QoS provision for haptic communication over the Tactile Internet

Antonakoglou, Konstantinos

Awarding institution:
King's College London

University of London
King's College London
Department of Informatics

QoS provision for haptic communication over the Tactile Internet

Konstantinos Antonakoglou

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Telecommunications of King's College London,
Department of Informatics, 26 April 2019

Abstract

As the networking infrastructure and its components evolve within the telecommunications ecosystem, future networks will be able to accommodate services and applications of the upcoming 5G Tactile Internet. After analysing and experimenting with 5G enablers, this thesis focuses on haptic communication over the future communication systems. More specifically, the thesis addresses the issues related to Quality of Service (QoS) provisioning for Ultra Reliable Low Latency Communication (URLLC) traffic and specifically for bilateral teleoperation applications. Assuming a dynamic priority queuing system serving URLLC traffic as high priority, under certain conditions, prioritisation causes resource starvation for traffic of lower priority. On the other hand, without guaranteed QoS, bilateral teleoperation system performance and operation can be negatively affected, especially in the presence of delay.

In this thesis, the aforementioned problem is addressed by proposing a novel framework that focuses on minimising the impact of delay on haptic communication while also minimising the effect of QoS overprovisioning. It also presents a possible 5G architecture design and suggests necessary architectural design components that enable the Tactile Internet. Simulation results verify that it is feasible to manage the balance of increasing teleoperation performance of high priority traffic and decreasing the impact of resource starvation of low priority traffic.

Acknowledgements

This work was supported financially by an EPSRC studentship.

Furthermore, I would like to extend my sincere gratitude to:

- My first supervisor, Dr. Toktam Mahmoodi, for the crucial support, the always meaningful guidance and productive collaboration during my PhD studies.
- My second supervisor, Prof. Mischa Dohler, for the precious support and mentorship from the first day of this long journey.
- Current and past members of KCL CTR, CORE and other institutions or companies with which I actively collaborated during my PhD studies to accomplish several parts of this work or other contributions and goals. Special thanks to Dr. Maria A. L. Rosas, Dr. Fragkiskos Sardis, Dr. Luis Sequeira, Ms. Maliheh Mahlouji, Prof. Marcelo A. C. Fernandes, Dr. Xiao Xu and others.
- Other current and past members of KCL CTR for their valuable support and companionship. Many thanks to Dr. Giorgos Chochlidakis, Dr. Christophoros Vlachos, Mr. Enric Pardo, Dr. Ehsan Ghoreishi, Dr. Massimo Condoluci, Dr. Omar Al Kadri, Dr. Gao Zheng and others.
- Last but not least, my parents, Sophia and Stavros, and my sister Joanna for their unlimited support and patience, as well as all my close friends.

Dedication

I dedicate this work first and foremost to my parents. I am also grateful to everyone else who supported and inspired me to achieve my goals thus far.

Contents

List of Tables

List of Figures

Glossary

A_T PGF of total number of arrivals 122

B_e Viscous damping coefficient of the environment 39

B_m Viscous damping coefficient of master manipulator 38, 117

B_p Perceived damping of manipulator 117

B_s Viscous damping coefficient of slave manipulator 38

$E[d]$ Mean value of delay of any arbitrary packet 122

$E[d_H]$ Mean value of delay in high priority queue 122, 132, 133

$E[d_L]$ Mean value of delay in low priority queue 122, 132

$E[p_T]$ Total number of packets in the system 122

F_e Environment force 39

F_h Force exerted from the user 39

F_m Output force of the master manipulator 38, 39

F_s Output force of the slave manipulator 38, 39

F_{mc} Master manipulator controller force 39

F_{sc} Slave manipulator controller force 39

F_{st} Time-delayed output force of the slave manipulator 39, 41

G_c Control system parameter 41

K_e Environment stiffness 39, 117

K_p Perceived environment stiffness 117

Y Recursive function (numerically computed) 122

Δ Drift error 128

β jump probability for HOL-MBP 122, 123, 128, 131–133, 135, 136

δ Tracking error 128

λ_1 Arrival rate of high priority traffic 120, 121

λ_2 Arrival rate of low priority traffic 120, 121

λ_T Total arrival rate 120–122

λ_{TT} Second partial derivative of arrival process PGF 122

v_m Master manipulator mass 38, 117–119

v_m Perceived mass of manipulator 117

v_m Master manipulator velocity 38

v_s Slave manipulator mass 38

v_s Slave manipulator velocity 38, 39

x_m Master position 40–42

x_s Slave position 39–41

x_{mt} Time-delayed master position 39–41

x_{st} Time-delayed slave position 40

Acronyms

2CH two-channel 35, 36, 45

3GPP 3rd Generation Partnership Project 12, 67, 68, 74, 80–83, 100, 107, 109, 141

4CH four-channel 35, 36, 45

5G fifth generation 9, 12, 15, 17, 67, 68, 82, 100, 102, 104

5QI 5G QoS Indicator 78, 80

ACE-LPC Algebraic Code-Excited Linear Prediction coding 33

ADAMS Adaptive Mulsemedia Delivery Solution 64, 66

Admux Adaptive Multiplexer 65, 66

ADPCM Adaptive Differential Pulse-Code Modulation 28

AES Advanced Encryption Standard 66

ALPHAN Application Layer Protocol for HAptic Networking 65, 66

AMF Access and Mobility Management Function 74, 107

AMFC Adaptive Motion/Force Control 48, 117, 118, 126

BE Best Effort 94–97, 99, 100, 102–104, 141

C-HAVE Collaborative-Haptic Virtual Environments 24, 64

C-MTC Critical Machine-type Communication 68

CDOB Communication Disturbance Observer 49, 50

CN Core Network 11, 84, 99, 100, 102, 106

CPLF Control Performance Loss Function 118, 123–129, 131, 138, 142, 143

CPS Cyber-Physical Systems 66, 67

DCT Discrete Cosine Transform 28, 32

DiffServ Differentiated Services 60, 76–78, 80

DoF Degrees of Freedom 23, 29, 34, 85

ETP Efficient Transport Protocol 61–63

ETSI European Telecommunications Standards Institute 83, 106, 107

F/T Force/Torque 90

FDPA Fair Dynamic Priority Assignment 79

FOLP First-order linear predictor 31

FR Force Reflection 38–41, 49, 117, 118, 125, 127, 130, 131

HMTP Hybrid Multicast Transport Protocol 62

HoIP Haptics over Internet Protocol 65, 66

HOL-MBP Head-Of-Line Merge-By-Probability 79, 80, 120, 121, 123, 138, 143

IntServ Integrated Services 76, 78, 80

IoT Internet of Things 9, 18

IPG Interpacket Gap 62

IPSec IP Security protocol 66

IRTP Interactive Real-Time Protocol 61

IS Intrinsically Stable 37, 49, 118, 126

ISS Input-to-State Stability 54

ITP Interoperable Telesurgical Protocol 65, 66

ITU International Telecommunication Union 67

JND Just Noticable Difference 29–31, 33, 53, 65

KPI Key Performance Indicator xi, 67–70, 100

LTE Long-Term Evolution 72, 78

LTI linear time-invariant 36

M-MTC Massive Machine-type Communication 68

MANO Management and Orchestration 108, 110

MEC Mobile Edge Computing 75

MIS minimally invasive surgery 118

MMTA Model-Mediated Teleoperation Approach 48, 50–53, 55–57, 75, 118

MOS Mean Opinion Score 33

MPC Model Predictive Control 44

NFV Network Function Virtualization 71, 74, 83, 104, 106, 107, 109, 110, 140

NFVI Network Functions Virtualization Infrastructure 108

NS Network Service 108

ODL OpenDaylight 93, 102

Ofcom UK Office of Communications 68

OVSDB Open vSwitch Database 101

PAHCP Perception-based adaptive haptic communication protocol 64–66

PC Predictive Control 49, 117, 118, 131

PCF Policy Control Function 74, 106, 107

PCO Passivity Controller 46, 47

PCP Predictive Control with Passivity 49, 117, 118, 131

PD Proportional Derivative 44, 56, 57

PDB Perceptual Deadband 55

PDF Probability Density Function 97, 98

PDV Packet Delay Variation 58

PE Position Error 40, 117, 125, 127, 130, 131

PHB per-hop behaviour 77

PO Passivity Observer 46, 47, 56

PP Probabilistic Priority 79

PS Possibly Stable xiv, 37, 40, 123, 125–127, 130

QCI QoS Class Identifier 78, 80

QoC Quality of Control 20

QoE Quality of Experience 10, 19, 25, 59, 63, 68, 81, 82, 86, 88, 110, 124, 142, 144

QoS Quality of Service i, 11–15, 34, 58, 59, 65, 66, 75–78, 80, 83, 87–89, 92, 96, 98, 100, 101, 103, 106–109, 115, 139–141, 144, 145

RAN Radio Access Network 11, 67

RMTP Reliable Multicast Transport Protocol 62

ROS Robot Operating System 99

RTNP Real-Time Network Protocol 61

RTP Real-Time Transport Protocol 64

RTT Round-Trip Time 68

SCTP Synchronous Collaboration Transport Protocol 61–63, 65

SDN Software Defined Networking 12, 70, 72–74, 77, 83, 84, 89, 100–102, 104, 107–109, 111, 140, 141

SIP Session Initiation Protocol 64

SMF Session Management Function 74, 107, 108

SP Strict Priority 78, 79

SRMP Scalable Reliable Multicast Protocol 62

SRTP Secure Real-time Transport Protocol 62

SSR-UDP Secure and Statistically Reliable UDP 66

STRON Supermedia TRansport for teleoperations over Overlay Networks 61–63

TC Traffic Class 77

TCP Transmission Control Protocol 60–63, 79, 93

TD Tactile Device 104–106

TDPA Time-Domain Passivity Approach 43, 46, 47, 55, 56

TDPN Time Delay Power Network 47, 48

TE Tactile Edge 104, 105

TFRC TCP-Friendly Rate Control 61

TI Tactile Internet 9, 10, 13, 17, 18, 80, 81, 89, 104–106, 111, 141

ToS Type of Service 77

TPTA Telepresence and Teleaction 20, 34, 43, 57, 58, 64

TSN Time-sensitive Networks 76

TTI Transmission Time Interval 71, 72

UDP User Datagram Protocol 60–65, 93, 96, 102

UPF User Plane Function 74, 100

URLLC Ultra Reliable Low Latency Communication i, 9, 10, 12, 13, 68, 83, 137, 139–141

VNF Virtual Network Function 73, 75, 83, 107–110, 141

WV Wave Variable 43, 55, 56

ZOH Zero order hold 30, 31, 55, 56, 65

Chapter 1

Introduction

One of the main goals of the forthcoming fifth generation (5G) of wireless and mobile networks is to allow the realisation of the Tactile Internet (TI) and the various use cases that surround it, integrating the necessary enablers to meet the demanding requirements of a plethora of application domains. Healthcare, transportation, education and of course telecommunications, are only a few of the sectors that will benefit from such advances.

According to the International Telecommunication Union, the predecessor of TI, the Internet of Things (IoT), allowed the distribution of data, collected or generated by interconnected smart devices over the Internet enabling delay-tolerant use cases. The next step, to be taken by TI, is envisioned to include the use of current and future network enablers of use cases that demand extremely low-latency, high availability, high reliability and security [1].

The transition from 4G to 5G has been characterised not just by an enhancement of previous features, but a redesign and rethinking of the overall system design due to the diverse types of traffic of distinct categories of services it has to accommodate. One of these categories is URLLC and its corresponding services require mainly very low latency and very high availability [2]. Incorporation of technologies and components that present features such as network softwarisation, dynamic network reconfiguration and orchestration will provide improved performance and efficient resource management. In this way, applications that generate URLLC and other types of traffic will be able to meet their full potential.

With the arrival of the TI era, it is expected that users will be able to use communication systems that offer improved immersion to remote environments. Such systems will be capable of multi-modal communication, where audition and vision will be complemented by the sense of touch by means of bilateral teleoperation, also known as haptic communication, a use case of URLLC. Apart from teleoperation of remote robotic systems such as robotic arms and unmanned aerial or ground vehicles, other use cases of URLLC are monitoring and control in smart grids, wireless systems for industrial automation [3].

Evidently, the scale and high requirements of such applications entail a revolutionary approach in materialising the 5G infrastructure, not only with the improvement of telecommunication system performance but doing so by enriching the network with application-specific functionalities.

1.1 Motivation and Objectives

Even though bilateral teleoperation is a well-known concept for decades now, delivering high Quality of Experience (QoE) to the users of such systems is still a challenging task due to the high requirements. In bilateral teleoperation terminology, QoE can be defined using the concept of transparency, which Lawrence [4] describes as “the degree of telepresence of the remote site available to the operator through the teleoperator device”. Furthermore, transparency can be measured objectively [4] and subjectively [5].

Currently, even though highly competent robotic devices with haptic feedback capabilities are available, with the introduction of latency in the communication channel that links these devices, the haptic communication system may underperform or become unstable. The solution to this problem has been detected decades ago in the design of bilateral teleoperation control methodologies that maintain the stability of the system under undesirable delay in the communication channel.

Presently, it is impossible for the current network infrastructure to meet the requirements of applications of long-distance haptic communication due to the restrictions imposed by the speed of light and the delays caused by the elements that compose the end-to-end network [6]. However, previous generations of networks do not possess the essential end-to-end network performance, flexibility and scalability capabilities that 5G networks are designed to offer, for haptic communication services to be successfully delivered [7].

Current enablers of 5G networks, both in Radio Access Network (RAN) and Core Network (CN) allow the separation of control and user plane as well as the softwarisation and system-level management of network, storage and computation functions. As such, an abstraction layer between physical network infrastructure and the user application can be provided, which allows for greater flexibility in creating optimised network slices for different use cases [8].

The aforementioned application-specific optimisation of network performance can only become reality by making the network more application-aware. However, there are two main challenges that need to be addressed. Firstly, the joint consideration and collaboration of application-specific methodologies of haptic communication and the underlying network infrastructure. Essentially, even though the network infrastructure and the user application using it share the same goal, there is limited research activity on this topic. Secondly, while critical and low-latency traffic should be prioritised, this prioritisation should impose minimal impact on traffic with a lower level of requirements, so that the network as a whole remains scalable.

1.2 Thesis statement

The design of future networks that will enable the Tactile Internet is aiming to satisfy requirements of various diverse use cases, such as haptic communication. While guaranteeing the appropriate QoS for each use case is a major goal, it is also important to examine, evaluate and manage the coexistence of diverse data traffic in order to avoid QoS overprovisioning.

1.3 Original contributions of the thesis

Considering all of the above, for the rest of this chapter this work presents the following contributions in order of significance.

1) **Dynamic traffic prioritisation for haptic communication**: This contribution presents a framework for dynamic traffic prioritisation for haptic communication. The proposed framework optimises traffic prioritisation between URLLC traffic for bilateral teleoperation and lower priority URLLC traffic and enables the selection of the most suitable teleoperation control scheme, to maintain the stability and transparency of the teleoperation system in desirable levels.

2) **A novel analysis and description of the efforts towards implementing haptic communication over 5G networks**: This contribution includes an investigation and discussion of the state-of-the-art focusing on bilateral teleoperation and the enablers of haptic communication over 5G networks. It describes the importance of elements of communication, computation and teleoperation control should jointly address the issues of long-distance bilateral teleoperation.

3) **Orchestration of a possible 5G architecture for haptic communication**: In this contribution, a possible 5G core network architecture that orchestrates 3rd Generation Partnership Project (3GPP) components implemented as VNFs by using physical and virtual nodes of Software Defined Networking (SDN) controller instances to enable URLLC is presented and experimentally verified. This 5G architecture aims to satisfy the requirements of URLLC as well as other types of traffic by means of traffic differentiation.

4) **Experimental frameworks for 5G enabler evaluation focusing on URLLC:** This contribution is the creation of experimental frameworks to evaluate haptic communication and 5G enablers, that prove and evaluate how URLLC QoS can be guaranteed, as well as measure the impact of QoS overprovisioning. This evaluation focuses on network performance metrics of latency, packet loss and jitter. In the experiments both indicative

and novel haptic communication systems are used. Furthermore, two of the experiments are prototype health-care focused haptic communication systems.

5) **Contribution to IEEE P1918.1 architecture standards**: I contributed to the IEEE P1918.1 by participating in the effort of the IEEE P1918.1 Standards Working Group during the initial definition of the proposed architecture elements within IEEE P1918.1. This included participation in the definition of the architecture entities/domains, physical and logical interfaces and the functional capabilities that are considered within the Tactile Internet.

1.4 Structure of the thesis

This section reveals the flow of the thesis. To begin with, Chapter 2 introduces and describes haptic communication and 5G enablers, providing an analysis of the state-of-the-art methods and technologies. The chapter starts with the broad topics of bilateral teleoperation, data processing and networking for haptic communication. After examining 5G networks, relevant 5G use case scenarios and their requirements, the chapter narrows the focus on traffic differentiation and dynamic scheduling. The latter provide the bases for the experiments and methodologies presented later on.

Next, Chapter 3 includes experimentation frameworks for haptic communication systems that use 5G enablers. The purpose of these frameworks is to evaluate the performance of 5G enablers in terms of providing the necessary QoS to haptic communication systems, but also examine how QoS provision affects the coexistence of differentiated traffic flows in the network. The chapter also presents and discusses standardisation efforts on the generic architecture design for the implementation of the TI. Additionally, it presents a 5G system architecture, composed of 3GPP virtualised components, managed by a network orchestration system.

Taking into account the discussion and results of the previous chapter, Chapter 4 focuses on dynamic traffic prioritisation for haptic communication. Specifically, this chapter presents a novel framework that incorporates a network queuing approach that accommodates URLLC traffic flows with different priorities. More importantly, it proposes a model for balancing QoS provision between high priority bilateral teleoperation traffic and lower priority traffic. It also enables the selection of the most suitable teleoperation control scheme, to maintain the stability and transparency of the teleoperation system.

Finally, Chapter 5 is the conclusion which contains the thesis summary, a discussion on the main contributions, a reflection on future avenues of research and concluding remarks.

Chapter 2

Background

2.1 Overview

As stated in the previous chapter, in order for 5G networks to provide efficient service for haptic communication to achieve optimum performance, they need to become more application-aware. On the other hand, the system must also take into account the QoS provision of other types of traffic. This requires a systematic evaluation of both application-specific methodologies as well as the network infrastructure enablers and technologies.

This chapter, as part of the contributions of this thesis, presents a novel analysis and description of the efforts towards implementing haptic communication over 5G networks. The contribution presented is an investigation and discussion of the state-of-the-art focusing on bilateral teleoperation and the enablers of haptic communication over 5G networks. Furthermore, the main goals are to introduce and describe the importance of working on the evolution of networks capable of hosting haptic communication, to present the state-of-the-art or most prominent methods and technologies from the subject areas of networking and bilateral teleoperation as presented in [9]. This chapter also focuses on the efforts for haptic communication in networked teleoperation systems over the Tactile Internet and examines in detail the advancements in teleoperation over long distances. As shown in Figure 2.1, networked haptic communication is an interdisciplinary field composed of three main domains:

- Robotics
- Networking
- Data processing

In this chapter, three areas within this domain are studied: *(i)* the communication network from the perspective of providing reliable (guaranteed) low-latency communications, *(ii)* intelligent data processing to compensate for the communication latency and for reducing bandwidth usage, and finally *(iii)* stability control schemes implemented at the teleoperation devices to reduce the impact of potential latency. Finally, dynamic network scheduling disciplines for improving fairness, in terms of latency, are presented in the final section of this chapter.

Nevertheless, with regard to the research objectives of this work, Figure 2.1 also illustrates the area of interest (dotted lines) within which the main components of the contributions as presented (in red) in Section 1.3. The author of this thesis is fully responsible for the work presented in this chapter, as seen in [9], except for content on joint control scheme and data reduction methods in which he partially contributed.

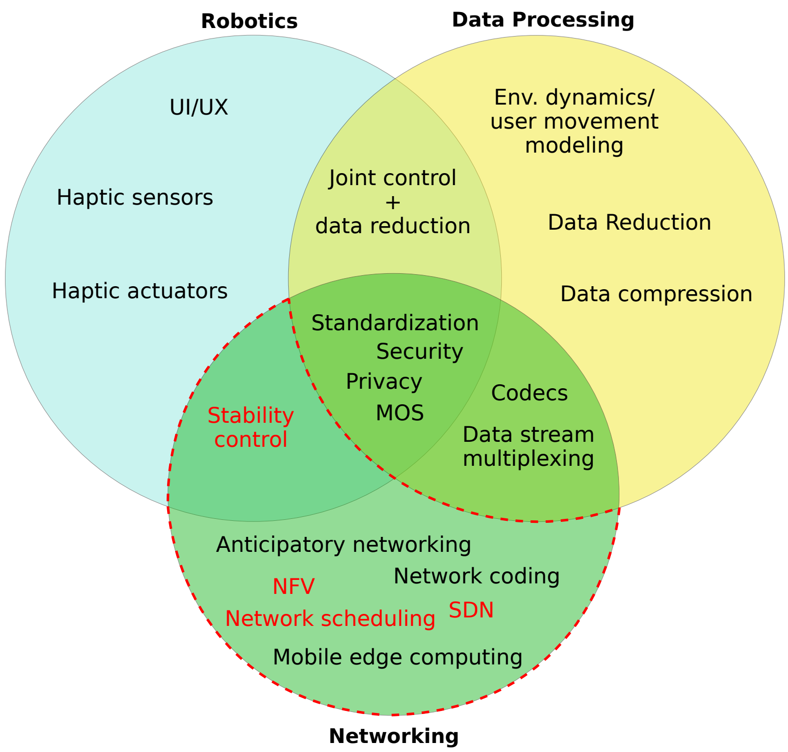


Figure 2.1: The main challenges in networked haptic communication.

2.2 Multi-modal telehaptic systems

Great part of ongoing research on the 5G of mobile networks is focused on meeting the requirements of the Tactile Internet [10]. A major design challenge here is to provide ultra-low delay communication over the network which would enable real-time interactions across wireless networks. This, in turn, will empower people to wirelessly control both real and virtual objects. It will undoubtedly add a new dimension to human-machine interaction and lead to an unprecedented revolution in almost every segment of society with applications and use cases like mobile augmented video content, road traffic/autonomous driving, healthcare, smart grid, remote education, and remote immersion/interaction among others [11].

One specific application domain of the TI is teleoperation which allows for remote immersion, including remote touch. Traditional remote interaction solutions such as voice or video conferencing, remote teaching, etc., have reached a high level of sophistication and widespread use thanks to the growth and progress of audio-visual communications.

A scenario that can be given as an example for the need to change traditional solutions for remote interaction is medicine. As of today, medicine relies very much on minimally invasive surgery, and concepts like laparoscopy or products like the da Vinci robot are well accepted in both public and private healthcare systems, and have proven to improve outcomes by reducing hospital stay, recovery time, pain, and post-operative impact. Specifically in the remote care context, the use of robots for surgery or consultation is particularly interesting, since it combines the benefits of minimally invasive interventions with the decentralisation of hospitals. The TI is a key enabler of remote operation or consultation, where the skills of the healthcare professional are captured, transmitted and reproduced in the remote end.

With the benefits of such technologies from robotics and telecommunications, users experience an improved virtual presence, immersing in remote environments. With current advances in the communication infrastructure, it has been foreseen that, in the near future, a complete remote immersion can be realised with the ability of physical interaction with the remote environment. This is achieved by the exchange of multi-modal information, such as the combination of audio, video and haptic information, over the Internet. Such immersion will be feasible for commercially acceptable use, with real-time applications such as teleoperation with haptic feedback (referred to as teleoperation) or haptic data broadcasting in virtual environments [12].

Haptics refer to both kinesthetic perception (information of forces, torques, position, velocity, etc. sensed by the muscles, joints, and tendons of the body) and tactile perception (information of surface texture, friction, etc. sensed by different types of mechanoreceptors in the skin) [13]. It must be noted that the previously mentioned term ”tactile” refers to its literal meaning, i.e., the human perception of touch. When used in the term TI, it signifies the feature of ultra-low delay communication over the Internet which is a necessity for many 5G use cases including haptic communication. As one of the applications of the TI, haptic communication using networked teleoperation systems has specific requirements, the most demanding being the efficient and timely exchange of kinesthetic or tactile information while synchronously providing the user with auditory and visual information.

Different from the communication of audio and video signals, haptic signals in bilateral teleoperation systems are bidirectionally exchanged over the network. It involves human users and closes a global control loop between the human users and the actuators/teleoperators. Thus, system stability and teleoperation quality are very sensitive to communication delay [4].

Use cases of TI, which highlight its importance, can be found in the medical, industrial, education and entertainment sectors. These include remote medical examination or surgery, industrial teleoperation in e.g., construction sites, mines or factories, tele-mentoring and gaming to name a few.

The benefits of the realisation of TI will revolutionise our way of living and increase the safety and efficiency of various tasks. Nonetheless, there are hindrances to be overcome and the previously mentioned requirements to be met.

Concepts and technologies around the IoT, 5G and the Tactile internet overlap each other, as indicated in [14], requiring very low latency and high reliability communication channels, high-bandwidth low-latency and secure infrastructure as well as bringing the intelligence of the network closer to the edge of the network.

As described in [15], one of the challenges in 5G mobile networks development is the provision of low-latency communications with acceptable QoE for the users. Since evaluating QoE in haptic-based applications with force feedback over the Internet is a process that has only recently taken its first steps, the way to resolve this open issue is still under investigation.

The delay requirements of haptic communication for networked teleoperation systems are heavily dependent on the application scenarios. Taking into account the latest achievements on haptic communication, as illustrated in Figure 2.2, the less dynamic the remote environment, the more the interaction between a user and the remote environment is increased. Consequently, different application scenarios arise in accordance with each level of dynamics and the corresponding range of time delay that is considered as acceptable for feasible interaction.

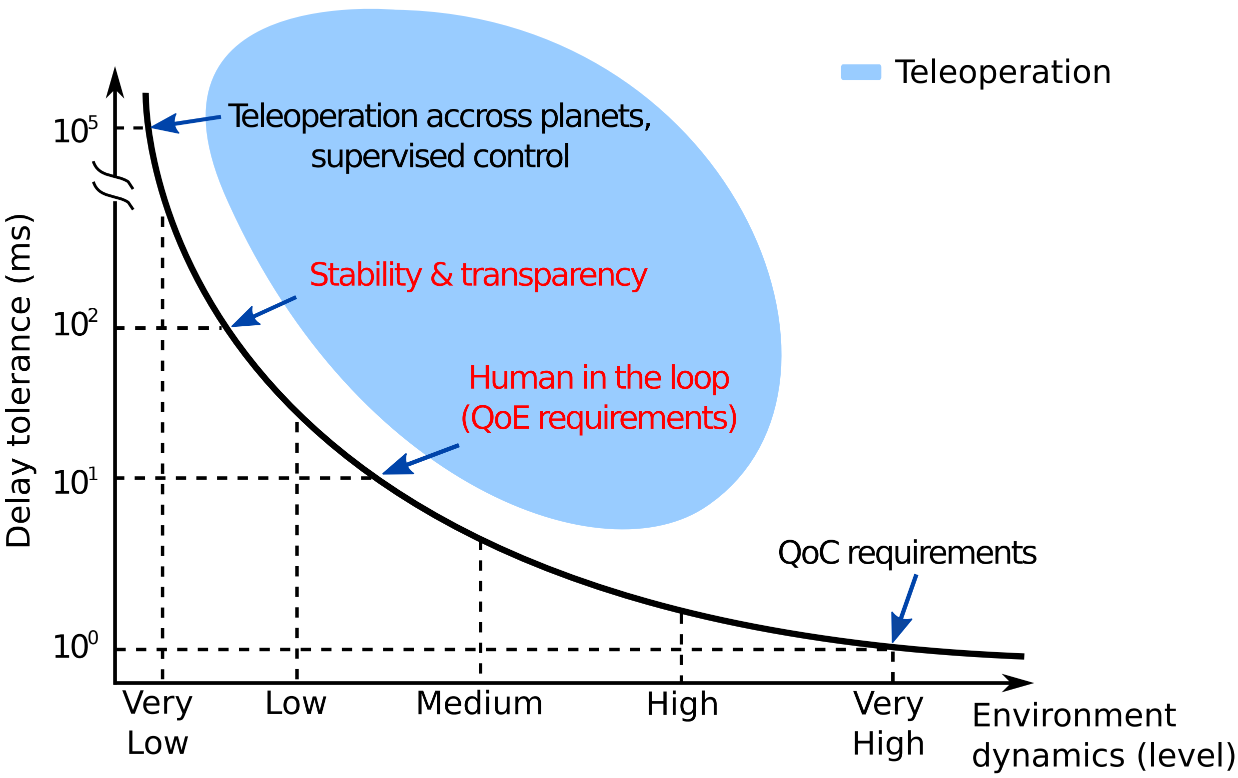


Figure 2.2: Delay requirements on different applications of immersive perception (from [9]).

Applications which can tolerate delays over $1ms$ are within the scope of teleoperation (the blue circle of Figure 2.2); a broad range of applications that can be divided into three categories of teleoperation, wherein each scenario is associated to a level of dynamics of the remote environment the user is interacting with and the corresponding delay tolerance. This leads to teleoperation applications with different degrees of immersive perception that range from space teleoperation to remote steering of automobiles, demonstrating different levels of abstraction between the user and the remote environment.

The case of highly dynamic environments, where a latency of under $1ms$ is needed, is out of the scope of teleoperation as only control systems can undertake the completion of tasks with such latency requirements because humans are underqualified for this kind of interaction. Specifically, for completing such tasks high Quality of Control (QoC) is needed, i.e., good performance of closed-loop control [16] of systems where no human user interacts with the loop. Examples would be a magnetic levitation system that keeps a running train floating in midair, a fully automatic driving system that precisely platoons vehicles and zips the vehicles through intersections without traffic lights, or a real-time simultaneous localisation and mapping (SLAM) with autonomous-controlled cameras [17]. As a result, these cases will not be examined in this chapter. On the other hand, this section focuses on remote environments of low and intermediate dynamics (red text in Figure 2.2). Within this range of dynamics there are a variety of applications such as remote surgery (low dynamics) or collaboration of users in virtual or real environments (intermediate dynamics).

2.2.1 Teleoperation systems

Multi-modal Telepresence and Teleaction (TPTA) systems for haptic telemanipulation, also known as *telehaptic* systems [18], usually consist of one human operator using a haptic interface (master device) on one end, a communication channel and one teleoperator (slave actuator) on the other end (Figure 2.3). For short distance applications the communication channel can be a direct wired or wireless communication channel without the need for network infrastructure. On the other hand, long distance applications benefit from packet-switched network infrastructures, and transmit their data as packets. The scale of long distance applications can range from teleoperation over local area networks to teleoperation over the Internet.

The goal of TPTA systems, as implied by their name, is to provide to the user the feeling of presence in the remote environment where the teleoperator exists. It is a goal which can be achieved due to the ongoing improvement of the relevant hardware and software for providing the human users with multi-modal (visual, auditory, and haptic) feedback.

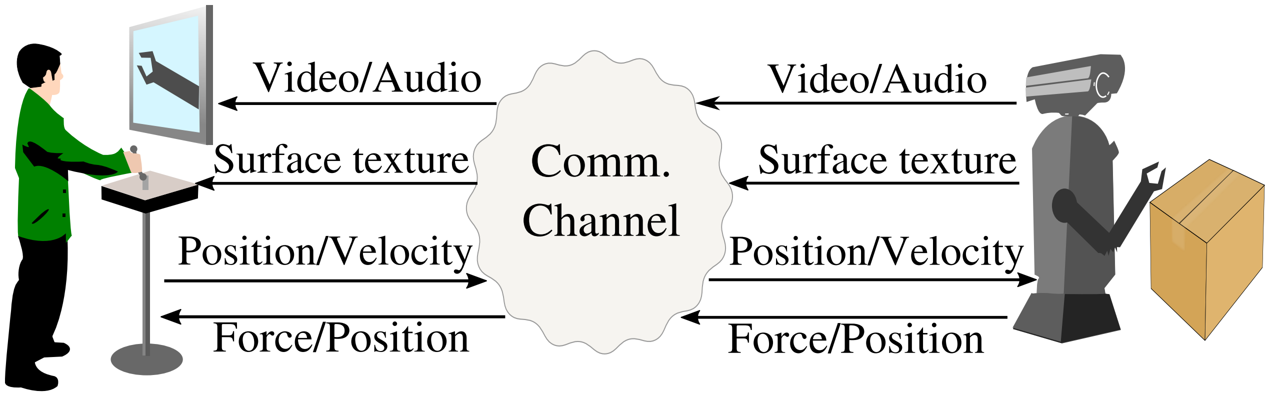


Figure 2.3: An example of a haptic communication system. In this case, the master device (user) sends position and/or velocity data while the slave device (robot) transmits the haptic feedback data, audio and video data streams.

2.2.2 Classification of teleoperation systems

Nowadays, extensive research has been done in bilateral and multilateral telehaptic systems [19, 20]. An approach for classifying teleoperation systems can be based on the different communication delays and interaction levels a user may experience and results in two main categories, *Direct control* systems and *Supervisory control* systems as described in [21]. As shown in Figure 2.4, we subdivide each of these categories further into subcategories:

1. *Direct control*: The human operator interacts in real-time with the environment while the master and slave devices communicate using position/force signals.

   (a) *Closed-loop with negligible delay*: In this case, the communication channel presents minimum delay and therefore the user is restricted to be in close proximity to the slave device.

   (b) *Time-delayed closed loop*: The most common form of teleoperation for digital closed-loop control systems. Similarly to the previous subcategory, the master device controls the slave actuator but the user is less restricted in terms of distance from the device (e.g. transatlantic teleoperation). The remote side is not autonomous, however, an internal control loop which processes the command signals from the master device is included in the teleoperator. In this case, the communication channel (e.g. the Internet), may introduce variable delays [4].

2. *Supervisory control*: The teleoperator is (a) autonomously or (b) semi-autonomously controlled and receives high-level commands from the master. It is also referred to as task-based teleoperation. Examples are teleoperation across planets or teleoperated robots with autonomous functionalities [22].

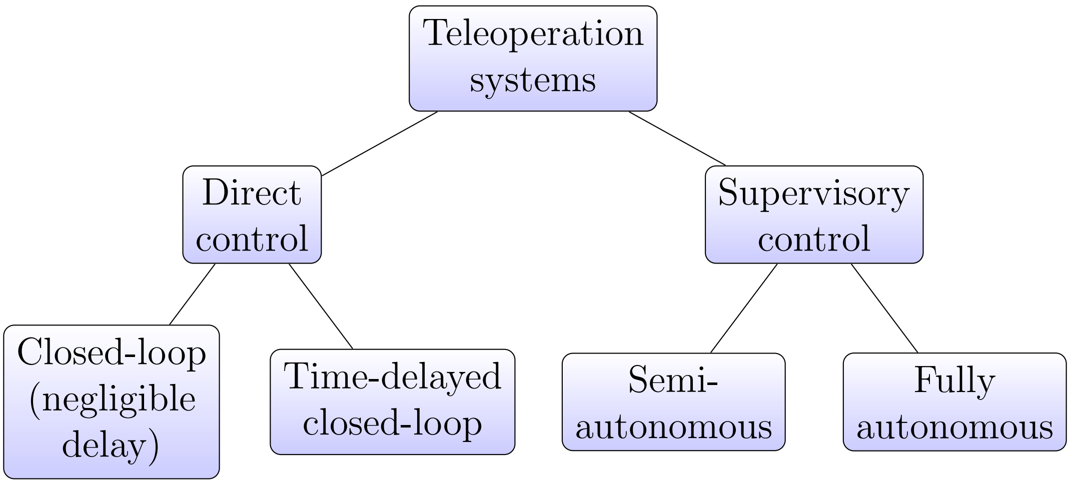


Figure 2.4: A classification of teleoperation systems

2.2.3 Master and slave subsystems

Typically, at the master subsystem of a haptic bilateral communication system, a human operator interacts with a haptic interface which uses sensors and transmits motion data (position or velocity data which are previously packetised) over a communication channel, to the slave subsystem. In return, the latter will respond with the force reflection/feedback of the remote environment, in the form of kinesthetic or vibrotactile force feedback data [20] while in some cases, such as in the concept of virtual fixture, position data may also be transmitted.

Haptic devices which are used as master teleoperation interfaces, also called haptic manipulators, consist of actuators and sensors which form the kinesthetic and tactile device subsystems.

Such haptic devices may be able to reproduce and process kinesthetic (kinesthetic interfaces), tactile (tactile interfaces) or both types of haptic data (haptic interfaces). Such devices have been created either as commercially available products or prototypes for academic research.

In [23], the authors discuss the topic of haptic devices and haptic actuators in relation to haptic communication over the Tactile Internet, making the important point that there is a need for ungrounded haptic devices with which the user does not need to stay in a specific area, contrary to the current state of haptic devices which are grounded. A list of hand-held kinesthetic devices as well as a performance evaluation was presented in [24]. As stated the most popular haptic interface is the Geomagic Touch (formerly known as Phantom Omni). These devices present specific technical characteristics such as the Degrees of Freedom (DoF) they support (either for sensing position or exerting force), the maximum force or torque they can output, the usable space they can operate in and their rotation capabilities (if their DoF specification allows them).

Many haptic interfaces, such as CyberGrasp [25] (an exoskeleton device), may also be entirely wearable or have wearable components in order to provide tactile feedback more effectively. It is possible to use more than one actuator for each finger. A variety of such interfaces are called tactile displays and make use of tactile actuator arrays using various technologies. Examples of such tactile devices are TPad [26], which is applied to the screen of mobile phones and Gloveone [27], a glove that provides tactile feedback to the fingers and palm.

The hardware design parameters of haptic devices (e.g., sampling frequency) and the number and type of sensors and actuators determine the amount of data the device will output or needs as input. They also determine the limitations of the interaction between a user or an object and the device. A recent detailed review of tactile sensors has been made in [28].

The slave haptic subsystem can be either a physical device which interacts with a physical remote environment or a virtual pointer of any form (e.g., a virtual hand) that operates in a virtual environment. A key difference between physical and virtual environments is that the control laws that govern a physical environment are of continuous nature whereas a virtual is of discrete nature. Virtual environments, even though it is not feasible to perfectly replicate a physical environment, have the advantage of allowing, in some cases, multiple users to interact with each other in a virtual space over a local network or the Internet. By employing the tactile or kinesthetic modalities these systems are called Collaborative-Haptic Virtual Environments (C-HAVE) [29].

A system diagram of a bilateral teleoperation system is shown in Figure 2.5 where a user is interacting with the master haptic manipulator using the master controller of the device. On the other side of the communication network the slave controller and manipulator are interacting with the environment.

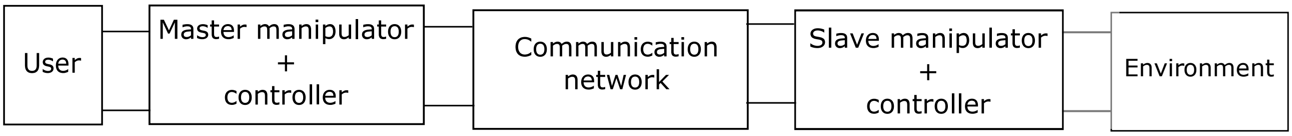


Figure 2.5: Bilateral teleoperation system diagram

2.2.4 Challenges of teleoperation systems

Communication of haptic information for teleoperation systems imposes strong demands on the communication network. This presents two main challenges for designing a reliable teleoperation system.

First, haptic sensor readings from kinesthetic devices are typically sampled, packetised and transmitted at a rate of 1 kHz or even higher [4, 30, 31] to maintain stability and transparency of the system (further discussed in 2.4.1). It must be noted that this is not a strict requirement, however, according to the stability analysis in [32, 33, 34] there is a relationship between the sampling rate, the maximum displayed stiffness and the system damping for ensuring system stability. A teleoperation system operating with lower values of sampling rate may still work and the user may be able to complete a task. Nonetheless, the maximum displayed stiffness, while guaranteeing system stability, is smaller than that of a higher sampling rate and therefore the system may require larger damping for stabilising a hard contact.

Communication of kinesthetic information for teleoperation systems, hence, requires a thousand or more haptic data packets per second to be transmitted between the master and the slave devices. Such a high packet rate may lead to the consumption of a large amount of network resources in combination with the transmission of audio and video data, and leads to inefficient data communication (see Section 2.3). Therefore, haptic data reduction, or packet rate reduction, is required in teleoperation systems. Moreover, tactile information, especially in the form of complicated texture surfaces, requires data compression.

Second, teleoperation systems are very sensitive to data loss and latency [4]. Concerning the latter, a haptic communication system device usually needs to transmit and receive a packet every millisecond, otherwise stability cannot be guaranteed. Consequently, an important question can be raised concerning the amount of latency compared to the amount of data loss that a system can tolerate. As it has been shown in [35], a 90% reduction can be attained, whereas, even a small amount of delay, can disrupt the stability of a bilateral teleoperation system. Even for a small communication delay or packet loss rate, teleoperation systems may show stability issues making degradation of teleoperation quality and task performance. With the introduction of a communication channel such as the Internet over mobile networks, this issue is inevitable.

Therefore, to guarantee system stability and improve QoE are key objectives of telemanipulation systems [36, 37]. QoE can be quantified by measuring the task completion time due to simplicity, but other performance measures are the Sum of Squared Forces (SOSF), peak forces, task error/failure rate, the haptic device trajectory, range of motion and velocity [38].

On the other hand, this also implies that the network infrastructure itself, if improved to the point of meeting all requirements, should be able to provide adequate resources and quality of communication for the best possible QoE and decrease the dependence on altering haptic information.

In addition, haptic communication systems usually need to provide to the user visual and audio feedback from the slave subsystem. High packet rate, packet loss and variable delay can cause the management and synchronisation of the data streams to become a challenging problem. In this case, packet-switched network frameworks and protocols are needed for synchronising the data streams [39], for measuring the network conditions and managing the Quality of Service

[40].

Summing up, we detail three main solution spaces to improve haptic communication:

- Due to the finite number of network resources, buffering and processing delays accumulate as the complexity of the end-to-end network increases. The *communication network* solution space targets to improve several aspects to either reduce the delay or the impact of delay.
- *Data processing* solutions to reduce data transmission using perceptual thresholds or prediction methods in order to compensate the incurred delay by long distance communications.
- *Stability control* solutions to reduce the effect of extra delay and provide stability for the control loop.

Improvements in all solution spaces of haptic communication are under development and research. Main contributions to these solution spaces will be presented in the next chapters. Individual or joint improvement of the communication channel, control components and signal processing will guarantee high teleoperation quality, system stability and scalability. While current research studies address mainly these solution spaces independently (few studies address two of these spaces jointly), the ultimate solution for enabling haptic communication should be based on joint optimisation of these three solution spaces. A discussion of future challenges on haptic communication over 5G exists in Section 5.4.

2.3 Haptic data reduction and compression

Haptic sensor readings, especially when reading the kinesthetic signals, as previously mentioned in Section 2.2.4, have a sampling rate of 1 kHz or even higher. In order to minimise the processing delay, haptic samples are packetised and transmitted instantly. As a result, the communication of haptic data in teleoperation systems requires 1000 or more packets per second to be transmitted. For vibrotactile signals (touch emulated with vibrations), the sampling frequency greatly depends on the type of interaction of the user with the remote environment. For tasks of low precision, it requires a feedback frequency of 20 Hz to 30 Hz, while high precision tasks require a feedback frequency of 5 kHz to 10 kHz [41, 42, 43]. Such high rate of packet transmission incurs substantial data overhead due to the transmission of packet header and, thus, results in increasing latency [44, 45, 46].

Future teleoperation systems which will be able to provide full body immersion will use a large number of sensors and actuators increasing proportionally to the number of Degrees-of-Freedom required by the haptic applications. As an example, the CyberGrasp exoskeleton system makes use of 22 sensors and 5 actuators found to transmit up to 324 kbps downlink and 415 kbps uplink (combining all 3 subsystems of the glove) [47]. This is only for acquiring the position and motion of 5 fingers and the hand (only one hand).

Even though 5G networks will have data rate capabilities which can easily cover the needs of a haptic data transmission of a user, it is necessary to have in mind that the standard update frequency of these devices can 1KHz as well as the additional transmission of audio and video data which complement the haptic data stream. Therefore, it is required to explore and improve haptic data reduction and compression methods, in order for future networks to be able to accommodate multiple haptic communication users. Understanding the characteristics of haptic data traffic is also important for considering accurate haptic data traffic models for the analysis of haptic communication systems.

Haptic data reduction techniques, either for kinesthetic data reduction or tactile data reduction, can be considered as lossy data reduction/compression schemes as full recovery of the original raw data is not possible. These techniques can be applied in the Application Layer since they rely on processing the data as acquired by the haptic devices. On the other hand, network throughput reduction can be achieved by other means such as Physical Layer Network Coding (NC) which will be discussed in Section 2.6 viewed from the scope of the 5G infrastructure. In this section, data reduction will be related to the processing of haptic data only.

2.3.1 Kinesthetic data reduction

Kinesthetic data reduction techniques are mainly based on two approaches of statistical and perceptual schemes [48]. The former one normally uses the statistics of the haptic signals to compress the packet size, while the latter one mainly focuses on reducing the packet rate over the communication network. Since the packet header overhead is significant as kinesthetic data packet payload size is small, reducing the frame rate seems to be an obvious choice for reducing the total amount of data.

Statistical schemes

Early attempts with respect to signal sampling employ predictive models to reduce data redundancy. Quantization techniques (e.g., Adaptive Differential Pulse-Code Modulation (ADPCM)) for kinesthetic data reduction are presented in [49]. In [50], kinesthetic data are 32-bit IEEE floating-point values. After the master and slave device have exchanged enough raw data, a simple position prediction method was proposed. Compression was achieved by performing an exclusive-or operation between the predicted and the previously predicted value and the result being reduced to 8 important bits.

Apart from prediction, lossy kinesthetic data compression and decompression has also been achieved by using Discrete Cosine Transform (DCT) [51], similarly to the JPEG codec, in a teleoperation system with force feedback with a compression ratio of 20%. Finally, another compression method that has been tested on 1-DoF haptic data is Wavelet Packet Transform (WPT) [52]. In this case, decompression is accomplished with the Inverse WPT (or IWPT).

Perceptual schemes

The first proposal that targets packet rate reduction for networked control systems can be found in [44]. In this work only samples that contain changes more than a given/fixed threshold are transmitted. The receiver reacts to a missing sample by holding the value of the most recently received sample. The approach in [44], however, ignores that the human operator comes with strong limitations in terms of perceivable signal changes.

State-of-the-art methods of perceptual data reduction have shown that it is possible to exploit the limitations of human operators and how they perceive haptic signals [53] towards achieving more efficient data reduction [54]. Such works mainly rely on a concept from psychophysics that is the difference threshold, otherwise known as Just Noticable Difference (JND), which is the minimum amount of change in stimulus intensity needed for a perceptible increment in sensory experience. This threshold is formulated by Weber's law [55]:

$$\frac{\Delta I}{I} = c \tag{2.1}$$

where I is the stimulation intensity, ΔI is the difference of stimulation intensity to be perceived (the JND) and c is a constant, also known as the Weber's fraction. Difference thresholds, also known as discrimination thresholds, are defined both for haptic system parameters and quantities such as stiffness, velocity and force. These thresholds also differ depending on the movement scenario and the muscles involved [56]. For example, the JND when a human operator perceives force feedback to the index finger is approximately 10% [57]. It must be noted that the JND results for human perception are statistical and evaluated based on psychophysical observations of human users. Therefore, JND results are affected by various factors that may influence each user such as motivation or different perception abilities [58]. Furthermore, a reinterpretation of the JND model has been made in [59] presenting a Bayesian approach aiming to improve the modelling of magnitude estimation.

Implementing kinesthetic data reduction showed up to 90% decrease in packet rate in [45]. Perceptual kinesthetic data reduction schemes have also been implemented for position and velocity signals using distance metrics (the Euclidean distance) between haptic data vectors (position vectors) [60]. This approach, however, needs further investigation because in psychophysics there is no result that shows Weber's law also applies to positions. This methodology also applies to orientation data and has also been extended to six DoF.

In comparison to other sampling methods such as the level crossings method (that incorporates absolute differences instead of percentages between samples), the perceptual-based kinesthetic data reduction schemes are proven to have good but similar accuracy [61]. Nonetheless, in [62] it is stated that the *level crossings sampler* outperforms the sampling method based on Weber's law. It has also been shown that the JND decreases with increase of the rate of kinesthetic force stimuli.

Perceptual schemes with predictive coding

Prediction models of haptic signals can be used to estimate future haptic samples from previous data. This is able to achieve further reduction of haptic packet rate. As illustrated in Figure 2.6, the same predictors can run in parallel at both the master and slave sides. At the sender side, the predictor generates the predicted haptic signal at every sample instant. If the prediction error is smaller than the corresponding JND, no update is triggered. Otherwise, the input sample is transmitted to the other side and the transmitted sample is used for updating the prediction model. At the receiver side, if a packet is received, it is directly applied as the output and the received haptic signal is used for updating the prediction model. Otherwise, the predictor generates a predicted haptic signal as the current output.

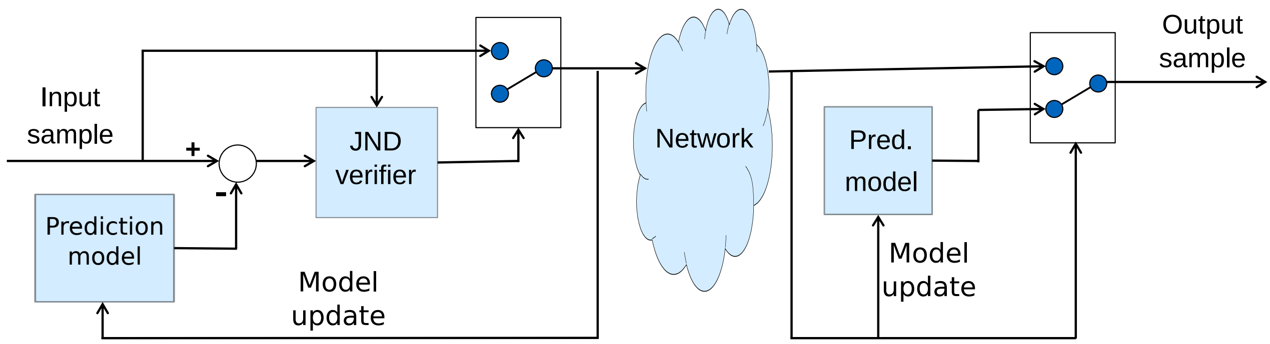


Figure 2.6: Overview of the perceptual kinesthetic data reduction with predictive coding

The simplest but also the least efficient prediction method is the Zero order hold (ZOH) predictor. When no data has been transmitted from the sender the receiver holds the last value of the sample it previously received.

Different kinds of predictors can be used to estimate the future haptic samples. For example in [63, 46], a linear predictor of the first order was adopted, namely a First-order linear predictor (FOLP). This simple predictor can lead to a significantly decreased packet rate up to 90-95% without deteriorating the immersiveness of the system. The velocity signal approximation used in the prediction model, however, is very sensitive to noise, even more than force signals. Therefore, haptic samples need to be filtered by using a low-pass filter to minimise the undesirable effects of measurement noise. An augmented version of this framework, presented in [64], introduces noise reduction by employing a scalar Kalman filter on the input signals.

Yet another prediction model that takes the prediction error into account was employed in [65] for three-dimensional position and force data, a third-order autoregressive model. According to the method's algorithm, after an initialisation and training process, the adaptive coefficients of the model are computed so that the predicted values are produced. Afterwards, taking into account the JND threshold, the algorithm decides whether the training values need to be updated either from the predicted data or the current real data.

Contrary to transmitting sample values obtained from haptic devices or their derivatives, regression analysis also allows the transmission of only the model parameters. Such a method can be found in [66] where samples are first fitted in a quadratic curve.

Using a more complex predictor, a geometry-based prediction model was proposed in [67]. The remote environment is modelled either as a plane or a sphere according to the historical interaction, without taking friction, slave inertia and time delay of the network into account. The future haptic samples were predicted based on the interaction with the geometry model. Taking into account a total of four predictors (ZOH, FOLP, plane and sphere predictors), a predictor selection method is also proposed in [67] which identifies the predictor with the least prediction error. Psychophysical tests on human subjects showed that the hybrid approach performed as well as the best predictor (sphere predictor).

Prediction of signal samples and perceptual coding of the signal was also proposed in [68] by implementing the Particle Filtering method. This framework uses the probability distribution function (PDF) of the user's motion or force to predict future or lost samples.

Even though the purpose of the aforementioned data reduction schemes is to reduce the data rate of the haptic components in the teleoperation system, there has been investigation on the impact of data rate reduction on the performance of the haptic communication system. An indicative example can be seen in [69] where transparency (a main teleoperation performance indicator explained in detail later in the following section) is evaluated against data rate.

2.3.2 Tactile data reduction

While data reduction on kinesthetic signals is widely investigated as discussed in the previous subsections, the number of studies on the compression of vibrotactile texture signals is limited. The kinesthetic signals involve large amplitude low-frequency force feedback and were found lacking in realism due to the absence of high-frequency transients (e.g., tapping on hard surfaces [70]) and small-scale surface details (e.g., palpation of textured surfaces [71]). Transmission of vibrotactile signals for increasing fidelity of real-time teleoperation systems and its storage for later playback necessitate data compression.

The necessary step before compressing the vibrotactile signals is to model them. Significant research has been devoted toward modelling tactile texture signals [72, 73, 74]. The vibrotactile signals are raised from coarse regularly patterned textures by decaying sinusoids [72]. In [73], Kuchenbecker et al. use a linear predictor to model the texture signals. In [74], the authors segment the recorded real-world vibrotactile texture signals based on their physical surface feature. These segmented signals are fitted and the corresponding filter parameters are stored. Then, a virtual visual-haptic model representing the previously extracted surface features is constructed and haptic rendering is performed based on this model. The above works, however, are not optimised for compression.

Towards the compression of vibrotactile signals, Okamoto and Yamada presented a frequency-domain texture compression algorithm loosely based on the knowledge of human vibrotactile perception [75]. The textured surfaces are scanned and the surface height is represented by a waveform. This waveform is transformed to the temporal frequency domain using DCT, and the DCT coefficients are thresholded and quantized according to the knowledge of frequency-domain amplitude-JND for vibrotactile stimuli [76]. The authors of [75] showed a 75% compression of the texture data with guaranteed perceptual transparency. Unfortunately, this algorithm works only offline, which means prior knowledge about the surface must be known (e.g., pre-scanning procedure). Further research on JNDs of vibrotactile perception has been made in [77] by studying the JNDs with low-intensity reference stimuli, starting at 5 Hz, close to the sensory absolute threshold. In [78] three experiments were carried out, first an experiment showing that acceleration is not a vibration property that affects humans due to the nature of the human tissue, second an experiment to determine JNDs showing it is unimportant to subjectively optimise tactile displays and third an experiment that showed the impact of using different devices in low frequency vibrations (starting at 100 Hz).

The first online compression of vibrotactile signals can be found in [79] for bilateral teleoperation. The compression algorithm is inspired by the similarities observed between texture signals and speech signals. Thus, a well-developed speech coding technique, the Algebraic Code-Excited Linear Prediction coding (ACE-LPC) [80], is adapted for developing a perceptually transparent texture codec. The authors of [79] reported a compression rate of 8:1 with a very low bit rate (4 kbps) on data transmission. An extended version of this compress algorithm was proposed in [81], in which the masking phenomenon in the perception of wide-band vibrotactile signals was applied to further improve the efficiency of the texture codec. The masking phenomenon [82] implies that humans can tolerate larger errors in high-energy frequency bands, and smaller ones in low-energy frequency bands. Therefore, for encoding (compressing) the texture signals, the bit rate should be allocated more in the low-energy frequency bands compared to the high-energy ones. In [81], the authors experimentally showed that the masking for haptics is very similar to its auditory equivalent. With the help of the experimental results, the bit rate of the codec output can be driven down to as low as 2.3 kbps without distorting the subjective perception.

Last but not least, it must be noted that there is currently no objective quality metric, such as Mean Opinion Score (MOS), for the evaluation of vibrotactile signals (with the exception of [83] on the effect of delayed kinesthetic and 3D video data on the user). Nonetheless, the similarities shared between audio and tactile signals will allow the design of tactile codecs in a similar fashion as with audio codecs [84].

2.4 Haptic control system approaches

Even though the network infrastructure and mechanisms are improving, there are physical barriers, as in the case of long distance communications, that can introduce a minimum latency which can make certain teleoperation applications impossible. As previously mentioned, latency can disrupt the stability of a bilateral teleoperation system. Although this is true, there are stability control architectures and methods that can minimise the impact of latency. Therefore TPTA systems will not solely rely on 5G network infrastructure for optimising the QoS of the communication channel to the standards of each application, but will be able to compensate for delay to a certain extent.

The foundation of teleoperation system control analysis is a model that best characterises the interaction between the human and the remote environment. This model is usually in the form of a mass-spring-damper system which portrays the behaviour of the master and slave subsystems. A common way to describe this behaviour is the Euler-Lagrange equations of motion for the joint-space nonlinear dynamic model of an m-DoF master and slave device [85].

In this section, the focus is on robust stability control methods and concepts that offer the possibility of jointly using data reduction along with their background (subsections 2.4.1-2.4.2, 2.4.3). Methods for closed loop teleoperation implementations that model time delay, system plants and the robot devices in order to compensate for any delays (subsection 2.4.2) are mentioned together with other control methods for the sake of completeness (subsection 2.4.2). Furthermore, an analysis of control performance is presented in this section.

It must be mentioned that there are numerous stability control schemes which could be mentioned in this section. However, in this chapter the aim is not to summarise all existing control schemes, but to survey the work that can capture several important aspects of haptic communication such as:

- The impact of delay on each stability control category
- The significance of choosing among these stability control categories in relation to critical performance criteria.
- The ability to jointly address stability control and haptic data reduction

2.4.1 Control performance

The different control architectures that permit signals to be exchanged between a human operator and the remote environment can be classified according to the arrangement of the control system building blocks. In order for the system to meet the objectives for teleoperation of acceptable quality, adaptive control subsystems can be introduced in the teleoperation system design. This results in a wide range of different mechanisms and architectures that attempt to tackle the issues of telemanipulation [86]. A comparison of different control schemes was presented in [87] stating that all schemes have advantages and disadvantages and that it is at the discretion of the system designer to choose which is the best one for the intended application. This plays a key role in designing the end-to-end haptic communication system as well as how the underlying network infrastructure will need to react according to the different requirements imposed by the different haptic applications.

Bilateral control teleoperation system classification can be based on whether the system targets to compensate for communication delay, focuses on estimating the operator and environment model, is responsible for handling internal and external disturbances of the subsystems, or, provides a combination of the aforementioned tasks. Another approach for classifying teleoperation systems states that the information processed in the system for controller gain adaptation is focused on the *environment*, the human *operator* or the *task* to be accomplished, therefore calls these controllers EOT-adapted controllers [88].

The two most common generic control architectures, based on the number of communication channels the system uses, are the two-channel (2CH) and the four-channel (4CH) architectures. In the former one, the master and slave manipulators need to establish only one channel for each direction of the bilateral communication, whereas in the latter one, both velocity and force information is exchanged by using two different channels for each direction. In the absence of time delays 4CH outperforms 2CH. On the other hand, in the presence of delay, system performance, stability and transparency are negatively affected. There are also other possible schemes with one human operator and multiple slave devices [89] or multiple human operators and multiple slave devices [90]. A human operator may also communicate with a virtual environment, instead of a haptic device, where computational delay must also be taken into account (no delay, constant and time-varying delay) [91, 92].

Transparency and stability

Transparency and stability are key aspects of a haptic teleoperation system and also the main focus of system control techniques. A fully transparent system is a system in which telepresence is a flawless and seamless experience. To achieve transparency, the system also needs to be stable for an expected (bounded) behaviour of the operator and the remote environment. In practice, there is a conflict between transparency and stability and a compromise needs to be made [4].

A stable system must always have bounded output for a bounded input. Bounded signals are those which do not exceed a finite value over time. Transparency of a bilateral control system can be defined in many ways, the most popular being the mechanical impedance approach. In this approach, maximum transparency is achieved when the impedance of the operator is the same as the impedance of the environment [93], known as impedance matching.

Based on linear time-invariant (LTI) dynamics in the Laplace domain, maximum transparency is achieved when the human and environment impedances are matched. The human impedance is defined by the ratio of force applied by the teleoperator to the velocity of the master device, and the environment impedance is the ratio of the force slave device receives from the environment to the velocity of the slave device.

In [94], the notions of reproducibility and operationality, which complement stability, are investigated as two goals that when achieved the condition for transparency is satisfied. Reproducibility is referring to the reproduction of the environmental impedance from the master manipulator, whereas to achieve ideal operationality the operational force (additional undesirable force produced by the system controllers due to inaccuracies) should not be felt by the human operator and therefore must be zero.

Furthermore, theoretically, any non-zero value of delay leads to instability. However, the damping of the haptic device, slave dynamics and human arm movement contribute to the stabilisation of the system. As a result, there is some tolerance, which varies for different system settings and teleoperation tasks [95].

2.4.2 Analysis of stability and control performance

The first and foremost criterion for any controller is stability, a necessary requirement for any control scheme, which, in teleoperation, is highly dependent on the communication delay [96]. A stable system can be Intrinsically Stable (IS), i.e. stable for all values of communication delay (d) and control parameters, or can be PS, i.e. stable for all delay values under specific control parameters and stable for delay values less than a maximum value, i.e. d_{max}, for other control parameters [87].

Performance of control schemes in terms of *transparency* and *tracking precision* can be captured through the following metrics [87, 97]:

- *Inertia (M_p) and damping (B_p)* perceived by the master when no force is applied to the slave manipulator by the environment; M_p and B_p affect transparency.
- *Tracking error (δ)* of the master manipulator position versus the slave manipulator when no force is applied by the environment to the slave manipulator; δ affects tracking precision.
- *Stiffness (K_p)* perceived by the master when force is exerted to the slave manipulator by the environment; K_p also affects transparency.

- *Position Drift (Δ)* when a force is exerted to the slave manipulator by the environment, which impacts tracking precision.

As mentioned above, the first three metrics (M_p, B_p and δ) can be considered during movement without interaction with the remote environment, whereas the latter two parameters play a role while interacting with the remote environment.

The M_p, B_p, and K_p affect transparency of the teleoperation system; in other words quality of telepresence of the operator (master) at the remote (slave) side. For perfect telepresence, M_p, and B_p should be equal to the remote environment inertia and damping (assuming no environmental force), and K_p should be the same as the environment's stiffness. On the other hand, δ and Δ, affect the tracking precision and represent precision of the teleoperation system. For a high precision teleoperation system, tracking error δ and position drift Δ should be negligible [4].

To better understand how these parameters are formulated in [87], in the following subsection, an indicative example for the Force Reflection (FR) control scheme will be presented. More specifically, in this example, an analysis of M_p and B_p will be provided. Formulations regarding other control schemes and their performance metrics, can be made following the same process for their respective transfer functions which are constructed according to their respective human, environment and control system dynamics.

For the overall analysis presented in this thesis, regarding teleoperation system dynamics, the master and slave manipulators considered have identical mass-damper dynamics defined as :

$$\begin{aligned} F_m &= v_m s v_m + B_m v_m \\ F_s &= v_s s v_s + B_s v_s \ , \end{aligned} \tag{2.2}$$

where s is the Laplace transform variable, v_m, B_m, v_s and B_s are mass and viscous damping coefficients of master and slave manipulator respectively. Given the identical assumption above, $v_m = v_s$ and $B_m = B_s$. In the case of force exertion from the environment on the slave manipulator, the environment is modelled as a spring-damper, as described in Equation (2.3).

$$F_e = -(B_e + \frac{K_e}{s})v_s \ , \tag{2.3}$$

where B_e is the environment's viscous damping coefficient and K_e is the stiffness of the remote environment. By capturing the environment force, the slave device transmits relevant information (e.g., force, position or velocity) to the master device which needs to accurately represent it to the user in order to perceive the environment's stiffness.

Furthermore, the force applied on the slave and master manipulators is:

$$\begin{aligned} F_m &= F_h - F_{mc} \\ F_s &= F_e + F_{sc} \ , \end{aligned} \tag{2.4}$$

where F_mc and F_sc is the forces exerted by the master and slave controllers respectively.

Force Reflection & Position Error control schemes

The FR also known as kinesthetic force feedback scheme uses the position-force architecture. In this architecture, position is sent from master to slave and force is sent back from slave to master to have a sense of the slave environment and for the user to adapt the position of master manipulator based on the force feedback from slave manipulator. This scheme is one of the simplest control schemes for teleoperation and is explained in a number of research works [98, 99, 100].

In the FR, the "master controller" entity in Figure 2.5 comprises of the delayed version of F_{sc} (reflecting the communication delay) multiplied by a control parameter (that is a control gain). The "slave controller" entity is shaped by another control parameter (again a control gain) multiplied by the position error, i.e. the difference between delayed position of master manipulator (x_{mt}) and the position of slave manipulator (x_s). Therefore, a human operator can feel the time-delayed force feedback (F_{st}) from the slave manipulator (generated by the environment) and intuitively adapt the applied human force (F_h) in order to achieve the desired position. However, the force feedback might destabilise the system even in the presence of small communication delays, since the delayed feedback might mislead the human operator and thus the slave controller. Nevertheless, having force feedback is beneficial in providing telepresence.

The FR scheme is PS, i.e. choosing the slave controller gain sufficiently low will keep the teleoperation system stable for a wide range of delay values in the expense of transparency and tracking precision [97]. On the other hand, no knowledge of delay or the environment is required. In terms of sensitivity to delay, perceived inertia and tracking error are linearly proportional to d, and position drift, perceived damping and perceived stiffness are independent of d.

The Position Error (PE) scheme uses the position-position architecture in which position is sent from master to slave and vice versa [101]. Hence, inside the master and slave controller blocks, as seen in Figure 2.5, the difference of positions between master and slave manipulators is multiplied by a control parameter (gain). However, knowing the exact position of master and slave at the same time is an ideal case. Due to communication time delay, only the difference between master position (x_m) and *time-delayed* slave position (x_{st}) is available in the master side. The same applies to the slave side i.e. x_s and time-delayed x_{mt} are available. Therefore, large communication delay will destabilise the system and also will adversely affect transparency and performance.

It has been shown in the literature that position-position architecture, as used in PE scheme, offers lower transparency compared to position-force [102]. While stability improves in PE compared to FR, PE is still possibly stable [97]. In addition, in PE scheme, inertia perceived in the master side is proportional to the transmission delay d and its square, while damping perceived is linearly proportional to d. Perceived stiffness, position drift and tracking error are independent of d. This method neither requires knowledge of d nor the remote environment model.

At this point, it is of interest to present how it is possible to infer the formulas of the performance metrics, taking as an indicative example the FR control scheme. Specifically for the case of

FR, we can elaborate more on F_{mc} and F_{sc} with the following formulas:

$$
\begin{aligned}
F_{mc} = G_c F_{st} \ , \ F_{st} = F_{sc} e^{-sd} \ , \\
F_{st} = F_{sc} e^{-sd} \ , \ x_{mt} = x_m e^{-sd} \ .
\end{aligned}
\tag{2.5}
$$

where G_c is aBy using equations for F_m in (2.2) and (2.4), we can write:

$$
(M_m s^2 + B_m s) x_m = F_h - F_{mc}. \tag{2.6}
$$

Then we can replace F_{mc} from the equations in (2.5):

$$
(M_m s^2 + B_m s) x_m = F_h - G_c F_{st} \ , \tag{2.7}
$$

which, using F_{st}, F_{sc} and x_{mt} according to the equations in (2.5), leads to:

$$
(M_m s^2 + B_m s + G_c K_c e^{-2sd}) x_m - G_c K_c x_s e^{-sd} - F_h = 0 \ , \tag{2.8}
$$

For the slave, using the slave dynamics equations we can formulate the following:

$$
F_s = F_e + K_c (x_{mt} - x_s) \tag{2.9}
$$

where K_c is an important control parameter that determines the delay tolerance of the system in terms of stability. For high values of K_c, transparency of the system increases, but delay tolerance decreases. Using a similar analysis as with the master we get:

$$
(M_s s^2 + B_s s + K_c) x_s - K_c x_m e^{-sd} - F_e = 0 \ , \tag{2.10}
$$

To facilitate the analysis for the performance metrics more easily, we will use equations (2.8) and (2.10) that correspond to a system of two equations with two unknowns, to formulate the following variables A and B:

$$
C = M_m s^2 + B_m s + G_c K_c e^{-2sd} D \quad = M_m s^2 + B_s s + K_c \tag{2.11}
$$

By using the aforementioned system of equations and substituting with variables A and B, we get:

$$CDx_m - G_cK_c^2e^{-2sd}x_m - DF_h = 0 \tag{2.12}$$

Now we have reached the point where we have all the tools for calculating the performance metrics. Specifically, according to [87], inertia M_p and damping B_p can be obtained using the following transfer function (using C and D in 2.11):

$$G_1(s) = \left(\frac{x_m}{F_h}\Big|_{F_e=0}\right)^{-1} = \frac{CDx_m - G_cK_c^2e^{-2sd}}{D} \tag{2.13}$$

In order to linearise the transfer function, the delay term e^{-2sd} can be approximated as a Taylor series. In this case, the series is expanded up to the fourth order term because terms of higher order do not affect the perception of inertia or damping from the user (according to the order of the polynomials of the system dynamics):

$$\begin{aligned} e^{-2sd} &= 1 + \frac{1}{1!}(-2sd) + \frac{1}{2!}(-2sd)^2 + \frac{1}{3!}(-2sd)^3 + \frac{1}{4!}(-2sd)^4 \\ &= 1 - 2sd + 2s^2d^2 - \frac{4}{3}s^3d^3 + \frac{3}{2}s^4d^4 \end{aligned} \tag{2.14}$$

In order to infer M_p and B_p one needs to consider that $G_1(s) = M_ps^2 + B_ps^2 + G_1^*(s)$ where $G_1^*(s)$ are negligible terms of third or higher order [87]. Therefore, after replacements and relevant operations it is possible to infer:

$$\begin{aligned} M_p &= (1 + G_c)M_m - G_cBm(\frac{B_m}{K_c} + 2T) \\ B_p &= (1 + G_c)B_m \end{aligned} \tag{2.15}$$

Passivity-based control

Built upon the idea that bilateral control systems must be passive and therefore stable by Anderson and Spong [103], passivity-based control methods have been applied to haptic communication systems in order to compensate for time delays or data loss. Due to its effectiveness in non-linear control systems it has been thoroughly studied in teleoperation systems.

Nonetheless, the passivity condition applies only if all the system components are or are assumed to be passive (i.e., subsystems that do not produce energy), as any arrangement of passive components results in a passive system. With regard to teleoperation systems, the previous statement also applies to teleoperation systems assuming that the human operator and the remote environment behave as passive elements along with the existence of an ideal communication channel. Concerning the human operator, this assumption is only valid for the sake of simplicity, otherwise it does not hold for all kinds of tasks as stated in [104].

In general, a teleoperation system can be modelled in various ways, such as the two-port network model [105, 98] or the port-Hamiltonian system approach [106]. Focusing on the two-port network model, all subsystems between the human operator and the environment can be represented by a two-port network where energy flows through its inputs and outputs. From an electrical domain point-of-view, this can also be viewed as a transmission line system that ideally is needed to be lossless (with perfect impedance matching). In this domain, force is represented as voltage, position as current and therefore the product of the two is power. A two-port element inside a teleoperation system can be characterised as passive when the energy (integral of power over time) of the output of the two-port element is greater than the energy of the input [107].

An alternative analysis in [108] investigates bilateral control system stability with a non-passive human operator or teleoperation environment using Mobius transformations.

Since the communication network of a TPTA framework introduces delays, which can be represented in a control architecture by active elements, it is were passivity-based control needs to be applied. A detailed description of the theoretical base of passivity-based control has been made in [109]. Wave Variable (WV) control and Time-Domain Passivity Approach (TDPA) are such methods and will be discussed in the next subsections. Furthermore, augmented versions of these methods have also been proposed.

Passivity-based approaches have been proposed both for linear and non-linear teleoperation systems, i.e., systems that use control models for linear dynamics or non-linear dynamics with respect to the operator and teleoperator devices as well as the environment and human user interaction [110]. Alternative teleoperation control methods to the passivity-based approach are Proportional Derivative (PD) or PD-like control [111], PD control for stochastic stabilisation [112], for adaptive time-delay compensation [113] or without, $\mathcal{H}_\infty$-control and μ-synthesis [114, 115], the computationally complex but constantly improved Model Predictive Control (MPC) [116] methods, fuzzy logic system approaches [117, 118] or the more recently proposed immersion and invariance (I&I) observer methods [119]. Passivity-based control methods can be applied on each DoF of a teleoperation system, but the system becomes more conservative [120].

Acceleration-based bilateral control methods have also been proved to provide robust stability even when the system is under time delay both for two-channel [121] and four-channel architectures [122, 123].

Network delay, especially when it is considered time-varying, becomes a hindrance for the synchronisation of master and slave positioning and the transmission of the human operator's movement trajectory or the remote environment's force feedback. For each teleoperation framework, several augmented versions that attempt to optimally solve the position tracking issue have been proposed such as the sliding-mode controller architecture [124].

Using the representation of the two-port network model for teleoperation in the frequency domain and Llewellyn's absolute stability criterion [125], it is possible to define the scattering approach which examines how scattered waves (output of the communication network) differ to their original form (input of the communication network) [98].

By creating abstraction layers for transparency and passivity, a haptic system that transmits mixed feedback of kinesthetic and tactile information was described in [126], also providing additional tactile force feedback when the passivity layer disrupts the kinesthetic force feedback in order to preserve passivity.

Wave-variable control methods

The previously mentioned work of Anderson and Spong which combined scattering transformation, network theory and passivity control, led to the concept of wave-variables (i.e. wave-variable transformation) by Niemeyer and Slotine [127, 128], used in haptic communication systems by algorithms created to ensure stability and transparency between the master and slave device when time delay is introduced [129]. Viewing the system from a virtual transmission line point-of-view the wave-variables represent the incident and reflected waves respectively and the wave (or virtual) impedance can be used to control the behaviour of the system to preserve passivity.

A quantitative comparison of the performance between the two-channel and the four-channel wave-based control schemes revealed that the four channels of the 4CH architecture can be reduced to three and also achieve better performance than the 2CH architecture. Even so, the 2CH scheme is able to achieve similar performance with better stability robustness, while being less complex to implement [130]. Wave-based bilateral control has also been applied to micro-teleoperation systems in which the slave device operates on soft/fragile objects [131]. Furthermore, wave variables can also be used in multiple-DoF teleoperation systems by adopting more general equations that incorporate impedance matrices, also called scaling matrices [132]. The scattering transformation also allows the power transmitted from one side of the teleoperation system to the other to be scaled, a characteristic of a passive two-port system.

With the adoption of the scattering transformation, converting power variables to wave variables raises important issues. On one hand, power variables preserve passivity, on the other hand, they also introduce desynchronisation and the phenomenon of wave reflection which disrupts the system transparency. The position tracking error, also known as position drift, and the force tracking error between master and slave is caused because of the time delay introduced by the system's communication channel and the fact that it is impossible to perfectly model the environment in which the slave device is operating. To resolve this issue, several attempts have been made either for constant time delay [133] or varied time delay [134] in the communication channel. An augmented version of the wave-based control architecture is recommended in

[135]. Other methods propose several techniques and schemes such us the transmission of wave integrals [136] along with wave energy [128], predictors [137] to compensate for network delays or even communication blackouts [138] or the utilisation of neural network theory for enabling improved modelling of the system which is considered as nonlinear [139]. In [140], Munir and Book proposed a method that corrects the position tracking error taking time-varying delay into account. This method employs a modified Smith predictor, a Kalman filter and an energy regulator. An improved version of this method was suggested in [141]. Several alternatives have been proposed as well in the scope of wave-variable control [142, 143, 144, 145] .

Time-domain passivity control

TDPA was defined by Hannaford and Ryu [146] for haptic interfaces and extended to apply to teleoperation systems [147]. The approach has gained interest during the past few years due to its simplicity and robustness to communication delays.

As illustrated by Figure 2.7, the basic concept behind time-domain passivity control is to monitor the energy flowing to and from the master, the slave side or both in real time using a Passivity Observer (PO) which can be placed in series or in parallel to the communication channel. In the series arrangement, velocity is chosen as an input whereas, in the parallel arrangement, force is used as input to the PO. If the PO decides that the passivity condition is not satisfied, meaning that the system generates energy and therefore is active, then, a Passivity Controller (PCO) has the responsibility to retain the system's passivity by using adjustable damping elements [148]. Besides being applied to 1-DoF applications, TDPA has also been applied to 6-DoF systems [149].

Following a relevant arrangement in [150], after the acquisition of the environment parameters (related to velocity and force data of the slave device) and transmission through the communication channel, a model of the environment is created on the master side and according to this model the damping coefficients of a PCO are adjusted according to a PO's output.

Bounding energy signals [151] or control signals [152] of TDPA systems has shown improvement of the method's effectiveness. In [153], a different scheme, as in Figure 2.7, is proposed where the segment of the control system on the teleoperator side including the communication block is modelled as a two-port network that receives position and provides force feedback. A method that combines time domain passivity control with perceptual data reduction is introduced in [154].

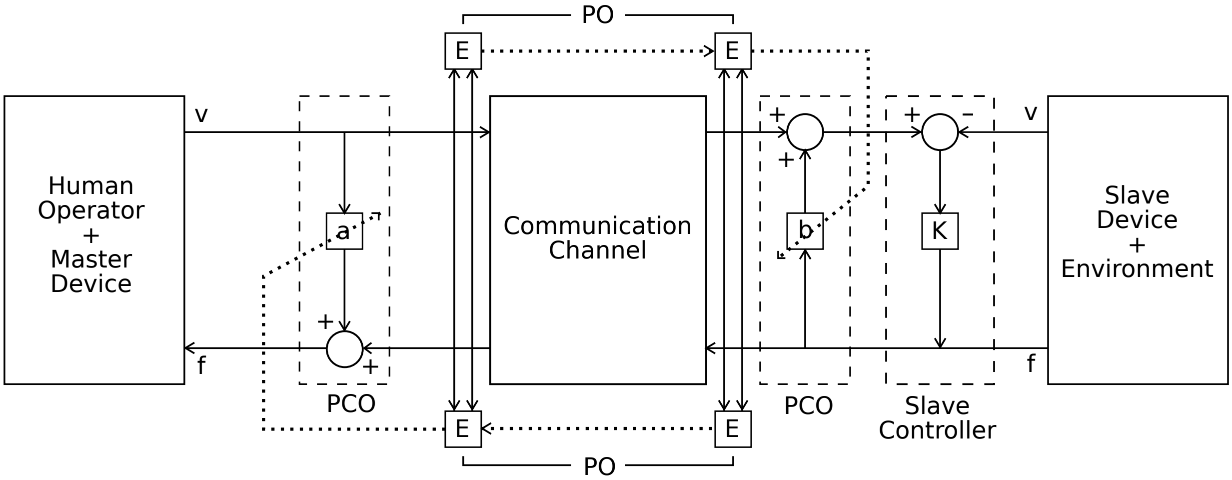


Figure 2.7: A bilateral teleoperation system using the time-domain passivity control architecture (from [155]). The PO entities compute the energy flows for both directions and provide input to the PCO on each side.

An augmented version of TDPA was proposed in [156] based on the framework of network-based analysis of passivity-based teleoperation systems in [157]. Modelling the teleoperation system using an electrical representation, rather than a mechanical one, is beneficial due to its simplicity. The electrical representation employs ideal flow (velocity) and effort (force) dependent sources as the analogous system elements to the motion commands of the human operator and reflected force of the teleoperator. Furthermore, the sources of the power network can be modelled as delayed dependent, meaning that they provide a delayed version of the current (i.e. velocity).

The communication channel equivalent is called Time Delay Power Network (TDPN) and it is in the form of a two-port subsystem that can be coupled with a PCO. Another differentiation of this framework lies in the possible structures of the proposed architecture as it further disambiguates the network channel representation, with regard to the energy flows.

Furthermore, TDPN modelling can also be applied to four-channel architecture systems [155].

In another approach, a system segmented in such a way as to provide three types of force feedback is presented in [158]. Further improvement of the TDPN method has been proposed in [159] with respect to position drift and by suggesting a different feedback scheme where the measured force from the environment is directly sent to the master.

Adaptive Motion/Force Control (AMFC)

In adaptive control, some coefficients are automatically being adjusted online to get the desired performance, and two main approaches are considered [86]: (i) adaptive-gain control and (ii) disturbance accommodating adaptive control. The Adaptive Motion/Force Control (AMFC) ([160]) is a well-used example of adaptive control schemes that allows adaptive estimation of controller parameters. Nevertheless, this scheme requires knowledge of the human and environment dynamics which are incorporated in the master and slave controller dynamics. Accurate selection of these models allows for improved delay compensation.

According to [97], AMFC is possibly stable. Perceived inertia and damping in the master side are linearly proportional to the communication channel delay (d). Perceived stiffness, tracking error and position drift are independent of d. Implementing this method does not require knowledge of communication delay but the master and slave models have to be accurately estimated. Furthermore, stability is maintained for any time delay when there is no external force (free motion). This is also true during contact with the environment for a certain range of stiffness and damping [87].

Predictive control and Communication Disturbance Observers

In the previous section, Model-Mediated Teleoperation Approach (MMTA) is using the environment model to compensate for the delays allowing the user to interact only with the estimated model of the remote environment. In the Predictive Control (PC) schemes [161], the remote environment model and estimation of communication delay is used to predict the current output based on the delayed input, in order to remove the delay term from the denominator of the closed loop transfer function. This control scheme provides stability and better performance when communication delay is large, compared to the other control schemes [162]. Smith Predictor [163], is one of the predictive control mechanisms which can be considered for delay compensation in teleoperation systems. using a plant model transfer function and a time delay model to compensate for the network delays in the communication. This has been extended with the addition of neural networks to better deal with the nonlinear nature of the remote environment [164].

The Predictive Control (PC) schemes, using position-force architecture, have similar performance as the FR scheme with the exception that with prediction of current force feedback, communication delay can be compensated. Finally, PC is Possibly stable and instability is caused due to the interaction with the environment and unknown F_e (environment force). In terms of performance, inertia, damping and stiffness perceived, as well as position drift are all independent of d, while tracking error is linearly proportional to d [97]. Another well known but closed loop control scheme, the Smith predictor, a predictive control mechanism is

To improve stability of passivity-based control, predictive control with passivity was proposed a new control scheme, a combination of prediction with wave variables, and designed a regulator to guarantee its passivity [140]. The advantage of this method compared to simple PC is that it can tolerate higher delay values while remaining stable. However, similar to PC, it needs the knowledge of the remote environment as well as the communication delay. In the Predictive Control with Passivity (PCP), wave variables are transmitted on the communication network between master and slave instead of power variables i.e force and velocity.

In PCP, tracking error and position drift are linearly proportional to d, stiffness perceived is proportional to $1/d$ and inertia and damping perceived is independent of d [97]. In addition, PCP requires the knowledge of communication delay and remote environment. Since PCP scheme maintains passivity, it is IS.

The concepts of network disturbance (ND) and Communication Disturbance Observer (CDOB) have been proposed in order to compensate for time delays. This approach uses the transfer function model of the robots, in order to estimate the communication disturbance [165]. The influence of the controller parameters on a CDOB system's transparency is analyzed in [166]. This method has been extended in order to work with variable delay in [122].

Augmented versions of the CDOB can also be used for the four-channel architecture [167]. Since works based on CDOB mostly focus on position control rather than force control, in [168], the authors propose a method for compensating in the presence of network disturbance in the force feedback channel. A comparison of different CDOB implementation has been shown in [169] along with a CDOB control scheme that integrates fuzzy control theory and neural network modelling.

Model-mediated teleoperation approach

As previously discussed, stability and transparency are conflicting objectives in passivity-based teleoperation design. This means that the system gains stability at the cost of degraded transparency. For example, the perceived stiffness of the remote objects decreases with the increase of communication delay [170].

To guarantee both the system stability and transparency at the same time in the presence of arbitrary communication delay, the concept of MMTA has been proposed. The main concept is illustrated in Figure 2.8 where it is shown that rather than directly sending back the haptic (force) signals, the parameters of the object model which approximate the remote environment are estimated and transmitted back to the master in real time during the slave's interaction with the remote environment. The model parameters include the surface geometry and physical properties of the remote objects. On the master side, a copy of this object model is maintained according to the received model parameters, and the haptic feedback is computed on the basis of the local model without any delay. The MMTA was first presented in [98] and afterwards extended in [171].

The MMTA opens the control loop between the master and slave and leads to two decoupled control loops, one on the master and one on the slave side. The stability of the MMTA system can be determined using the stability of the human-master local model closed loop and the slave-environment closed loop [88, 172]. If the estimated model is an accurate approximation of the remote environment, then both stable and transparent teleoperation can be achieved. This approach is of interest as it shifts from the problem of latency to the problem of creating accurate models which is closely related to the computation aspects of haptic communication.

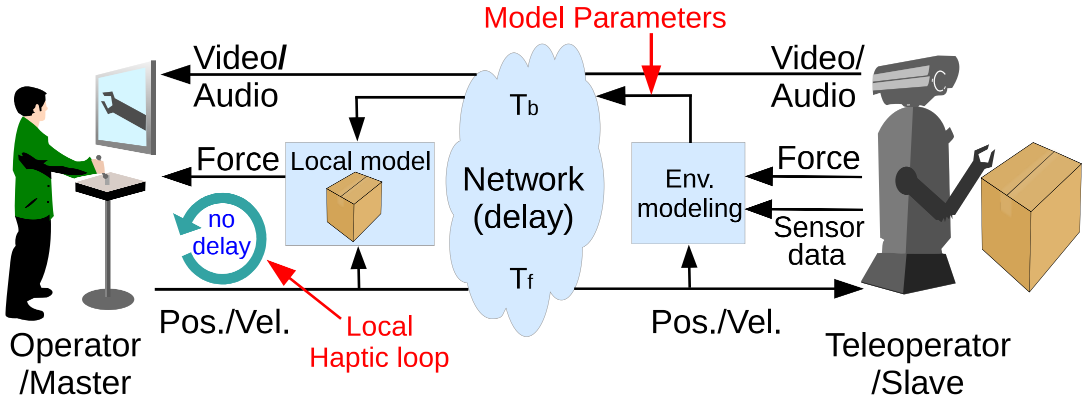


Figure 2.8: MMTA: The user interacts with the remote environment indirectly, i.e. using a model which receives model parameters from the slave device. At the same time the user's movement data are transferred through the network to the slave device [171].

More specifically, in [98], the local control loops at one side of the haptic communication system aim to simulate the impedance observed at the opposite side. Later on, in [171], in contrast to transmitting position or force values, an abstraction layer was introduced, but implemented for a 1-DoF application. The suggested algorithm replicates the remote environment at the master side and issues commands through the communication channel to the slave device.

When the master receives new model parameters from the slave side, an update of the local model according to the received model parameters is required. Ideally, the parameters of the local model need to be updated to the correct ones as quickly as possible. However, improper update schemes, e.g., a sudden change in stiffness or model position, result in a suddenly changed force that is displayed to the human user. This is called the model-jump effect [173]. To allow for a moderate model update which guarantees stability, passivity-based schemes were developed [171, 173, 174]. In [171, 173], the model position is updated only if no energy is injected into the local model system after the update. In [174], the authors used an adaptive damper to dissipate the energy injected into the local model system during the model update. This allows for a quicker model update and a higher subjective preference rate compared to the scheme proposed in [171, 173].

In general, MMTA has the benefit of being simultaneously stable and transparent in 1D or simple 3D real environments compared to the passivity-based control approaches. However, due to the limitations of existing online model-estimation algorithms, the MMTA cannot work efficiently in complex or completely unknown environments.

There is no doubt that obtaining a precise object model for complex environments (both object geometry and physical properties) is the most important task and also the main challenge for the MMTA, since a perfect match between the local model and the environment enables stable and transparent teleoperation in the presence of arbitrary communication delays. Early attempts employ a predefined model for MMTA systems [105]. This requires the master system to have rich knowledge about the remote environment. In practice, there are situations in which we have limited knowledge about the remote environment, especially when the slave enters a new environment or interacts with dynamic (movable or deformable) objects. Therefore, online environment modelling and model updating are inevitable. In recent decades, online environment modelling (parameter identification) for teleoperation systems has been widely investigated, e.g., for estimating linear [175, 176] / non-linear [177, 178] environment models, rigid [67] / deformable [179] / movable [180] objects, and for estimating unknown environment models using online neural network approaches [181, 164].

Instead of modelling the environment, an alternative architecture of the MMTA is to model the behaviour of the human operator. The estimated model parameters on the master side are transmitted to the slave to guide the slave's motion. The slave is thus not controlled by the delayed master motion commands, but performs specific tasks in complete autonomy based on the received human behaviour model. Similarly, if the model as well as the model parameters can accurately approximate the human behaviour, the slave can behave like a human user and a complete skill transfer can be realised [182, 183, 184]. The modelling of human behaviour, however, is quite challenging and the model of human behaviour has not been fully studied yet. Most of the MMTA are thus based on the modelling of remote environments, but not the modelling of human behaviour.

The estimated model parameters need to be transmitted back to the master for building/updating the local model. The transmission happens normally when the slave enters a new environment, the environment the slave is interacting with changes, or the parameter estimation is not precise. Once the estimates converge to the true values, there will be no updates required and thus the system achieves zero transmission in the backward communication channel. For real teleoperation systems, however, the estimates can vary over time due to measurement noise, natural tremble of human arm movement, etc. Obviously, to transmit every estimate is a waste of the network resources. Thus, an efficient data reduction scheme is needed to selectively transmit the estimated model parameters. Verscheure et al. [185] presented an event-triggered estimation scheme. The estimation and transmission are activated only when special conditions are satisfied, e.g. sufficiently large force/velocity of the slave, or sufficiently large displacement from the last estimation.

An alternative approach has been followed by other methods proposing a perceptual MMTA scheme where a prediction model is employed at the master and the slave side resulting in local closed-loop control on each side to ensure high fidelity. The models on both sides are updated in order to be in sync if the predicted values exceed the JND threshold. This combination of perceptual and statistical methods has been made in [186] by first applying the JND threshold and then a double exponential smoothing prediction algorithm to fill in the values not transmitted due to the threshold. Xu et al. [187] also applied the perceptual deadband approach to the estimated model parameters to reduce the transmission rate. The proposed framework incorporates 3D sensors to produce a point cloud model of a static rigid object's surface in the remote environment. The depth images are processed with a median filter and then with a temporal averaging filter to reduce noise and fill holes in the depth image. Afterwards, the depth image vectors, which consist of a 2D position and the corresponding depth value, are transformed from pixel coordinates to real world coordinates. This enables the object's geometry modelling while the slave device is in free space (not touching the object). Physical properties of the object (friction coefficient and stiffness) are also computed. Extrapolation is used when the slave device needs to operate outside the area produced by the point-cloud model, although issues are very likely to emerge. The authors of [187] also reported a data reduction of about 90% with guaranteed (significantly high) subjective quality of teleoperation. Augmented feedback information was also considered in [188] again with the use of a stereo camera.

Other control schemes

The previously mentioned control schemes are passivity-based approaches for solving the instability caused by delays in the transmission of information between master and slave. In this subsection, we will refer to other bilateral teleoperation control schemes which are not based on the passivity of the system.

Recently, Jafari et al. [189] have proposed the Input-to-State Stability (ISS) approach to guarantee the stability of teleoperation systems. It allows a bigger output energy and is less conservative compared to the passivity-based control schemes.

The ISS approach is able to generate a bounded amount of energy in the teleoperation systems while still guaranteeing stability. It has also been extended for bilateral haptic teleoperation systems in the presence of communication delays [190]. Although the ISS approach is not fully developed compared to the passivity-based approaches, it shows great potential to improve transparency due to its less conservative design.

In [191], there is a recent review of several predictive control methods with comparison and shows that a control scheme can be chosen over others for certain conditions and tasks.

2.4.3 Joint control scheme and data reduction

The aforementioned data reduction approaches for teleoperation systems in Section 2.3 have been initially developed without considering the stability issues and control scheme. In the presence of communication delays, however, the data compression schemes have to be combined with stability-ensuring control schemes. In this subsection, we briefly review the research works that studied haptic data reduction in combination with control schemes. Table 2.1 gives an overview of the efforts in the combination of control schemes and haptic data reduction approaches.

Currently, the control schemes for WV control, the TDPA and the MMTA can be combined with data reduction methods. Combining other control schemes with data reduction approaches is an interesting work for future investigation.

Table 2.1: Overview of the combination of teleoperation control architectures with data reduction schemes for different communication assumptions.

	known const. delay	unknown const. delay	time-varying delay
WV + data reduction	[5]	[69]	-
TDPA + data reduction	[154]	[154]	[154]
MMTA + data reduction	[187, 185]	[187, 185]	-

Haptic data reduction + wave-variable control architecture

The Perceptual Deadband (PDB) packet rate reduction scheme has been combined with the WV control scheme in [5, 69] for dealing with constant communication delay. The PDB approach is applied either on the wave variables [69] or on the time domain signals [5] (force and velocity). In order to modify the control schemes and to incorporate data reduction schemes, the passive PDB schemes, such as the energy supervising transmission [69] and the passive ZOH reconstruction scheme [5] were developed. In [69], the authors experimentally found the subjectively best deadband parameter for interacting with a rigid wall. On the contrary, in [5], the authors showed that applying the PDB approach on time-domain signals leads to better performance on both system transparency and data reduction compared to applying the PDB approach on wave variables.

Haptic data reduction + TDPA

Xu et al. [154] have recently combined the PD approach with the TDPA control scheme to reduce the packet rate over the communication network while preserving system stability in the presence of time-varying and unknown delays. On both master and slave sides the signals are processed with the deadband method to regulate the transmission rate of the velocity, force, and energy signals based on the PD approach discussed previously. In order to incorporate the control scheme with the PD approach, the energy calculation in the PO is modified. At each sampling instant, if no update is received, the PO outputs the same energy as the most recently received one (ZOH reconstruction) for the subsequent computation.

Compared to the existing WV-based haptic data reduction approaches, the TDPA-based haptic data reduction scheme presented in [154] can robustly deal with time-varying delays and does not require the use of the passive PD approach. This is because the deadband controllers and reconstructors are set in between the two POs, and the TDPA is capable of ensuring passivity of any two-port networks between the POs on the master and slave side. Experiments show that the TDPA-based haptic data reduction scheme is subjectively more transparent compared to the WV-based schemes. In addition, it is able to reduce the packet rate by up to 80%, without significantly distorting the user's experience for the tested communication delays of up to $100\,\text{ms} \pm 30\,\text{ms}$.

Haptic data reduction + MMTA

Similarly, a perception-based model update scheme is also incorporated into a MMTA architecture [187]. The environment model as well as its physical properties (stiffness and surface friction coefficient) are estimated at the slave side in real-time and transmitted back to the master for building/updating the local model. The transmission happens normally when the slave enters a new environment, the environment the slave is interacting with changes, or the parameter estimation is not precise. Once the estimates converge to the true values, no updates are required and thus the system achieves zero transmission in the backward communication channel.

However, for real teleoperation systems, the estimates can vary over time due to measurement noise, natural tremble of human arm movement, etc. Obviously, to transmit every estimate is a waste of the network resources. Thus, an efficient data reduction scheme is needed to selectively transmit the estimated model parameters. Verscheure et al. [185] presented an event-triggered estimation scheme. The estimation and transmission are activated only when special conditions are satisfied, e.g. sufficiently large force/velocity of the slave, or sufficiently large displacement from the last estimation. Xu et al. [187] applied the PD approach to the estimated model parameters to reduce the transmission rate. The authors also reported a data reduction of about 90% with guaranteed subjective quality of teleoperation.

The aforementioned perceptual or event-trigger control schemes for the MMTA avoid the transmission of irrelevant updates to reduce the packet rate on the network. System stability and transparency are verified in the presence of a round-trip communication delay of up to 1000 ms.

2.5 Networked haptic communication

With the increase of network-enabled devices all over the world, networks play an important role as the medium for the exchange of data. Independently of whether the access to the network is wired or wireless, the increase of intermediate network nodes in an end-to-end communication system, inevitably increases the delay of the communication. In this way, system stability and transparency will be easier to maintain.

In this section, firstly, the fundamental networking infrastructure that has already been used for haptic communication over networks is discussed. Additionally, this section focuses on network protocols of the transport and application layers as well as frameworks for the synchronization of multi-sensorial data streams. Furthermore, works on security aspects of haptic communication are mentioned.

2.5.1 Haptic communication protocols and frameworks

Internet-based TPTA systems implement closed-loop control schemes over a real-time communication framework that allows interaction between a human operator and a remote environment using sensors and actuators. A system which can be described in this way is referred to as *networked-based control system* (NBCS) [192].

Since one of the core modules of such systems is the communication channel, several network protocols have been used or created for all teleoperation frameworks mentioned in previous chapters and their implementations over the Internet, supporting their efficient functionality either for virtual environment applications [193] or physical systems.

As with every networking system, the correct functionality of NBCS is liable to several obstacles which negatively affect their performance and can also be viewed as performance indicators which allow the comparison between different protocols and the quantification of the QoS they can deliver which is of critical importance to applications such as telesurgery. Considering that teleoperation systems are NBCSs, it is natural to inherit these performance aspects with respect to requirements of TPTA systems. Common performance parameters [194]:

- *Network delay* is the average time needed for a packet to travel from the input of the communication channel to its output. A survey that thoroughly lists and discusses the main sources of network delay as well as the solutions that can be currently implemented is [195].

- *Jitter* is the result of the influence delay has on packet arrival times, formally known as Packet Delay Variation (PDV) which changes the periodicity of the packet transmission and may affect the packet sequence. A common way to avoid the effects of jitter (e.g., instability of control loop) force feedback is to use packets with sequence numbers or timestamps, nonetheless this presumes the use of buffers which in turn will increase the overall delay of the communication.

- *Packet loss* is a consequence of the network traffic congestion or interference (in the case of wireless transmission). As a result, the master and slave side of a TPTA system need to operate even with lack of information due to missing packets. Ways to overcome packet loss are substituting the missing values with null values, hold the last value or use interpolation (e.g. using a prediction method).

- *Data rate* of the communication channel. This can be affected by the sampling frequency, the sample resolution and the protocol overhead.

Additionally, alternative factors that may affect the system performance are signal quantization and other sources of noise.

The study in [196] focuses on the effects of packet loss and latency on the temporal perception of visual and haptic events. Although the two experiments in the study focused on packet loss with or without latency, the study states that visual signals need to be presented approximately 50ms earlier than haptic signals in order for both to be perceived as synchronous. When a packet loss rate greater than 10% is introduced in both haptic and visual signals the visual-haptic events were perceived as asynchronous. Nonetheless, according to [197] the effects of packet loss can be managed by using mechanisms which increase the communication reliability, but, as a result these mechanisms will increase the total latency. Therefore, it is a matter of balancing the trade-off between reliability and delay.

A detailed description of QoS control methods has been given in [198], referring to traffic management, flow control, error control, Δ-causality control, other types of control and last but not least the aspect of synchronisation of media streams. It is worth mentioning that the authors of the paper also specify methods for estimating QoE. All protocols try to satisfy the aforementioned QoS requirements and therefore encompass characteristics and mechanisms, such as QoS control methods that focus on haptic data stream or also take the other modalities (sound, video) into account, towards a reliable but also transparent haptic communication. A list of these characteristics has been defined in [199].

Network-based control haptic systems (NCHS) can be divided into two main classes: those that implement the client-server architecture and those that implement a peer-to-peer one, the latter being the most popular choice due to its support to parallel computation, scalability and also being less sensitive to negative networking conditions.

Despite the recent advances in telecommunication infrastructure, choosing a communication protocol for a teleoperation system needs thinking of how the network conditions might affect haptic applications, as some tasks have higher requirements than others in order to provide high QoE. With respect to the Internet protocol suite networking model, transport and application layer protocols have mainly been developed [199]. Nonetheless, network layer solutions, such as the Differentiated Services (DiffServ) architecture [200] (discussed later on in this section) and different network coding strategies (discussed later on in Section 2.6) have also been examined. Of course other approaches, also mentioned later on, are taking the 7-layered Open Systems Interconnection (OSI) model into account.

Transport layer

In the transport layer, the most common protocols used in the research literature for haptic communication over the Internet are the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). Even so, other protocols have been developed in pursuance of keeping the system stable and for effectively reaching greater transparency. These protocols have either been tested for the physical interaction between human operators and remote environments or have been used for communication between physical devices and virtual environments that allows the manipulation of virtual objects.

According to [201], a survey made in 2012, a total of ten different transport and application layer protocols are reported. Since then, other protocols have emerged as well. Other sources [202] can be used to extend the list of the survey with other protocols as well.

Iterating through the previously mentioned performance parameters, we can classify existing protocols according to the parameter or parameters they try to optimally improve.

TCP employs several mechanisms that attempt to mainly minimise packet loss using TCP Slow Start, Congestion Avoidance, Fast Retransmit and Fast Recovery algorithms [203], however, these mechanisms also prevent TCP from being used as a real-time protocol. This is because of buffering of data by the aforementioned mechanisms that increase delay and jitter. Evidently, it is the least suitable protocol for haptic communication applications. A modified version of TCP with Nagle's Algorithm Invalidation [204] avoids one of TCP's mechanisms in which the sender must continue buffering unless the Maximum Segment Size (MSS) is exceeded by the accumulated data also taking into account the receiver's capacity capability, i.e., the window size. This mechanism was introduced to avoid congestion over slow links, but by avoiding it time delays are decreased.

Regarding the minimisation of delay and assuming acceptable network conditions of minimum network congestion and packet loss, the most suitable protocol is UDP as it does not offer any mechanism that may introduce delays. Nonetheless, UDP's simplicity does not meet the reliability requirements of most haptic applications especially in networks under network congestion. A more suitable solution that is built on UDP, the Smoothed Synchronous Collaboration Transport Protocol (SCTP) is mainly used in haptic virtual environment applications [205] as its predecessor SCTP and attempts to deal with jitter by employing a buffer at the receiver and handling packets according to a timestamp placed at the header of each packet. This method results in a fixed delay for all messages. Smoothed SCTP should not be confused with S-SCTP which stands for Secure SCTP.

Apart from dealing with jitter, SCTP, S-SCTP and Interactive Real-Time Protocol (IRTP) [206], also prioritize messages according to their significance. Specifically for IRTP, it establishes a connection the same as TCP at first and for transmitting essential data. This makes it a connection-oriented protocol. To transmit less important data, IRTP employs UDP. It also addresses the issue of the non-optimised size of the packet header by proposing a redesigned structure of header fields.

A protocol called Supermedia TRansport for teleoperations over Overlay Networks (STRON) [207], was created to operate over overlay networks transmitting data using different network paths. STRON was compared against TCP and SCTP, showing that it performs significantly better in the case of a network that includes paths with heavy packet loss. To address interoperability with other protocols such as TCP and deal with competing network or Internet traffic, STRON uses TCP-Friendly Rate Control (TFRC) mechanism.

Another protocol called Real-Time Network Protocol (RTNP), created by Uchimura *et al.*, was developed for use on UNIX environments in order to eliminate time delay caused by the specific multitasking operating system [208], therefore, this protocol cannot be implemented on other platforms. Timely execution of the protocol handler tasks with real-time interrupts allows for more immediate transmission of haptic data packets. Furthermore, the Efficient

Transport Protocol (ETP) [209], aims to reduce round-trip delay time which is related to the Interpacket Gap (IPG). By monitoring the transfer rate, it is possible to optimise IPG by setting it to a minimum value in order to maintain stability and maximum performance of the haptic application.

A hybrid solution, a protocol that tries to leverage the advantages of others such as Scalable Reliable Multicast Protocol (SRMP), Secure Real-time Transport Protocol (SRTP), Reliable Multicast Transport Protocol (RMTP) and SCTP is the Hybrid Multicast Transport Protocol (HMTP) [210] and is mainly used for realizing haptic collaboration in virtual environments. In order to investigate interoperability of HMTP, it would be required to investigate each of the protocols that it uses. Concerning SCTP is encapsulated in UDP packets but employs its own reliability and congestion control mechanisms. RMTP's performance has been reported in [211]. According to the SRTP Internet Draft [212], the protocol's intended use is on wide area networks and "any host upon which SRTP is run must provide an Internet protocol stack with UDP, IP,and IGMP enabled". Regarding SRMP, an analysis of the congestion control capabilities shows that the protocol performs similarly to TCP [213].

A comparative evaluation of the performance of these protocols for haptic applications does not exist to the best of our knowledge, therefore it is not possible to conclude about which one would be more suitable. Nonetheless, we understand that with the exception of TCP and UDP which represent maximum reliability and minimum packet header overhead respectively, all other protocols need to balance the trade-off between reliability and latency.

Table 2.2 shows a qualitative comparison among the previously discussed protocols. Evidently, all protocols based on UDP inherit UDP's transmission of packets in a connectionless mode in comparison to the ones based on TCP. Three of the protocols listed (ETP, STRON and HMTP) were created as haptics-specific protocols whereas only one of them (HMTP) has been created for the purpose of being used in virtual environments for haptic collaboration. Only HMTP implements a security feature for user authentication when joining a session in a virtual environment.

Table 2.2: Comparison of transport layer protocols

	Connectionless	Haptics-specific	Security	Virtual Env. only
TCP				
UDP	✓			
IRTP [206]				
Smoothed SCTP [205]				
STRON [207]		✓		
ETP [209]	✓	✓		
HMTP [210]		✓	*Partial*	✓
RTNP [208]	✓			

Application layer

Even though selecting a suitable transport layer protocol, can be beneficial in meeting the QoS requirements of haptic communication applications, there are still aspects of haptic communication affecting QoE, such as stream synchronisation, haptic data reduction (to reduce packet rate) as well as stability control that can be addressed by application layer protocols.

Apart from only streaming haptic data through the communication channel, as previously described, the QoE requirements demand the synchronized transmission of both audio and video data without leaving the scope of real-time interaction of the haptic interface user with the remote environment. A system that can provide the user with such services is included in the multi-modal or multi-sensorial media (mulsemedia) systems category. In the highest of all layers, an important aspect of the application layer protocols is the aggregation and management of streams of video, audio and haptic data in order to be transported using a single data stream.

Temporal management of the data streams is a key objective of mulsemedia systems in order to provide synchronisation of all media. An investigation on how synchronisation errors affect mulsemedia systems has been made in [214, 215]. It needs to be noted that in this work we focus on data streams for the visual, audio and haptic modalities, as some mulsemedia systems in general may also support other modalities and sensations such as scent or air flow (for emulating wind). Furthermore, kinesthetic, tactile, audio and video data are sent in separate data streams. A multiplexing scheme was presented in [39].

Several attempts have been made for synchronising haptic, video and audio data streams by using different protocols, codecs and procedures for establishing the connection between two or more communication terminals. After the first stage of packetizing data of each stream, all packets are aggregated in a single stream by a multiplexing unit. Existing frameworks that use application layer protocols and frameworks with such capabilities will be further discussed in the following paragraphs together with other synchronisation practices.

A framework for adaptively controlling the data rate of different mulsemedia streams according to the human perception limits, the Adaptive Mulsemedia Delivery Solution (ADAMS), is based on a client-server architecture. The server consists of several modules that take into account various information sent from the client (e.g. network conditions) and decides on the amount of quality reduction that needs to be made on the multimedia and mulsemedia data streams [216].

Again, based on the client-server scheme, in [217] haptic communication is achieved by employing the Session Initiation Protocol (SIP) on the application layer in order to establish a teleoperation session and to manage haptic transport streams that use the Real-Time Transport Protocol (RTP) which encapsulates the haptic data in UDP packets. In this case, SIP allows for having an abstraction layer in order to incorporate encoded data in the packets using a haptic codec. Another protocol, based on RTP, is the "RTP for Distributed Interactive Media" (RTP/I is an application layer protocol focused on media beyond audio and video, as stated in [218]. Therefore, a generic interactive media model that covers the spectrum of interactive media applications in which TPTA applications are included is also introduced. A protocol created to surpass the disadvantages of RTP [219], the *MPEG Media Transport*, is an application layer transport protocol used in [220] for the purpose of multi-modal data transmission on 3D tele-immersion environments (3DTI).

In the multi-modal communication framework of Perception-based adaptive haptic communication protocol (PAHCP) [48], which is concerned with C-HAVE applications (not physical ones), data synchronisation is implemented using the Network Time Protocol (NTP) while graphics and haptic data are transmitted with Virtual Network Connection (VNC) and PAHCP respectively. PAHCP enables perception-based data reduction implementations. Based on UDP, this protocol is a "modified version of the smoothed SCTP".

Another protocol mainly focused on interactive haptic virtual environments is the Application Layer Protocol for HAptic Networking (ALPHAN). ALPHAN is built on top of the UDP for enhancing the latter's characteristics which are unable to meet the high-demanding C-HAVE conditions and exchanges the QoS parameters with the XML-based Haptic Application Meta-Language (HAML) file format [221]. HAML is also used by Adaptive Multiplexer (Admux) framework, a framework/protocol that implements statistical multiplexing at the application layer also focusing on synchronising the haptic, audio and video streams [222].

The authors of [223], focusing on telesurgery, have presented an application layer protocol, called the Interoperable Telesurgical Protocol (ITP), but in the experiments performed the communication was not bilateral, the users only had visual feedback. It should be noted, though, that the protocol could be extended to be used in implementations with data transforms such as the wave-variable transform (discussed later in Section 2.4).

In physical teleoperation systems with constant bit rate communication channels, a multiplexing scheme has been proposed for transmitting video and haptic data with the application of perceptual data reduction using the ZOH method [39]. While multiplexing, if no force data are to be sent then the video data are prioritised. By assuming a constant bit rate connection, packet delay can be computed and used for correctly demultiplexing the data stream. Another framework that employs the JND method, explored later on in this paper, is the Haptics over Internet Protocol (HoIP). Implemented in C++, HoIP is using the unreliable UDP and a multiplexing algorithm that enables the packetization of either haptic and audio data or haptic and video data. The header of each packet allows the estimation of QoS parameters, the use of adaptive sampling by employing the JND method and also a flow control mechanism [224].

Furthermore, other frameworks that exist perform QoS management over overlay networks. In [225] methods for error correction and optimal path selection are applied for efficient data stream transmission. Finally, in [40], a module measures the complexity of a task to dynamically adjust the number of active paths and network resources for each media stream.

Last but not least, there are two application layer protocols with a focus on security which is a major concern for e.g. telesurgery applications. These are Secure ITP [226], a version of the previously mentioned ITP extended to be more secure and the Secure and Statistically Reliable UDP (SSR-UDP). The first one implements user authentication and authorisation as well as data encryption using the Advanced Encryption Standard (AES). In the latter one, the system design includes transmitting data with what the authors call a "privacy scheme" as well as a feedback channel for sending acknowledgement packets back to the master device [227].

In Table 2.3, there is a qualitative comparison of all aforementioned application layer protocols. Two of the listed protocols (ALPHAN and Admux) are used specifically for virtual environments, while three of them (PAHCP, HoIP and ALPHAN) implement data reduction. HoIP, Admux and ALPHAN multiplex audio and video and haptic data but none of the protocols except Secure ITP and SSR-UDP incorporate security mechanisms.

Table 2.3: Comparison of application layer protocols

	Data Reduction	Mux (Audio-Video)	Security	Virtual Env. only
PAHCP [48]	✓			
HoIP [224]	✓	✓		
ALPHAN [221]	✓			✓
Admux [222]		✓		✓
ADAMS [216]		✓		
ITP [223]				
Secure ITP [226]			✓	
SSR-UDP [227]			✓	

2.5.2 Network Security for Haptic Communication

Security is yet another cornerstone of the Tactile Internet, considering teleoperation sessions often represent critical communication scenarios. Security has indeed been the subject of recent research under the umbrella of Cyber-Physical Systems (CPS) to address the needs of emerging sensor networks [228]. In [229], a list of threats and possible attacks on teleoperation systems are presented, along with a protocol that deploys mechanisms to meet QoS requirements of bilateral teleoperation and at the same time respect the IP Security protocol (IPSec) standard.

In addition, as mentioned in [230], nowadays security threats of CPSs are not focused on communication standards only. All layers of communication, from physical to application, can be targeted. Despite its importance, in this work, we devise our attention to the enablers of reliable low-latency communications and hence security is not within context.

Nonetheless, it needs to be mentioned that methodologies for enhancing security of communication have a negative impact on the end-to-end latency. This is another trade-off to be taken into consideration when designing a haptic communication system. Therefore, it is a challenge to integrate security in such systems.

2.6 Haptic communication over 5G mobile networks

Providing the services mentioned above in remote geographical areas and in an on-demand manner, where high bandwidth and dedicated networking infrastructure is not available, is yet another crucial aspect, which can be addressed by mobile networks. Furthermore, in comparison to fixed broadband networks, mobile networks have the advantage of having the ability to be deployed e.g. in case of emergency, a lot more rapidly.

Such scenarios become technically feasible due to progress anticipated with the 5G technology. Nonetheless, 5G will provide more than that. The transition from 4G to 5G is based on KPI values, such as latency, peak data rate (per user) and reliability among others, which define the challenges and targets towards 5G and that need to be improved in order for e.g. haptic communication to be realized.

Standardisation of the next generation 5G wireless communication systems has recently been initiated. Within the on-going 3GPP RAN 5G study item, also known as *New Radio*, technical components are being identified for a 5G radio interface and the next generation network architecture. 3GPP agreed to develop the 5G system specification in two phases, which correspond to 3GPP releases 15 and 16; a full system specification needs to be finalised and submitted to International Telecommunication Union (ITU) by end of 2019. On-going work in both ITU and 3GPP define, at a high level, use case categories, resulting in requirements and evaluation methodologies for 5G system design. While earlier generations of mobile networks focused on mobile broadband services (targeting services for people), it has already been identified that 5G should, in addition, address the two new areas of Massive Machine-type Communication (M-MTC) and Critical Machine-type Communication (C-MTC), where services are provided to things and objects. C-MTC is, in 3GPP parlance, also referred to as URLLC. These two latter areas address the successive transformation of our society into a networked society.

According to [231] and based on data provided by the UK Office of Communications (Ofcom), the average Round-Trip Time (RTT) for 3G is 63.5 ms and in 4G it is reduced to 53.1 ms. RTT in this case is considered the time between sending a packet of data to a server and receiving a response. In the US, according to [232], presented in 2012, median RTT for 4G is 69.5 ms by measuring in a similar fashion the time difference between a SYN and SYN-ACK packet. As mentioned in previous sections these latency values are unacceptable within the scope of bilateral teleoperation with high QoE, as even with the application of stability control methods there will be a decrease of transparency.

2.6.1 5G use cases and requirements

It is obvious that the 5G network capabilities are determined by the requirements of the use cases which will need to utilise the network effectively. Essentially, we need to iterate through the use cases and extract those requirements. This is something that has already been done by 3GPP mainly in [233] with further information in [234]. Table 2.4 lists the use cases using the first classification used in [233] along with a number of examples and briefly showing the main KPIs and requirements that need to be satisfied for the users to have good QoE.

Teleoperation is mainly related to the first three use case categories, but since a broad spectrum of applications exists, the different requirements can be grouped into many different classes. As seen in Table 2.4, the names of the use case families are self-descriptive as they include some of the main KPIs mentioned or a combination of them. These KPIs were selected to better demonstrate the main similarities and differences among the use cases. These include:

Table 2.4: Classification 5G use cases with examples and their corresponding requirements for the main KPIs

Use case family	**Traffic scenario examples**	**Main KPIs**	**Requirements**
Higher reliability, availability & lower latency	• Medical treatment in ambulance • Low-latency industrial applications • Telemedicine cloud applications	e2e latency Reliability Availability Mobility Data rate	$\leq$ 1 ms $\geq$ 99, 999% $\approx$ 100% $\geq$ 120 km/h 10s of Mbps per device
Very low latency	Human interaction, Immersive VR, Remote healthcare Telementoring	e2e latency	1 ms one-way
Mission critical services	Prioritised access when: (i) the network is congested (ii) simpler access procedures or guaranteed QoS are needed	e2e latency Reliability Security	down to 1 ms $\approx$ 100% max. confidentiality & integrity
Higher reliability & lower latency	• Unmanned Aerial Vehicles (UAVs) & Ground-based Vehicles • VR/AR applications • Cloud robotics • Industrial applications/ Power plants	e2e latency Reliability Data rate Energy efficiency	1 ms min. 99, 999% 250 Mb/s max. Various or NA
Higher accuracy positioning	• Outdoor positioning (high speed moving) • Indoor/Outdoor positioning (low speed moving) • UAV positioning for critical applications	Accuracy e2e latency Mobility	$\leq$ 3 m for 80% of occasions $\leq$10 ms to 15 ms two-way $\approx$ 280 km/h
Higher availability	• Secondary connectivity for emergencies (mobile-to-satellite)	Coverage	Service continuity

- *End-to-end latency* (e2e latency): The time it takes for data to be transferred from source device to destination (in milliseconds).
- *Reliability*: The number of packets successfully received by one end node divided by the total number of packets sent (percentage).
- *Availability*: The amount of time the communication system can provide service to the

user divided by the total amount of time which is expected to deliver the services.

- *Mobility*: The speed at which the user is requesting services from the network provider. One example is telesurgery with the patient inside an ambulance moving with high speed.
- *Data rate*: The amount of data that the network can deliver in one second.
- *Coverage*: The area in which a network provider can offer services.
- *Positioning accuracy*: The accuracy at which a user's location can be tracked.
- *Security*: Maintaining the integrity of the data, in many cases, is a basic requirement. In Table 2.4, we also mention the relevant concept of confidentiality, which also relies on the network operator's discretion.
- *Service continuity*: Even when there is a change in the way a service is delivered to the user, this needs to happen in a seamless manner. This change can be a different access technology (e.g. satellite).
- *Energy efficiency*: The amount of bits per Joule of energy consumed.

We need to mention that most of these KPIs behave differently in case the user is in an indoor or an outdoor environment.

2.6.2 Realising the Tactile Internet

For the rest of this section, we will discuss recent progress in mobile networks towards delivering reliable low-latency communication for realisation of the 5G Tactile Internet. Such developments are:

- *Software Defined Networking (SDN)*: By decoupling the control from the data plane, and providing logically centralised control, SDN will be one of the key components of the 5G network. More importantly physical and virtual SDN controllers will allow dynamic changes in the interconnections among physical and virtual components within the network also enabling network slicing (i.e., virtual end-to-end differentiated connectivity). The centralised control allows for easier management of traffic within the network [235, 236], while taking advantage of the abstraction, mobility can be handled more reliably and with incurring less latency [237]. Furthermore, the software-based nature and the programmability enable delivery of QoS based on granular and flow-based policies [238, 239].

- *Network Function Virtualisation (NFV)*: The virtualisation and softwarization of network functions drastically decreases the dependency on hardware and therefore increases network scalability and reliability. Scalability is increased in the sense of increased interoperability and easier cross-administration as well as the reduced cost on deploying network components on generic purpose devices [240]. Furthermore, reliability of softwarised components is easier to be guaranteed, considering that they are easier to maintain, upgrade, back up or migrate, even though there are still ongoing efforts for optimising performance considering redundancy [241, 242]. It is also easier to share resources among different network functions and also transfer network functions across the network in order to optimise a service's performance in terms of latency [243].

- *Mobile Edge Computing (MEC)*: While allowing mobility, a remotely located network of servers, either physical or both physical and virtual, is responsible for processing and storing data from a mobile device, enhancing the capabilities of a service or application, as well as acting as a computation offloading mechanism for the mobile device [244]. A relevant expansion of MEC related to haptic communication is cloud computing for mobile robotics. In this case, the cloud is used for off-loading computations (e.g., for stability control) from the remote robot [245]. The previously mentioned Network Function Virtualization (NFV) is a complimentary technology to Mobile Cloud Computing inside the 5G technology framework which allows optimal distribution of "intelligence" inside the network.

- *New Radio*: The new radio standards will enable services with diverse latency requirements. This will be primarily implemented by allowing a scalable Transmission Time

Interval (TTI) and a redesign of the sub-frame (SF) making it easier to support a variety of services. Long-Term Evolution (LTE) standards can currently offer 10 ms to 20 ms round trip time (between air interfaces only) using a 1 ms TTI. Nonetheless, 5G requirements demand a user plane end-to-end latency of less than 1 ms [246].

Furthermore, the deployment of Massive MIMO will ensure that the bit-error-rate (BER) will be kept at minimum for reliable low-latency communication[247].

- *Dual Connectivity*: Extra reliability in heterogeneous networks will be provided by decoupling uplink (UL) and downlink (DL) connections [248].

Radio resource allocation for haptic devices in LTE-A systems has been proposed in [249], in the scope of optimising power and resource block allocation for both UL and DL channels. In [250], by taking into consideration the traffic patterns of haptic communication systems, soft resource reservation is proposed in order to reduce latency caused by the LTE scheduling request (SR) procedure in the UL channel.

In the case of SDN, methodologies have been developed for predicting performance by modelling the underlying network using queuing theory and network calculus (either stochastic or deterministic) making use of the network monitoring capabilities that SDN has. In this way, it is possible to perform traffic shaping and path optimisation based on the application requirements. Such mathematical tools have been presented in a survey on the analysis and modelling of SDN [251].

Furthermore, as described in [252], QoS can be managed within the SDN paradigm through the OpenFlow protocol, the most consolidate communication interface between the control and infrastructure layers, and it is mainly used to access the infrastructure layer and modify the switch's flow table. The flow tables kept in an SDN-capable switch contain the rules to apply to each flow, which are programmed by the SDN controller and pushed to the infrastructure using OpenFlow. Queues and meters are two OpenFlow features useful for QoS management and traffic differentiation. The mechanism used is illustrated in Figure 2.9. In more detail, packets arriving at a switch can be identified as a particular flow and set to a queue which has a

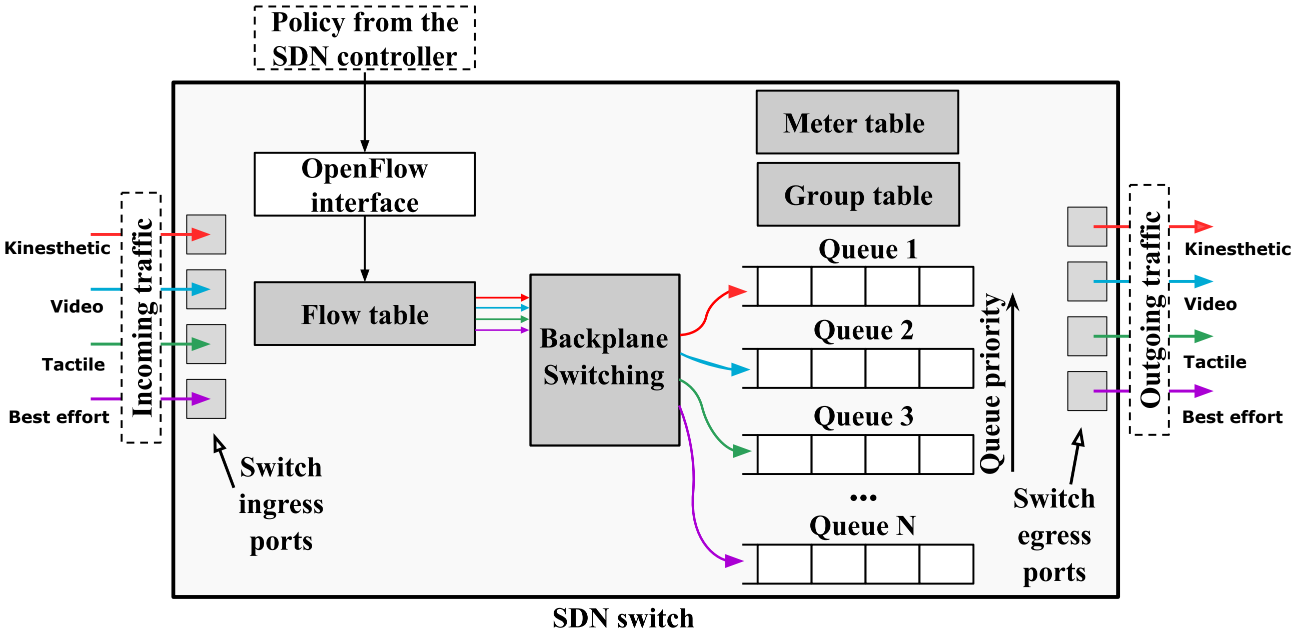


Figure 2.9: QoS mechanism using queues and meters in SDN

predefined configured transmission rate on an output port. Meters allow to set a transmission rate threshold that can trigger other functions once the threshold is exceeded. One packet arriving at the switch may be assigned to multiple meters which trigger different functions once the rate is exceeded.

Prediction, in the context of anticipatory mobile networking, can offer benefits in other various areas as well, such as improving mobility management, decreasing latency and improving reliability with optimised resource allocation. This will offer the possibility of high mobility scenarios to become reality [253].

An implementation of a network coding strategy deployed using Virtual Network Function (VNF) instances in combination with an SDN controller was shown in [254]. The authors claim that random linear network coding not only increases the reliability of the communication but also positively affects the reduction of latency, although even in the case of lossless 3-hop communication network and an 8 Mb/s channel rate, the minimum latency achieved is 100 ms.

Since bilateral teleoperation can be classified as a latency-sensitive application, the aforementioned technologies will be used to provide low-latency connectivity to users of various types of applications with low latency requirements.

2.6.3 3GPP architecture

As discussed in the previous subsections, NFV and SDN are enablers of a softwarised 5G network that can accommodate URLLC use cases of the Tactile Internet. Nevertheless, apart from meeting the requirements of URLLC, the design of such a network needs to rely on a service-based architecture that can offer a scalable and flexible network infrastructure.

A distinct feature of 5G system-level architecture defined by 3GPP is the modular principle in the network function design, which brings the needed flexibility to run network functions as well as enables the concept of network slicing. The modularity in the 5G architecture is designed to support deployments using both NFV and SDN technologies. The main network functions (for the purposes of this thesis), illustrated in Figure 3.10, as defined in 3GPP TR 23.501 [255] are:

- *User Plane Function (UPF)* is in charge of handling the user plane path of a data session and will provide an interface to the data network (outside the 3GPP domain).
- *Session Management Function (SMF)* is in charge of the establishment, modification, and release of the session. Amongst the multiple tasks of the SMF are: policy control, QoS enforcement, data plane routing information to the UPF.
- *Access and Mobility Management Function (AMF)* controls the access decisions as well as handles all related mobility procedures. Based on multiple inputs and metrics, software-defined rule enforcement allows the AMF to decide on the best access point for the user.
- *Policy Control Function (PCF)* is in charge of providing the policy rules to the relevant control plane functions and supports a unified policy framework across the network.

2.6.4 Latency reduction and traffic management for URLLC

This section discusses a selection of representative works on the topic of URLLC and its enablers within the end-to-end network infrastructure. The main goals of the following proposed methods and frameworks are latency reduction and efficient traffic management.

An important aspect of 5G networks is the ability to bring computational capabilities closer to the user, Mobile Edge Computing (MEC), thus reducing latency compared to previous centralised architectures. This can be particularly helpful, for example in the case of MMTA control scheme as the models required for successful teleoperation can be computationally prohibiting for end devices. In [256], a VNF management framework that optimises certain functionalities within Virtual Machine instances where VNFs are deployed, for the purpose of reducing delays within the MEC domain. Latency reduction can be boosted further by deploying VNFs with network coding capabilities [257].

During their operation, networks serve a variable amount of data traffic. In many cases, during certain time periods, traffic congestion needs to be managed to reduce its latency-inducing effects and provide the required QoS to traffic flows. In [258], the authors present a method that multiplexes bilateral teleoperation traffic and sensor traffic of lower priority modelled as deterministic and stochastic traffic respectively. In this approach QoS overprovisioning of high priority traffic is averted by allowing high priority traffic frames to be overwritten by low priority traffic frames assuming a certain reliability threshold is not exceeded. Nevertheless this approach is valid only for certain scenarios.

Furthermore, the framework proposed in [259] takes into consideration the relatively small capacity of buffers inside commodity 10Gbps switches used in data centres and potential high congestion conditions that would increase latency in the expense of URLLC traffic. This framework makes use of simulated queues that emulate congestion conditions and the DCTCP protocol to notify end nodes that data transmission reduction is needed. This work reports similar performance with strict priority queuing. On the radio domain, a solution for multiplexing URLLC and eMBB but focusing on radio resource allocation and link adaptation is proposed in [260].

Within the Cloud-Radio Access Network (C-RAN), the authors of [261] present an optimisation problem that minimises transmit power by adjusting queuing delays generated within a C-RAN, assuming pairs of URLLC users are served by the same C-RAN infrastructure. A framework for managing prioritisation with priority jumps for traffic in cellular networks for Device-to-Device communications has been presented in [262].

The aforementioned efforts have contributed in QoS provision for URLLC traffic, nevertheless, in a survey on Time-sensitive Networks (TSN) and Deterministic Networking standards [263], it is stated that future efforts need to focus towards the direction of dynamically managing flow prioritisation.

Finally, with regard to haptic communication, it is evident that the above methodologies and frameworks do not take advantage of the capabilities of bilateral teleoperation to compensate for the delays induced within the different network domains. In the upcoming sections, a framework that combines bilateral teleoperation control and network scheduling that attempts to avoid overprovision of high priority traffic will be presented.

2.7 QoS provision in the network

By itself, Internet operates on the best-effort basis by treating all packets in the same way, therefore not guaranteeing QoS. Number of approaches have been proposed to provide QoS including: relative priority marking, service marking, label switching, static per-hop classification, Integrated Services (IntServ) and DiffServ. Each of these categories have been implemented in different ways, but because of relevance, we explain IntServ and DiffServ further here, while the latter being the most scalable and thus more preferred solution currently. Last but not least, the efforts on QoS provision over networks is examined and further discussed under the prism of dynamic network scheduling.

The IntServ architecture [264] relies on the storage of information in all routers of the network in terms of flows that will pass through them. A preallocation of resources is done using the relevant signalling protocol in order for the data stream to travel end-to-end. The downside of IntServ is mainly its scalability as supporting a large network such as the Internet can easily become too complex. Also, the periodic information update concerning each flow can increase the traffic significantly.

On the other hand, DiffServ is not providing QoS to flows individually. Instead, it classifies flows by labelling the data streams. DiffServ is implemented in IPv4 and IPv6 [265] as a field inside the IP header of a packet called Type of Service (ToS) and Traffic Class (TC) respectively which determines how the network should manage each packet in a per-hop behaviour (PHB). With a total amount of 64 different classes (6 bits of the octet) available, DiffServ allows the aggregation of different flows into a single class. It is important to mention that DiffServ is completely transparent to all Layer 2 mechanisms as it operates exclusively on Layer 3.

Two important mechanisms of DiffServ are the Expedited Forwarding PHB and the Assured Forwarding PHB (AF PHB). The first one, is highly related to haptic communication as it provides queue prioritisation for applications/services with high requirements in terms of packet loss, latency, jitter and data rate. On the other hand, AF PHB offers a framework for providing different drop rates which depend on a predefined table of drop rate classes.

Furthermore, using the aforementioned protocol features is not an easy task since monitoring and dynamic management of resources which requires the configuration (and re-configuration) of all network nodes is demanding. Hence, flexible traffic management while simultaneously supporting strict and dynamic QoS requirements is yet critical and challenging in the Internet. On the other hand, the effective and quick adaptation of resources to the actual traffic demand is one of the main features expected to be effectively handled by next generation networks, that will also be a key enabler for the Tactile Internet. A step forward in introducing flexibility in network management is represented by SDN, where control and data planes are decoupled and split into logically centralised network intelligence and an underlying abstracted infrastructure [266]. Among the various key features of SDN, the programmability and agility of the network (re)configurations can significantly ease management of diverse QoS requirements, while the logically centralised control enables scalability. Using features such as queues and meters per flow basis across the end-to-end path, for example, provides granular QoS and can allow for prioritisation of more latency demanding flows such as teleoperation flows that can be dynamically reconfigured depending on the volume of traffic [267, 268].

These features can be applied in a multi-domain end-to-end network either in terms of multi-technology infrastructure or multi-operator administrative domains with the use of end-to-end management, orchestration and network slice co-ordination [269]. In this case, the multi-domain network can be divided in sub-domains, where each of the VNFs can be interconnected using sub-domain SDN controllers [270].

Regarding wireless access in mobile networks, in LTE, to assigned different QoS levels to different traffic classes, QoS Class Identifier (QCI) was introduce. In the 3GPP Technical specification 23.501, in 5G systems, a new scalar with 21 different values (i.e., classes) called 5G QoS Indicator (5QI) will replace QCI. According to [271], usage of IntServ and DiffServ is limited as the first one is not scalable and the second requires its integration in all nodes of the network. Furthermore, QCI appears to be commercially unattractive so its usage is even more limited. Regarding 5QI, the proposed indicator is currently not commercially launched.

At this point it should be mentioned that flow prioritisation and net neutrality are contradictory concepts but this discussion is out of the scope of this work and will not be further discussed.

2.7.1 QoS provision with dynamic network scheduling disciplines

The increasing number of services that current telecommunication networks need to support suggests that in order to provide appropriate QoS to each service a flexible and efficient networking infrastructure is needed. One of the ways to address this difficult task that networks need to undertake is multi-class traffic prioritisation. Nevertheless, this mechanism has a major drawback which will be discussed in this section.

This section is focused on the aspect of network scheduling and specifically dynamic traffic prioritisation on queuing systems. However, to better understand dynamic priority schedulers it is useful to examine how regular traffic prioritisation works. A well-known and widely implemented traffic prioritisation discipline is Strict Priority (SP) queuing. The SP scheduling algorithm has been proposed for service differentiation in the Internet [272], but with a major shortcoming. In SP, packets of low-priority queues are served only if higher priority queues are empty, which results in starvation of low-priority traffic.

To reduce the effects of starvation, dynamic priority scheduling has been studied in the literature [273] and several works have been proposed with the aim of providing adaptive fairness in systems SP-based queuing systems. One of the strategies to minimise starvation is Head-of-Line Priority Jump (HOL-PJ), first presented in [274]. The main concept is to alter the priority of packets entering the system during execution of the queuing system, instead of using absolute priorities. Effectively, this results in a parameterised moderation of system performance with a focus on fairness among different classes of traffic being served by the system. Many implementations use the concept of priority upgrades also known as priority jumps by changing priority levels of each packet waiting in one or more queues depending on the system model.

For the rest of this subsection, we present indicative works on the topic of dynamic priority scheduling non-preemptive disciplines including the discipline that is used in our proposed framework. We will also focus on disciplines which are not time-dependent, i.e., disciplines that do not use time-stamping to monitor each packet for changing its priority, tasks which increase computation requirements.

The authors of [275] proposed a scheduling algorithm called Fair Dynamic Priority Assignment (FDPA) with a focus on fair bandwidth sharing for TCP-based traffic, using arrival rate on each queue as a metric for managing priorities of queues. In [276], the scheduling algorithm monitors the mean queuing time of each class to select the class that will be served next, based on the concept of relative priorities.

Even though many performance indicators can be used for dynamically changing the service of prioritised traffic, probabilistic disciplines offer more flexibility as the parameterisation can also be user-controlled, allowing the system designer to select the criterion or criteria that will affect performance of the queues using the same queuing discipline. In [277], the queue-based Probabilistic Priority (PP) discipline is proposed. In PP system model, based on the p_i-persistent protocol, every queue is assigned a user-selected probability which affects the choice of the next queue to be served, also known as probabilistic service. The Head-Of-Line Merge-By-Probability (HOL-MBP) scheduling discipline, lies within the packet-based category of disciplines. For simplicity, assuming a single-server system of two priority queues (one for high priority traffic and one for low priority traffic), packets of low priority traffic queue to the high priority queue according to a user-defined probability [278]. This last method is the one used in this work as user-controlled parameterisation fits the purposes of the proposed framework in Chapter 4. However, HOL-MBP is a simplified version of more complex disciplines proposed in [279] but operating under the same theme of priority jumps. The main benefits of this type of queuing models are twofold. First, there is no need of tracking the delay of each packet as in the case of models similar to e.g., Weighted Fair Queuing [280], therefore there is no time-dependency as previously described. The latter also removes the need for using a timing mechanism. Second, their performance analysis is kept analytically tractable.

2.8 Summary

This chapter surveys, evaluates and analyses enablers of haptic communication by examining the fields of teleoperation robotics and telecommunications, as well as discusses current and state-of-the-art methodologies and technologies. This chapter also presents a literature review related to the contributions presented in the following chapters of the thesis. The flow of this chapter's content is depicted in Figure 2.10. The first three sections discuss solutions around the three solution spaces and enablers of haptic communication. Then we draw our attention towards 5G mobile networks discussing the 5G enablers, including a section on QoS provision with traffic differentiation, the main topic of this thesis.

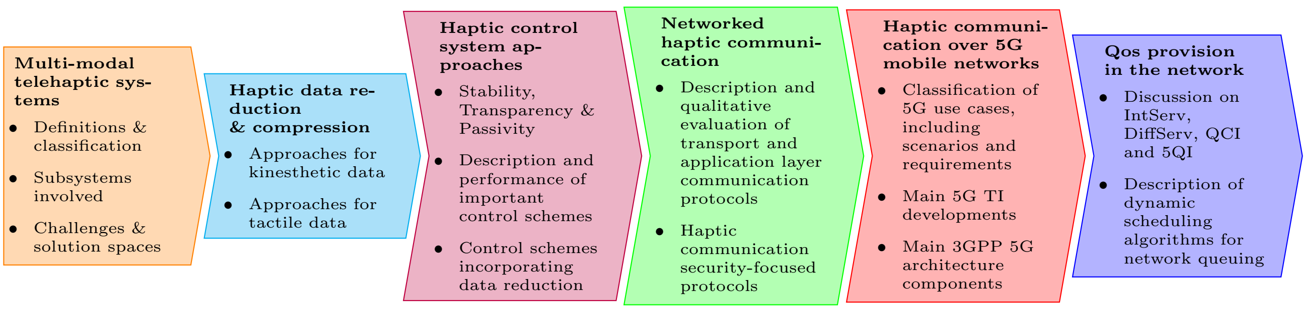


Figure 2.10: Flow diagram of Chapter 2 content

The chapter starts with the description of haptic communication systems, along with a classification of these systems and definitions of components and subsystems involved. Furthermore, it includes the challenges and the three solution spaces (communication network, data processing and robotics) which are explored in the next sections.

Section 2.3 includes details on the behaviour and types of data transmitted by haptic devices, as well as methodologies of data reduction schemes for kinesthetic and tactile data with an emphasis on the perceptual aspect of data reduction and compression. Understanding the way data is generated is important for identifying requirements with respect to the generated network traffic.

The rest of the sections presented in this chapter play a key role to contributions presented later on in the thesis. In section 2.4, the concepts of stability, transparency and passivity are essential in understanding how telepresence can be achieved and maintained with good QoE. In this respect, important stability control schemes and their performance are discussed. In addition, existing combinations of control schemes and data reduction methodologies are presented. More importantly, in this section the analysis of control performance is made.

Transitioning from the robotics to the communication network solution space, section 2.5 includes a description and qualitative evaluation of the transport and application layer communication protocols. Also, existing protocols with security features are also presented for completeness.

Section 2.6 presents 5G use cases and their classification, including relevant traffic scenario examples and their corresponding requirements. Moreover, the main enablers of 5G TI as well as the main 3GPP 5G architecture components (relevant to the purposes of this thesis) are enlisted and discussed.

Last but not least, the last section focuses towards a more specific networking concept, QoS provision, firstly by discussing past, current and future traffic differentiation methodologies and then by examining dynamic scheduling within the network, a topic related to the main contribution of the thesis.

Chapter 3

Experimentation frameworks and an architecture for the Tactile Internet

3.1 Overview

After the previous chapter's thorough discussion on main aspects of haptic communication, the 5G enablers and the 3GPP 5G architecture, this chapter has two main objectives.

The first objective is to assess the behaviour of haptic communication systems regarding:

- The impact of delay on haptic communication QoE.
- The impact of packet loss on haptic communication performance.
- The impact of network slicing on haptic communication performance under congestion conditions.
- The co-existence of best effort and mission-critical traffic on the same network under congestion conditions.

The first three sections of this chapter include experimentation frameworks for evaluating indicative and prototype haptic communication systems using 5G enablers, to examine the effectiveness of 5G enablers to provide appropriate QoS, but also the impact of QoS provisioning when accommodating differentiated traffic flows. These frameworks cover a variety of use cases such as virtual haptic systems emulating simple tasks for preliminary study on control performance, medical examination [281] and mission critical services [282].

It needs to be noted that a common mechanism used in the packet loss, network slicing and traffic co-existence experiments is the use of dynamic QoS management, leveraging the SDN capabilities of network configuration.

The second objective of this chapter is to discuss the standardisation efforts that focus on the generic architecture design for the implementation of the Tactile Internet for efficient and scalable URLLC, as well as propose a 5G system architecture that uses the aforementioned enablers and fulfils the requirements of haptic communication. In this respect, the contributions made include:

- One possible 5G core network architecture that orchestrates 3GPP components running as VNFs by using physical and virtual nodes of SDN controller instances to enable URLLC.
- The definition of architecture entities/domains, physical and logical interfaces and the functional capabilities that are considered within the IEEE P1918.1 Tactile Internet standards, which was a task as part of my participation in the Standards Working Group meetings.

More specifically, the proposed 5G architecture allows the orchestration 3GPP's 5G System Architecture (described at length in [255]), but introduces it as part of the global European Telecommunications Standards Institute (ETSI) NFV architecture for virtualisation[1]. The 3GPP components are running as VNFs over an SDN-capable network by using physical and virtual nodes of SDN controller instances to enable URLLC. In particular, all network functions are considered to be pieces of software, which can run in standard hardware and may in principle

[1] ETSI, "Network Functions Virtualisation (NFV); Architectural Framework," ETSI GS NFV 002 V1.1.1, October 2013.

be moved around across different locations subject to the requirements of the communications provider.

Furthermore, the role of softwarisation from the end-to-end perspective is also addressed, including:

1. The software-defined traffic steering decisions in the access part.
2. The role of SDN in the core part.
3. How it should be interfaced with the 3GPP architecture.

In general, software-defined rules have to be enforced in the physical infrastructure to satisfy a certain level of QoS, starting from accurate selection of access nodes and up to managing traffic in the transport network where the CN resides. In this work, mission-critical services are considered; hence the use of softwarisation in policy and QoS enforcement is crucial to ensure appropriate traffic differentiation as well as correct delivery of user-plane data, which allows for prioritisation in the transport network if required [282].

3.2 Preliminary study on control performance

In order to evaluate the impact of latency in teleoperation systems, which include either only physical components or both physical and virtual ones, two platforms for preliminary study of control performance were created. The first one involved the connection of two Phantom Omni devices using an SDN-capable switch. The second one involved the use of only one Phantom Omni device interacting with a virtual environment with haptic interaction.

3.2.1 Impact of delay in a multimodal virtual environment

To first evaluate the impact of delay in bilateral teleoperation, with an indicative haptic communication system experiment using one haptic device (Phantom Omni) which allows the user to touch objects on a 3D virtual environment [2]. The experiment demonstrated the impact of latency on teleoperation when performing a simple task. The task was to move a cylinder across a wire (1 DoF movement) with two red spheres connected at the edge of the wire acting as objects that terminated the cylinder's trajectory. To move the cylinder, a small grey sphere was used as a 3D cursor manipulated by the haptic device as shown in Figure 3.1.

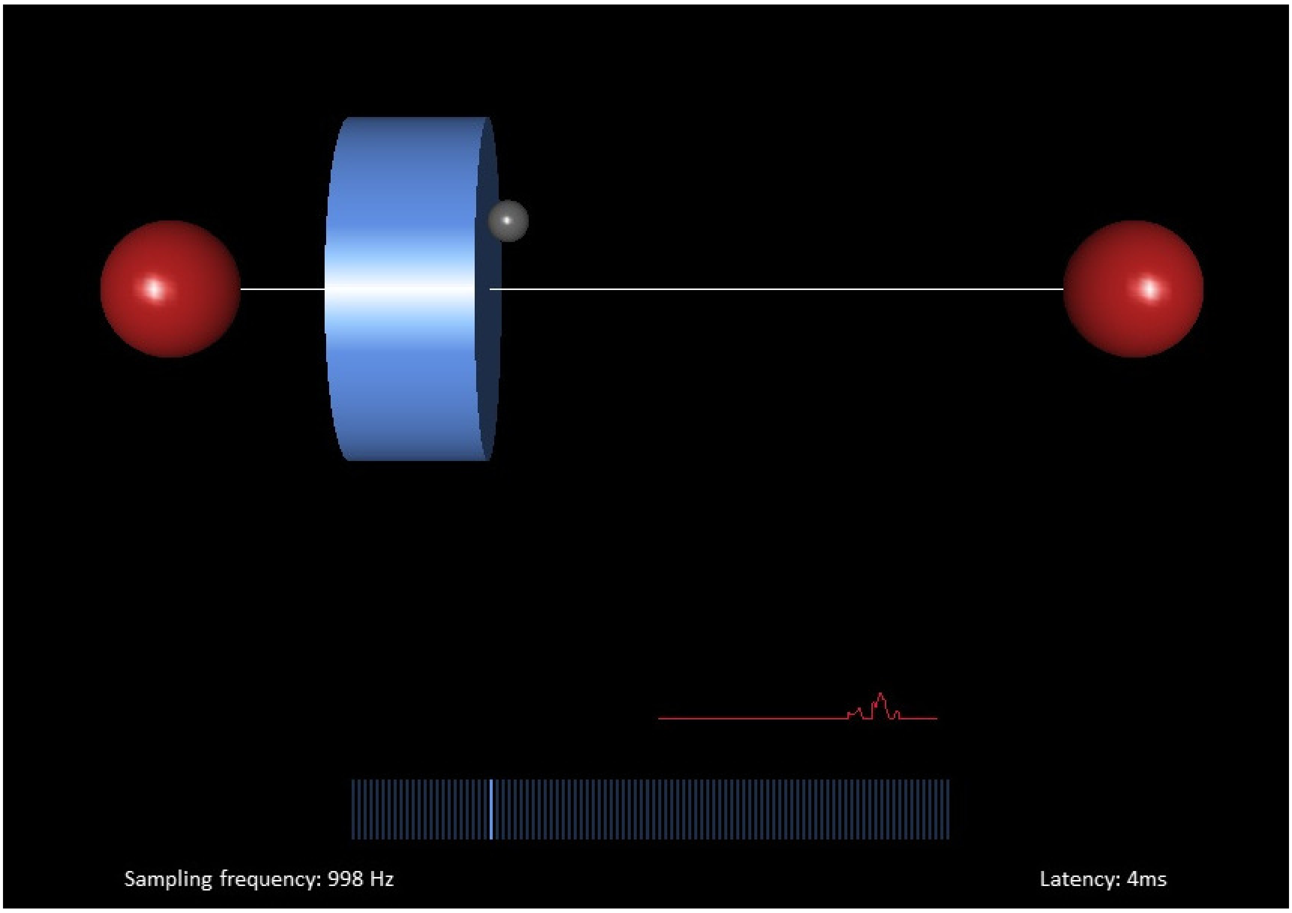


Figure 3.1: Impact of delay when performing the task of moving a cylinder from one edge of a wire to the other (screenshot from the application).

The main feature of this application is that it allows the increase or decrease of latency in the communication between the haptic device and the 3D cursor. This was implemented as a buffer of adjustable size between the haptic device and the cursor, inside the program that realises the haptic-capable virtual environment. The '+' and '-' buttons of the keyboard were used to manually increase or decrease the latency to a desired fixed value both for the information sent from the device to the virtual environment and reversely. This allowed to manually verify the

[2]A demonstration video is available on https://vimeo.com/349916619 and source code is available on https://github.com/constanton/DelayedTeleoperationDemo

latency threshold at which the haptic system stability and QoE decreased noticeably. For this specific task, it was found that by increasing the latency over 10 ms the QoE was starting to noticeably deteriorate as it became a lot more difficult for the haptic device to stabilise itself to the desired position. Therefore, the user was unable to perform the task efficiently as latency increased.

3.2.2 Master-slave experimentation framework

As described in [283], two Phantom Omni devices were connected to one PC each. Each PC was connected on the same network via an SDN-capable switch used for managing the traffic flow of the communication of the two haptic devices. The two Phantom Omni devices were used in a teleoperation set-up where one of the devices acted as a master and the other as a slave. Two force sensors were connected on the slave device in order to provide haptic feedback back to the master when the sensors (i.e., the slave device) were touching an object.

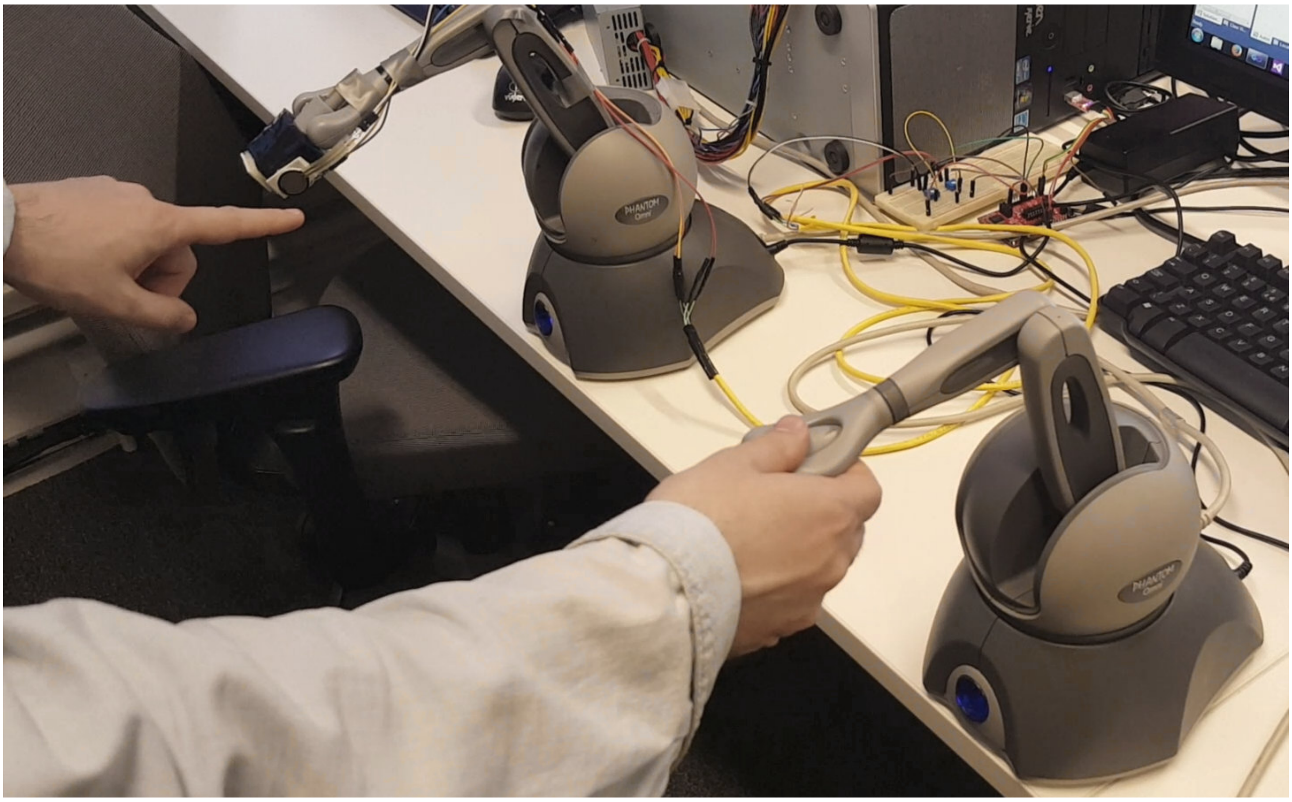


Figure 3.2: Two Phantom Omni devices (as master and slave) operating in position-force configuration over a network using an SDN-capable switch.

Computation of the kinematic equations to produce the force feedback information was made using an FPGA device at the slave side in order to reduce the computation time as much as possible. Another key element of this study is the use of an SDN-enabled switch. The experiment examined two cases of high QoS and low QoS.[3] In the case of high QoS, communication is achieved without issues as all teleoperation requirements are met in order to perform simple tasks of touching objects with the slave device's force sensors. On the other hand, when low QoS is emulated using the SDN-enabled switch by increasing packet loss to 90%, it is evident that, not only the reaction times of the master device are slower but also the perceived stiffness of the remote object is not equal to the object stiffness as measured by the slave device. The packet loss increase was emulated with OpenFlow meters. Even though the use of meters is not restricted to this functionality or purpose, in order to emulate packet loss the meters dropped packets when a certain data rate of incoming traffic exceeded a desired threshold.

3.3 A healthcare-oriented Tactile Internet application

Tele-surgery or remote surgery, is a natural evolution of tele-mentoring with the use of surgical robots. In tele-mentoring, on-site healthcare professionals are guided by another in a remote location, and the level of mentoring may vary from verbal guidance to remotely controlling the robotic arm. In the case of full remote surgery, the primary surgeon is at a site remote from the patient and all surgical tasks are performed by a robot controlled remotely. To enable this, and carry out the surgical procedure, the computer console (i.e., control console of a robot like the Da Vinci [284]) and the remote surgical device are connected by a high-speed ultra-reliable communications network [285]. Remote operation or consultation brings huge advantages to the healthcare system, as it allows the decentralisation of hospitals, and provides enhanced solutions for remote care. The main drawback of such operations is that by substituting the doctor's hands with current commercially available robotic arms, the surgeon loses the sense of touch, which is essential for palpation and for example, locating hard tissue or nodules (e.g. tumours), only relying on verbal and visual information.

[3] A demonstration video is available on https://vimeo.com/331503689

Current research in the robotics field is focused in the design of high-precision force or stiffness sensors, that can recreate the sense of touch when manipulating a robot [286]. In an attempt to combine the advantages of both remote medical practice and the enhanced haptic feedback for minimally invasive surgery, in this practical implementation, the overall operation system is decoupled. On one hand, there is the master domain, where the healthcare professional controls the operating side with the use of a wearable device. On the other hand, there is the slave domain, where a robotic probe performs a palpation task and senses the level of stiffness. In particular, when operating the robotic probe (or robotic finger), the doctor receives stiffness information of the palpation area, which is reproduced in the wearable device. With this, the doctor will identify accurately and in real time localisation of the hard nodules within the soft tissue, thanks to the stiffness information. All of this is done through a reliable high-speed communications network.

3.3.1 Sensory Perception and 5G Networks

As discussed in Chapter 2, 5G networks will support multi-modal communication with high QoE by enabling the exchange of audio, video, kinaesthetic and tactile data among a multitude of devices in real time, Among these four modalities, vision and kinaesthesia have the highest requirements in terms of bandwidth and low latency respectively. A high-definition 1080p video stream can reach up to more than 80 Mbps when minimally compressed, and a complex movement of a human palm touching an object may require latency down to 1 ms. Aside from the key performance indicators another important requirement is the synchronisation of all the data streams. Of course, as with current audio-visual communication systems, data compression and perceptual data reduction can be applied to both kinaesthetic and tactile information.

In the proposed experimentation framework, by combining video, audio and touch in a multi-modal communication system, a user is able to control the movements of the robotic finger using a haptic device, and can receive both visual and tactile feedback. Multi-modal communications require a per-flow QoS management, since every traffic type has different requirements: the real-time nature of the use case requires ultra low latency for all flows, but in terms of capacity, video and tactile traffic are very different. Also, the robot control information and the tactile feedback are considered critical traffic, and should be treated with higher priority.

To satisfy the QoS and ensure that critical traffic is always delivered, SDN is employed to differentiate the different flows on the network and dedicate a slice of the bandwidth to each flow, along with a high priority queue for minimising latency, when needed. In the following, the specific hardware and testbed configurations for the TI proof of concept is presented.

The purpose of the proof of concept described and evaluated in the next sections is to show two major aspects of 5G:

- Remote sensing with the use of low latency networks: the prototype presented in this article shows the potential of adding the sense of touch in a teleoperation system,via state-of-the-art sensing techniques and using a tactile Glove as the main controlling engine which is also capable of reproducing the tactile information.
- Network Slicing with the use of SDN: we set-up the transmission of multiple flows, in this case control, vision and touch, through separate communication channels. SDN allows systems to dynamically control their independent QoS depending on the network status and demand of each one of these flows.

3.3.2 Creating the Prototype

Controlled or Slave Domain: This proof of concept demonstration uses a state-of-the-art robotic probe platform developed in [286] as the haptic sensing device. The design of this robotic probe is largely inspired by the human finger and it abstracts its components to represent the way humans sense force. The robotic probe consists of two parts, a controllable stiffness joint, which represents the human metacarpophalangeal joints, and a force sensor at the base, which represents the function of a human tendon.[4]

[4]Demonstration video is available on https://youtu.be/L4nGXopLK8w

For the purpose of this demonstration, the stiffness of the joint is kept constant and the force felt at the base is measured using an ATI Nano17 Force/Torque (F/T) transducer. The F/T transducer connects to the computer via a data acquisition card. The sampling rate of the force measurement is taken at 1000 Hz. To allow the robotic probe to move in space, it is mounted on a *XY Table*, that guides the movement along the three axis (x,y and z), the XY-linear stage ANT130 (Aerotech Inc., resolution of 1nm), which is connected to the computer via Ethernet. Finally, the software for moving the probe and measuring the force data is programmed in LabView.

The soft tissue sample used in the demo was made of artificial materials replicating the tumour and biological tissue. We use artificial materials because the physical properties of real biological tissue can change over time. In general, the contrast in the stiffness between the malignant tumour and healthy fibroglandular breast tissue is quite prominent. Therefore, we can use the soft silicone and solid plastic material to approximately replicate the healthy tissue and the tumour respectively. The soft silicone phantom with an embedded hard nodule was made by inserting the hard nodule in between two layers of the phantom.

Master Domain: The proposed system made use of a tactile device that can act both as a tactile actuator and as a touch sensor. The tactile device is a commercial glove (Neurodigital GloveOne[27]) and can be connected to a PC using a USB cable or via a Bluetooth connection. It consists of two parts: a vibration subsystem which uses piezoelectric vibrators, placed under the fingertips and the palm, mainly used to recreate the sense of touch, and four conductive areas located at the centre of the palm, thumb, index and middle finger, that when in contact with each other create closed circuits. These closed circuits are represented by different gestures as shown in Figure 3.3, which we use to communicate the different movement commands to the probe.

For the haptic application, we set up a set of hand gestures using the glove to control the overall system. The first set of gestures is designed to control the movements of the robotic probe through the XY Table, each unique gesture makes the robotic probe move backwards or forwards over the soft tissue sample. The gesture information is sent using UDP packets to

(a) Forward (b) Backwards (c) Reset

Figure 3.3: Glove gestures for robot commands. Red circles show the contact points of each gesture.

the LabView software, which moves the robot accordingly and returns readings captured by the robotic probe back to the haptic application, which translates them into vibrations on the glove.

The second set of gestures is designed to control the configuration of the system. In particular, two messages are sent: the first one is sent to LabView and resets the robot's position, while the second message is directed at a Python script that changes the configuration of the SDN controller. Detailed information on the configuration of the SDN controller is given in the following section.

Visual Feedback: Two camera feeds are also included in the testbed with the main purpose of sending visual feedback of the robotic probe to the master domain. The main camera feed is full HD at 30fps while the secondary feed is 720p, also at 30fps. The cameras we use are Microsoft LifeCam HD webcams that feature a built-in hardware encoder and connect to the computer via USB. This reduces processing delays as the computer does not need to encode the video stream. For low latency video streaming, we use UltraGrid[5] [287] and it is configured to transmit the raw video as it receives it from the cameras, to avoid increased delay in software compression.

Figure 3.4 depicts the high level application functionalities of the system and details the message

[5] *UltraGrid* is a software implementation of high-quality low-latency video and audio transmissions using commodity hardware

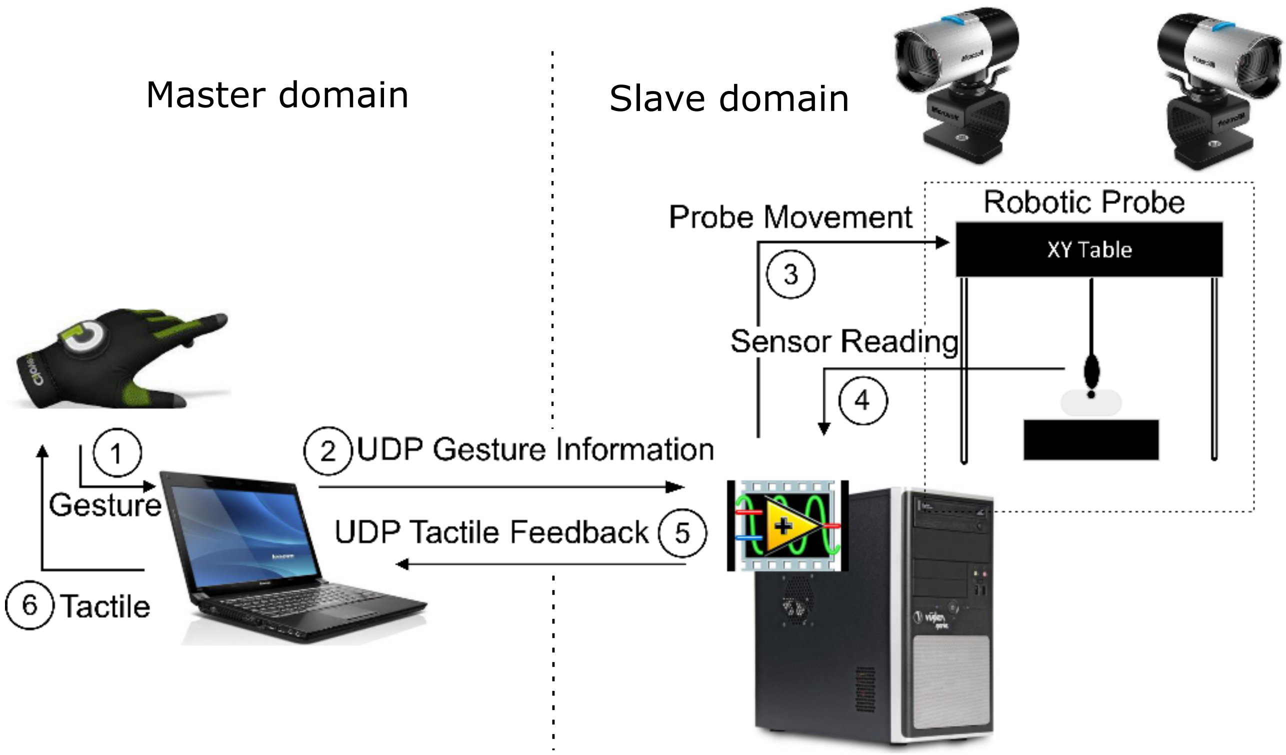


Figure 3.4: Illustration of the prototype demo detailing the application level functionalities

exchange from the two sides of the communication, the master domain and the slave domain, as described previously. This figure presents the end-to-end system for the communication of the haptic glove and the robotic probe, detailing the application level functionalities.

3.3.3 Connecting with SDN

The network that this proof of concept is emulating is one where network nodes exchange different types of traffic that the network operator is aware of. According to the requirements of each type of traffic the network can take decisions for QoS provision, i.e., on how to manage the traffic flows in order to satisfy the desired application requirements. This requires that the network operator has control over the entirety of the network which is a main assumption of this experiment. An example of such a network can be an industrial private network.

The main objective of this proof of concept is to implement traffic differentiation so that independent QoS metrics can be applied on a per-flow basis. According to the NGMN Alliance (Next Generation Mobile Networks), a network slice instance is a set of network functions and resources to run these functions that may be fully or partially separated from other network slice instances [288]. Hence, by differentiating the traffic and allowing it to be treated independently with different forwarding policies, we create a proof of concept of the physical and logical resource reservation aspect of a network slicing instance [289]. At this point, the concept of network slicing for network functions is left out of scope of this work and we focus on the resources for network slicing on a sub-network instance.

In this setup, an OpenFlow switch is used along with an OpenDaylight (ODL) SDN controller. ODL is a production quality controller supported by the Linux Foundation and has a modular design that allows us to load only the set of functions we wish to use. For this proof of concept, we create a minimal setup that allows us to push flows to the switch and also enables a representational state transfer (REST) interface for pushing configurations onto the controller.

The setup consists of a PC running LabView for controlling the robot over the network, along with Ultragrid for streaming video from the two cameras. We also run a Linux virtual machine (VM) for ODL which directly binds to one of the network interfaces of the PC for OpenFlow communication with the switch. The second network interface of the PC is used for the traffic coming from the camera, LabView and glove flows. Additionally, a virtual 'host only' network connects the Linux VM to the host operating system (OS) for passing configuration messages to the SDN controller. Finally, we use a laptop for running the haptic software and receiving the camera feeds. The physical setup along with the different connections is illustrated in Figure 3.5.

Traffic is distinguished between different applications based on the source and destination IP addresses as well as the TCP/UDP ports. Since our implementation uses traffic shaping on each slice, this setup can be considered as a use case of an edge switch where inbound traffic is placed on a slice and a set of configurations is applied in order to guarantee a minimum rate, and also implement a cap so that it won't interfere with other traffic. The minimum rate is configured using strict priority queuing where the highest priority queue is served first until it is empty, and subsequently the same process is repeated for all the other queues. OpenFlow meters are configured in each slice with a "Drop" action, once a predefined threshold

Table 3.1: Slice setup on physical ports, queues and meters

Nodes	Flow Direction	Port/Queue	Meter/Action
PC to Robot	LabView to XY	P1/Q7	512Kbps/Drop
Robot to PC	XY to LabView	P2/Q7	512Kbps/Drop
PC to Laptop	CamA to UltraGrid	P3/Q7	80Mbps/Drop
PC to Laptop	CamB to UltraGrid	P3/Q7	60Mbps/Drop
Laptop to PC	Glove to LabView	P2/Q1	–
Laptop to PC	Glove to ODL	P2/Q2	–
PC to Laptop	LabView to Glove	P3/Q1	–

is exceeded, limiting the amount of bandwidth that a configured slice can use. Since queues control the egress rate while meters control the ingress rate, this setup allows us to control both aspects of the traffic per physical port. The slicing configuration is presented in Table 3.1, along with the physical switch ports that each node is connected to. In order to configure the minimum and maximum rates for each slice, we performed measurements for each application using Windows performance monitor and a legacy gigabit Ethernet (GbE) switch in order to understand the bandwidth requirements of each type of traffic generated by the cameras, the software controlling the glove, the robotic probe as well as the software controlling the XY table. All physical links corresponding to these components are shown in Figure 3.5. Based on the collected performance results, we set up the queues rates as shown in Table 3.2. This configuration environment is emulating a scenario of congestion on an edge switch which apart from the critical traffic also serves Best Effort (BE) traffic. Therefore the rates and link capacity used throughout the experiment are chosen so that congestion conditions are reached.

The Python script runs on the host OS and communicates with the ODL via the REST interface over the virtual network between the host and the VM. There are two configurations stored on the Python script, which are quite similar apart from the secondary camera's flow. This flow is restricted to a lower maximum rate on the OpenFlow meter, which results in disrupting its feed. The change in configuration is triggered by a gesture of the glove in the master domain, and upon repeating the gesture on the glove, the original configuration is restored by the Python script and the camera feed is restored. This change in the SDN controller configuration reprograms the switches rules for a certain flow, allowing to show how the traffic is completely unaffected while one flow might be completely disrupted, the QoS of the other

Table 3.2: Queue configuration per egress port

Port/Queue	Min Rate	Max Rate	Burst
P1/Q7	1Mbps	1.5Mbps	188Kbps
P2/Q7	1Mbps	1.5Mbps	188Kbps
P3/Q7	150Mbps	150Mbps	28Mbps
P3/Q7	150Mbps	150Mbps	28Mbps
P2/Q1	512Kbps	512Kbps	–
P2/Q2	512Kbps	512Kbps	–
P3/Q1	512Kbps	512Kbps	–

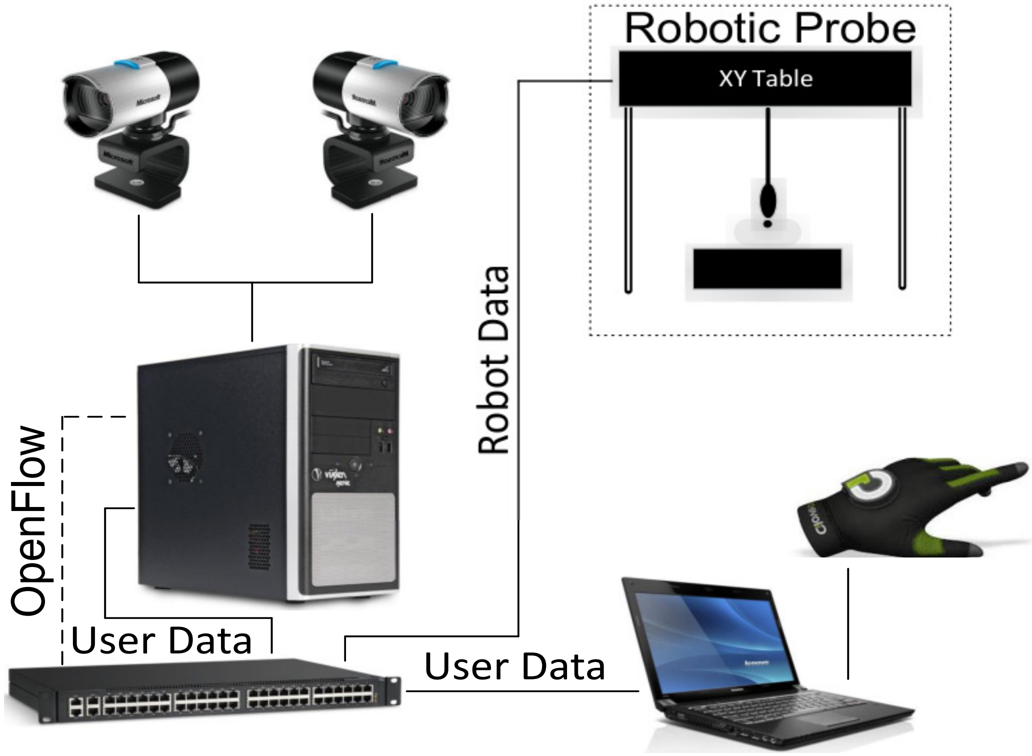


Figure 3.5: Physical links among components of the prototype demo setup

data feeds remains unchanged.

3.3.4 Experimental Setup and Performance Discussion

For testing purposes, we set up three different physical hosts with one of them acting as a traffic sink (receiver) and two of them acting as traffic generators (transmitters). We configure *iperf3* to generate background traffic at 400 Mbps from one of the hosts while the second host generates traffic at 80 Mbps, which represents one of the camera feeds. The BE traffic data rate was chosen to be large enough to create congestion conditions considering the 1Gpbs

Table 3.3: Testbed configuration

Parameter	Value
BE Traffic	400 Mbps
Video Traffic	80 Mbps
Traffic type	UDP
Switch capacity	1 Gbps

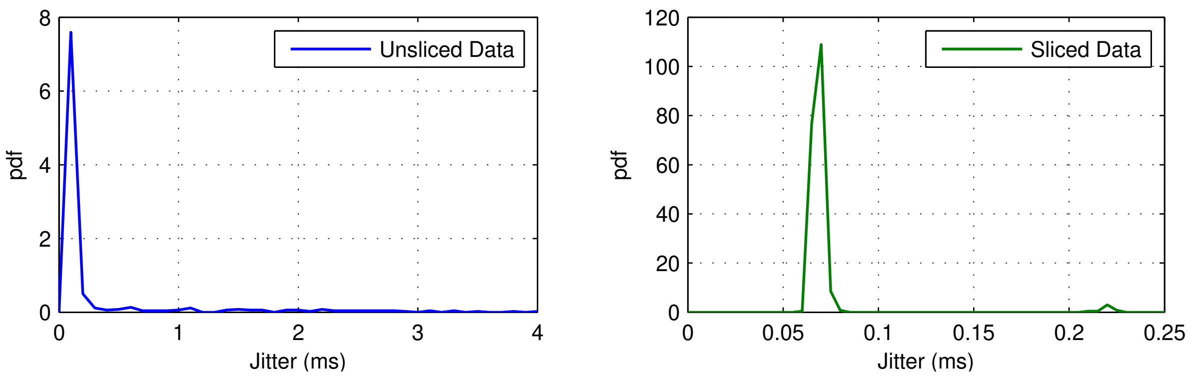
switch capacity and the expected queuing performance of the switch. A 50% maximum link utilisation limit was considered in this case, which was empirically verified before executing the experiment. The reason for analysing the camera traffic is that it requires a high bandwidth as well as the lowest possible latency to synchronise with the tactile traffic. In particular, to allow for a correct functioning of the system and a correct overall user experience both visual and tactile feedback should have similar end-to-end latencies. To verify the functionality of traffic differentiation through network slicing and the proposed QoS management implementation, key performance indicators of latency, jitter and packet loss are measured for both un-sliced and sliced network set-ups. Furthermore, we consider one switch in our testing scenario, however the results obtained can be easily extrapolated to a network with a higher number of switches.

The experiment consists of 6 sets of measurements for two different configurations. In the first configuration, the switch assigns the critical traffic flow to a high priority queue, whereas in the second configuration it doesn't. The experiment lasts for 90 seconds for each configuration with link statistics reported every second. These sets of measurements are then used to create one set with their average. This generates a data set of 540 measurements for each set-up. Both traffic generators transmit UDP packets towards the sink host. Since the camera traffic is considered to be critical for the purpose of this use case, the traffic flow is placed on the higher priority queue in the sliced network scenario. The rest of the traffic generated is treated as BE. Table 3.3 summarises the test-bed configuration parameters for the performance evaluation.

The observations from benchmarking show that slicing can provide a benefit even in scenarios where there is a single switch on the network. Statistics such as one-way latency and jitter are given by iperf3 which takes measurements during the communication of the iperf3 UDP client and server. The one-way latency measured in the camera traffic was very much consistent:

8.5 ms for the unsliced scenario and 0.7 ms for the sliced scenario. This difference in latency is given because of the nature of the traffic treatment, in the unsliced configuration all traffic goes to the same pool and it is pushed out of the switch without any differentiation, having the same forwarding rules despite being BE or critical. When slicing is configured, the critical traffic is always mapped to the highest priority queue, which is served always first, as pictured in Figure 2.9.

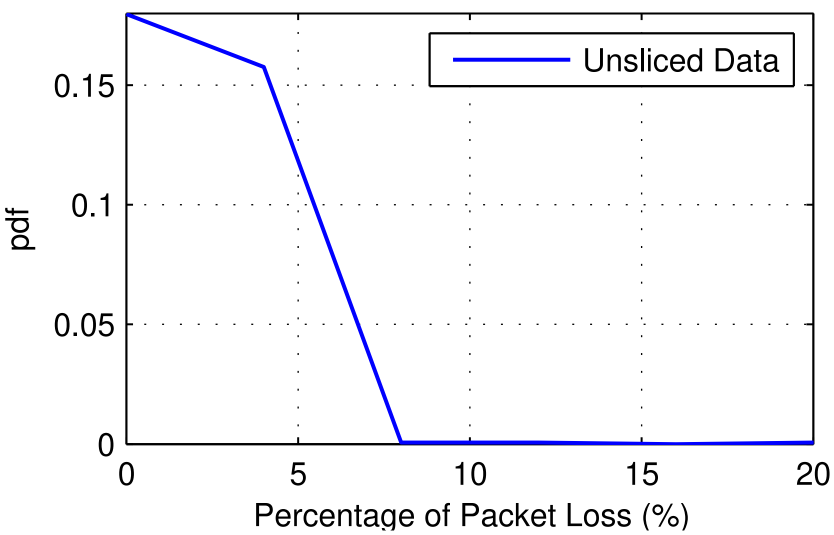
The same rationale applies to the jitter. Figure 3.6 compares the Probability Density Function (PDF) of the measured jitter before and after the slicing. From the results it is appreciated that both average and standard deviation of the experimental jitter is reduced. Notably, the average jitter is reduced by 88%, and the standard deviation is reduced by 98%.



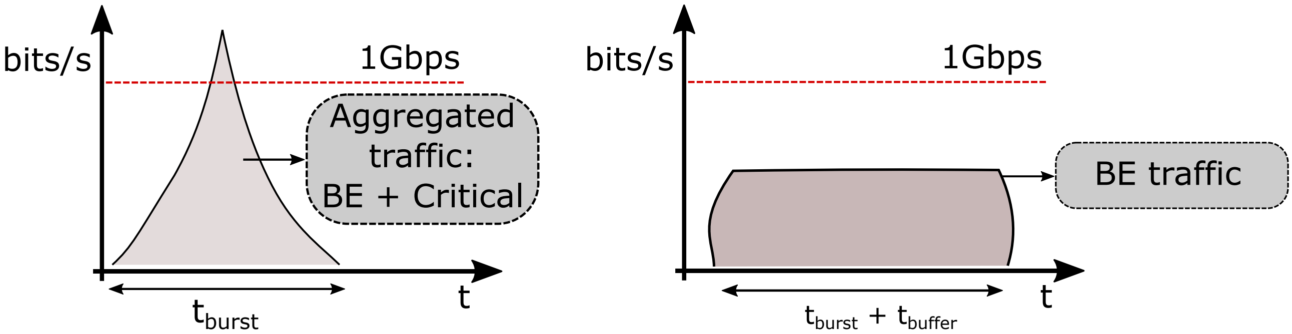
(a) PDF of jitter in unsliced configuration (b) PDF of jitter in sliced configuration

Figure 3.6: Jitter comparison

When comparing packet loss there is a similar effect when going from an unsliced to sliced configuration; Figure 3.7a shows the PDF of the packet loss in the unsliced configuration. In particular, when traffic is not differentiated and no QoS management is performed, the bursts of traffic that ingress the switch can be aggregated at some point in time, resulting in traffic bursts higher than the switch capacity, as sketched in Figure 3.7b. Since there is no traffic matching, the switch will discard packets to satisfy its maximum allowed data rate. When slicing traffic, the camera flow is always placed in the highest priority queue, and the BE traffic is forced to wait until higher priority queues are emptied to be served. As a result, there is no packet loss in the camera flow, but the BE traffic suffers an increase in the queuing latency, (t_{buffer}). According to these results, when operating under congestion conditions, which can



(a) PDF of packet loss in unsliced configuration



(b) Traffic burst and packet loss

Figure 3.7: Packet loss in experimental setup

lead to packet loss for all types of traffic without congestion control, traffic differentiation can guarantee good QoS for critical traffic. On the other hand, due to the OpenFlow priority queuing mechanism, QoS provision for critical traffic may result in reduced QoS for non-critical traffic.

3.3.5 Upgraded version of the haptic application

An upgraded version of this application was created and demonstrated at Mobile World Congress 2017. While the network functionality of this demonstration remained the same, the hardware components and relevant software used made the interaction of the user with the remote environment while using the glove more intuitive and natural providing improved telepresence and immersion in the remote environment.

In this new version, on the master side, the user was equipped with an upgraded version of the tactile glove used in the older version, called Neurodigital Avatar. The new glove has accelerometers and gyroscopes placed in the fingers and the centre of the palm. Apart from the sensors on the glove the user has to wear three more sensors on. One sensor around the forearm, one around the arm and one around the chest. Together with the glove all components are connected to a PC running software that can extract the 3D position of the user's hand as well as the user's fingers.

As the user's hand is moving and according to the speed of the movement on each of the six axis (up, down, front, back, left and right), once a speed threshold is exceeded, a position command was sent to a Universal Robots UR3 robot arm controlled by a Linux machine running Robot Operating System (ROS). The position commands directed the robotic arm to move towards the desired direction of the user.

Furthermore, the UR3 robot arm was extended with a robotic device with the capability of sensing stiffness by touching the surface of human organs using an optical sensor [290]. This instrument was used to generate the tactile feedback information sent back to the glove. The glove's vibrators became active with a variable vibration level depending on the value of the stiffness observed in the remote environment.[6]

3.4 Experimental Study of Softwarized 5G: Effects of Mission-Critical Traffic on CN

In this section, an experiment based on a mission-critical traffic use case is presented. This part solely concerns the impact of CN on the BE traffic generated by an ambulance in which a patient's condition is assessed by paramedics and doctors located in a hospital by means of a haptic communication system which uses multiple radio access technologies (including a 5G

[6]Demonstration video is available on https://youtu.be/GJFLZDskJmk

mmWave cellular system) to access the 5G network.

Reliable and timely delivery of mission-critical data over wireless alone is not sufficient to guarantee the required end-to-end reliability, since the CN part has to also be considered. Particularly, the transport network plays an important role in the successful delivery of KPIs, such as reliability. Both mission-critical and BE traffic will have to travel through a complex network while competing for resources during transmission, buffering, and computing. In order to achieve the desired levels of end-to-end reliability, the mission-critical traffic has to be properly marked, identified, and prioritised at all the network elements.

3.4.1 Representing softwarised 5G data networks

For the proposed experiment a network comprising SDN-capable switches connected to a number of servers is considered. These servers are running various 5G VNFs in both the user and the control plane. These functions are the 3GPP 5G architecture functions described in Section 2.6.3. The physical network topology is illustrated in Figure 3.8. The purpose of the emulated network is the management of mission-critical traffic and BE traffic coexisting within the network.

For the proposed experiment, a network of softwarized functions is considered as shown in 3.8 in order to emulate the management of mission-critical traffic and BE traffic coexisting within the network. These functions are the 3GPP 5G architecture functions described in Section 2.6.3.

To make sure the QoS is aligned across the entire user plane path, the SDN controller has to receive the traffic shaping configuration from the control plane function in charge of the traffic steering and policy/QoS control (according to the considered architecture, the SMF). Between the access network and the UPF, switches interconnect VMs (running VNFs) to form the service chain (as shown previously in Figure 3.10). The SDN controller will then configure all these switches, so that these are able to differentiate between the various types of traffic, including the mission-critical one, accordingly. In other words, SDN controller reserves a slice of the transport link capacity, such that the network can satisfy the minimum guaranteed bit rate at all times as well as ensure no packet losses.

While the OpenFlow controls the application of traffic forwarding rules, the Open vSwitch Database (OVSDB) protocol is used for the configuration and management of the OpenFlow switches. OVSDB provides an interface for the centralised controller to configure the underlying hardware of switches and thus meet the network requirements. The parameters can be configured from the SDN controller, which includes configuring data-paths and assigning ports to them, as well as managing link speeds, queue configuration, and policing.

As mentioned in the previous chapter, SDN is able to differentiate and manage the different flows arriving from the access network and apply the necessary forwarding rules to ensure that the QoS is maintained throughout the network. The SDN controller can modify the forwarding rules for each flow that are kept at the infrastructure layer in the form of a flow table. In order

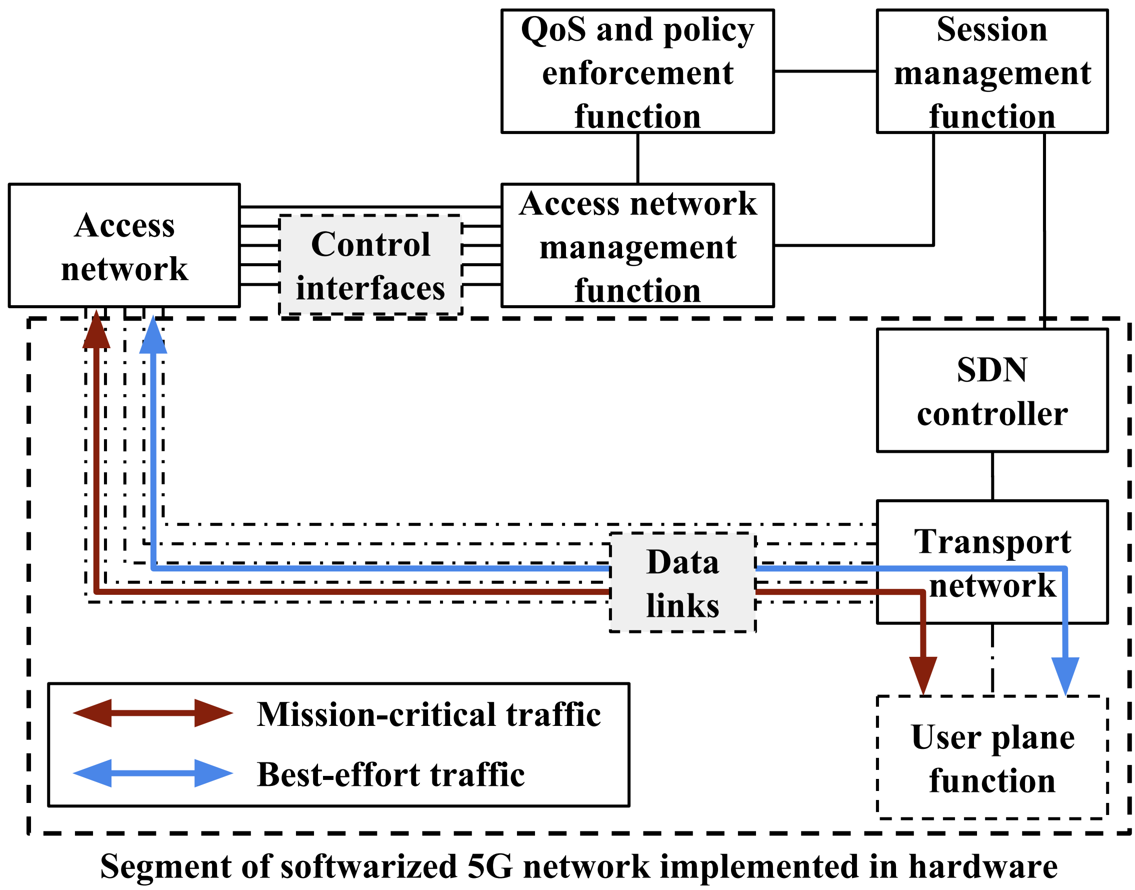


Figure 3.8: Softwarized 5G network and its segment used for the study on the data flows coexistence that was implemented in hardware.

to maintain the priority of the critical flows with respect to other flows (e.g., BE traffic), the SDN controller maps a flow onto a priority queue. For the purposes of this study an OpenFlow switch is utilised operating as described in Chapter 2 and illustrated in Figure 2.9.

3.4.2 Implementing softwarized 5G data networks

For this test implementation, equipment similar to that utilised in a practical 5G CN is employed. To demonstrate the end-to-end approach advocated in this section, a simple but representative proof-of-concept implementation is presented using a Pica8 SDN-capable switch and five Linux servers connected to it.

In this test implementation, one of the Linux machines acts as the SDN controller, while others are hosts used to either generate or receive data traffic. Thus generated data flows of UDP packets represent either mission-critical or BE traffic, which is created with iperf. The latter tool also allows measuring bandwidth, jitter, and packet loss. An L2-switch module is additionally installed inside the controller, making the switch capable of handling ICMP packets, to be able to use ping in order to assess the round-trip time (RTT).

Further, the ODL SDN controller is employed to manage the SDN-capable switch. The communication between the ODL controller and the switch is based on the OpenFlow protocol, and in this case has been implemented via the REST API in order to apply the appropriate flow configuration to each switch. The controller's role is to assign the data flows initiated by the Linux hosts to the queues. To distinguish the critical and the BE traffic, the destination IP is identified as encapsulated in the packets. The queues prioritise the data flows at the egress interface of the switch as displayed in Table 3.4. The minimum (guaranteed) and maximum service rate need to be maintained in each of the queues to avoid disturbance from the BE traffic toward the critical traffic, since the sum of both data flows tends to reach the full capacity of the link.

The developed platform is used to obtain the numerical results in the following section.

Table 3.4: The experiment settings

Nodes	Traffic type	Queue	Queue min and max rates
APa to UPFa	Critical	Q1	35Mpbs and 50Mbps
APb to UPFb	Best-effort	Q0	20Mbps and 55Mbps

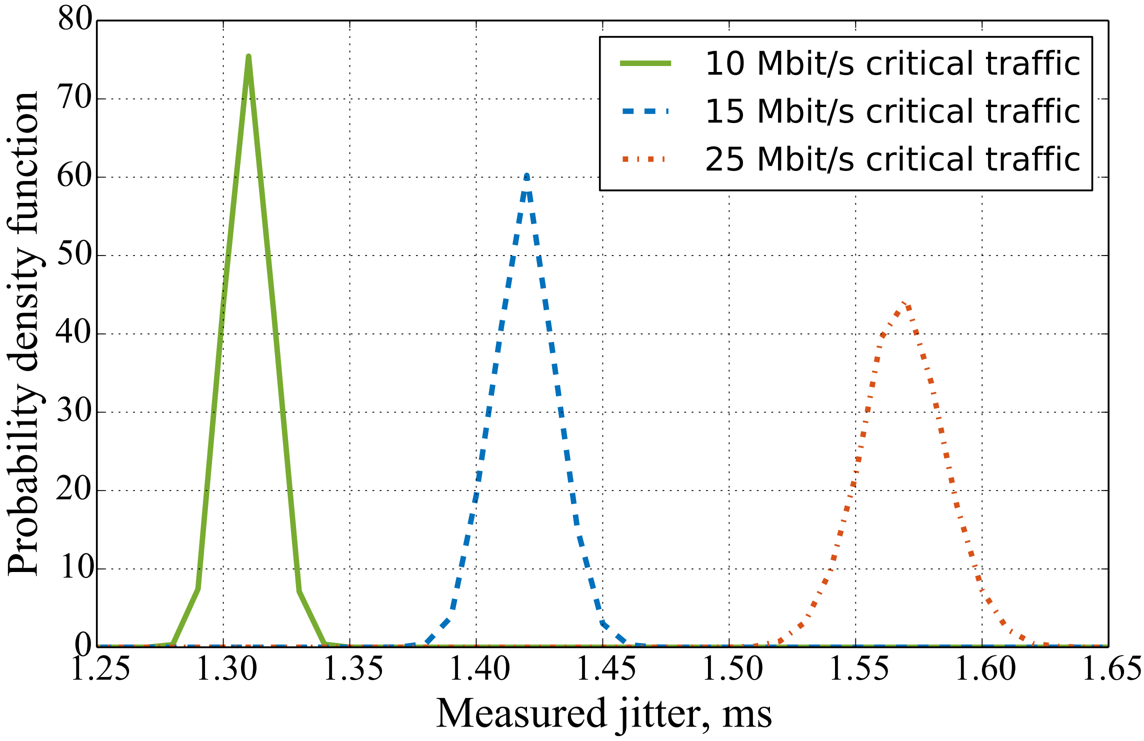


Figure 3.9: Measured jitter when best-effort and mission-critical traffic coexist in CN.

3.4.3 Numerical Results

Similarly to the previous experiment conducted in Section 3.3, the rates and link capacity used throughout the experiment are chosen so that congestion conditions are reached. More specifically, for critical traffic the rates used were 10, 15 and 25 Mbps, whereas BE traffic was kept fixed at 50 Mbps. These numbers were chosen considering the requirements of URLLC in a mission critical scenario of medical treatment in an ambulance. The mission-critical scenario requirements correspond to the first row of Table 2.4.

In this experiment, good QoS for the critical-traffic is guaranteed and therefore we are interested in examining the impact of such QoS provision on the BE traffic. In this case, according to the configuration of the queues of the SDN-capable switch in Table 3.4, packet loss is not expected due to the chosen traffic rates and the rates the switch is configured to satisfy. Therefore, in this experiment we examine the jitter of BE traffic.

Figure 3.9 presents an empirical pdf for the jitter of the BE traffic. Comparing the results for 10 Mbit/s, 15 Mbit/s, and 25 Mbit/s data rate of the mission-critical traffic, we note that both the average value and the variance of jitter become considerably worse at higher loads. We also notice that the absolute values of the averages for 10 Mbit/s and 25 Mbit/s are on the order of 0.2 ms, with the relative difference of over 20%. Since only a segment of the real network has been used for our measurements, the ultimate performance degradation caused by serving mission-critical traffic for the jitter of the BE data can become even more severe. To avoid the effects of jitter, data buffers are used which in turn increase latency which can hinder a service with non-prioritised traffic, e.g., traffic of a Voice-over-IP service.

3.5 IEEE Tactile Internet Standards working group

As verified from the experimental frameworks proposed in the previous sections, the 5G enablers, such as SDN and NFV will impact the implementation and performance of the TI use cases. Nevertheless, for successful deployments of 5G system implementations globally, even though incorporating such enablers is necessary, there is also the need for a standardisation framework regarding the TI.

The necessary standardisation efforts for the TI have already been initiated and are under development within the IEEE P1918.1 "Tactile Internet" Standards Working Group. These efforts include a detailed description of TI use cases and the definition of a system-wide functionalities and architecture for the TI to support use cases of human-in-the-loop systems for physical interaction over URLLC-capable communication infrastructure. Furthermore, P1918.1 includes the "Haptic Codecs for the TI" task group, also known as IEEE 1918.1.1, which concerns the specification of haptic communication codecs for the exchange of kinesthetic and tactile information [291].

Within the P1918.1 architecture, two main domains exist, the Tactile Edge (TE) domain and the network domain. A TE includes important components such as the Tactile Device (TD) and other components which allow exchange of information with other TEs via the network domain which is responsible for providing communication resources. It needs to be noted that the proposed architecture allows for the exchange of information between TEs and the network domains by means of interfaces that allow interconnection of TI architecture components. The aforementioned interfaces allow connectivity :

- Between the TE and the network domain (Access Interface).
- Among the entities within the TE and if needed with the network gateway controller that might reside within the TE (Tactile Interface).
- With elements that provide edge cloud services to the TE (Open Interface).
- For the network control plane elements (Service Interface).
- Among the network domain elements (Network Side Interface)

Interfaces allow interactions that enable the operation of the tactile applications and the adjustment of the end-to-end components to satisfy the requirements of applications.

Furthermore, the proposed architecture is agnostic to the communication infrastructure technologies that can possibly be used, presenting entities and their interfaces designed to enable three main TI communication paradigms. These paradigms present different KPIs and procedures for instantiating the TI services:

1. **Omnipresent**: this paradigm involves constant activation of all components of the TI architecture, to ensure quick coupling of a newly instantiated TDs to the TI infrastructure and bootstrapping of the end-to-end communication. The large scale of the infrastructure involved in this paradigm makes it very difficult to manage and maintain.
2. **Ad hoc**: this case concerns close proximity of TDs where a local, small scale instantiation of TI components is required. An advantage of this paradigm is the minimisation of resource requirements, but compared to the omnipresent paradigm, it has an increased setup time.

3. **Hybrid**: a solution to the high resource management requirements of the omnipresent paradigm is to involve special TI components to act as meeting point entities that will manage and handle TI sessions of TDs as well as the setup of TI components that will be needed for each of the sessions.

3.6 Network architecture for end-to-end services using SDN and NFV

Network virtualisation and softwarisation are actively studied within the research community and improvements in both topics will impact the implementation and performance of the TI communication paradigms presented in Section 3.5.

Several examples in the context of H2020 EU projects exist, providing architectural solutions that can satisfy the requirements of different use cases (e.g., 5GCAR); contributing to softwarisation of radio access and core networks (e.g., 5G COHERENT); and focusing on the integration of network services in an orchestration platform (e.g., SONATA). While there are numerous existing approaches to construct a flexible 5G network – [292, 293, 294] among dozens of others – the proposed architecture combines the components, which are key in delivering QoS-enabled services in well-established solutions (such as, e.g., 3GPP and ETSI NFV). It combines both RAN- and CN-related aspects in the context of QoS and policy management, with the objective of providing overall end-to-end reliability, ensuring consistency between the enforcement of policies across the entire chain of the provided network service.

The methodology presented in this section considers the 5G system-level architecture as its baseline, where distributed software-defined decisions are made independently for both the access and the core network. The end-to-end QoS/Policy framework is thus a combination of distributed functions that monitor network performance and enforce QoS policies based on the session's context-related information as well as a centralised PCF that ensures an efficient coordination and alignment among all these distributed functions.

Within the 3GPP community, there has been a lot of discussion on how to ensure adequate alignment between the QoS enforcement rules in the User Plane network functions (such as interactions between the PCF, SMF, and AMF) and the QoS enforcement rules in the transport network. To this end, in this section, the proposed mechanism is the interfacing of the SMF with the virtual SDN controller ("vSDN Ctrl" in Figure 3.10). This allows the 3GPP system to control the forwarding rules in the network infrastructure and thus satisfy the service requirements. In a network with multiple flows that require different levels of QoS, SDN allows to dynamically control their respective KPIs depending on the effective demands of each of these flows. In this sense, once the UE/Access Network (as seen Figure 3.8) initiates a service with a particular QoS requirement, the SMF will be in charge of enforcing those rules in the physical network as well as ensuring that there is consistency between radio access, core, and transport networks. As an example, QoS enforcement rules can be applied on the virtualised SND-switches by the virtual SDN controller differentiating incoming traffic from the Access Network. The communication between the virtual SDN controller and the switches can be made, e.g., using the OpenFlow protocol which can enable the use of its priority queuing framework among all switches.

In more detail, Figure 3.10 outlines the standard ETSI NFV architecture, where the bottom layer comprises a set of virtualised resources running on top of the physical infrastructure. Network functions are then introduced as Virtual Machines (VMs) or Containers running inside the virtual infrastructure. In this work, all the network functions are considered to be running as VNFs within the vitalised infrastructure i.e., cloud infrastructure. While there are multiple ways to encapsulate VNFs, including containers [295] and unikernels [296], in this work VMs are discussed as an illustrative example, since paravirtualization is an established technology and OpenStack is widely used in telco provider clouds [297]. These network functions can perform various operations, such as load balancing, firewalling, switching, 3GPP-defined network functions, as well as act as a virtualised SDN controller responsible for managing VNF instances of OpenFlow switches.

As shown in Figure 3.10, the box drawn to contain the 3GPP components shows the network service as a logical abstraction where interconnected components form the service chain, illustrated by a VNF forwarding graph. The QoS management and policy enforcement functions within the user plane are controlled by the SMF, which is also in charge of the traffic steering at the UPF to route data to its intended destination. To ensure appropriate mapping of QoS onto adequate forwarding rules in the transport network, the SMF talks to the vSDN controller during the session initiation. The vSDN controller configures the transport network elements with the forwarding rules.

The software instances, as seen in the Network Service (NS) dotted box, use virtual resources enabled by the virtualisation software running in the Network Functions Virtualization Infrastructure (NFVI) part of the figure. Of course all these are managed by the Management and Orchestration (MANO) software as directed by the relevant service descriptor.

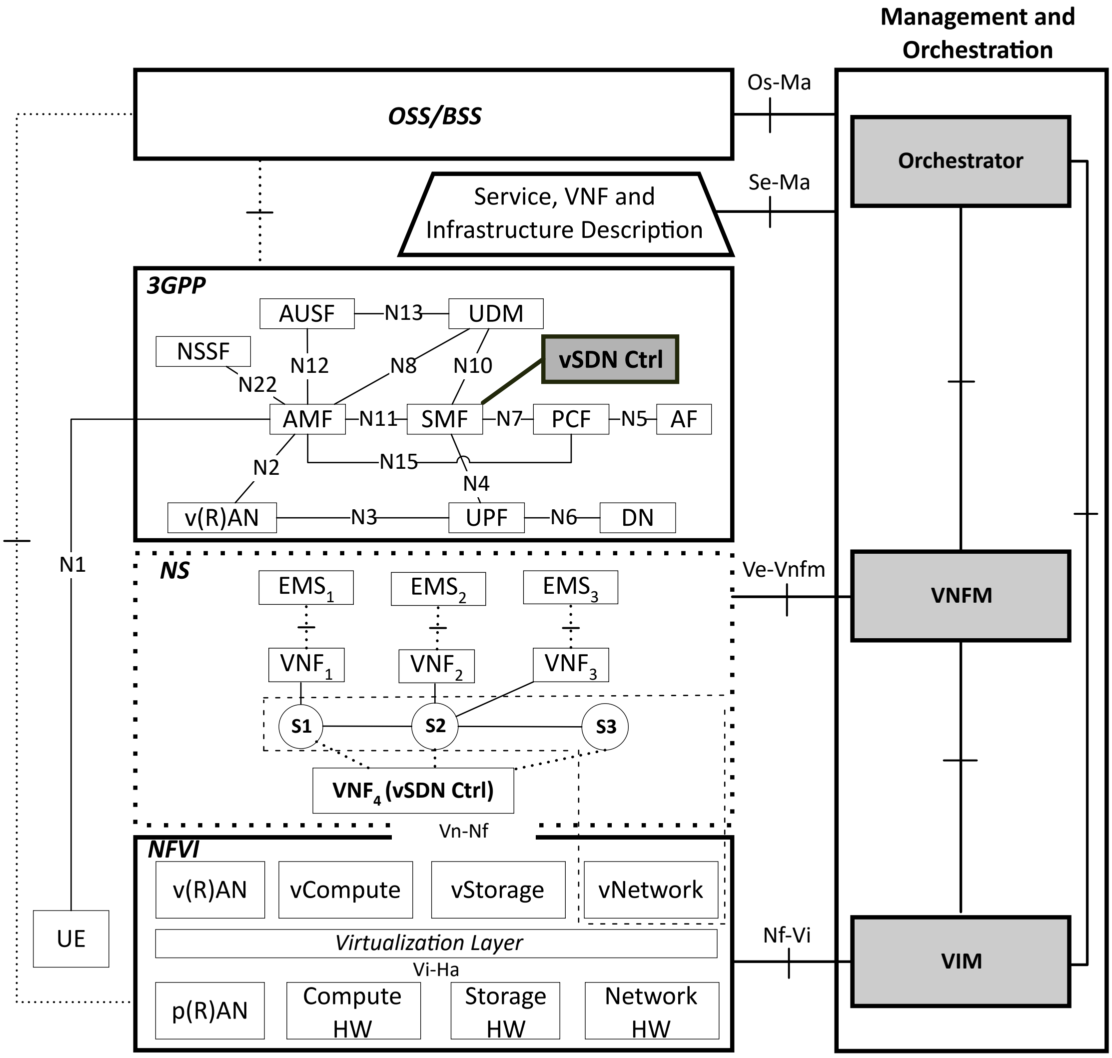


Figure 3.10: Considered architecture of a softwarised 5G network from [282].

The envisaged system employs SDN virtualisation as a means of offering control of the virtualised network infrastructure to network services, such as 3GPP architecture, on top of the physical network infrastructure, which is managed by an SDN controller. An example of SDN virtualisation can be found in [293], where the authors present an architecture that allows virtualised SDN infrastructures to be controlled by virtualised SDN controllers. The latter reside on top of the physical SDN infrastructure, which is managed by an operator's SDN controller.

Within the proposed architecture, all 3GPP network functions are running as VNFs; hence, a virtualised SDN controller becomes an important network element that can be contacted by any of the 3GPP functions [298]. More specifically, the SMF can instruct the SDN controller to adjust flow rules within the virtualised 3GPP network for the QoS management purposes. This enables full control of the virtualised network resources by the 3GPP system as opposed to having a static configuration that is dictated by the NFV operator's SDN controller. The existing solutions developed in the context of intent APIs, such as the ONOS Intent Framework and the OpenDaylight Network Intent Composition, offer comfortable configurability of the network interfaces but, at the same time, are featured by the limited functionality, which is insufficient to support the intended compound solution. Therefore, the proposed architecture suggests that the physical SDN controller manages the physical infrastructure that connects the physical nodes of the Virtual Infrastructure Manager between them as well as to external networks. In its turn, the virtualised SDN controller manages virtualised switches created as VNFs to support the virtualized 3GPP functions deployed within the tenant's slice.

There is currently no interface defined for the virtualised SDN controller to communicate with the physical SDN controller. The said interface is useful for the purposes of communicating its requirements to the physical infrastructure and thus influencing its configuration according to the 3GPP network requirements. This approach introduces benefits in terms of resource utilisation efficiently within the virtual environment and can therefore reduce the operating costs of a virtualised 3GPP system as well as enhance its manageability and flexibility. Indeed, it would be possible to create an interface between the virtualised SDN controller and the physical SDN controller; however, such an approach would break the separation between the control plane and the user plane of the NFV architecture and could potentially introduce security problems for the NFV operator.

At the higher layer of the architecture, network service descriptors as per the NFV specifications are utilised to describe the network services and network functions, while the OSS/BSS is used by the NFV operator for management and billing purposes. On the right-hand side of Fig. 3.10, the MANO layer spans across all the layers of the NFV architecture and is responsible for managing the infrastructure, monitoring the status of the network functions, coordinating their life-cycle, and finally maintaining a catalogue of descriptors for the network services and functions that can be deployed in the system.

To implement a 3GPP system, the NFV operator needs to provide VNF descriptors to the Orchestrator. These descriptors contain configuration and deployment information, such as the number of CPU cores, RAM, storage, and network interfaces for each VM that will host a VNF. The descriptors can include additional information that is specific to the VNF, such as 3GPP configuration options for a Packet Gateway (PGW). The Operator will need to on-board the VNFs to the VM by adding the images of these VNFs into the virtual storage pool. At this point, a Network Service descriptor can be used to create a complete network service by including all of its component VNFs and the virtual network infrastructure that will connect them together to deliver a functional 3GPP system.

3.7 Summary

In Sections 3.2-3.4, four different experimentation frameworks were presented. The first experiment focused on evaluating the impact of constant delay on haptic communication QoE by emulating latency using data buffers between the haptic device and a virtual haptic environment.

In the second experiment, an SDN switch was used to emulate packet loss conditions by using OpenFlow meters. The meter of the haptic data traffic was configured to allow a certain data rate. When this threshold is exceeded the switch drops the packets. Results show acceptable performance of the haptic system for up to 90% packet loss.

The third one, examined the behaviour of packet loss and jitter when traffic differentiation is applied using an SDN switch that prioritised critical haptic communication traffic over best-effort traffic under congestion conditions. The results of this indicative methodology of network slicing using SDN verify the hypothesis that network slicing can guarantee good performance of haptic communication.

The fourth experimental framework used a similar methodology and experimentation equipment as the third experiment. The experiment was emulating the condition and network of a remote healthcare use case where an ambulance is accessing the network and requires guaranteed QoS for its mission-critical URLLC traffic. The experiment evaluated the coexistence of such traffic with best effort traffic and presented results on the impact of jitter.

The experimentation methodology used in these frameworks, particularly the use of SDN-capable switches and the relevant OpenFlow configuration, can generalise beyond this work as SDN can be used to slice the network and apply QoS rules on the network in order to meet the service requirements as shown in the experiments.

Furthermore, another contribution of this chapter is the proposal and analysis of an architecture of virtualised 3GPP VNFs interconnected and controlled by a virtual SDN controller. In this way traffic differentiation and management is possible within the 5G core network. Finally, this chapter includes a description of the main architecture components of the IEEE P1918.1 standard which targets the joint consideration of control, communication and computation aspects and is important for the design of future TI systems.

Chapter 4

Dynamic traffic prioritisation for haptic communication

4.1 Overview

The experimental frameworks presented in Chapter 3 show that it is possible for networks to accommodate more and more promising applications with impact in a wide range of industries, from manufacturing to education and healthcare. This large amount and variety of applications generate traffic that can be divided into three different classes: Ultra Reliable Low Latency Communication (URLLC), enhanced Mobile BroadBand (eMBB) and massive Machine-Type Communication (mMTC).

Even though bilateral teleoperation, a use case of URLLC is a well-known concept for decades now, its implementation over packet-switched networks, is not trivial because of the latency generated within the communication channel. As explained in Section 2, some teleoperation systems rely on having the knowledge of communication delay in order to provide stability, and hence require high precision estimation of the communication delay, while some rely on having the knowledge of the local or remote environment or both, in order to predict and compensate for communication delay. On the other hand, a wide range of control schemes have been presented in the literature that guarantee stability, nevertheless, by losing transparency and precision as a trade-off.

However, as with every long-distance communication system in packet-switched networks, packets need to be routed and travel through different nodes of the network while stored in queues to be directed according to the desired destination. Therefore, queuing and processing delays become of essence due to the increasing number of data flows and network nodes. The effects of queuing become even more apparent in the case of multi-class communication networks. In this case, priority queuing disciplines are applied, therefore, the important phenomenon of starvation of resources for low-priority queues needs to be taken into account, especially during traffic surges.

In this chapter, the focus is on one of the enablers of URLLC, traffic prioritisation and specifically prioritisation with priority jumps (also known as priority upgrades). The goal is to create a framework that allows queues within the network to meet the requirements of a haptic communication system in a competitive data flow traffic scenario without overprovisioning. Therefore, performance of teleoperation systems is modelled for a wide range of control schemes, also using a dynamic traffic prioritisation discipline with priority upgrades modelled to serve a high-priority and low-priority class of URLLC traffic. More specifically, the dynamic traffic prioritisation discipline used in this study concerns a queuing system composed of two queues. These queues are serving the two classes of URLLC flows of packets with each flow assigned to a different priority queue. The system is able to dynamically change the priority of each packet depending on a user-defined variable which can be optimized to serve the requirements of each flow.

An example of high-priority URLLC traffic can be bilateral teleoperation for remote engineering by a human operator while smart grid control traffic can have the role of the lower priority URLLC traffic. The framework proposes the appropriate priority queuing system configuration to accommodate both URLLC flows with minimum effects of starvation for the low-priority queue, while at the same time is able to suggest the control scheme option to achieve maximum teleoperation performance.

A schematic diagram that includes the main functionalities of the proposed system is shown in

Figure 4.1. In this figure, the main components of the bilateral teleoperation system are the same as in Figure 2.5, namely the user interacting with the master manipulator and controller as well as the slave manipulator and controller interacting with the environment. However, Figure 4.1 also depicts a scenario where two URLLC traffic flows are accommodated by the same communication network. One from the bilateral teleoperation exchange of haptic data, and another which is assumed to be of lower priority.

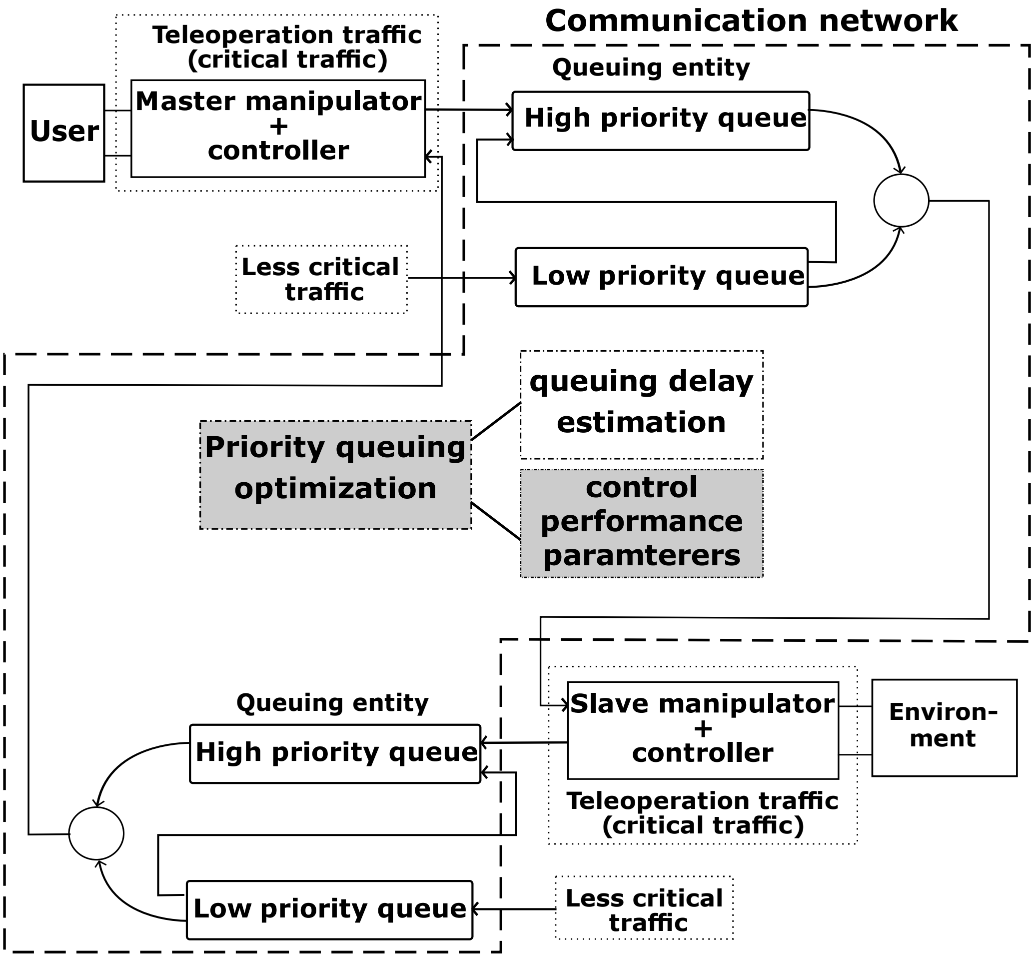


Figure 4.1: Schematic diagram of the proposed system with a communication channel accommodating two URLLC flows. Traffic for one flow is generated by a bilateral teleoperation system.

The structure of this chapter is as follows. In Section 4.2, the motivation and contributions of this chapter are shown. Next, an analysis of the mechanisms used within the proposed framework is presented in Section 4.3. Based on the previous, a control scheme selection methodology and its results are presented in Section 4.4. Finally, Section 4.5 includes the formulation, description and discussion of results of the optimisation problem proposed. All mathematical symbols in this Chapter have been summarised in the Glossary of the thesis.

4.2 Motivation & Contribution

Success of bilateral teleoperation not only depends on ensuring control stability is maintained, but also on delivery of high resolution interaction, between the operator and remote environment in terms of transparency and precision of task movement. Whilst stability is essential for any teleoperation system, these metrics of teleoperation system performance also define which applications can be supported. Applications that require high levels of transparency and high precision in operation are only possible when high resolution teleoperation can be offered.

As explained later in this section, the formulation of performance of teleoperation occurs by considering loss in transparency and precision between master controller and slave controller. By using the proposed control performance evaluation formula, not only it is possible to identify the most appropriate control scheme for a given communication performance, but also to configure the communication system, e.g., offer lower delay and improve control performance. However, such configuration can potentially have an inverse impact on the overall performance of a communication system without mechanisms that ensure fairness with respect to how the rest of the traffic is managed. As demonstrated in Section 3.4 of Chapter 3, QoS provisioning for one traffic flow can result in hindering the communication system performance for serving other flows of traffic. Therefore, the analysis of both control and communication performance for delivery of bilateral teleoperation as a potential URLLC service is important.

Hence, one needs to quantify how delay of the communication network affects performance of the teleoperation system. In order to understand the impact of delay on the performance of teleoperation systems, in this chapter teleoperation performance is formulated with movement precision and transparency captured as a function of delay. To this end, the contributions are in threefold:

1. A formulation of a polynomial function that captures teleoperation performance in terms

of transparency and precision as a function of delay. This model is generic, i.e. can be applied to any control scheme, and has the ability to accommodate future control schemes, with delay dependency. The formulation provides a mapping between order of magnitude of reduction/increase in communication delay and the order of magnitude of improved/degraded control performance respectively, and hence allows for prioritisation of competing flows with the most significant impact on the teleoperation performance.

2. A demonstration of using the formulated polynomial function by employing a dynamic priority scheduling discipline, based on the model of [299], for two different priority flows of highly critical and less critical teleoperation applications.

3. Combining the previous points, we propose a framework which allows comparison of bilateral teleoperation control schemes for different application scenarios as well as optimisation of the dynamic priority queuing system configuration to maximise performance of higher priority haptic communication with minimum effects on the lower priority flow.

As explained in Chapter 2, stability and transparency are conflicting teleoperation system design goals. In this work, stability is assumed to be guaranteed by the chosen control schemes and to enable the comparison using teleoperation performance indicators, the operator and environment model parameters are assumed to be the same among all control schemes. The teleoperation performance model that is formulated in this chapter, takes into account performance indicators that reflect two important aspects.

First, the perception of the interaction of the user with the remote environment (both in free motion and during contact), i.e., what the user eventually perceives because of the unavoidable degradation of system performance due to network delay, compared to what the user should have felt. Even though perceptual data reduction is not included in the performance model in this study, haptic communication performance can still be noticeably degraded independently of the different perception capabilities of human users. Nevertheless this requires an expansion of the model, not presented in this study.

The second aspect is the movement accuracy which can be evaluated using the tracking error and position drift related to the movement of the master and slave manipulators. By minimising the difference between the actual and perceived model parameters as well as the accuracy errors, teleoperation performance is maximised.

Furthermore, in this study, perceptual data reduction is not included in the teleoperation performance model.

It will be thus possible for teleoperation systems to incorporate multiple control schemes and switch between them for achieving the best performance with minimum cost to the overall performance of the communication platform. The grey boxes in Figure 4.1 represent the components proposed in this work.

4.3 Control Performance and Queuing Delay Analysis

4.3.1 Generic delay-based control performance model

Based on the analysis presented in Section 2.4.2, an ideal bilateral teleoperation system would satisfy the followings,

$$\begin{aligned} &v_m - v_m = 0, \\ &B_m - B_p = 0, \\ &K_e - K_p = 0, \\ &\delta = 0,\, \Delta = 0 \end{aligned} \tag{4.1}$$

Of course, in reality teleoperation systems do not operate under ideal conditions and therefore the differential values in (4.1) are always non-zero in the presence of delay.

After reviewing control schemes FR, PE, PC, PCP and AMFC, as presented in Section 2, along with their respective performance metrics, control performance can be expressed as a loss function. The control schemes chosen present features but also limitations that differentiate one from the other. Control schemes FR and PE present no stabilisation features in case of delay, nevertheless they are the basic type of control schemes therefore easy to implement. The main practical difference among these two schemes is that FR requires the use of force sensors on the slave device. In the case of PC, this scheme requires prior knowledge of delay to regulate the force feedback. The more complex PCP control scheme extends PC but unlike PC, it is IS therefore can maintain its stability for any amount of delay. Lastly, AMFC is a control scheme that requires accurate knowledge of the human and environment model, similarly to the MMTA approach.

The Control Performance Loss Function (CPLF) reveals the loss in perceived inertia (M_p) versus the manipulator's inertia (M_m), perceived damping (B_p) versus the manipulator's damping (B_m), perceived stiffness (K_p) versus environment stiffness (k_e), tracking error (δ) and position drift (Δ). As shown in Equation (4.1), the best performance of teleoperation system is achieved when $M_m = M_p$, $B_m = B_p$, $k_e = k_p$, $\delta = 0$, and $\Delta = 0$. In this equation, weight values w_i are used to normalise the contribution of the different loss values depending on the teleoperation scenario. For example, a teleoperation system used in underwater inspection will be more tolerant in movement accuracy than a system used for minimally invasive surgery (MIS), therefore in the MIS case the contribution of the movement accuracy loss values should be greater and therefore the corresponding weights should be increased.

$$CPLF = w_1(v_m - M_p)^2 + w_2(B_m - B_p)^2 + w_3(k_e - k_p)^2 + w_4\delta^2 + w_5\Delta^2 \ . \tag{4.2}$$

To further clarify Equation (4.2), we elaborate the loss function for one of the control schemes, explained earlier in Section 2.4. In the PCP control scheme, the control parameters of inertia, damping, stiffness, tracking error, and drift error can be expressed as follows [87],

$$\text{Perceived inertia } (M_p)\text{: } 2v_m - \frac{B_m^2}{K_c} \tag{4.3}$$

$$\text{Perceived damping } (B_p)\text{: } 2B_m \tag{4.4}$$

$$\text{Tracking error } (\delta)\text{: } \frac{B_m + K_c d}{2B_m K_c} \tag{4.5}$$

$$\text{Stiffness } (K_p)\text{: } \frac{K_e K_c (B_i + B_m)}{(K_e + K_c)(B_i + B_m) + 2K_e K_c d} \tag{4.6}$$

$$\text{Drift error } (\Delta)\text{: } \frac{B_i + B_m + 2K_c d}{(B_i + B_m)K_c} \tag{4.7}$$

where B_i is the impedance of the channel and K_c a control parameter. It can be seen that while the primary two parameters are delay-independent, the latter three parameters have dependency to d, d^{-1}, and d respectively. In order to capture the relationship between the loss function, CPLF, and the communication delay in a generic form, we formulate a polynomial function, as seen in Eq. (4.8). This formula generalises the result of using Eq. (4.2) and substituting the corresponding control parameters of the chosen control scheme, capturing the impact of delay on CPLF. The coefficients of d are the result of substitution of the relevant loss parameters and the polynomial degrees are considered to be between -1 and 2 based on the observations of control parameter formulas from control schemes under discussion, e.g., as in Equations (4.3)-(4.7), presented in Table 4.1. Clearly $l = 0$ capture the effect of those loss values that are not delay-dependent.

$$CPLF(d) = \sum_{l} \phi_l d^l \ , \ l \in \{-1, 0, 1, 2\} \tag{4.8}$$

In the next sections, we first present a methodology through which we can select the most appropriate control scheme to be used under certain communication network conditions and given a teleoperation application domain. Finally, an optimisation problem is formulated for delivering maximum control performance with minimum impact on the resource utilisation.

Table 4.1: Comparison of control schemes with respect to delay

Scheme	Stability	Known d	M_p	B_p	K_p	δ	Δ
FR	PS	No	$\propto d$	-	-	$\propto d$	-
PC	PS	No	-	-	-	$\propto d$	-
PE	PS	No	$\propto d, d^2$	$\propto d$	-	-	-
AMFC	PS	No	$\propto d$	$\propto d$	-	-	-
PCP	IS	Yes	-	-	$\propto 1/d$	$\propto d$	$\propto d$

4.3.2 Delay analysis of priority queue with jumps

As discussed in Section 2.7.1, probabilistic prioritisation queuing disciplines offer flexibility in terms of selecting which criteria to influence the behaviour of the queuing system. Furthermore, because of the characteristics of bilateral teleoperation traffic, i.e., small payload and high rate of packet generation, the prioritisation model should allow simple per-flow distribution of different priority queues, with fine granularity characteristics, affecting traffic on a per-packet basis. Hence, we choose to use the HOL-MBP model, due to the properties mentioned above as well as the tractability of analysis based on probability generating functions (PGFs). PGFs offer important properties (such as the calculation of moments) which can be used to infer performance metrics of the system. For the rest of this section we will present the analysis of such a queuing model.

The system model, as seen in Figure 4.1, shows two level of priority for traffic, the high priority (class 1), corresponding to critical URLLC, and one low priority (class 2), corresponding to non-critical URLLC, each directed to First Come First Serve (FCFS) queues. The queuing system is a discrete-time system and therefore time is divided in fixed-length intervals called slots. The service time of one packet equals to one time slot n which is therefore defined as $n = \frac{\text{packet size}}{\text{available transmission link capacity}}$.

A total arrival process with arrival rate λ_T is composed of two arrival processes, one for each type of packets, with arrival rates λ_1,λ_2. These arrival processes are modelled as Poisson processes (which is reasonable for URLLC traffic in a discrete-time model as shown in . Furthermore, queues are of infinite capacity, nevertheless, the model used allows the estimation of the amount of packets in each queue for each time slot. Moreover, the system is work-conserving, meaning that the system provides service always at full capacity and it is never idle. The last two assumptions of infinite capacity and the work-conservation allow the study of the queuing system when stable which is true for a total arrival rate $\lambda_T = \lambda_1 + \lambda_2 < 1$.

Though with the SP discipline, the low priority queue would be served only when high priority queue is empty, in HOL-MBP, during time slot n, packets from low priority queue (all waiting packets plus those which arrive during slot n) may jump to the high priority queue with a probability of β. This jump can occur both when high priority is empty or non-empty. The equations that describe how the content of each queue evolves in time are shown in [278] (Section 3, equations [1]-[4]).

As mentioned in Chapter 2, the HOL-MBP model was chosen as it presents two important features. The first feature is that compared to other fair queuing models (e.g., the Weighted Fair Queuing scheduling algorithm) there is no need to keep track of the delay of each packet. The second one, is that the model is analytically tractable. Even though in the proposed system there are only two queues, it must be noted that the HOL-MBP model can be applied on an N-input and N-output switch as described in [299]. Due to the increasing numerical work needed for the analysis of the model as the traffic classes increase, the assumed traffic classes are kept at only two in all cases. Nevertheless, it is also possible to extend the model to allow more than two traffic classes with a matrix-analytic solution.

The relevant equations for modelling the queuing system, among those discussed in [278], are the expressions for the mean values of packet delay for each of the queues. These expressions can be obtained after the system analysis and inference of the corresponding PGFs. By calculating the moments of each PGF, we get the following expressions:

- the total number of packets in the system:

$$E[p_T] = \lambda_T + \frac{\lambda_{TT}}{2(1-\lambda_T)} \tag{4.9}$$

- the mean value of waiting time in high priority queue:

$$E[d_H] = 1 + \lambda_2 - \frac{1-\lambda_1}{\beta} + \frac{\lambda_{TT}\lambda_1 + \lambda_{11} - \lambda_{11}\lambda_T}{2(1-\lambda_T)\lambda_1} - \frac{(1-\lambda_T)A_T(Y(\beta,1))}{Y(\beta,1) - A_T(Y(\beta,1))} \tag{4.10}$$

- the mean value of delay of any arbitrary packet:

$$E[d] = \frac{\lambda_1}{\lambda_T}E[d_H] + \frac{\lambda_2}{\lambda_T}E[d_L] \tag{4.11}$$

where, λ_T is the total arrival rate ($\lambda_T = \lambda_1 + \lambda_2$), λ_{11} and λ_{TT} the second partial derivatives of the PGF of the joint arrival process, A_T the PGF of the total number of arrivals and Y a factor which can be obtained numerically (in this work bisection approximation method is used with initial values 0 and 1).

Since directly calculating the mean value of waiting time in low priority queue is complex, and by using Little's theorem it has been proven that $E[p_T] = \lambda_T E[d]$ [279], combining three equations of (4.9)-(4.11), one can calculate $E[p_T]$, as follows,

$$E[d_L] = \frac{\lambda_T}{\lambda_2}(\lambda_T E[p_T] - \frac{\lambda_1}{\lambda_T}E[d_H]) \tag{4.12}$$

Equation (4.10) shows that both low and high priority delay values, i.e., d_H and d_L, depend on the parameter β. The optimisation framework will, hence, define parameter β to meet the minimum requirements of low priority traffic.

4.4 Choice of Control Scheme

By combining the previously mentioned network scheduler and proposed control performance model of Sections 4.3.2 and 4.3.1, this section presents a methodology that focuses on satisfying the requirements of a high priority URLLC flow served by a HOL-MBP system. This methodology concerns the selection of the most suitable control scheme based on prior knowledge of the delay requirements of a low priority URLLC flow. In the next section, a bi-objective optimisation framework will be presented for maximising the performance of a chosen control scheme (which has first been evaluated in this section) as well as minimising the low priority URLLC flow delay.

To begin with, it is necessary to calculate the possible queuing delays for both URLLC flows for $\beta \in [0, 1]$ using Equations 4.10 and 4.12. Furthermore, assuming knowledge of the delay requirements for low priority URLLC traffic, a fixed upper threshold value t_L on low priority queuing delay can be considered to find the corresponding β value (based on the calculations previously mentioned). In this way, since both high priority and low priority queuing delays are linked by the same β, it is possible to find the corresponding high priority queuing delay which is essential for calculating CPLF.

As shown in Equation (4.2), weights play an important role in the outcome of the calculation. These weights can be determined based on the teleoperation scenario. Of course, several other teleoperation and control scheme configuration parameters need to be known and initialised before the calculation of CPLF, which will be discussed further in this section. Most importantly, calculation of CPLF for all desired control schemes can make possible the quantitative comparison of each control scheme's performance not only under different delay values but also under different application scenarios. In this way, it is possible to select the most suitable control scheme.

It needs to be noted that for PS control schemes (as defined in Section 2.4.2) the control scheme parameters determine the delay tolerance of the control scheme in order for the teleoperation system to maintain its stability. This means that CPLF loses its meaning if the delay used to calculate CPLF exceeds this stability threshold since, by definition, in an unstable system there is no meaning of measuring or evaluating QoE performance. As a result, for values of delay beyond such stability thresholds, the CPLF control schemes available for comparison are reduced, narrowing down the available options.

4.4.1 Simulation

In this section a performance evaluation of the proposed control scheme selection framework is presented, for a diverse set of scenarios that represent different teleoperation application classes.

First, the simulation setup, the assumptions and values used are explained. Afterwards, the numerical results are presented.

Simulation Setup

Two simulation models in Matlab were set up, one for the dynamic priority network scheduler as presented in Section 4.3.2 and one for CPLF of five control schemes for different weight values of the CPLF formula in Equation (4.2).

For the queuing system simulation, it is assumed that the available transmission rate for both URLLC flows is 2 Mbps, with a total utilisation of 95% and the ratio α of the two arrival rates $\alpha = \frac{\lambda_1}{\lambda_T} = 0.5$. These parameter values are chosen to agree with the model assumptions for providing a tractable analysis and emulate a scenario where both types of traffic are served by a network of switches under congestion, in a similar way as with the experimentation frameworks presented in Chapter 3. Therefore, all packets are of equal size of 500 Bytes (the 500 Bytes packet size is chosen based on work presented in [300]).

Additionally, similar to 4.3.2, one time slot is equal to the transmission time of a packet. Hence, calculating transmission time knowing the size and the available transmission rate is intuitive. As previously described in Section 4.3.2, it is further assumed that the queuing system is stable. For the purpose of demonstrating the control scheme selection for a given low priority URLLC

Table 4.2: Simulation values

Communication parameter		Value
Packet size		500B
Available transmission rate		2 Mbps
Total link utilization λ_T		0.95
Traffic balance α		0.5
Teleoperation parameter		**Value**
Mass M_m (kg)		10
Damping B_m (kg/s)		1
Stiffness K_e (N/m)		10
Gain G_c		1
Control Scheme	**Parameter**	**Value**
FR (Configuration 1)	Gain K_c	10
FR (Configuration 2)	" K_c	100
PE (Configuration 1)	" K_c	10
PE (Configuration 2)	" K_c	100
PC	" K_c	100
PCP	" K_c	100
	Impedance B_i	1
AMFC	A	0.006
	C	50
	Gain K_m	$10 \cdot M_m$
	Λ	1

flow delay requirement, delay value d_L, is set to 44 ms. A possible scenario for this latency requirement is for the monitoring of process automation in an industrial scenario (maximum latency allowed is 60ms as shown in a list of URLLC use cases in 3GPP TS 22.261 [301]).

With regard to the CPLF simulation, CPLF equations for five different control schemes (FR, PE, PC, AMFC and PCP as described in Section 2.4) are implemented with teleoperation parameters as detailed in Table 4.2. As previously discussed, we can only evaluate control performance within the stable region of the control scheme, i.e., within the range of delay values where stability is guaranteed by the control scheme. To better illustrate this point, for control schemes FR and PE, two different configurations of control parameter K_c have been used in the simulation, to show how this change can affect their delay tolerance, i.e. their stability threshold, because they are PS control schemes. Additionally, the simulated PCP control scheme is IS, hence there is no delay threshold. The delay threshold for PC which is PS is 120 ms whereas AMFC is tolerant to any delay for certain parameter values as explained in Section 2.4.2. Finally, it is assumed that the parameters of the remote environment are constant during the simulation, therefore delay thresholds remain the same throughout the simulation.

Simulation Results

We first look into the values of d_L and d_H and how they are affected by changing of β. Figure 4.2 shows the two values of d_L, and d_H, versus β. It can, for example, be seen that the low priority delay value of $d_L = 44$ ms corresponds to a value of $\beta = 0.3$. This value of β corresponds to a high priority delay value of $d_H = 35$ ms.

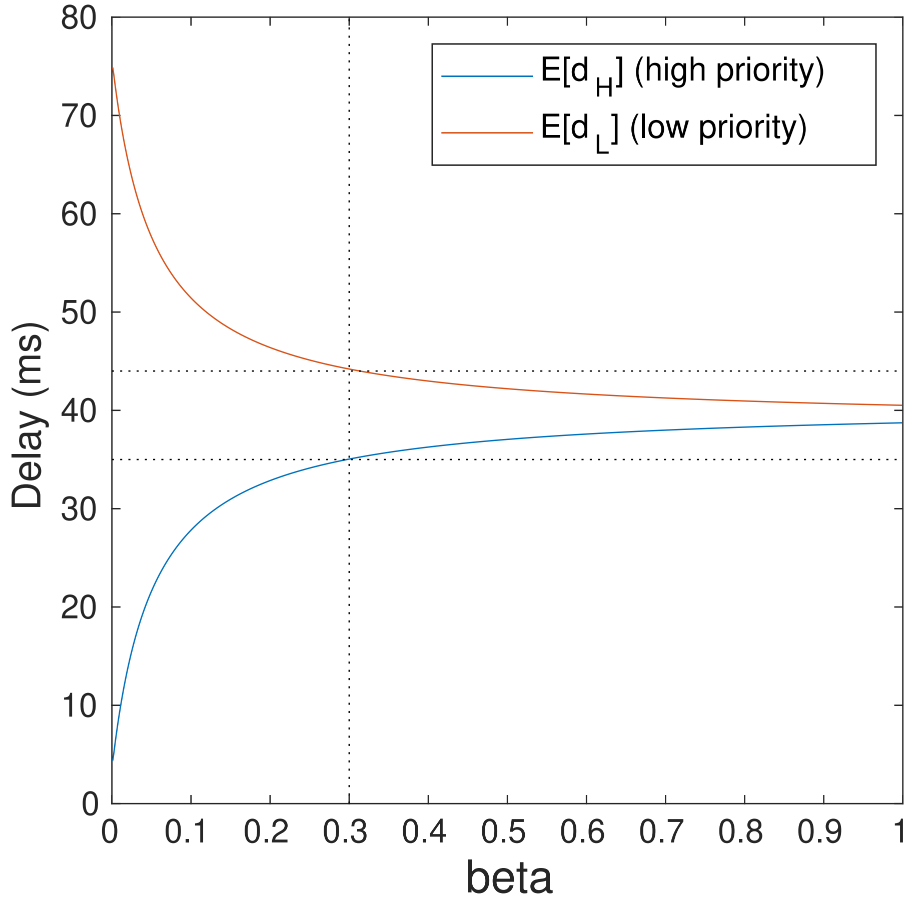


Figure 4.2: Mean value of packet delays versus beta

After calculation and normalisation of CPLF values for each control scheme, in each plot of Figure 4.3, we show the behaviour of normalised CPLF of each control scheme as a function of delay $d \in [0, 0.15]$ in seconds. Regarding the weights of CPLF, four different weight settings have been proposed which correspond to four different teleoperation scenarios. Furthermore, as shown in Table 4.2, for control schemes FR and PE, two different configurations of control parameter K_c have been used in the simulation. All scenarios have been simulated two times, one for each of the two different configurations for FR and PE shown in Table 4.2. This is because we use FR to normalise CPLF values (only to be able to visualise CPLF in a meaningful way) and therefore all normalised CPLF values have to be recomputed.

Even though the delay value range in Figure 4.3 is between 0 and 150 ms, according to the results of d_H in Figure 4.2, the delay value range of interest is 4 to 38 ms. This range of values is the high priority delay that the queuing system generates according to the queuing system and communication parameters of Table 4.2.

It needs to be noted that the normalised CPLF values show the percentage of decrease of a control scheme's CPLF compared to the CPLF values of the FR scheme according to:

$$CPLF(i)_{norm} = \frac{CPLF(FR) - CPLF(i)}{CPLF(FR)} \cdot 100 \ , \tag{4.13}$$

where i $\in$ {PE, PC, AMFC, PCP}.

A normalised CPLF value shows how much better this scheme performs compared to the FR scheme. Therefore, a positive value (i.e., positive normalised CPLF percentage) means that higher performance than FR and a negative value (i.e., negative normalised CPLF percentage) means worse performance than FR. Therefore, it is easy to compare the performance of control schemes for a given value of delay.

In each plot, to illustrate how the queuing system affects CPLF, the vertical black line β is used to indicate how much is the high priority delay value d_H for a corresponding β value. Furthermore, the stability thresholds of the PS control schemes are shown using dotted vertical lines which indicate the delay value of the threshold (top horizontal axis).

Based on the aforementioned numbers and assumptions we demonstrate the control scheme selection for the two different configurations of control parameters, Configurations 1 and 2 in

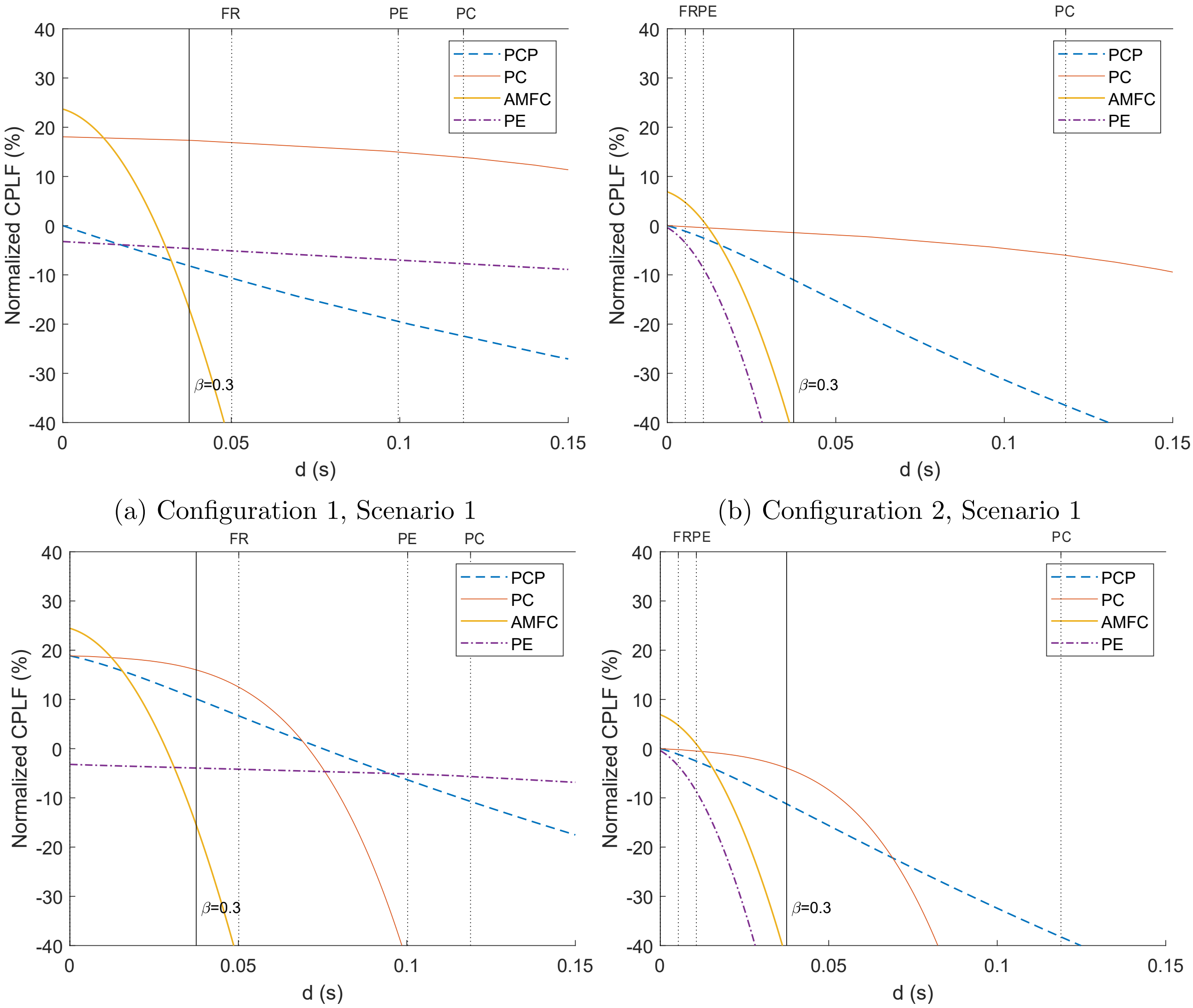
Figure 4.3, also as shown in Table 4.2. We also present results for four different cases of CPLF weights corresponding to four different teleoperation scenarios:

- *Scenario 1*: Equal participation of all performance metrics with weights $w_1, w_2, w_3, w_4, w_5 = 1$. This case refers to applications where tracking capabilities are not as important such as bilateral teleoperation for underwater inspection or other systems with semi-autonomous capabilities.
- *Scenario 2*: Increased participation of tracking and drift error by choosing weights $w_1, w_2, w_3 = 1$ and $w_4, w_5 = 100$, can be relevant to applications such as minimal invasive surgery.
- *Scenario 3*: Increased participation of tracking and drift error with reduced participation of inertia, damping and stiffness in the calculation of CPLF by choosing weights $w_1, w_2, w_3 = 0.5$ and $w_4, w_5 = 100$. Possible applications can be certain cases of micro-assembly, e.g. rotational movements.
- *Scenario 4*: Participation of inertia, damping and tracking only with weights $w_1, w_2 = 1$, $w_4 = 100$ and $w_3, w_5 = 0$. This choice of weights corresponds only to free-space movement applications.

It must be noted that the case of participation of performance metrics with weights $w_1, w_2, w_3 > 1$ and $w_4, w_5 \leq 1$, meaning increased participation of inertia, damping and stiffness but decreased participation of tracking and drift error, provides virtually the same results as in Scenario 1. Furthermore, it has been empirically observed that, after calculating CPLF parameters δ and Δ (according to each control scheme's formulas), their absolute value is $0 < \delta, \Delta < 1$ and therefore their squared values, as in Equation (4.2), are orders of magnitude smaller than the rest of the performance metrics. This finding has been taken under consideration for the appropriate selection of weights to normalise CPLF, i.e., to adapt the contribution of δ and Δ in the loss function.

As previously discussed, $d_L = 44$ ms is an acceptable low priority traffic latency which corresponds to $\beta = 0.3$. This β value is used to calculate d_H using Equation (4.10) and therefore it is possible to compare CPLF values of control schemes for this specific delay, for the four application scenarios.

CPLF Comparison

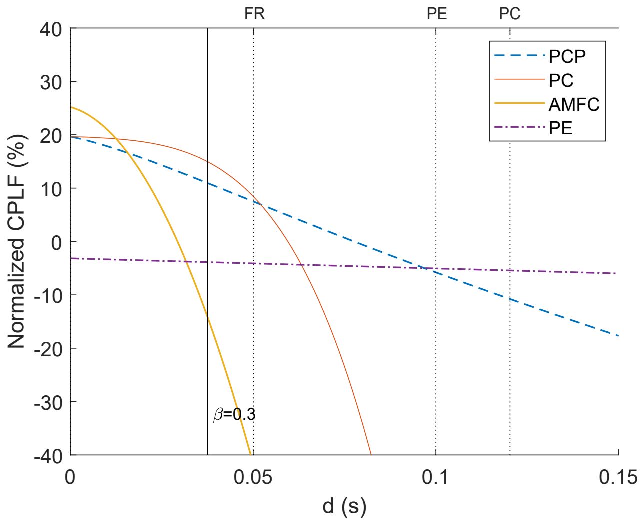


(a) Configuration 1, Scenario 1

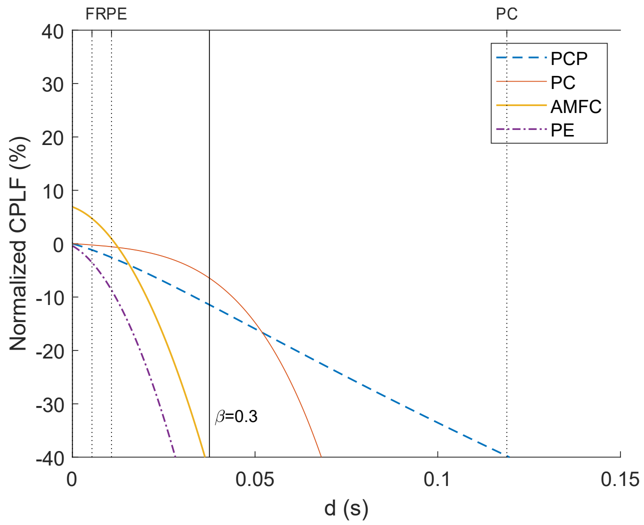
(b) Configuration 2, Scenario 1

(c) Configuration 1, Scenario 2

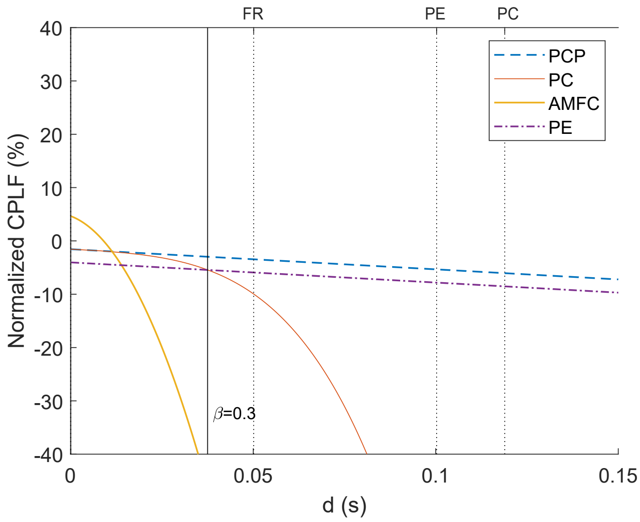
(d) Configuration 2, Scenario 2



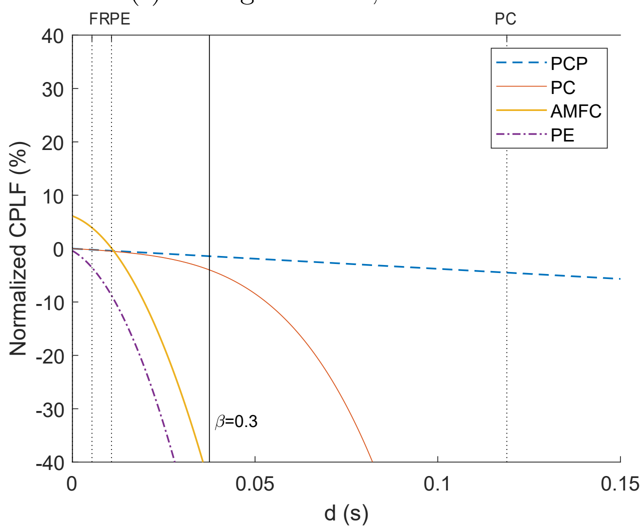
(e) Configuration 1, Scenario 3



(f) Configuration 2, Scenario 3



(g) Configuration 1, Scenario 4



(h) Configuration 2, Scenario 4

Figure 4.3: Comparison of CPLF among different control schemes in four different scenarios (i.e., different weight choices) and two different PS control scheme configurations.

Regarding the results presented in Figure 4.3, we can first discuss the behaviour of the plots of Configuration 1 (left column plots) compared to those of Configuration 2 (right column plots). As a reminder, the difference between these groups of plots lies in the recalculation of CPLF, due to changing control parameter K_c for control schemes PE and FR. The latter control scheme is also used for normalising the CPLF values of all other control schemes as shown in 4.13. As discussed in Section 2.4.2, due to the increase of control parameter K_c, from $K_c = 10$ in Configuration 1 to $K_c = 100$ in Configuration 2, there is an increase increase of performance in FR therefore lower values of normalised CPLF can be observed in Configuration 2 than in 1. Nevertheless, due to the trade-off of stability and transparency, this performance increase comes with a cost of a reduced stability threshold, which, as illustrated by the dotted line for FR, is 5 ms in Configuration 2 from 50 ms in Configuration 1. Similarly, the stability threshold for PE also decreases by increasing K_c, from 100 ms to 10 ms.

Focusing more on the comparison among the control schemes illustrated by lines of different colour, we take into account the β line, which corresponds to the latency under which the critical traffic flow operates. We use the β line to see, for this specified delay, which control scheme has the highest CPLF value and if this value is positive.

Specifically, for Configuration 1, in Scenarios 1, 2 and 3, as shown by Figures 4.3a, 4.3c and 4.3e respectively, the control scheme that performs best is PC. In these three cases, PC has the highest value among all four control schemes at the point indicated by the β line and its value is also positive, therefore, it has a greater value than FR. For Scenario 4 (Figure 4.3g), the highest value in the plot is the one of PCP, nevertheless, the value is negative therefore the best option in this case is FR.

In Configuration 2, it is important to notice the change of the values of the stability threshold lines of FR and PE. This change can determine whether we are observing CPLF performance in the stability region of each control scheme. As it can be observed, in Scenarios 1, 2 and 3 (Figures 4.3b, 4.3d and 4.3f), PC has highest value among all other control schemes. Even though this value is negative and therefore FR seems to outperform PC, the best choice is PC because in this case the comparison point indicated by the β line is past the stability threshold of FR. For the same reason, in Scenario 4 in Fig. 4.3h, the best choice is PCP.

Regarding all plots, it can be seen that for lower values of high priority delay (and therefore lower β) control scheme choice becomes more competitive. Finally, it needs to be noted that in bilateral teleoperation the choice of the control scheme is an important system design decision that is affected by several factors (e.g., the type of interaction that it enables the user to have with the remote environment) therefore in many cases the best control scheme may not only be selected based on the highest CPLF, but also taking into account other system design factors.

4.5 Optimisation of Priority Queue Performance

In this section, a bi-objective optimisation problem is presented, that maximises control performance of the system for a high priority teleoperation traffic, i.e., critical URLLC, while minimising communication delay for the lower priority non-critical URLLC traffic.

The first objective (problem P_1) of the optimisation problem is to maximise control performance, i.e., to minimise CPLF. The control performance function of CPLF represents the loss in transparency and movement precision as a result of delay $E[d_H]$ on communication between master and slave controller. The second objective (problem P_2) is the minimisation of delay $E[d_L]$ for the non-critical URLLC traffic. Both of these delays change values as a function of β and therefore the corresponding objective functions can be modelled as functions of β.

After the selection of a control scheme and since the delay of the critical URLLC traffic is $E[d_L]$, the objective function for problem P_1 can be formulated by using the corresponding CPLF equation replacing d with $E[d_H]$. Therefore, based on the polynomial in Equation (4.8) P1 can be formulated as:

$$P_1(\beta) = CPLF(\beta) = \sum_l \phi_l E[d_H]^l \ , \ l \in \{-1, 0, 1, 2\}, \tag{4.14}$$

with $E[d_H]$ as defined in Equation (4.10).

The objective function for Problem P_2, which concerns the minimisation of $E[d_L]$ that is defined as in Equation (4.12), can be formulated by also expanding $E[d_H]$ using Equation (4.10):

$$\begin{aligned} P_2(\beta) &= \frac{\lambda_T}{\lambda_2}(\lambda_T E[p_T] - \frac{\lambda_1}{\lambda_T} E[d_H]) \\ &= \frac{\lambda_T}{\lambda_2}\left(\lambda_T E[p_T] - \frac{\lambda_1}{\lambda_T} - 1 + \lambda_2 - \frac{1-\lambda_1}{\beta}\right. \\ &\quad \left. + \frac{\lambda_{TT}\lambda_1 + \lambda_{11} - \lambda_{11}\lambda_T}{2(1-\lambda_T)\lambda_1} - \frac{(1-\lambda_T)A_T(Y(\beta,1))}{Y(\beta,1) - A_T(Y(\beta,1))}\right). \end{aligned} \tag{4.15}$$

This combination of P_1 and P_2 can find the right balance between delay incurred to the low priority traffic and the performance of high priority teleoperation traffic. To this end, we formulate a problem that configures the value of β for maximising teleoperation performance, i.e., lowest value of loss as captured in $CPLF(d)$ in Equation (4.14), while also ensuring the lower priority traffic achieves the lowest possible latency, as captured in Equation (4.15). The method used to formulate the problem is the Weighted Sum Scalarisation method (as in Equation (3) in [302]). Therefore, the optimisation problem can be formulated as:

$$\begin{aligned} &\underset{\beta}{\text{minimise}} && F(\beta) = \gamma_1 \frac{P_1(\beta)}{sf_{1,0}(\beta)} + \gamma_2 \frac{P_2(\beta)}{sf_{2,0}(\beta)} \\ &\text{subject to} && 0 > \beta > 1, \end{aligned} \tag{4.16}$$

where $F(\beta)$ is the master problem objective function and $sf_{1,0}(\beta)$ and $sf_{2,0}(\beta)$ are normalisation factors for transforming the objective functions so that they have similar orders of magnitude and adapt units, so that they are both comparable. Another important aspect of this formulation are the scaling parameters that have a sum of 1, therefore by choosing γ_1 for P_1 we get $1 - \gamma_1 = \gamma_2$ for P_2. Nevertheless, the objective function of Equation 4.16 can be reformulated without normalisation and in this case the weights chosen can have a sum different than 1. The use of weights is significant for revealing the relative importance between P_1 and P_2.

The proposed problem can be expanded by adding an additional constraint on β. Due to the stability thresholds of each control scheme, as presented in the previous section, in certain cases where the stability threshold is within the range of delay values of $E[d_H]$, an change of the value of β should not allow the $E[d_H]$ value to exceed the threshold. Nevertheless, such a case is not studied in the formulation of this problem. Excluding the case of considering a stability threshold, it can be observed that all possible values of $P_1(\beta)$ and their corresponding values $P_2(\beta)$ for the same β, are also all the possible solutions of the master problem $F(\beta)$. Therefore, for this problem there are no dominated solutions.

The optimisation problem as formulated in Equation 4.16 using the Weighted Sum method is reduced from a bi-objective problem to a single objective problem which can be solved with the. An important aspect of multi-objective problem solving is the principle of Pareto optimality or else Pareto efficiency [303], which will be further explained in the next section.

4.5.1 Numerical Results

In this subsection, we present numerical solutions and their interpretations for the optimisation problem formulated in the previous section. The results presented in this section are associated with Pareto efficiency that will be discussed further.

First, in the case of a bi-objective optimisation problem, the objective function space, more formally known as the criterion space, can be created by plotting the feasible solutions provided by the objective functions. An example can be seen in Figure 4.4 where each axis represents an objective function and the points inside the plot are the feasible solutions. In this case we consider a utopia point at $(0, 0)$, which is an ideal solution. The small circles inside the plot represent the feasible solutions. The set of feasible solutions that are better than others in at least one objective and not worse than others in any objective are called Pareto efficient and form the Pareto front (blue line). In this case, the Pareto front is convex and the solutions at the area above the line are called dominated (whereas those on the line are the dominant solutions).

Figure 4.5, presents two sets of plots of possible P_1 solutions against possible P_2 solutions, based on the simulation parameter values for CPLF and the priority queuing model set in Section

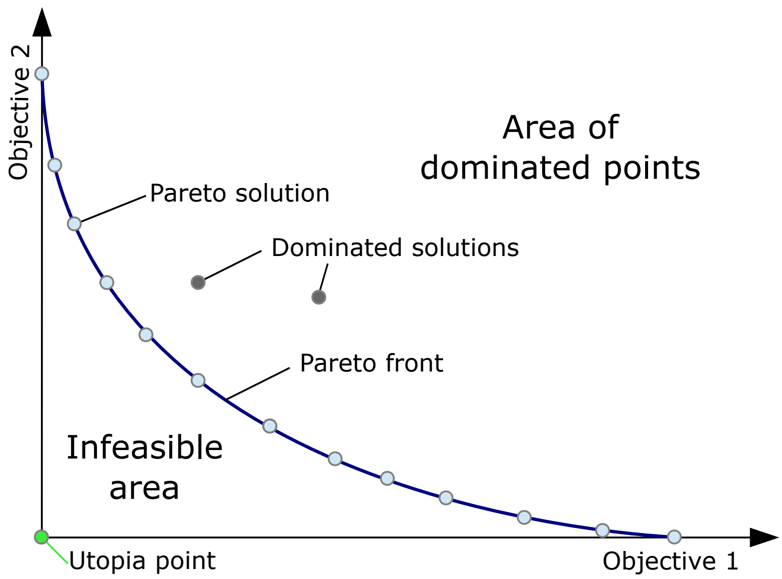


Figure 4.4: An example of a graphical representation of a bi-objective optimisation problem in the criterion space

4.4.

Moreover, the solutions of both objectives of the optimisation problem presented in this section concern PCP control scheme in the four different teleoperation scenarios enlisted in Section 4.4.1 as well as for two different different values for the scaling parameters γ_1 and γ_2 of Equation (4.16).

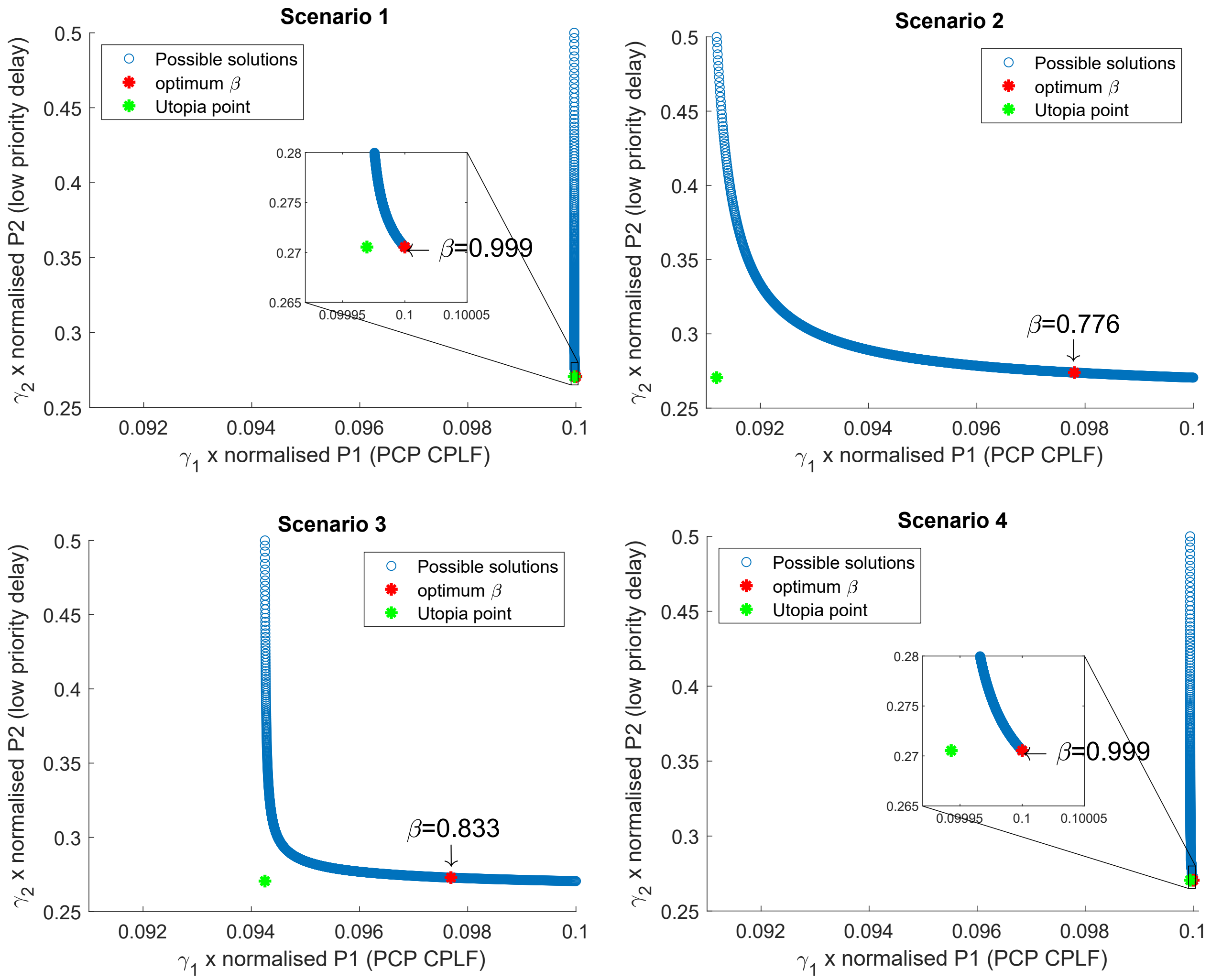
Both sets of plots in Figures 4.5a and 4.5b, display the Pareto front of solutions for P_1 and P_2. The solutions in the criterion space of the optimisation problem are the calculated values of CPLF that correspond to the values of $E[d_L]$ according to $\beta \in [0, 1]$ and are all possible non-dominated solutions (i.e., they all belong on the blue line as depicted in Figure 4.4). For each of these plots, the optimal solution is the one with the minimum Euclidean distance from the cross-section point of the axes, which is considered as a Utopian point P_i^U for $i = 1, 2$ (where i represents the two axis [303]). More specifically, the coordinates of the Utopian point in each plot are the minimum values of each objective. Furthermore, each objective solution is normalised according to:

$$P_i^{norm} = \frac{P_1(\beta) - P_i^U}{P_i^{max} - P_i^U} \tag{4.17}$$

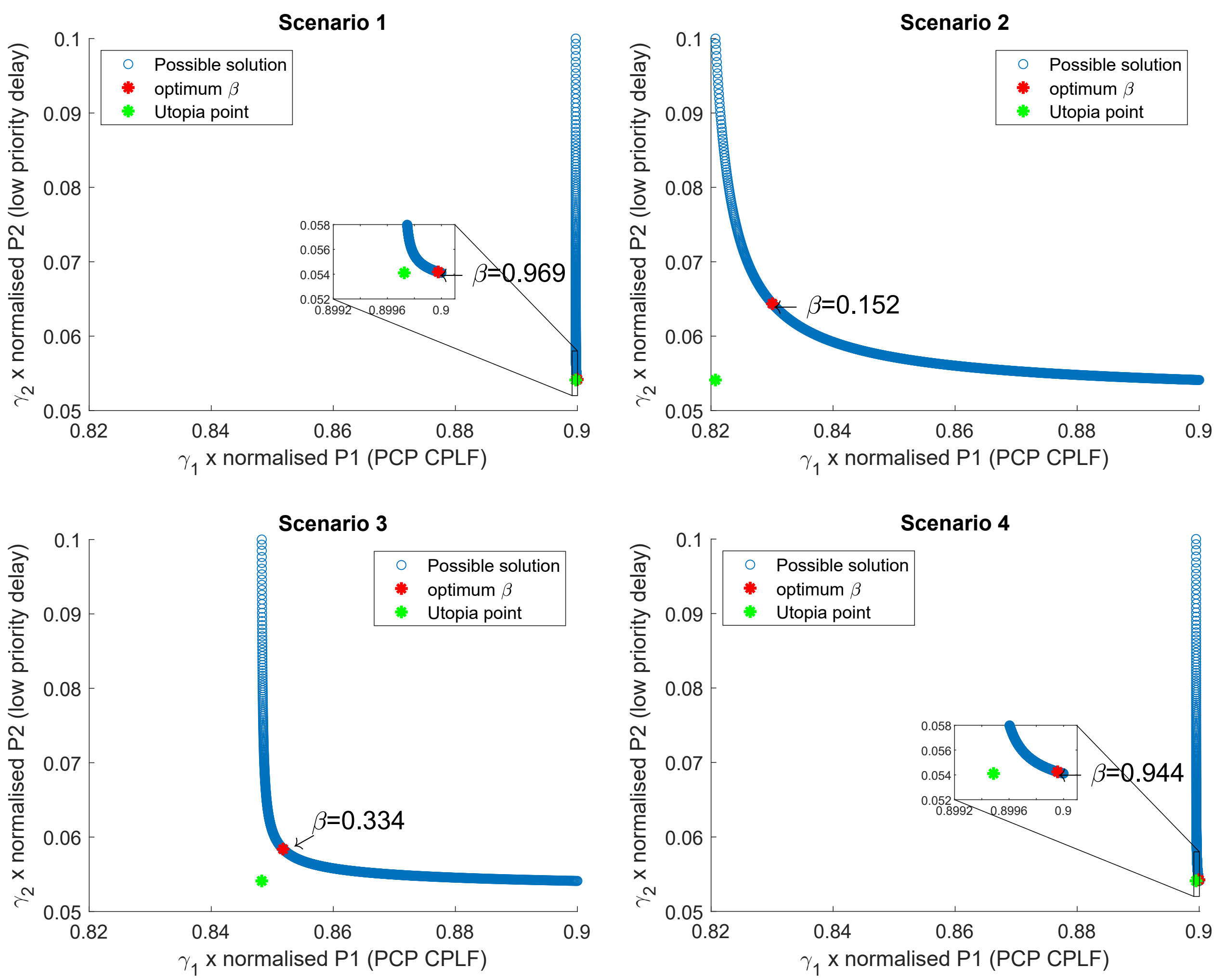
In order to demonstrate the behaviour of the optimisation problem, extreme cases of weights γ_1 and γ_2 are used demonstrating the impact of preference towards one of the two objective functions. The same network and control simulation parameters are used as in the previous section, as variability of network conditions is not examined.

The first set of plots in Figure 4.5a depicts the use of scaling parameters $\gamma_1 = 0.1$ and $\gamma_2 = 0.9$. This means that the master problem objective function should act in favour of the P_2 problem. Another important thing to pay attention to is the behaviour of the optimisation of β depending on the scenario. For Scenario 1, the case of equal weights, as well as for Scenario 4, the free space movement scenario, the optimal solution is in favour of fully minimising the low priority queue latency as $\beta = 0.999$. This is due to the very small change of objective $P1$ values, i.e., the CPLF of PCP doesn't change a lot for the corresponding values of d_H. For the cases of the other two scenarios minimisation of both objective functions is more evident as β is reduced but still close to 0.9 favouring low priority delay.

The second set of plots, presents the results for scaling parameters $\gamma_1 = 0.9$ and $\gamma_2 = 0.1$. This means that the master problem objective function should act in favour of the P_1 problem. Comparing the plots of Figure 4.5a to the plots of Figure 4.5b, the preference change is evident. In all scenarios the value of β is smaller. It is also important to note that, between Figures 4.5a and 4.5b, β behaves in agreement to each scenario. Scenarios 1 and 4 always have higher β than Scenarios 2 and 3. This can be justified considering that Scenarios 1 and 4 concern applications with lower delay requirements than Scenarios 2 and 3.



(a) $\gamma_1 = 0.1$ and $\gamma_2 = 0.9$



(b) $\gamma_1 = 0.9$ and $\gamma_2 = 0.1$

Figure 4.5: Optimisation of β for four different application scenarios and scaling parameter values γ_1 and γ_2.

4.6 Summary

Currently, as explained earlier in Chapter 2 there are very few attempts in the literature for communication and teleoperation control methods that jointly address haptic communication requirements. This is one of the two main motivations behind this chapter. The other one is the management of coexistence of critical URLLC traffic with non-critical traffic.

To this end, the contribution presented in this chapter focuses on a system that enables traffic prioritisation for URLLC using a priority queuing mechanism with priority upgrades. Two URLLC traffic flows are being served by the system, one critical (i.e., of high priority) and one non-critical (i.e., of low priority). A key component of this system is that its performance can be controlled by using a parameter. In this case, this parameter allows the control of queuing latency of low and high priority URLLC traffic. The appropriate tuning of this parameter is the main goal of the proposed system.

Furthermore, in this contribution, a bi-objective optimisation problem is proposed and formulated where one objective is the minimisation of low priority URLLC traffic latency and the other objective is the maximisation of performance of a bilateral teleoperation generating the critical URLLC flow. The first objective is enabled by the use of the analytically tractable queuing system HOL-MBP as seen in [278]. It must be noted that this is a theoretical model of a network scheduler which has not been implemented in real-life network systems, therefore generalisation of the results requires the further development of the model to better reflect such systems. The second objective is enabled by the main contribution of this chapter, the CPLF polynomial that quantifies the performance of control schemes for bilateral teleoperation as a function of delay.

Results were presented for a variety of indicative control schemes. A methodology of selecting the best possible control scheme given certain network delay conditions. It needs to be noted that CPLF is intended to be a proof-of-concept framework that allows the haptic communication system designer to narrow down the available control scheme options, even though such decision can be made based on multiple factors and not just the performance of the control scheme under different delay values. Furthermore, the framework is deterministic and the variability of the model parameters is not examined in this work.

Apart from the prior determination of the teleoperation and control parameters used in the modelling of the system, an important aspect of this paper is the selection of appropriate weights for using CPLF (as in Equation (4.2)) in various haptic communication scenarios, as well as the selection of appropriate scaling parameters for solving the bi-objective problem (as in Equation (4.16)). The optimisation scaling parameters define the preference towards each objective.

Chapter 5

Conclusion

5.1 Overview

The Tactile Internet will be enabled by future networks, aiming to satisfy requirements of various diverse use cases, such as haptic communication. The thesis approached the topic of haptic communication over the 5G Tactile Internet from the point-of-view of QoS provisioning and included an analysis and experimental evaluation of URLLC and haptic communication enablers. It also provided contributions regarding the 5G architecture and QoS provisioning for haptic communication.

This chapter includes a summary of the thesis and reflections on the presented contributions, analyses, methodologies and experiments. Furthermore, there is a discussion on the main contributions presented in this thesis as well as an exploration of future avenues of research.

5.2 Thesis Summary

With the continuous improvements of the network infrastructure capabilities, it will be possible for future networks to accommodate URLLC use cases, such as haptic communication, as we move towards the era of the 5G Tactile Internet. In this thesis, a systematic evaluation of the current state-of-the-art enablers of haptic communication reveals the need for addressing haptic communication issues, such as haptic system instability caused by latency, with the use of joint communication and teleoperation control methodologies. Furthermore, results of experimentation frameworks for evaluation of haptic communication over 5G networks indicated that it is possible to prioritise URLLC traffic to guarantee high QoS and therefore good bilateral teleoperation performance. Nevertheless, under certain conditions this results in resource starvation for traffic of lower priority. With respect to network performance, the outcome of the experiments showed an increase in packet loss and jitter for low priority traffic. On the other hand, without guaranteed QoS, the performance of a high priority bilateral teleoperation system decreases and the system becomes unstable.

To address the necessary changes in the 5G infrastructure for accommodating URLLC traffic, the thesis first proposes a possible 5G architecture using virtualised 3GPP components enabled by SDN and NFV. It also suggests the necessary architectural design entities and their interfaces that enable the Tactile Internet. Furthermore, the main contribution of this thesis is a novel framework that focuses on minimising the effects of QoS overprovisioning in terms of latency, assuming a dynamic priority queuing system. The framework optimises the balance between maximising teleoperation performance of high priority URLLC traffic and minimising resource starvation of low priority URLLC traffic. A key component of this framework is a proposed analytical model for measuring the impact of delay on teleoperation performance.

Regarding the analytical model, the simulated performance of selected teleoperation control methods was evaluated and results revealed the impact delay and teleoperation model parameters have on each teleoperation control method. It was also shown that it is possible to compare control methods and assist the system designer in selecting the best possible method. Moreover, the analytical model can also be used under different haptic communication scenarios. The overall optimisation framework was also simulated and evaluated and was shown that the computed optimum solutions depend on the haptic communication scenario and the selection of optimisation parameters.

5.3 Main Contributions

This section includes a discussion on the main contributions of the thesis. More specifically these contributions are mentioned here in order of appearance in this thesis to also reflect the thought process that the thesis flow depicts.

Firstly, the thesis surveys, evaluates and analyses the enablers of haptic communication over the TI infrastructure. Key contributions of this effort is the collection and definition of URLLC use case requirements as well as the evaluation of current efforts in joint communication and teleoperation control methodologies.

Second, experimental frameworks for evaluating haptic communication and 5G enablers are proposed. This contribution showed and evaluated how URLLC QoS can be guaranteed, as well as measure the impact of QoS overprovisioning. This evaluation focused on network performance metrics of packet loss and jitter. In the experiments both indicative and novel haptic communication systems were used. Results showed and verified the effectiveness of 5G enablers to guarantee QoS for URLLC, as well as the need for traffic prioritisation mechanisms to avoid QoS overprovisioning and resource starvation for BE traffic served by the network.

Relying on the above results and due to the fact that the TI requires a 5G architecture capable of satisfying the requirements of URLLC and other types of traffic by means of traffic differentiation, this thesis proposes a 5G architecture that orchestrates 3GPP 5G architecture components. Each component is implemented as a VNF over a network with SDN capabilities for haptic communication. Furthermore, the required description of main architecture design components and their interfaces is made with respect to the standardisation efforts in IEEE P1918.1.

Considering all the above, the thesis makes its main contribution regarding dynamic traffic prioritisation for haptic communication. This contribution presents a framework for optimisation of traffic prioritisation between URLLC traffic for bilateral teleoperation and lower priority URLLC traffic. This framework also enables the selection of the most suitable teleoperation control scheme, to maintain the stability and transparency of the teleoperation system at desirable levels. Results showed how this framework is able to evaluate teleoperation control schemes based on the impact of delay in their performance and how it is possible to optimise the fairness of a priority queuing system to satisfy the requirements of two prioritised flows.

5.4 Future avenues of research

In this section, future development of this work as well as future research directions regarding haptic communication and the Tactile Internet are discussed.

5.4.1 Future development of the thesis's work

Regarding Chapter 2, the analysis of the control schemes presented in section 2.4 can be further extended and enhanced. First of all, the extension of the analysis would offer a larger variety of control schemes to be used by CPLF. Second, the models of human user and remote environment used are simple. Even though these models are good enough to prove the main points in this work, it would be of interest to provide an analysis with more accurate modelling of the dynamics. As a result this can lead to more accurate representation of the bilateral teleoperation performance metrics used in CPLF of Chapter 4.

An important aspect of evaluating bilateral teleoperation systems is the determination of QoE that the system can offer. In this study, the experimental frameworks present results of jitter and packet loss regarding the co-existence of traffic. Even though these results are essential metrics of network performance and it is important to know how much they can be improved by the enablers of haptic communication, it is also important to evaluate the improvement in terms of assessing the QoE that the end-to-end system can offer by means of subjective tests.

A more practical aspect of the experiments in Chapter 3 is the network protocols used. In essence, the network performance measurement tool used is iperf, which can provide testing for UDP and TCP traffic it generates. Evaluation of how other protocols affect the performance of the haptic communication system would be also of interest.

In Chapter 4, even though the theoretical model HOL-MBP is analytically tractable, its analysis is not simple. This resulted in using a model of a single system of two queues that can accommodate only two flows of high and low priority traffic. This limits how accurately the system can emulate the desired network conditions. It is possible to use the model to emulate an NxN switch that serves only one high and one low priority flow, but this is not a representative system. Although, according to the authors of [278], the model can be extended to using multiple prioritised flows by using a matrix-analytic methodology.

Furthermore, another interesting future work can be the emulation of a switch that is implementing the network scheduling algorithm of HOL-MBP, or any other that offers tuning their performance using parameters. It needs to be noted that currently, physical switches are unable to implement the queue jumps which requires accessing the content of the low priority queues of a switch and re-routing packets towards the high priority queue.

Last but not least, it is not necessary to combine CPLF with HOL-MBP for managing QoS overprovisioning. CPLF offers the possibility of integrating it in other systems that would benefit from prior knowledge of bilateral teleoperation performance under different values of delay.

5.4.2 Long-term future research directions

A list of long-term future research avenues is presented below:

Joint communication and teleoperation control

Being the primary goal of the main contribution of this thesis, joint communication and control approaches for haptic communication should be a goal for the research and standardisation community. High-quality bilateral teleoperation requires the joint orchestration of control and communication approaches to cope with limitations such as restricted transmission capacity, time-varying delay and random or bursty packet losses. So far, the number of studies that jointly consider stability-ensuring control and haptic data communication (including data reduction) is limited.

Network protocols

Regarding network protocols with respect to haptic communication, there is a need for protocols with a greater focus on haptic communication performance concerning application, transport or other layers involved. This is made even more complicated by the fact that haptic communication in many cases needs to be secure, therefore, new and secure methodologies that will have minimum negative impact in the QoE of the user are required.

Data reduction and teleoperation control

With respect to the combination of haptic data reduction and teleoperation control, plurality of combinations of systems that combine haptic data reduction and control algorithms is encouraged. Having a set of different control, data reduction and communication approaches to implement teleoperation systems is important, because they vary in their robustness towards certain QoS parameters (e.g. delay, delay variation or packet loss).

Task performance evaluation

Nevertheless, different control and communication approaches lead to different types of artefacts. The performance of such systems also varies among tasks (e.g. free space versus contact, soft objects versus rigid surface, etc.). To date, there is no common understanding about the preferred architecture for certain QoS parameters or tasks.

Haptic interfaces

The development of haptic interfaces and actuators will allow the natural and more precise execution and replication of the desired user movement, but also improve force feedback experience. According to current technology trends, haptic interfaces are to be used from devices connected to mobile networks. Therefore, it is essential to explore further how teleoperation systems can be optimally integrated into the next generation mobile networks. This includes the optimisation of the communication network by investigating the mobile network infrastructure and the development of new protocols and evaluation metrics based on precise traffic models.

Privacy

Last but not least, it is obvious that in the future generations of the Internet, operators will have a more active role in acquiring and processing user data, especially since data prediction will play a major role in optimising the QoS offered by the network. Furthermore, prioritisation of network traffic raises issues of net neutrality which need to be further addressed so that appropriate legislation can be implemented.

5.5 Concluding remarks

The ongoing evolution of haptic communication and 5G enablers shows that the path is set for revolutionary applications to become part of our everyday lives. This work is highlighting the need of an integrated computing, communication and control framework, with its components distributed over the 5G network infrastructure.

Bibliography

[1] ITU-T, "Itu-t technology watch report: The tactile internet," tech. rep., ITU-T, Geneva, Aug. 2014.

[2] 3GPP, "3gpp release 15 description," tech. rep., 3GPP, September 2019. TR 21.915, Release 15.

[3] Chih-Ping Li, Jing Jiang, W. Chen, Tingfang Ji, and J. Smee, "5g ultra-reliable and low-latency systems design," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[4] D. Lawrence, "Stability and transparency in bilateral teleoperation," *Robotics and Automation, IEEE Transactions on*, vol. 9, pp. 624–637, Oct 1993.

[5] S. Hirche and M. Buss, "Transparent data reduction in networked telepresence and teleaction systems. part ii: Time-delayed communication," *Presence: Teleoperators and Virtual Environments*, vol. 16, pp. 532–542, Sep 2007.

[6] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5g cellular networks," *CoRR*, vol. abs/1510.02826, 2015.

[7] 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.

[8] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, “Nfv and sdn—key technology enablers for 5g networks,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2468–2478, 2017.

[9] 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, pp. 3034–3059, Fourthquarter 2018.

[10] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, “5g-enabled tactile internet,” *IEEE Journal on Selected Areas in Communications*, vol. PP, no. 99, pp. 1–1, 2016.

[11] G. Fettweis, “The tactile internet: Applications and challenges,” *Vehicular Technology Magazine*, vol. 9, pp. 64–70, Mar 2014.

[12] J. Cha, Y. S. Ho, Y. Kim, J. Ryu, and I. Oakley, “A framework for haptic broadcasting,” *IEEE MultiMedia*, vol. 16, pp. 16–27, July 2009.

[13] S. Lederman and R. Klatzky, “Haptic perception: A tutorial,” *Attention, Perception, & Psychophysics*, vol. 71, no. 7, pp. 1439–1459, 2009.

[14] 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, pp. 138–145, May 2016.

[15] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5g wireless networks: A comprehensive survey,” *IEEE Communications Surveys Tutorials*, vol. 18, pp. 1617–1655, thirdquarter 2016.

[16] P. Marti, J. Yepez, M. Velasco, R. Villa, and J. M. Fuertes, “Managing quality-of-control in network-based control systems by controller and message scheduling co-design,” *IEEE Transactions on Industrial Electronics*, vol. 51, pp. 1159–1167, Dec 2004.

[17] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera,” in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV ’03, (Washington, DC, USA), pp. 1403–, IEEE Computer Society, 2003.

[18] X. Shen, J. Zhou, A. El Saddik, and N. D. Georganas, "Architecture and evaluation of tele-haptic environments," in *Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004. Eighth IEEE International Symposium on*, pp. 53–60, Oct 2004.

[19] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," *Automatica*, vol. 42, no. 12, pp. 2035 – 2057, 2006.

[20] J. Lee and S. Payandeh, *Haptic Teleoperation Systems: Signal Processing Perspective.* Springer International Publishing, 2015.

[21] T. B. Sheridan, "Space teleoperation through time delay: review and prognosis," *IEEE Transactions on Robotics and Automation*, vol. 9, pp. 592–606, Oct 1993.

[22] N. Pavon-Pulido, J. Lopez-Riquelme, J. Pinuaga-Cascales, J. Ferruz-Melero, and R. Morais Dos Santos, "Cybi: A smart companion robot for elderly people: Improving teleoperation and telepresence skills by combining cloud computing technologies and fuzzy logic," in *Autonomous Robot Systems and Competitions (ICARSC), 2015 IEEE International Conference on*, pp. 198–203, April 2015.

[23] D. V. D. Berg, R. Glans, D. D. Koning, F. A. Kuipers, J. Lugtenburg, K. Polachan, P. T. Venkata, C. Singh, B. Turkovic, and B. V. Wijk, "Challenges in haptic communications over the tactile internet," *IEEE Access*, vol. 5, pp. 23502–23518, 2017.

[24] K. Zareinia, Y. Maddahi, C. Ng, N. Sepehri, and G. R. Sutherland, "Performance evaluation of haptic hand-controllers in a robot-assisted surgical system," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 11, no. 4, pp. 486–501, 2015.

[25] M. Aiple and A. Schiele, "Pushing the limits of the CyberGrasp™ for haptic rendering," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 3541–3546, May 2013.

[26] L. Winfield, J. Glassmire, J. Colgate, and M. Peshkin, "T-pad: Tactile pattern display through variable friction reduction," in *EuroHaptics Conference, 2007 and Symposium*

on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint, pp. 421–426, March 2007.

[27] NeuroDigital Technologies, "Gloveone." `https://www.neurodigital.es/`.

[28] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands — review," *Robotics and Autonomous Systems*, vol. 74, Part A, pp. 195 – 220, 2015.

[29] A. El Saddik, "C-HAVE: Collaborative haptic audio visual environments and systems," *Multimedia Systems*, vol. 13, no. 4, pp. 251–251, 2008.

[30] J. Colgate and J. Brown, "Factors affecting the z-width of a haptic display," in *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pp. 3205–3210 vol.4, May 1994.

[31] H. Tan, J. Radcliffe, B. N. Ga, H. Z. Tan, B. Eberman, M. A. Srinivasan, and B. Cheng, "Human factors for the design of force-reflecting haptic interfaces," 1994.

[32] J. E. Colgate and G. G. Schenkel, "Passivity of a class of sampled-data systems: Application to haptic interfaces," *Journal of Robotic Systems*, vol. 14, no. 1, pp. 37–47, 1997.

[33] J. J. Abbott and A. M. Okamura, "Effects of position quantization and sampling rate on virtual-wall passivity," *IEEE Transactions on Robotics*, vol. 21, pp. 952–964, Oct 2005.

[34] N. Diolaiti, G. Niemeyer, F. Barbagli, and J. K. Salisbury, "Stability of haptic rendering: Discretization, quantization, time delay, and coulomb effects," *IEEE Transactions on Robotics*, vol. 22, pp. 256–268, April 2006.

[35] E. Steinbach, S. Hirche, M. Ernst, F. Brandi, R. Chaudhari, J. Kammerl, and I. Vittorias, "Haptic communications," *Proceedings of the IEEE*, vol. 100, pp. 937–956, April 2012.

[36] J. Forsslund, S. Chan, J. Selesnick, K. Salisbury, R. G. Silva, and N. H. Blevins, "The effect of haptic degrees of freedom on task performance in virtual surgical environments.," in *MMVR* (J. D. Westwood, S. W. Westwood, L. Felländer-Tsai, R. S. Haluck, R. A.

Robb, S. Senger, and K. G. Vosburgh, eds.), vol. 184 of *Studies in Health Technology and Informatics*, pp. 129–135, IOS Press, 2013.

[37] V. Nitsch and B. Farber, "A meta-analysis of the effects of haptic interfaces on task performance with teleoperation systems," *IEEE Transactions on Haptics*, vol. 6, no. 4, pp. 387–398, 2013.

[38] A. Mora and A. Barrientos, "An experimental study about the effect of interactions among functional factors in performance of telemanipulation systems," *Control Engineering Practice*, vol. 15, no. 1, pp. 29 – 41, 2007.

[39] B. Cizmeci, R. Chaudhari, X. Xu, N. Alt, and E. Steinbach, "A visual-haptic multiplexing scheme for teleoperation over constant-bitrate communication links," in *Haptics: Neuroscience, Devices, Modeling, and Applications* (M. Auvray and C. Duriez, eds.), vol. 8619 of *Lecture Notes in Computer Science*, pp. 131–138, Springer Berlin Heidelberg, 2014.

[40] Z. Cen, M. Mutka, Y. Liu, A. Goradia, and N. Xi, "Qos management of supermedia enhanced teleoperation via overlay networks," in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pp. 1630–1635, Aug 2005.

[41] T. Brooks, "Telerobotic response requirements," in *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, pp. 113–120, Nov 1990.

[42] K. Shimoga, "A survey of perceptual feedback issues in dexterous telemanipulation. i. finger force feedback," in *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pp. 263–270, Sep 1993.

[43] K. Shimoga, "A survey of perceptual feedback issues in dexterous telemanipulation. ii. finger touch feedback," in *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pp. 271–279, Sep 1993.

[44] P. Otanez, J. R. Moyne, and D. Tilbury, "Using deadbands to reduce communication in networked control systems," in *American Control Conference, 2002. Proceedings of the 2002*, vol. 4, pp. 3015–3020 vol.4, 2002.

[45] P. Hinterseer, E. Steinbach, S. Hirche, and M. Buss, "A novel, psychophysically motivated transmission approach for haptic data streams in telepresence and teleaction systems," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, pp. ii/1097–ii/1100 Vol. 2, March 2005.

[46] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems," *Signal Processing, IEEE Transactions on*, vol. 56, pp. 588–597, Feb 2008.

[47] M. Abu-Tair and A. Marshall, "An empirical model for multi-contact point haptic network traffic," in *Proceedings of the 2nd International Conference on Immersive Telecommunications*, IMMERSCOM '09, (Brussels, BEL), ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

[48] Q. Nasir and E. Khalil, "Perception based adaptive haptic communication protocol (pahcp)," in *Computer Systems and Industrial Informatics (ICCSII), 2012 International Conference on*, pp. 1–6, Dec 2012.

[49] C. Shahabi, A. Ortega, and M. Kolahdouzan, "A comparison of different haptic compression techniques," in *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, vol. 1, pp. 657–660 vol.1, 2002.

[50] Y. You and M. Y. Sung, "Haptic data transmission based on the prediction and compression," in *Communications, 2008. ICC '08. IEEE International Conference on*, pp. 1824–1828, May 2008.

[51] H. Tanaka and K. Ohnishi, "Haptic data compression/decompression using dct for motion copy system," in *Mechatronics, 2009. ICM 2009. IEEE International Conference on*, pp. 1–6, April 2009.

[52] A. Kuzu, E. A. Baran, S. Bogosyan, M. Gokasan, and A. Sabanovic, "Wavelet packet transform-based compression for teleoperation," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 229, no. 7, pp. 639–651, 2015.

[53] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari, "Haptic data compression and communication," *Signal Processing Magazine, IEEE*, vol. 28, pp. 87–96, Jan 2011.

[54] V. Nitsch, B. Farber, L. Geiger, P. Hinterseer, and E. Steinbach, "An experimental study of lossy compression in a real telepresence and teleaction system," in *Haptic Audio visual Environments and Games, 2008. HAVE 2008. IEEE International Workshop on*, pp. 75–80, Oct 2008.

[55] E. Weber, *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae ...* Annotationes anatomicae et physiologicae, C.F. Koehler, 1834.

[56] J. Awed, I. Elhajj, and N. Slobodenyuk, "Haptic force perception in bimanual manipulation," in *Haptics: Perception, Devices, Mobility, and Communication* (P. Isokoski and J. Springare, eds.), vol. 7283 of *Lecture Notes in Computer Science*, pp. 1–6, Springer Berlin Heidelberg, 2012.

[57] S. Allin, Y. Matsuoka, and R. Klatzky, "Measuring just noticeable differences for haptic force feedback: implications for rehabilitation," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pp. 299–302, 2002.

[58] I. Jo, Y. Park, H. Kim, and J. Bae, "Evaluation of a wearable hand kinesthetic feedback system for virtual reality: Psychophysical and user experience evaluation," *IEEE Transactions on Human-Machine Systems*, vol. 49, pp. 430–439, Oct 2019.

[59] F. H. Petzschner, S. Glasauer, and K. E. Stephan, "A bayesian perspective on magnitude estimation," *Trends in Cognitive Sciences*, vol. 19, no. 5, pp. 285 – 293, 2015.

[60] N. Sakr, J. Zhou, N. D. Georganas, J. Zhao, and E. Petriu, "Robust perception-based data reduction and transmission in telehaptic systems," in *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pp. 214–219, March 2009.

[61] A. Bhardwaj, O. Dabeer, and S. Chaudhuri, "Can we improve over weber sampling of haptic signals?," in *Information Theory and Applications Workshop (ITA), 2013*, pp. 1–6, Feb 2013.

[62] A. Bhardwaj, S. Chaudhuri, and O. Dabeer, "Design and analysis of predictive sampling of haptic signals," *ACM Trans. Appl. Percept.*, vol. 11, pp. 16:1–16:20, Dec. 2014.

[63] P. Hinterseer, E. Steibach, and S. Chaudhuri, "Model based data compression for 3d virtual haptic teleinteraction," in *Consumer Electronics, 2006. ICCE '06. 2006 Digest of Technical Papers. International Conference on*, pp. 23–24, Jan 2006.

[64] P. Hinterseer, E. Steinbach, and S. Chaudhuri, "Perception-based compression of haptic data streams using kalman filters," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5, pp. V–V, May 2006.

[65] N. Sakr, N. Georganas, J. Zhao, and X. Shen, "Motion and force prediction in haptic media," in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 2242–2245, July 2007.

[66] F. Guo, C. Zhang, and Y. He, "Haptic data compression based on a linear prediction model and quadratic curve reconstruction," *Journal of Software*, vol. 9, no. 11, 2014.

[67] X. Xu, J. Kammerl, R. Chaudhari, and E. Steinbach, "Hybrid signal-based and geometry-based prediction for haptic data reduction," in *Haptic Audio Visual Environments and Games (HAVE), 2011 IEEE International Workshop on*, pp. 68–73, Oct 2011.

[68] S. Payandeh and J.-Y. Lee, "User-centered force signal processing for internet-based tele-manipulation: An overview," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pp. 3978–3983, Oct 2014.

[69] I. Vittorias, J. Kammerl, S. Hirche, and E. Steinbach, "Perceptual coding of haptic data in time-delayed teleoperation," in *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pp. 208–213, March 2009.

[70] K. J. Kuchenbecker, J. Fiene, and G. Niemeyer, "Improving contact realism through event-based haptic feedback," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, pp. 219–230, Mar. 2006.

[71] W. McMahan, J. Romano, A. Abdul Rahuman, and K. Kuchenbecker, "High frequency acceleration feedback significantly increases the realism of haptically rendered textured surfaces," in *Haptics Symposium, 2010 IEEE*, pp. 141–148, March 2010.

[72] A. Okamura, J. Dennerlein, and R. Howe, "Vibration feedback models for virtual environments," in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 1, pp. 674–679 vol.1, May 1998.

[73] K. J. Kuchenbecker, J. M. Romano, and W. McMahan, "Haptography: Capturing and recreating the rich feel of real surfaces.," in *Int. Symp. on Robotics Research (invited paper)*, vol. 70 of *Springer Tracts in Advanced Robotics*, pp. 245–260, Springer, 2009.

[74] V. Guruswamy, J. Lang, and W.-S. Lee, "Iir filter models of haptic vibration textures," *Instrumentation and Measurement, IEEE Transactions on*, vol. 60, pp. 93–103, Jan 2011.

[75] S. Okamoto and Y. Yamada, "Perceptual properties of vibrotactile material texture: Effects of amplitude changes and stimuli beneath detection thresholds," in *System Integration (SII), 2010 IEEE/SICE International Symposium on*, pp. 384–389, Dec 2010.

[76] J. Craig, "Difference threshold for intensity of tactile stimuli," *Perception & Psychophysics*, vol. 11, no. 2, pp. 150–152, 1972.

[77] H. Pongrac, "Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions," *Multimedia Systems*, vol. 13, no. 4, pp. 297–307, 2008.

[78] C. Hatzfeld and R. Werthschützky, "Just noticeable differences of low-intensity vibrotactile forces at the fingertip," in *Haptics: Perception, Devices, Mobility, and Communication* (P. Isokoski and J. Springare, eds.), vol. 7283 of *Lecture Notes in Computer Science*, pp. 43–48, Springer Berlin Heidelberg, 2012.

[79] R. Chaudhari, B. Çizmeci, K. J. Kuchenbecker, S. Choi, and E. Steinbach, "Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation," in *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, (New York, NY, USA), pp. 409–418, ACM, 2012.

[80] I. T. Union and I. T. U. T. S. Sector, *Coding of Speech at 8 Kbit/s Using Conjugate-structure Algebraic-code-excited Linear-prediction (CS-ACELP): Reduced complexity 8 kbit/s CS-ACELP speech codec. Annex A.* ITU-T recommendation, International Telecommunication Union, 1996.

[81] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, "Perceptual and bitrate-scalable coding of haptic surface texture signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, pp. 462–473, April 2015.

[82] E. Zwicker and R. Feldtkeller, *Das Ohr als Nachrichtenempfänger.* Monographien der elektrischen Nachrichtentechnik, Hirzel, 1967.

[83] A. Tatematsu, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Qoe assessment in tele-operation with 3d video and haptic media," in *2011 IEEE International Conference on Multimedia and Expo*, pp. 1–6, July 2011.

[84] X. Liu, M. Dohler, T. Mahmoodi, and H. Liu, "Challenges and opportunities for designing tactile codecs from audio codecs," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[85] C. Aldana, E. Nuno, and L. Basanez, "Control of bilateral teleoperators in the operational space without velocity measurements," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 5445–5450, Nov 2013.

[86] L. Chan, F. Naghdy, and D. Stirling, "Application of adaptive controllers in teleoperation systems: A survey," *Human-Machine Systems, IEEE Transactions on*, vol. 44, pp. 337–352, June 2014.

[87] P. Arcara, C. Melchiorri, "Control schemes for teleoperation with time delay: A comparative study," 38 (2002) 49–64, Robotics and Autonomous Systems, 2002.

[88] C. Passenberg, A. Peer, and M. Buss, "A survey of environment-, operator-, and task-adapted controllers for teleoperation systems," *Mechatronics*, vol. 20, no. 7, pp. 787 – 801, 2010. Special Issue on Design and Control Methodologies in Telerobotics.

[89] D. Lee, O. Martinez-Palafox, and M. W. Spong, "Bilateral teleoperation of multiple cooperative robots over delayed communication networks: Application," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, April 18-22, 2005, Barcelona, Spain*, pp. 366–371, 2005.

[90] S. Sirouspour and P. Setoodeh, "Adaptive nonlinear teleoperation control in multi-master/multi-slave environments," in *Control Applications, 2005. CCA 2005. Proceedings of 2005 IEEE Conference on*, pp. 1263–1268, Aug 2005.

[91] T. Maneewarn and P. Tothawornyuenyong, "Effects of system delay in passivity-based haptic system," in *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, pp. 1–6, Dec 2006.

[92] N. Yasrebi and D. Constantinescu, "Passive wave variable control of haptic interaction with an unknown virtual environment," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 919–924, May 2011.

[93] K. Hashtrudi-Zaad and S. Salcudean, "Analysis and evaluation of stability and performance robustness for teleoperation control architectures," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 4, pp. 3107–3113 vol.4, 2000.

[94] W. Iida and K. Ohnishi, "Reproducibility and operationality in bilateral teleoperation," in *Advanced Motion Control, 2004. AMC '04. The 8th IEEE International Workshop on*, pp. 217–222, March 2004.

[95] R. Chaudhari, C. Schuwerk, V. Nitsch, E. Steinbach, and B. Färber, "Opening the haptic loop: Network degradation limits for haptic task performance," in *Haptic Audio Visual Environments and Games (HAVE), 2011 IEEE International Workshop on*, pp. 56–61, Oct 2011.

[96] Jie Chen, "Corrections to "on computing the maximal delay intervals for stability of linear delay systems"," *IEEE Transactions on Automatic Control*, vol. 45, pp. 2198–2198, Nov 2000.

[97] P. Arcara, "Control of Haptic and Robotic Telemanipulation Systems," phd thesis, Università degli Studi di Bologna, 2014.

[98] B. Hannaford, "A design framework for teleoperators with kinesthetic feedback," *Robotics and Automation, IEEE Transactions on*, vol. 5, pp. 426–434, Aug 1989.

[99] W. R. Ferrell, "Delayed Force Feedback," *Human Factors*, vol. 8, no. 5, pp. 449–455, 1966. PMID: 5966936.

[100] W. S. Kim, B. Hannaford, and A. K. Fejczy, "Force-reflection and shared compliant control in operating telemanipulators with time delay," *IEEE Transactions on Robotics and Automation*, vol. 8, pp. 176–185, April 1992.

[101] G. J. Raju, G. C. Verghese, and T. B. Sheridan, "Design issues in 2-port network models of bilateral remote manipulation," in *International Conference on Robotics and Automation*, pp. 1316–1321 vol.3, May 1989.

[102] J. Kim, P. H. Chang, and H.-S. Park, "Transparent teleoperation using two-channel control architectures," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1953–1960, August 2005.

[103] R. Anderson and M. Spong, "Bilateral control of teleoperators with time delay," *Automatic Control, IEEE Transactions on*, vol. 34, pp. 494–501, May 1989.

[104] M. Dyck, A. Jazayeri, and M. Tavakoli, "Is the human operator in a teleoperation system passive?," in *World Haptics Conference (WHC), 2013*, pp. 683–688, April 2013.

[105] F. Buzan and T. Sheridan, "A model-based predictive operator aid for telemanipulators with time delay," in *Systems, Man and Cybernetics, 1989. Conference Proceedings., IEEE International Conference on*, pp. 138–143 vol.1, Nov 1989.

[106] C. Secchi, S. Stramigioli, and C. Fantuzzi, "Digital passive geometric telemanipulation," in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, vol. 3, pp. 3290–3295 vol.3, Sept 2003.

[107] J. Slotine and W. Li, *Applied Nonlinear Control.* Prentice Hall, 1991.

[108] A. Jazayeri and M. Tavakoli, "Bilateral teleoperation system stability with non-passive and strictly passive operator or environment," *Control Engineering Practice*, vol. 40, pp. 45 – 60, 2015.

[109] E. Nuño, L. Basañez, and R. Ortega, "Passivity-based control for bilateral teleoperation: A tutorial," *Automatica*, vol. 47, no. 3, pp. 485 – 495, 2011.

[110] D. Lee and P. Li, "Passive bilateral control and tool dynamics rendering for nonlinear mechanical teleoperators," *Robotics, IEEE Transactions on*, vol. 21, pp. 936–951, Oct 2005.

[111] E. Nuño, R. Ortega, N. Barabanov, and L. Basanez, "A globally stable pd controller for bilateral teleoperators," *Robotics, IEEE Transactions on*, vol. 24, pp. 753–758, June 2008.

[112] J. Wu, Y. Shi, J. Huang, and D. Constantinescu, "Stochastic stabilization for bilateral teleoperation over networks with probabilistic delays," *Mechatronics*, vol. 22, no. 8, pp. 1050 – 1059, 2012.

[113] K. Abidi, Y. Yildiz, and B. E. Korpe, "Explicit adaptive time-delay compensation for bilateral teleoperation," in *Control and Decision Conference (CCDC), 2015 27th Chinese*, pp. 3379–3383, May 2015.

[114] G. M. H. Leung, B. A. Francis, and J. Apkarian, "Bilateral controller for teleoperators with time delay via µ-synthesis," *IEEE Transactions on Robotics and Automation*, vol. 11, pp. 105–116, 1997.

[115] A. Alfi, A. Khosravi, and A. Lari, "Swarm-based structure-specified controller design for bilateral transparent teleoperation systems via µ synthesis," *IMA Journal of Mathematical Control and Information*, vol. 31, no. 1, pp. 111–136, 2014.

[116] J. Sheng and M. Spong, "Model predictive control for bilateral teleoperation systems with time delays," in *Electrical and Computer Engineering, 2004. Canadian Conference on*, vol. 4, pp. 1877–1880 Vol.4, May 2004.

[117] Z. Li, X. Cao, and N. Ding, "Adaptive fuzzy control for synchronization of nonlinear teleoperators with stochastic time-varying communication delays," *Fuzzy Systems, IEEE Transactions on*, vol. 19, pp. 745–757, Aug 2011.

[118] Y. Yang, C. Hua, and X. Guan, "Synchronization control for bilateral teleoperation system with prescribed performance under asymmetric time delay," *Nonlinear Dynamics*, vol. 81, no. 1-2, pp. 481–493, 2015.

[119] I. Sarras, E. Nuño, M. Kinnaert, and L. Basañez, "Output-feedback control of nonlinear bilateral teleoperators," in *American Control Conference (ACC), 2012*, pp. 3490–3495, June 2012.

[120] A. Jafari, M. Nabeel, and J. H. Ryu, "Multi degree-of-freedom input-to-state stable approach for stable haptic interaction," in *World Haptics Conference (WHC), 2015 IEEE*, pp. 293–298, June 2015.

[121] H. Morimitsu and S. Katsura, "Circle theorem-based realization of nonlinear force control for teleoperation under time delay," in *Mechatronics (ICM), 2015 IEEE International Conference on*, pp. 562–567, March 2015.

[122] K. Natori, T. Tsuji, K. Ohnishi, A. Hace, and K. Jezernik, "Time-delay compensation by communication disturbance observer for bilateral teleoperation under time-varying delay," *Industrial Electronics, IEEE Transactions on*, vol. 57, pp. 1050–1062, March 2010.

[123] A. Suzuki and K. Ohnishi, "Frequency-domain damping design for time-delayed bilateral teleoperation system based on modal space analysis," *Industrial Electronics, IEEE Transactions on*, vol. 60, pp. 177–190, Jan 2013.

[124] J. H. Park and H. C. Cho, “Sliding-mode controller for bilateral teleoperation with varying time delay,” in *Advanced Intelligent Mechatronics, 1999. Proceedings. 1999 IEEE/ASME International Conference on*, pp. 311–316, 1999.

[125] F. Llewellyn, “Some fundamental properties of transmission systems,” *Proceedings of the IRE*, vol. 40, pp. 271–283, March 1952.

[126] C. Pacchierotti, A. Tirmizi, G. Bianchini, and D. Prattichizzo, “Enhancing the performance of passive teleoperation systems via cutaneous feedback.,” *IEEE transactions on haptics*, 2015.

[127] G. Niemeyer and J.-J. Slotine, “Stable adaptive teleoperation,” *Oceanic Engineering, IEEE Journal of*, vol. 16, pp. 152–162, Jan 1991.

[128] G. Niemeyer and J.-J. Slotine, “Towards force-reflecting teleoperation over the internet,” in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 3, pp. 1909–1915 vol.3, May 1998.

[129] D. Sun, F. Naghdy, and H. Du, “Application of wave-variable control to bilateral teleoperation systems: A survey,” *Annual Reviews in Control*, vol. 38, no. 1, pp. 12–31, 2014.

[130] A. Aziminejad, M. Tavakoli, R. Patel, and M. Moallem, “Wave-based time delay compensation in bilateral teleoperation: Two-channel versus four-channel architectures,” in *American Control Conference, 2007. ACC ’07*, pp. 1449–1454, July 2007.

[131] M. Boukhnifer and A. Ferreira, “Stability and transparency for scaled teleoperation system,” in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 4217–4222, Oct 2006.

[132] M. Alise, R. Roberts, D. Repperger, C. Moore, and S. Tosunoglu, “On extending the wave variable method to multiple-dof teleoperation systems,” *Mechatronics, IEEE/ASME Transactions on*, vol. 14, pp. 55–63, Feb 2009.

[133] Y. Ye and P. Liu, "Improving trajectory tracking in wave-variable-based teleoperation," *Mechatronics, IEEE/ASME Transactions on*, vol. 15, pp. 321–326, April 2010.

[134] E. Rodriguez-Seda and M. Spong, "A time-varying wave impedance approach for transparency compensation in bilateral teleoperation," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pp. 4609–4615, Oct 2009.

[135] H. Li and K. Kawashima, "Achieving stable tracking in wave-variable-based teleoperation," *Mechatronics, IEEE/ASME Transactions on*, vol. 19, pp. 1574–1582, Oct 2014.

[136] C. Zhang and K. T. Chong, "Passive bilateral teleoperation via the internet," in *SICE-ICASE, 2006. International Joint Conference*, pp. 3072–3078, Oct 2006.

[137] S. Ganjefar, H. Momeni, and F. Janabi-Sharifi, "Teleoperation systems design using augmented wave-variables and smith predictor method for reducing time-delay effects," in *Intelligent Control, 2002. Proceedings of the 2002 IEEE International Symposium on*, pp. 333–338, 2002.

[138] Y. Yokokohji, T. Tsujioka, and T. Yoshikawa, "Bilateral control with time-varying delay including communication blackout," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pp. 285–292, 2002.

[139] S. Yoo and K. Chong, "Adaptive wave variables for bilateral teleoperation using neural networks," *Neural Computing and Applications*, vol. 25, no. 6, pp. 1249–1262, 2014.

[140] S. Munir and W. Book, "Internet based teleoperation using wave variables with prediction," in *Advanced Intelligent Mechatronics, 2001. Proceedings. 2001 IEEE/ASME International Conference on*, vol. 1, pp. 43–50 vol.1, 2001.

[141] E. Mujcic and A. Oracevic, "Correction of position error in internet-based teleoperation," in *Telecommunications Forum (TELFOR), 2012 20th*, pp. 115–118, Nov 2012.

[142] N. Chopra, M. Spong, R. Ortega, and N. Barabanov, "On tracking performance in bilateral teleoperation," *Robotics, IEEE Transactions on*, vol. 22, pp. 861–866, Aug 2006.

[143] C. Secchi, S. Stramigioli, and C. Fantuzzi, "Compensation of position errors in passivity based teleoperation over packet switched communication networks," in *Proceedings of the 17th World Congress The International Federation of Automatic Control*, pp. 15648–15653, IFAC, July 2008.

[144] E. Nuno, L. Basanez, and R. Ortega, "Passive bilateral teleoperation framework for assisted robotic tasks," in *Robotics and Automation, 2007 IEEE International Conference on*, pp. 1645–1650, April 2007.

[145] A. F. Villaverde, A. Barreiro, and C. Raimúndez, "Passive position error correction in internet-based teleoperation," *Automatica*, vol. 46, no. 11, pp. 1884 – 1890, 2010.

[146] B. Hannaford and J.-H. Ryu, "Time-domain passivity control of haptic interfaces," *Robotics and Automation, IEEE Transactions on*, vol. 18, pp. 1–10, Feb 2002.

[147] J.-H. Ryu, D.-S. Kwon, and B. Hannaford, "Stable teleoperation with time-domain passivity control," *Robotics and Automation, IEEE Transactions on*, vol. 20, pp. 365–373, April 2004.

[148] J.-H. Ryu, J. Artigas, and C. Preusche, "A passive bilateral control scheme for a teleoperator with time-varying communication delay," *Mechatronics*, vol. 20, no. 7, pp. 812 – 823, 2010. Special Issue on Design and Control Methodologies in Telerobotics.

[149] C. Preusche, G. Hirzinger, J.-H. Ryu, and B. Hannaford, "Time domain passivity control for 6 degrees of freedom haptic displays," in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 3, pp. 2944–2949 vol.3, Oct 2003.

[150] H. Li, K. Tadano, and K. Kawashima, "Model-based passive bilateral teleoperation with time delay," *Transactions of the Institute of Measurement and Control*, vol. 36, no. 8, pp. 1010–1023, 2014.

[151] J.-P. Kim and J. Ryu, "Robustly stable haptic interaction control using an energy-bounding algorithm," *The International Journal of Robotics Research*, vol. 29, no. 6, pp. 666–679, 2010.

[152] L. Marton, J. Esclusa, P. Haller, and T. Vajda, “Passive bilateral teleoperation with bounded control signals,” in *Industrial Informatics (INDIN), 2013 11th IEEE International Conference on*, pp. 337–342, July 2013.

[153] Y. Hou and G. Luecke, “Time delayed teleoperation system control, a passivity-based method,” in *Advanced Robotics, 2005. ICAR ’05. Proceedings., 12th International Conference on*, pp. 796–802, July 2005.

[154] X. Xu, B. Cizmeci, C. Schuwerk, and E. Steinbach, “Haptic data reduction for time-delayed teleoperation using the time domain passivity approach,” in *IEEE World Haptics Conference, Chicago, USA*, Jun 2015.

[155] J. Rebelo and A. Schiele, “Time domain passivity controller for 4-channel time-delay bilateral teleoperation,” *Haptics, IEEE Transactions on*, vol. 8, pp. 79–89, Jan 2015.

[156] J. Artigas, J.-H. Ryu, C. Preusche, and G. Hirzinger, “Network representation and passivity of delayed teleoperation systems,” in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 177–183, Sept 2011.

[157] J.-H. Ryu, D.-S. Kwon, and B. Hannaford, “Stability guaranteed control: time domain passivity approach,” *Control Systems Technology, IEEE Transactions on*, vol. 12, pp. 860–868, Nov 2004.

[158] H. V. Quang, I. Farkhatdinov, and J.-H. Ryu, “Passivity of delayed bilateral teleoperation of mobile robots with ambiguous causalities: Time domain passivity approach,” in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 2635–2640, Oct 2012.

[159] V. Chawda and M. OMalley, “Position synchronization in bilateral teleoperation under time-varying communication delays,” *Mechatronics, IEEE/ASME Transactions on*, vol. 20, pp. 245–253, Feb 2015.

[160] W.-H. Zhu and S. Salcudean, “Teleoperation with adaptive motion/force control,” in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, vol. 1, pp. 231–237, IEEE, 1999.

[161] S. Ganjefar, H. Momeni, F. J. Sharifi, and M. T. H. Beheshti, “Behavior of Smith predictor in teleoperation systems with modeling and delay time errors,” in *IEEE Conference on Control Applications*, vol. 2, pp. 1176–1180 vol.2, June 2003.

[162] A. C. Smith and K. Hashtrudi-Zaad, “Smith Predictor Type Control Architectures for Time Delayed Teleoperation,” *Internatonal Journal of Robotic Research*, vol. 25, pp. 797–818, August 2006.

[163] G. Hirzinger, J. Heindl and K. Landzettel , “Closer Control of Loops with Dead Time,” tech. rep., Chemistry Engineering Progress, Vol. 53, No. 5, pp. 217-219., 1957.

[164] A. Smith and K. Hashtrudi-Zaad, “Neural network-based teleoperation using smith predictors,” in *Mechatronics and Automation, 2005 IEEE International Conference*, vol. 3, pp. 1654–1659 Vol. 3, 2005.

[165] Z. Zeinaly, A. Ramezani, and S. Ozgoli, “Robust stability for bilateral teleoperation systems under time-varying delay by communication disturbance observer,” in *2014 Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*, pp. 143–148, Oct 2014.

[166] K. Natori, R. Kubo, and K. Ohnishi, “Effects of controller parameters on transparency of time delayed bilateral teleoperation systems with communication disturbance observer,” in *2008 IEEE International Symposium on Industrial Electronics*, pp. 1287–1292, June 2008.

[167] A. Suzuki and K. Ohnishi, “Novel four-channel bilateral control design for haptic communication under time delay based on modal space analysis,” *Control Systems Technology, IEEE Transactions on*, vol. 21, pp. 882–890, May 2013.

[168] T. Tashiro, T. Mizoguchi, T. Shimono, and K. Ohnishi, “Time delay compensation for dob-based contact force control under time delay,” in *2017 IEEE International Conference on Mechatronics (ICM)*, pp. 214–219, Feb 2017.

[169] J. S. Lee, T.-C. Chien, D. H. Jian, and Y. H. Sun, "A comparative study of communication disturbance observers for time-delay systems," in *2014 9th IEEE Conference on Industrial Electronics and Applications*, pp. 616–620, June 2014.

[170] S. Hirche and M. Buss, "Human-oriented control for haptic teleoperation," *Proceedings of the IEEE*, vol. 100, pp. 623–647, March 2012.

[171] P. Mitra and G. Niemeyer, "Model-mediated telemanipulation," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 253–262, 2008.

[172] C. Passenberg, A. Peer, and M. Buss, "Model-mediated teleoperation for multi-operator multi-robot systems," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 4263–4268, Oct 2010.

[173] B. Willaert, H. Van Brussel, and G. Niemeyer, "Stability of model-mediated teleoperation: Discussion and experiments," in *Haptics: Perception, Devices, Mobility, and Communication* (P. Isokoski and J. Springare, eds.), vol. 7282 of *Lecture Notes in Computer Science*, pp. 625–636, Springer Berlin Heidelberg, 2012.

[174] X. Xu, C. Schuwerk, and E. Steinbach, "Passivity-based model updating for model-mediated teleoperation," in *Multimedia Expo Workshops (ICMEW), 2015 IEEE International Conference on*, pp. 1–6, June 2015.

[175] L. Love and W. Book, "Force reflecting teleoperation with adaptive impedance control," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, pp. 159–165, Feb 2004.

[176] L. Huijun and S. Aiguo, "Virtual-environment modeling and correction for force-reflecting teleoperation with time delay," *Industrial Electronics, IEEE Transactions on*, vol. 54, pp. 1227–1233, April 2007.

[177] N. Diolaiti, C. Melchiorri, and S. Stramigioli, "Contact impedance estimation for robotic systems," *Robotics, IEEE Transactions on*, vol. 21, pp. 925–935, Oct 2005.

[178] A. Haddadi and K. Hashtrudi-Zaad, "A new method for online parameter estimation of hunt-crossley environment dynamic models," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 981–986, Sept 2008.

[179] X. Xu and E. Steinbach, "Towards real-time modeling and haptic rendering of deformable objects for point cloud-based model-mediated teleoperation," in *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, pp. 1–6, July 2014.

[180] X. Xu, S. Chen, and E. Steinbach, "Model-mediated teleoperation for movable objects: Dynamics modeling and packet rate reduction," in *14th IEEE International Symposium on Haptic Audio-Visual Environments and Games*, (Ottawa, Canada), Oct 2015.

[181] H. Jin-quan and F. Lewis, "Neural-network predictive control for nonlinear dynamic systems with time-delay," *Neural Networks, IEEE Transactions on*, vol. 14, pp. 377–389, Mar 2003.

[182] P. Prekopiou, S. Tzafestas, and W. Harwin, "Towards variable-time-delays-robust telemanipulation through master state prediction," in *Advanced Intelligent Mechatronics, 1999. Proceedings. 1999 IEEE/ASME International Conference on*, pp. 305–310, 1999.

[183] I. B. France, "Predictive algorithms for distant touching," 2002.

[184] C. Weber, V. Nitsch, U. Unterhinninghofen, B. Farber, and M. Buss, "Position and force augmentation in a telepresence system and their effects on perceived realism," in *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pp. 226–231, March 2009.

[185] D. Verscheure, J. Swevers, H. Bruyninckx, and J. De Schutter, "On-line identification of contact dynamics in the presence of geometric uncertainties," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 851–856, May 2008.

[186] S. Clarke, G. Schillhuber, M. Zaeh, and H. Ulbrich, "Telepresence across delayed networks: a combined prediction and compression approach," in *Haptic Audio Visual Environments and their Applications, 2006. HAVE 2006. IEEE International Workshop on*, pp. 171–175, Nov 2006.

[187] X. Xu, B. Cizmeci, A. Al-Nuaimi, and E. Steinbach, "Point cloud-based model-mediated teleoperation with dynamic and perception-based model updating," *Instrumentation and Measurement, IEEE Transactions on*, vol. 63, pp. 2558–2569, Nov 2014.

[188] B. Willaert, J. Bohg, H. Van Brussel, and G. Niemeyer, "Towards multi-dof model mediated teleoperation: Using vision to augment feedback," in *Haptic Audio Visual Environments and Games (HAVE), 2012 IEEE International Workshop on*, pp. 25–31, Oct 2012.

[189] A. Jafari, M. Nabeel, and J.-H. Ryu, "Stable bilateral teleoperation with input-to-state stable approach," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), Hamburg, Germany*, 2015.

[190] A. Jafari, M. Nabeel, H. Singh, and J.-H. Ryu, "Stable and transparent teleoperation over communication time-delay: Observer-based input-to-state stable approach," in *IEEE Haptics Symposium 2016*, 2016.

[191] R. Uddin and J. Ryu, "Predictive control approaches for bilateral teleoperation," *Annual Reviews in Control*, vol. 42, pp. 82 – 99, 2016.

[192] T.-C. Yang, "Networked control system: a brief survey," *Control Theory and Applications, IEE Proceedings* -, vol. 153, pp. 403–412, July 2006.

[193] Q. Dang, B. Allouche, A. Dequidt, L. Vermeiren, and V. Dubreucq, "Real-time control of a force feedback haptic interface via ethercat fieldbus," in *Industrial Technology (ICIT), 2015 IEEE International Conference on*, pp. 441–446, March 2015.

[194] S. Tzafestas, *Web-Based Control and Robotics Education.* Intelligent Systems, Control and Automation: Science and Engineering, Springer Netherlands, 2009.

[195] B. Briscoe, A. Brunstrom, A. Petlund, D. Hayes, D. Ros, I. J. Tsang, S. Gjessing, G. Fairhurst, C. Griwodz, and M. Welzl, "Reducing internet latency: A survey of techniques and their merits," *IEEE Communications Surveys Tutorials*, vol. 18, pp. 2149–2196, thirdquarter 2016.

[196] Z. Shi, H. Zou, M. Rank, L. Chen, S. Hirche, and H. J. Muller, “Effects of packet loss and latency on the temporal discrimination of visual-haptic events,” *IEEE Transactions on Haptics*, vol. 3, pp. 28–36, Jan 2010.

[197] C. Seo, J.-P. Kim, J. Kim, H.-S. Ahn, and J. Ryu, “Robustly stable bilateral teleoperation under time-varying delays and data losses: an energy-bounding approach,” *Journal of Mechanical Science and Technology*, vol. 25, p. 2089, Sep 2011.

[198] P. Huang and Y. Ishibashi, “QoS control and QoE assessment in multi-sensory communications with haptics,” *IEICE Transactions on Communications*, vol. 96, no. 2, pp. 392–403, 2013.

[199] G. Kokkonis, K. Psannis, M. Roumeliotis, S. Kontogiannis, and Y. Ishibashi, “Evaluating transport and application layer protocols for haptic applications,” in *Haptic Audio Visual Environments and Games (HAVE), 2012 IEEE International Workshop on*, pp. 66–71, Oct 2012.

[200] K. M. Yap, A. Marshall, and W. Yu, “Providing qos for distributed haptic virtual environments in ip networks,” in *Proceedings of the First International Conference on Immersive Telecommunications*, ImmersCom ’07, (ICST, Brussels, Belgium, Belgium), pp. 13:1–13:6, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[201] G. Kokkonis, K. Psannis, M. Roumeliotis, and S. Kontogiannis, “A survey of transport protocols for haptic applications,” in *Informatics (PCI), 2012 16th Panhellenic Conference on*, pp. 192–197, Oct 2012.

[202] J.-y. Lee and S. Payandeh, “A layered protocol architecture for haptic teleoperation systems based on the bayesian approach,” in *Haptic Teleoperation Systems*, pp. 35–59, Springer International Publishing, 2015.

[203] W. R. Stevens, “Tcp slow start, congestion avoidance, fast retransmit, and fast recovery algorithms,” RFC 2001, RFC Editor, January 1997. `http://www.rfc-editor.org/rfc/rfc2001.txt`.

[204] N. Suzuki and S. Katsura, "Investigation of transport layer protocols for wireless haptic communication system," in *System Integration (SII), 2011 IEEE/SICE International Symposium on*, pp. 509–514, Dec 2011.

[205] S. Dodeller and N. Georganas, "Transport layer protocols for telehaptics update messages," in *Proc. 22nd Biennial Symposium on Communications, Queen's Univeristy, Canada*, pp. 191–193, 2004.

[206] L. Ping, L. Wenjuan, and S. Zengqi, "Transport layer protocol reconfiguration for network-based robot control system," in *Networking, Sensing and Control, 2005. Proceedings. 2005 IEEE*, pp. 1049–1053, March 2005.

[207] Z. Cen, M. W. Mutka, D. Zhu, and N. Xi, "Supermedia transport for teleoperations over overlay networks," in *NETWORKING 2005. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems* (R. Boutaba, K. Almeroth, R. Puigjaner, S. Shen, and J. P. Black, eds.), (Berlin, Heidelberg), pp. 1409–1412, Springer Berlin Heidelberg, 2005.

[208] Y. Uchimura, K. Ohnishi, and T. Yakoh, "Bilateral robot system on the real time network structure," in *Advanced Motion Control, 2002. 7th International Workshop on*, pp. 63–68, 2002.

[209] R. Wirz, M. Ferre, R. Marín, J. Barrio, J. Claver, and J. Ortego, "Efficient transport protocol for networked haptics applications," in *Haptics: Perception, Devices and Scenarios* (M. Ferre, ed.), vol. 5024 of *Lecture Notes in Computer Science*, pp. 3–12, Springer Berlin Heidelberg, 2008.

[210] A. Boukerche and H. Maamar, "An efficient hybrid multicast transport protocol for collaborative virtual environment with networked haptic," in *Haptic Audio Visual Environments and their Applications, 2006. HAVE 2006. IEEE International Workshop on*, pp. 78–83, 2006.

[211] S. Paul, K. K. Sabnani, J. C. . Lin, and S. Bhattacharyya, "Reliable multicast transport protocol (rmtp)," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 407–421, April 1997.

[212] V. Laviano, M. Pullen, and G. Mason, "Selectively reliable transmission protocol," ietf internet draft, Internet Engineering Task Force, March 1997.

[213] M. Pullen, F. Zhao, and D. Cohen, "Selectively reliable multicast protocol (srmp)," RFC 4410, RFC Editor, February 2006.

[214] H. Kim and J.-H. Ryu, "A study on the effect of haptic to video time-delay on teleoperation and a comment for improving the performance," in *Control, Automation and Systems (ICCAS), 2011 11th International Conference on*, pp. 1329–1332, Oct 2011.

[215] Z. Yuan, T. Bi, G.-M. Muntean, and G. Ghinea, "Perceived synchronization of mulsemedia services," *Multimedia, IEEE Transactions on*, vol. 17, pp. 957–966, July 2015.

[216] Z. Yuan, G. Ghinea, and G.-M. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *Multimedia, IEEE Transactions on*, vol. 17, pp. 104–117, Jan 2015.

[217] H. King, B. Hannaford, J. Kammerl, and E. Steinbach, "Establishing multimodal telepresence sessions using the session initiation protocol (sip) and advanced haptic codecs," in *Haptics Symposium, 2010 IEEE*, pp. 321–325, March 2010.

[218] M. Mauve, V. Hilt, C. Kuhmunch, and W. Effelsberg, "RTP/I-toward a common application level protocol for distributed interactive media," *Multimedia, IEEE Transactions on*, vol. 3, pp. 152–161, Mar 2001.

[219] Y. Lim, "MMT, new alternative to MPEG-2 TS and RTP," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*, pp. 1–5, June 2013.

[220] K. Venkatraman, S. Vellingiri, B. Prabhakaran, and N. Nguyen, "Mpeg media transport (mmt) for 3d tele-immersion systems," in *Multimedia (ISM), 2014 IEEE International Symposium on*, pp. 279–282, Dec 2014.

[221] H. Al Osman, M. Eid, R. Iglesias, and A. El Saddik, "ALPHAN: Application layer protocol for haptic networking," in *Haptic, Audio and Visual Environments and Games, 2007. HAVE 2007. IEEE International Workshop on*, pp. 96–101, Oct 2007.

[222] M. Eid, J. Cha, and A. E. Saddik, "Admux: An adaptive multiplexer for haptic-audio-visual data communication," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, pp. 21–31, Jan 2011.

[223] H. H. King, K. Tadano, R. Donlin, D. Friedman, M. J. H. Lum, V. Asch, C. Wang, K. Kawashima, and B. Hannaford, "Preliminary protocol for interoperable telesurgery," in *2009 International Conference on Advanced Robotics*, pp. 1–6, June 2009.

[224] V. Gokhale, S. Chaudhuri, and O. Dabeer, "HoIP: A point-to-point haptic data communication protocol and its evaluation," in *Communications (NCC), 2015 Twenty First National Conference on*, pp. 1–6, Feb 2015.

[225] Z. Cen, M. W. Mutka, D. Zhu, and N. Xi, "Supermedia transport for teleoperations over overlay networks," in *Proceedings of the 4th IFIP-TC6 International Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communication Systems*, NETWORKING'05, (Berlin, Heidelberg), pp. 1409–1412, Springer-Verlag, 2005.

[226] G. S. Lee and B. Thuraisingham, "Cyberphysical systems security applied to telesurgical robotics," *Comput. Stand. Interfaces*, vol. 34, pp. 225–229, Jan. 2012.

[227] M. E. Tozal, Y. Wang, E. Al-Shaer, K. Sarac, B. Thuraisingham, and B.-T. Chu, "On secure and resilient telesurgery communications over unreliable networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pp. 714–719, April 2011.

[228] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Systems*, vol. 35, pp. 20–23, Feb 2015.

[229] A. Abou El Kalam, A. Ferreira, and F. Kratz, "Bilateral teleoperation system using qos and secure communication networks for telemedicine applications," *Systems Journal, IEEE*, vol. PP, no. 99, pp. 1–12, 2015.

[230] A. Burg, A. Chattopadhyay, and K. Y. Lam, "Wireless communication and security issues for cyber-physical systems and the internet-of-things," *Proceedings of the IEEE*, vol. 106, pp. 38–60, Jan 2018.

[231] M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, J. Sachs, J. Markendahl, and M. Dohler, "Business case and technology analysis for 5g low latency applications," *CoRR*, vol. abs/1703.09434, 2017.

[232] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4g lte networks," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, (New York, NY, USA), pp. 225–238, ACM, 2012.

[233] 3GPP, "Feasibility study on new services and markeets technology enablers for critical communications; stage 1," tech. rep., 3GPP, October 2016. TR 22.862, Release 14.

[234] 3GPP, "Study on new services and markets technology enablers," tech. rep., 3GPP, September 2016. TR 22.891, Release 14.

[235] T. Mahmoodi and S. Seetharaman, "Traffic jam: Handling the increasing volume of mobile data traffic," *IEEE Vehicular Technology Magazine*, vol. 9, pp. 56–62, Sept 2014.

[236] T. Mahmoodi and S. Seetharaman, "On Using a SDN-based Control Plane in 5G Mobile Networks," in *Wireless World Research Forum, 32nd Meeting*, May 2014.

[237] A. C. Morales, A. Aijaz, and T. Mahmoodi, "Taming Mobility Management Functions in 5G: Handover Functionality as a Service (FaaS)," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2015.

[238] M. Amani, T. Mahmoodi, M. Tatipamula, and H. Aghvami, "Programmable policies for Data Offloading in LTE Network," in *IEEE International Conference on Communications (ICC)*, June 2014.

[239] M. Amani, T. Mahmoodi, M. Tatipamula, and H. Aghvami, "SDN-based Data Offloading for 5G Mobile Networks," *ZTE Communications*, July 2014.

[240] C. Bouras, A. Kollia, and A. Papazois, "Sdn nfv in 5g: Advancements and challenges," in *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, pp. 107–111, March 2017.

[241] N. Dinh and Y. Kim, "An efficient reliability guaranteed deployment scheme for service function chains," *IEEE Access*, vol. 7, pp. 46491–46505, 2019.

[242] L. Qu, C. Assi, K. Shaban, and M. Khabbaz, "Reliability-aware service provisioning in nfv-enabled enterprise datacenter networks," in *2016 12th International Conference on Network and Service Management (CNSM)*, pp. 153–159, Oct 2016.

[243] I. Giannoulakis, E. Kafetzakis, G. Xylouris, G. Gardikis, and A. Kourtis, "On the applications of efficient nfv management towards 5g networking," in *1st International Conference on 5G for Ubiquitous Connectivity*, pp. 1–5, Nov 2014.

[244] S. Shorgin, K. Samouylov, I. Gudkova, O. Galinina, and S. Andreev, "On the benefits of 5g wireless technology for future mobile cloud computing," in *Science and Technology Conference (Modern Networking Technologies) (MoNeTeC), 2014 International*, pp. 1–4, Oct 2014.

[245] A. A. Dyumin, L. A. Puzikov, M. M. Rovnyagin, G. A. Urvanov, and I. V. Chugunkov, "Cloud computing architectures for mobile robotics," in *Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), 2015 IEEE NW Russia*, pp. 65–70, Feb 2015.

[246] E. Shin and G. Jo, "Uplink frame structure of short tti system," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 827–830, Feb 2017.

[247] W. Tarneberg, M. Karaca, A. Robertsson, F. Tufvesson, and M. Kihl, "Utilizing massive mimo for the tactile internet: Advantages and trade-offs," in *2017 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*, pp. 1–6, June 2017.

[248] M. A. Lema, T. Mahmoodi, and M. Dohler, "On the Performance Evaluation of Enabling Architectures for Uplink and Downlink Decoupled Networks," in *IEEE Global Communications Conference (GLOBECOM) Workshops*, December 2016.

[249] A. Aijaz, "Towards 5g-enabled tactile internet: Radio resource allocation for haptic communications," in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6, April 2016.

[250] M. Condoluci, T. Mahmoodi, E. Steinbach, and M. Dohler, "Soft resource reservation for low-delayed teleoperation over mobile networks," *IEEE Access*, vol. 5, pp. 10445–10455, 2017.

[251] L. Girish and S. K. N. Rao, "Mathematical tools and methods for analysis of sdn: A comprehensive survey," in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 718–724, Dec 2016.

[252] M. Karakus and A. Durresi, "Quality of service (qos) in software defined networking (sdn): A survey," *Journal of Network and Computer Applications*, vol. 80, pp. 200 – 218, 2017.

[253] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Communications Surveys Tutorials*, vol. 19, pp. 1790–1821, thirdquarter 2017.

[254] D. Szabo, A. Gulyas, F. H. Fitzek, F. H. Fitzek, and D. E. Lucani, "Towards the tactile internet: Decreasing communication latency with network coding and software defined

networking," in *European Wireless 2015; 21th European Wireless Conference; Proceedings of*, pp. 1–6, May 2015.

[255] 3GPP, "System architecture for the 5G system," TR 23.501, v1.4.0, 3GPP, 2017.

[256] Z. Xiang, F. Gabriel, E. Urbano, G. T. Nguyen, M. Reisslein, and F. H. P. Fitzek, "Reducing latency in virtual machines: Enabling tactile internet for human-machine co-working," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 1098–1116, May 2019.

[257] J. A. Cabrera, R. Schmoll, G. T. Nguyen, S. Pandi, and F. H. P. Fitzek, "Softwarization and network coding in the mobile edge cloud for the tactile internet," *Proceedings of the IEEE*, vol. 107, pp. 350–363, Feb 2019.

[258] A. E. Kalør, R. Guillaume, J. J. Nielsen, A. Mueller, and P. Popovski, "Network slicing in industry 4.0 applications: Abstraction methods and end-to-end analysis," *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 5419–5427, Dec 2018.

[259] M. Alizadeh, A. Kabbani, T. Edsall, B. Prabhakar, A. Vahdat, and M. Yasuda, "Less is more: Trading a little bandwidth for ultra-low latency in the data center," in *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, (San Jose, CA), pp. 253–266, USENIX, 2012.

[260] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint link adaptation and scheduling for 5g ultra-reliable low-latency communications," *IEEE Access*, vol. 6, pp. 28912–28922, 2018.

[261] N. Gholipoor, S. Parsaeefard, M. R. Javan, N. Mokari, and H. Saeedi, "Cloud-based queuing model for tactile internet in next generation of ran," *CoRR*, vol. abs/1901.09389, 2019.

[262] J. Xin, Q. Zhu, G. Liang, and T. Zhang, "Performance analysis of d2d underlying cellular networks based on dynamic priority queuing model," *IEEE Access*, vol. 7, pp. 27479–27489, 2019.

[263] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. El-Bakoury, "Ultra-low latency (ull) networks: The ieee tsn and ietf detnet standards and related 5g ull research," *IEEE Communications Surveys Tutorials*, vol. 21, pp. 88–145, Firstquarter 2019.

[264] R. Braden, D. Clark, and S. Shenker, "RFC 1633 : Integrated Services in the Internet Architecture: an Overview," tech. rep., IETF, 1994.

[265] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the differentiated services field (ds field) in the ipv4 and ipv6 headers," tech. rep., IETF, United States, 1998.

[266] D. Kreutz, F. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[267] F. Sardis, M. Condoluci, T. Mahmoodi, and M. Dohler, "Can QoS be dynamically manipulated using End-Device Initialization?," in *IEEE International Conference on Communications Workshops (ICC)*, pp. 448–454, May 2016.

[268] R. Durner, A. Blenk, and W. Kellerer, "Performance study of dynamic qos management for openflow-enabled sdn switches," in *IEEE/ACM International Symposium on Quality of Service (IWQoS)*, 2015.

[269] R. Guerzoni, I. Vaishnavi, D. Perez Caparros, A. Galis, F. Tusa, P. Monti, A. Sganbelluri, G. Biczók, B. Sonkoly, L. Toka, A. Ramos, J. Melián, O. Dugeon, F. Cugini, B. Martini, P. Iovanna, G. Giuliani, R. Figueiredo, L. M. Contreras-Murillo, C. J. Bernardos, C. Santana, and R. Szabo, "Analysis of end-to-end multi-domain management and orchestration frameworks for software defined infrastructures: an architectural survey," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 4, p. e3103, 2017. e3103 ett.3103.

[270] N. F. S. de Sousa, D. A. L. Perez, R. V. Rosa, M. A. Santos, and C. E. Rothenberg, "Network service orchestration: A survey," *Computer Communications*, vol. 142-143, pp. 69 – 94, 2019.

[271] E. Obiodu, N. Sastry, and A. Raman, "Towards a taxonomy of differentiated service classes in the 5g era," in *2018 IEEE 5G World Forum (5GWF)*, pp. 129–134, July 2018.

[272] B. Davie, A. Charny, J. Bennet, K. Benson, J. L. Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis, "An expedited forwarding phb (per-hop behavior)," RFC 3246, RFC Editor, March 2002.

[273] J. J. Bae and T. Suda, "Survey of traffic control schemes and protocols in atm networks," *Proceedings of the IEEE*, vol. 79, pp. 170–189, Feb 1991.

[274] A. M. Chesoong Kim, Youngjin Oh, "A Space Merging Approach to the Analysis of the Performance of Queueing Models with Finite Buffers and Priority Jumps," *Industrial Engineering and Management Systems*, vol. 12, pp. 274–280, 2013.

[275] C. Cascone, N. Bonelli, L. Bianchi, A. Capone, and B. Sansò, "Towards approximate fair bandwidth sharing via dynamic priority queuing," in *2017 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, pp. 1–6, June 2017.

[276] M. Haviv and J. van der Wal, "Waiting times in queues with relative priorities," *Operations Research Letters*, vol. 35, no. 5, pp. 591 – 594, 2007.

[277] Y. Jiang, C.-K. Tham, and C.-C. Ko, "Delay analysis of a probabilistic priority discipline," *European Transactions on Telecommunications*, vol. 13, no. 6, pp. 563–577, 2002.

[278] T. Maertens, J. Walraevens, and H. Bruneel, "On priority queues with priority jumps," *Performance Evaluation*, vol. 63, no. 12, pp. 1235 – 1252, 2006.

[279] T. Maertens, J. Walraevens, and H. Bruneel, "Performance comparison of several priority schemes with priority jumps," *Annals of Operations Research*, vol. 162, pp. 109–125, Sep 2008.

[280] J.-P. Georges, T. Divoux, and E. Rondeau, "Strict priority versus weighted fair queueing in switched ethernet networks for time critical applications," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) - Workshop 2 - Volume 03*, IPDPS '05, (USA), p. 141, IEEE Computer Society, 2005.

[281] M. A. Lema, K. Antonakoglou, F. Sardis, N. Sornkarn, M. Condoluci, T. Mahmoodi, and M. Dohler, "5g case study of internet of skills: Slicing the human senses," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–6, June 2017.

[282] V. Petrov, M. A. Lema, M. Gapeyenko, K. Antonakoglou, D. Moltchanov, F. Sardis, A. Samuylov, S. Andreev, Y. Koucheryavy, and M. Dohler, "Achieving end-to-end reliability of mission-critical traffic in softwarized 5g networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, pp. 485–501, March 2018.

[283] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami, "Internet of skills, where robotics meets ai, 5g and the tactile internet," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[284] D. B. Camarillo, T. M. Krummel, and J. Salisbury, "Robotic technology in surgery: Past, present, and future," *The American Journal of Surgery*, vol. 188, no. 4, Supplement 1, pp. 2 – 15, 2004.

[285] B. J. Challacombe, L. R. Kavoussi, A. Patriciu, D. Stoianovici, and P. Dasgupta, "Technology insight: telementoring and telesurgery in urology," *Nature Clinical Practice Urology*, vol. 3, pp. 611–617, 2006.

[286] N. Sornkarn and T. Nanayakkara, "Can a soft robotic probe use stiffness control like a human finger to improve efficacy of haptic perception?," *IEEE Transactions on Haptics*, vol. 10, pp. 183–195, April 2017.

[287] Ultragrid, "Wiki." `https://www.sitola.cz/igrid/index.php/UltraGrid`.

[288] NGMN Alliance, "Description of Network Slicing Concept," tech. rep., 01 2016.

[289] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network Slicing Management and Prioritization in 5G Mobile Systems," in *European Wireless*, May 2016.

[290] J. Back, G. Zhang, S. Wang, and H. Liu, "Surgical gripper with embedded soft tactile array for notes," 09 2016.

[291] O. Holland, E. Steinbach, R. V. Prasad, Q. Liu, Z. Dawy, A. Aijaz, N. Pappas, K. Chandra, V. S. Rao, S. Oteafy, M. Eid, M. Luden, A. Bhardwaj, X. Liu, J. Sachs, and J. Araújo, "The ieee 1918.1 "tactile internet" standards working group and its standards," *Proceedings of the IEEE*, vol. 107, pp. 256–279, Feb 2019.

[292] M. Amani, T. Mahmoodi, M. Tatipamula, and H. Aghvami, "SDN-based data offloading for 5G mobile networks," *ZTE communications*, vol. 12, pp. 34–40, June 2014.

[293] A. Blenk, A. Basta, and W. Kellerer, "HyperFlex: An SDN virtualization architecture with flexible hypervisor function allocation," in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pp. 397–405, May 2015.

[294] I. Farris, T. Taleb, H. Flinck, and A. Iera, "Providing ultra-short latency to user-centric 5G applications at the mobile network edge," *Transactions on Emerging Telecommunications Technologies*, pp. 1–13, February 2017.

[295] A. Sheoran, X. Bu, L. Cao, P. Sharma, and S. Fahmy, "An empirical case for container-driven fine-grained VNF resource flexing," in *IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, November 2016.

[296] A. Ledjiar, E. Sampin, C. Talhi, and M. Cheriet, "Network function virtualization as a service for multi-tenant software defined networks," in *Fourth International Conference on Software Defined Systems (SDS)*, May 2017.

[297] S. Oechsner and A. Ripke, "Flexible support of VNF placement functions in OpenStack," in *1st IEEE Conference on Network Softwarization (NetSoft)*, April 2015.

[298] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN – Key technology enablers for 5G networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 2468–2478, November 2017.

[299] T. Maertens, J. Walraevens, and H. Bruneel, "Service differentiation through priority jumps," in *Traffic and Performance Engineering for Heterogeneous Networks* (D. D. Kouvatsos, ed.), vol. 1 of *River Publisher Series in Information Science and Technology*, pp. 175–198, River Publishers, 2009.

[300] G. Mountaser, M. Condoluci, T. Mahmoodi, M. Dohler, and I. Mings, "Cloud-ran in support of urllc," in *2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Dec 2017.

[301] 3GPP, "5g service requirements for next generation new services and markets," tech. rep., 3GPP, October 2018. TR 22.261, Release 15.

[302] I. Y. Kim and O. de Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Structural and Multidisciplinary Optimization*, vol. 29, pp. 149–158, 02 2005.

[303] J. Arora, *Introduction to Optimum Design*. Elsevier Science, 2004.