U. F. Qaisar, Z. Ding, N. Al-Dhahir, A. Nallanathan *et al.*, "Unveiling the potential of NOMA: A journey to next generation multiple access," *IEEE Communications Surveys & Tutorials*, 2024.

[148] B. Clerckx, Y. Mao, Z. Yang, M. Chen, A. Alkhateeb, L. Liu, M. Qiu, J. Yuan, V. W. Wong, and J. Montojo, "Multiple access techniques for intelligent and multifunctional 6G: Tutorial, survey, and outlook," *Proceedings of the IEEE*, 2024.

[149] M. Luo, B. Xu, J. Chen, L. Yang, Y. Zhou, M. Ren, and B. He, "Distributed control for semi-grant-free non-orthogonal multiple access," *Wireless Networks*, vol. 30, no. 1, pp. 309–320, 2024.

[150] M. C. Mayarakaca and B. M. Lee, "A survey on non-orthogonal multiple access for unmanned aerial vehicle networks: Machine learning approach," *IEEE Access*, 2024.

[151] P.-V. Mekikis, K. Ramantas, A. Antonopoulos, E. Kartsakli, L. Sanabria-Russo, J. Serra, D. Pubill, and C. Verikoukis, "NFV-enabled experimental platform for 5G tactile Internet support in industrial environments," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1895–1903, 2020.

[152] S. K. Sharma, I. Woungang, A. Anpalagan, and S. Chatzinotas, “Toward tactile Internet in beyond 5G era: Recent advances, current issues, and future directions,” *IEEE Access*, vol. 8, pp. 56 948–56 991, 2020.

[153] V. Gokhale, K. Kroep, V. S. Rao, J. Verburg, and R. Yechangunja, “TIXT: An extensible testbed for tactile Internet communication,” *IEEE Internet of Things Magazine*, vol. 3, no. 1, pp. 32–37, 2020.

[154] W. Na, N.-N. Dao, J. Kim, E.-S. Ryu, and S. Cho, “Simulation and measurement: Feasibility study of tactile Internet applications for mmWave virtual reality,” *ETRI Journal*, vol. 42, no. 2, pp. 163–174, 2020.

[155] M. F. Zhani and H. ElBakoury, “FlexNGIA: A flexible Internet architecture for the next-generation tactile Internet,” *Journal of Network and Systems Management*, pp. 1–45, 2020.

[156] D. A. Meshram and D. D. Patil, “5G-enabled tactile Internet for tele-robotic surgery,” *Procedia Computer Science*, vol. 171, pp. 2618–2625, 2020.

[157] X. Wei and L. Wu, “A new proposed sensor cloud architecture based on fog computing for Internet of Things,” in *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData).* IEEE, 2019, pp. 615–620.

[158] I. Budhiraja, S. Tyagi, S. Tanwar, N. Kumar, and J. J. Rodrigues, “Tactile Internet for smart communities in 5G: An insight for NOMA-based solutions,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3104–3112, 2019.

[159] J. Vora, S. Kaneriya, S. Tanwar, S. Tyagi, N. Kumar, and M. S. Obaidat, “TILAA: Tactile Internet-based ambient assistant living in fog environment,” *Future Generation Computer Systems*, vol. 98, pp. 635–649, 2019.

[160] S. Aggarwal and N. Kumar, “Fog computing for 5G-enabled tactile Internet: Research issues, challenges, and future research directions,” *Mobile Networks and Applications*, pp. 1–28, 2019.

[161] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “A cloud-based traffic flow framework for tactile Internet using SDN and fog computing,” in *29th International Telecommunication Networks and Applications Conference (ITNAC).* IEEE, 2019, pp. 1–6.

[162] M. Maier and A. Ebrahimzadeh, “Towards immersive tactile Internet experiences: Low-latency FiWi enhanced mobile networks with edge intelligence,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 4, pp. B10–B25, 2019.

[163] J. Arshad, M. A. Azad, K. Salah, R. Iqbal, M. I. Tariq, and T. Umer, “Performance analysis of content discovery for ad-hoc tactile networks,” *Future Generation Computer Systems*, vol. 94, pp. 726–739, 2019.

[164] J. Ren, G. Yu, Y. He, and G. Y. Li, “Collaborative cloud and edge computing for latency minimisation,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 5031–5044, 2019.

[165] A. Liberato, M. Martinello, R. L. Gomes, A. F. Beldachi, E. Salas, R. Villaca, M. R. Ribeiro, K. Kondepu, G. Kanellos, and R. Nejabati, “RDNA: Residue-defined networking architecture enabling ultra-reliable low latency datacenters,” *IEEE Transactions on Network and Service Management*, vol. 15, no. 4, pp. 1473–1487, 2018.

[166] J. A. Cabrera, R.-S. Schmoll, G. T. Nguyen, S. Pandi, and F. H. Fitzek, “Softwarisation and network coding in the mobile edge cloud for the tactile Internet,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 350–363, 2018.

[167] D. Perepelkin, M. Ivanchikova, V. Byshov, and I. Tsyganov, “Algorithm and software of virtual slices formation in software defined networks,” in *28th International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2018, Conference Proceedings, pp. 1–4.

[168] Y. Xiao and M. Krunz, “Distributed optimisation for energy-efficient fog computing in the tactile Internet,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2390–2400, 2018.

[169] A. A. Ateya, A. Muthanna, I. Gudkova, A. Abuarqoub, A. Vybornova, and A. Koucheryavy, “Development of intelligent core network for tactile Internet and future smart systems,” *Journal of Sensor and Actuator Networks*, vol. 7, no. 1, p. 1, 2018.

[170] S. Troia, A. Rodriguez, R. Alvizu, and G. Maier, “SENATUS: An experimental SDN/NFV orchestrator,” in *IEEE Conference on Network Function Virtualisation and Software Defined Networks (NFV-SDN)*. IEEE, 2018, Conference Proceedings, pp. 1–5.

[171] M. Gharbaoui, C. Contoli, G. Davoli, G. Cuffaro, B. Martini, F. Paganelli, W. Cerroni, P. Cappanera, and P. Castoldi, “Demonstration of latency-aware and self-adaptive service chaining in 5G/SDN/NFV infrastructures,” in *IEEE Conference on Network Function Virtualisation and Software Defined Networks (NFV-SDN)*. IEEE, 2018, Conference Proceedings, pp. 1–2.

[172] Y.-W. Ma, Y.-C. Chen, and J.-L. Chen, “SDN-enabled network virtualisation for Industry 4.0 based on IoTs and cloud computing,” in *19th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2017, Conference Proceedings, pp. 199–202.

[173] T. Theodorou and L. Mamatas, “CORAL-SDN: A software-defined networking solution for the Internet of Things,” in *IEEE Conference on Network Function Virtualisation and Software Defined Networks (NFV-SDN)*. IEEE, 2017, Conference Proceedings, pp. 1–2.

[174] K. Wang, Y. Wang, D. Zeng, and S. Guo, “An SDN-based architecture for next generation wireless networks,” *IEEE Wireless Communications*, vol. 24, no. 1, pp. 25–31, 2017.

[175] A. A. Ateya, A. Vybornova, K. Samouylov, and A. Koucheryavy, "System model for multi-level cloud based tactile Internet system," in *International Conference on Wired/Wireless Internet Communication*. Springer, 2017, pp. 77–86.

[176] Y. Feng, C. Jayasundara, A. Nirmalathas, and E. Wong, "IEEE 802.11 HCCA for tactile applications," in 27^{th} *International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE, 2017, pp. 1–3.

[177] J. Pilz, M. Mehlhose, T. Wirth, D. Wieruch, B. Holfeld, and T. Haustein, "A tactile Internet demonstration: 1ms ultra low delay for wireless communications towards 5G," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2016, pp. 862–863.

[178] N. Athmiya, K. Shobha, and V. Sarimela, "Feasibility study and implementation of OpenFlow-based SDN controller for tactical scenario," in *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, 2016, pp. 789–794.

[179] K. Intharawijitr, K. Iida, and H. Koga, "Analysis of fog model considering computing and communication latency in 5G cellular networks," in *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 2016, pp. 1–4.

[180] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, 2016.

[181] L. Tong, Y. Li, and W. Gao, "A hierarchical edge cloud architecture for mobile computing," in *IEEE INFOCOM–The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[182] N. B. Truong, G. M. Lee, and Y. Ghamri-Doudane, "Software defined networking-based vehicular adhoc network with fog computing," in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2015, Conference Proceedings, pp. 1202–1207.

[183] P. Iovanna and F. Ubaldi, "SDN solutions for 5G transport networks," in *International Conference on Photonics in Switching (PS)*. IEEE, 2015, Conference Proceedings, pp. 297–299.

[184] D. Szabo, A. Gulyas, F. H. Fitzek, and D. E. Lucani, "Towards the tactile Internet: Decreasing communication latency with network coding and software-defined networking," in *Proceedings of European Wireless; 21th European Wireless Conference*. VDE, 2015, Conference Proceedings, pp. 1–6.

[185] E. Cuervo, A. Balasubramanian, D.-k. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "MAUI: Making smartphones last longer with code offload," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, 2010, pp. 49–62.

[186] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for Internet of Things and analytics," in *Big data and internet of things: A roadmap for smart environments*. Springer, 2014, pp. 169–186.

[187] S. Igder, S. Bhattacharya, and J. M. Elmirghani, "Energy efficient fog servers for Internet of Things information piece delivery (IoTIPD) in a smart city vehicular environment," in *10th International Conference on Next Generation Mobile Applications, Security and Technologies (NGMAST)*. IEEE, 2016, Conference Proceedings, pp. 99–104.

[188] F. Jalali, K. Hinton, R. Ayre, T. Alpcan, and R. S. Tucker, "Fog computing may help to save energy in cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1728–1739, 2016.

[189] S. Delfin, N. Sivasanker, N. Raj, and A. Anand, "Fog computing: A new era of cloud computing," in *3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, Conference Proceedings, pp. 1106–1111.

[190] M. Al Yami and D. Schaefer, "Fog computing as a complementary approach to cloud computing," in *International Conference on Computer and Information Sciences (ICCIS)*. IEEE, 2019, Conference Proceedings, pp. 1–5.

[191] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwälder, and B. Koldehofe, "Mobile fog: A programming model for large-scale applications on the Internet of Things," in *Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, 2013, pp. 15–20.

[192] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "Toward a fog-based traffic flow framework for tactile Internet," *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 10 718–10 731, 2021.

[193] J. Sachs, L. A. Andersson, J. Araújo, C. Curescu, J. Lundsjö, G. Rune, E. Steinbach, and G. Wikström, "Adaptive 5G low latency communication for tactile Internet services," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325–349, 2018.

[194] Sciforce, "The tactile Internet - A new IoT?" Available: https://medium.com/sciforce/the-tactile-internet-a-new-iot-6844a117c420, (Accessed: Apr. 15, 2025).

[195] "Microsoft Azure," Available: https://azure.microsoft.com/en-gb/, (Accessed: Apr. 15, 2025).

[196] "Amazon Web Services (AWS)," Available: https://aws.amazon.com/, (Accessed: Apr. 15, 2025).

[197] "Google Cloud," Available: https://cloud.google.com/, (Accessed: Apr. 15, 2025).

[198] "IBM Cloud," Available: https://www.ibm.com/cloud, (Accessed: Apr. 15, 2025).

[199] "Alibaba Cloud," Available: https://www.alibabacloud.com/en?_p_lc=5, (Accessed: Apr. 15, 2025).

[200] A. N. Nair, M. Jacob, and A. Krishnamoorthy, “The multi server M/M/(s, S) queueing inventory system,” *Annals of Operations Research*, vol. 233, no. 1, pp. 321–333, 2015.

[201] M. Harchol-Balter, *Performance modelling and design of computer systems: Queueing theory in action*. Cambridge University Press, 2013, (Accessed: Apr. 15, 2025). [Online]. Available: https://docshare.tips/performance-modeling-and-design-of-computer-systems-queueing-theory-in-action_5871b5a4b6d87f2f098b4945.html

[202] M. Jia, W. Liang, Z. Xu, and M. Huang, “Cloudlet load balancing in wireless metropolitan area networks,” in *IEEE INFOCOM - The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[203] R. Pirmagomedov, A. A. Ahmed, and R. Glushakov, “Multi-level architecture for P2P services in mobile networks,” in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2019, pp. 415–423.

[204] S. K. Battula, S. Garg, R. K. Naha, P. Thulasiraman, and R. Thulasiram, “A micro-level compensation-based cost model for resource allocation in a fog environment,” *Sensors*, vol. 19, no. 13, p. 2954, 2019.

[205] S. V. Margariti, V. V. Dimakopoulos, and G. Tsoumanis, “Modelling and simulation tools for fog computing - A Comprehensive survey from a cost perspective,” *Future Internet*, vol. 12, no. 5, p. 89, 2020.

[206] H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, “iFogSim: A toolkit for modelling and simulation of resource management techniques in the Internet of Things, edge and fog computing environments,” *Software: Practice and Experience*, vol. 47, no. 9, pp. 1275–1296, 2017.

[207] R. Mahmud and R. Buyya, “Modelling and simulation of fog and edge computing environments using iFogSim toolkit,” in *Fog and Edge Computing: Principles and Paradigms*. John Wiley & Sons, Inc., 2019, pp. 433–465.

[208] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “TINetS3: SDN-driven network slicing enabling scenario-based applications in tactile Internet,” *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 4639–4654, 2024.

[209] M. Simsek, G. P. Fettweis, and I. Chih-Lin, “Tactile Internet [scanning the issue],” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 253–255, Feb. 2019.

[210] F. H. Fitzek, S.-C. Li, S. Speidel, T. Strufe, M. Simsek, and M. Reisslein, *Tactile Internet: With human-in-the-loop*. Academic Press, 2021.

[211] “The 3rd generation partnership project (3GPP), Release 15.” Available: https://www.3gpp.org/specifications-technologies/releases/release-15, Apr. 2019, (Accessed: Apr. 15, 2025).

[212] M. Awais, F. Ullah Khan, M. Zafar, M. Mudassar, M. Zaigham Zaheer, K. Mehmood Cheema, M. Kamran, and W.-S. Jung, “Towards enabling haptic communications over 6G: Issues and challenges,” *Electronics*, vol. 12, no. 13, p. 2955, Jul. 2023.

[213] S. A. Kazmi, L. U. Khan, N. H. Tran, and C. S. Hong, *Network slicing for 5G and beyond networks*. Springer, Oct. 2019, vol. 1.

[214] "Network slicing market size, share and growth opportunities 2030." Available: https://www.fortunebusinessinsights.com/network-slicing-market-107303, Mar. 2023, (Accessed: Apr. 15, 2025).

[215] "Software-defined networking market size, share and forecast 2027." Available: https://www.alliedmarketresearch.com/software-defined-networking-market, Sep. 2022, (Accessed: Apr. 15, 2025).

[216] E. Ivanova, J. Eden, S. Zhu, G. Carboni, A. Yurkewich, and E. Burdet, "Short time delay does not hinder haptic communication benefits," *IEEE Transactions on Haptics*, vol. 14, no. 2, pp. 322–327, 2021.

[217] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges," *Computer Networks*, vol. 167, p. 106984, Feb. 2020.

[218] R. Botez, A.-G. Pasca, A.-T. Sferle, I.-A. Ivanciu, and V. Dobrota, "Efficient network slicing with SDN and heuristic algorithm for low latency services in 5G/B5G networks," *Sensors*, vol. 23, no. 13, p. 6053, Jun. 2023.

[219] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of Things realization in 5G networks," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957–994, Mar. 2021.

[220] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[221] X. Li, M. Samaka, H. A. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, "Network slicing for 5G: Challenges and opportunities," *IEEE Internet Computing*, vol. 21, no. 5, pp. 20–27, Sep. 2017.

[222] S. Zahoor, I. Ahmad, M. T. B. Othman, A. Mamoon, A. U. Rehman, M. Shafiq, and H. Hamam, "Comprehensive analysis of network slicing for the developing commercial needs and networking challenges," *Sensors*, vol. 22, no. 17, p. 6623, Sep. 2022.

[223] S. Singh, C. R. Babu, K. Ramana, I.-H. Ra, and B. Yoon, "BENS-B5G: Blockchain-enabled network slicing in 5G and beyond-5G (B5G) networks," *Sensors*, vol. 22, no. 16, p. 6068, Aug. 2022.

[224] Y.-S. Chiu, L.-H. Yen, T.-H. Wang, and C.-C. Tseng, "A cloud native management and orchestration framework for 5G end-to-end network slicing," in *IEEE International Conference on Service-Oriented System Engineering (SOSE)*. IEEE, Oct. 2022, pp. 69–76.

[225] K.-L. Lee, C.-N. Lee, and M.-F. Lee, "Realizing 5G network slicing provisioning with open source software," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, Feb. 2021, pp. 1923–1930.

[226] A. Hefele and J. Costa-Requena, "SDN managed network slicing in mobile backhaul," in *International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, Apr. 2020, pp. 1–8.

[227] N. Saha, A. James, N. Shahriar, R. Boutaba, and A. Saleh, "Demonstrating network slice KPI monitoring in a 5G testbed," in *IEEE/IFIP Network Operations and Management Symposium*. IEEE, Jun. 2022, pp. 1–3.

[228] O. A. Latif, M. Amer, and A. Kwasinski, "Achieving linear scaling in provisioning end-to-end network slicing," in *IEEE Future Networks World Forum (FNWF)*. IEEE, Mar. 2023, pp. 108–112.

[229] H. Shen, T. Wang, Y. Heng, and G. Bai, "Joint resource slicing and vehicle association for drone-assisted vehicular networks," *Drones*, vol. 7, no. 8, p. 534, Aug. 2023.

[230] W. Wu, C. Zhou, M. Li, H. Wu, H. Zhou, N. Zhang, X. S. Shen, and W. Zhuang, "AI-native network slicing for 6G networks," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 96–103, Feb. 2022.

[231] S. A. AlQahtani, "Towards an optimal cloud-based resource management framework for next-generation Internet with multi-slice capabilities," *Future Internet*, vol. 15, no. 10, p. 343, 2023.

[232] A. J. Zharabad, S. Yousefi, and T. Kunz, "Network slicing in virtualized 5G core with VNF sharing," *Journal of Network and Computer Applications*, vol. 215, p. 103631, 2023.

[233] L. Han, Y. Bin, S. Zhu, and Y. Liu, "Dynamic selection slicing-sased offloading algorithm for in-vehicle tasks in mobile edge computing," *Electronics*, vol. 12, no. 12, p. 2708, 2023.

[234] R. Dangi and P. Lalwani, "Feature selection based machine learning models for 5G network slicing approximation," *Computer Networks*, p. 110093, 2023.

[235] "Harshicorp Vagrant," Available: https://www.vagrantup.com/, (Accessed: Apr. 15, 2025).

[236] "Mininet," Available: https://mininet.org/, (Accessed: Apr. 15, 2025).

[237] I. Rhee, L. Xu, S. Ha, A. Zimmermann, L. Eggert, and R. Scheffenegger, "CUBIC for fast long-distance networks," Available: https://www.rfc-editor.org/info/rfc8312, Feb. 2018, (Accessed: Apr. 15, 2025).

[238] Y. Xiao and M. Krunz, "Dynamic network slicing for scalable fog computing systems with energy harvesting," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 12, pp. 2640–2654, 2018.

[239] K. Polachan, B. Turkovic, T. Prabhakar, C. Singh, and F. A. Kuipers, "Dynamic network slicing for the tactile Internet," in *ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2020, pp. 129–140.

[240] S. Venkatapathy, T. Srinivasan, H.-G. Jo, and I.-H. Ra, “An E2E network slicing framework for slice creation and deployment using machine learning,” *Sensors*, vol. 23, no. 23, p. 9608, 2023.

[241] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, “A study of downlink power-domain non-orthogonal multiple access performance in tactile Internet employing sensors and actuators,” *Sensors*, vol. 24, no. 22, p. 7220, 2024.

[242] “Ericsson mobility report, Q4 2023 update,” Available: https://www.ericsson.com/4923b9/assets/local/reports-papers/mobility-report/documents/2023/emr-q4-2023-update.pdf, Feb. 2024, (Accessed: Apr. 15, 2025).

[243] “Cisco annual Internet report (2018 - 2023),” Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf, 2023, (Accessed: Apr. 15, 2025).

[244] M. Series, “IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond,” *Recommendation ITU*, vol. 2083, pp. 1–21, 2015, (accessed on 26 October 2024).

[245] “The 3rd generation partnership project (3GPP), Release 16.” Available: https://www.3gpp.org/specifications-technologies/releases/release-16, Jul. 2020, (Accessed: Apr. 15, 2025).

[246] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, O. Kucur *et al.*, “A tutorial on nonorthogonal multiple access for 5G and beyond,” *Wireless communications and mobile computing*, vol. 2018, 2018.

[247] O. M. K. Al-Dulaimi, A. M. K. Al-Dulaimi, M. O. Alexandra, and M. K. H. Al-Dulaimi, “Strategy for non-orthogonal multiple access and performance in 5G and 6G networks,” *Sensors*, vol. 23, no. 3, p. 1705, 2023.

[248] A. Akbar, S. Jangsher, and F. A. Bhatti, “NOMA and 5G emerging technologies: A survey on issues and solution techniques,” *Computer Networks*, vol. 190, p. 107950, 2021.

[249] M. Abd-Elnaby, G. G. Sedhom, E.-S. M. El-Rabaie, and M. Elwekeil, “NOMA for 5G and beyond: Literature review and novel trends,” *Wireless Networks*, vol. 29, no. 4, pp. 1629–1653, 2023.

[250] B. Safia and K. Abdellatif, “Performance analysis of the NOMA network under joint hardware impairment and imperfect channel state information,” *Telecommunications and Radio Engineering*, vol. 82, no. 10, 2023.

[251] M. Hussain and H. Rasheed, “Nonorthogonal multiple access for next-generation mobile networks: A technical aspect for research direction,” *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–17, 2020.

[252] S. Tanwar, S. Tyagi, I. Budhiraja, and N. Kumar, “Tactile Internet for autonomous vehicles: Latency and reliability analysis,” *IEEE Wireless Communications*, vol. 26, no. 4, pp. 66–72, 2019.

[253] N. Gholipoor, S. Parsaeefard, M. R. Javan, N. Mokari, H. Saeedi, and H. Pishro-Nik, "Cloud-based queuing model for tactile Internet in next generation of RAN," in *IEEE 91st Vehicular Technology Conference (VTC - Spring)*. IEEE, 2020, pp. 1–6.

[254] N. Ye, X. Li, H. Yu, A. Wang, W. Liu, and X. Hou, "Deep learning aided grant-free NOMA toward reliable low-latency access in tactile Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2995–3005, 2019.

[255] T. Van Truong, A. Nayyar, M. Bilal, and K. S. Kwak, "System design and optimization of mobile edge computing in the NOMA wireless wactile Internet of Things network," *Alexandria Engineering Journal*, vol. 73, pp. 737–749, 2023.

[256] B. S. K. Reddy, K. Mannem, and K. Jamal, "Software defined radio based non-orthogonal multiple access (NOMA) systems," *Wireless Personal Communications*, vol. 119, no. 2, pp. 1251–1273, 2021.

[257] H. Yahya, A. Ahmed, E. Alsusa, A. Al-Dweik, and Z. Ding, "Error rate analysis of NOMA: Principles, survey and future directions," *IEEE Open Journal of the Communications Society*, 2023.

[258] S. A. H. Mohsan, Y. Li, A. V. Shvetsov, J. Varela-Aldás, S. M. Mostafa, and A. Elfikky, "A survey of deep learning based NOMA: State of the art, key aspects, open challenges and future trends," *Sensors*, vol. 23, no. 6, p. 2946, 2023.

[259] D. Sarkar, S. S. Yadav, V. Pal, N. Kumar, S. K. Patra *et al.*, "A comprehensive survey on IRS-assisted NOMA-based 6G wireless network: Design perspectives, challenges and future directions," *IEEE Transactions on Network and Service Management*, 2024.

[260] A. Apiyo and J. Izydorczyk, "A survey of NOMA-aided cell-free massive MIMO systems," *Electronics*, vol. 13, no. 1, p. 231, 2024.

[261] S. A. H. Mohsan, M. Sadiq, Y. Li, A. V. Shvetsov, S. V. Shvetsova, and M. Shafiq, "NOMA-based VLC systems: A comprehensive review," *Sensors*, vol. 23, no. 6, p. 2960, 2023.

[262] R. O. Ogundokun, J. B. Awotunde, A. L. Imoize, C.-T. Li, A. T. Abdulahi, A. B. Adelodun, S. N. Sur, and C.-C. Lee, "Non-orthogonal multiple access enabled mobile edge computing in 6G communications: A systematic literature review," *Sustainability*, vol. 15, no. 9, p. 7315, 2023.

[263] F. Khennoufa, K. Abdellatif, and F. Kara, "Bit error rate and outage probability analysis for multi-hop decode-and-forward relay-aided NOMA with imperfect SIC and imperfect CSI," *AEU-International Journal of Electronics and Communications*, vol. 147, p. 154124, 2022.

[264] G. Taricco, "Fair power allocation policies for power-domain non-orthogonal multiple access transmission with complete or limited successive interference cancellation," *IEEE Access*, vol. 11, pp. 46 793–46 803, 2023.

[265] ——, "Optimum power allocation for HARQ aided NOMA with proportional fairness on fading channels," *IEEE Access*, 2024.

[266] A. Nasser, A. Celik, and A. M. Eltawil, “Joint user-target pairing, power control, and beamforming for NOMA-aided ISAC networks,” *IEEE Transactions on Cognitive Communications and Networking*, 2024.

[267] Z. Ding, R. Schober, and H. V. Poor, “Design of downlink hybrid NOMA transmission,” *arXiv preprint arXiv:2401.16965*, 2024.

[268] Y. Liu, M. Jin, Q. Guo, and J. Yao, “Secure beamforming for NOMA-ISAC with system imperfections,” *IEEE Communications Letters*, 2024.

[269] A. A. Brovkin, Y. V. Kryukov, D. A. Pokamestov, E. V. Rogozhnikov, G. N. Shalin, and A. S. Shinkevich, “Scheduling PD-NOMA with round-robin, equal-rate and proportional-fair strategies,” in *International Siberian Conference on Control and Communications (SIBCON)*. IEEE, 2022, pp. 1–6.

[270] J. He, Z. Tang, Z. Tang, H.-H. Chen, and C. Ling, “Design and optimization of scheduling and non-orthogonal multiple access algorithms with imperfect channel state information,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 800–10 814, 2018.

[271] “MATLAB,” Available: https://au.mathworks.com/products/matlab.html, (Accessed: Apr. 15, 2025).

Appendix A

A List of Abbreviation and their Explanation

Table A.1 lists the abbreviations used in this thesis and their explanations.

Table A.1 A list of abbreviations and their explanations.

Abbreviation	**Explanation**
4G	Fourth-Generation
3GPP	3rd Generation Partnership Project
5G	Fifth-Generation
AR/VR	Augmented Reality/Virtual Reality
B5G	Beyond Fifth-Generation
BER	Bit Error Rate
BS	Base Station
CAGR	Compound Annual Growth Rate
CSI	Channel State Information
DL	Downlink
E2E	End-to-End
eMBB	Enhanced Mobile Broadband
FC	Fog Computing
H2M	Human-to-Machine
IoT	Internet of Things
KPI	Key Performance Indicators
MA	Multiple Access
ML/AI	Machine Learning/Artificial Intelligence
mMTC	Massive Machine-Type Communications
NFV	Network Function Virtualisation
NOMA	Non-Orthogonal Multiple Access
NS	Network Slicing
OMA	Orthogonal Multiple Access
OVS	Open Virtual Switch
PA	Power Allocation
PD	Power Domain
QoE	Quality of Experience

Table A.1 *Cont.*

Abbreviation	Explanation
QoS	Quality of Service
RB	Resource Block
RTT	Round-Trip Time
SDN	Software-Defined Networking
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference Noise Ratio
SLA	Service Level Agreement
SNR	Signal to Noise Ratio
TI	Tactile Internet
UE	User Equipment
URLLC	Ultra-Reliable Low-Latency Communication

***