ITUPublications

International Telecommunication Union

Kaleidoscope Academic Conference

Proceedings

2022

ITU KALEIDOSCOPE
ACCRA 2022

14TH ITU ACADEMIC CONFERENCE

ITU KALEIDOSCOPE
ACCRA 2022

Extended reality – How to boost quality of experience and interoperability

7-9 December
Accra, Ghana

Proceedings

Hosted by

Technically co-sponsored by

Organized by

ISBN:
978-92-61-36151-8 (electronic version)

Disclaimer

The opinions expressed in these Proceedings are those of the paper authors and do not necessarily reflect the views of the International telecommunication Union or of its membership.

Foreword

Dr Chaesub Lee
Director
ITU Telecommunication Standardization Bureau

Innovation to match the world's growing metaverse ambitions was in focus this year at Kaleidoscope 2022: *Extended reality − How to boost quality of experience and interoperability.*

The conference provided a forward-looking perspective on the future development and widespread adoption of extended realities.

Now in its fourteenth edition, this key academic event from the International Telecommunication Union (ITU) supports productive dialogue between academics and standardization experts on how standards can help boost quality of experience and interoperability in extended realities.

Kaleidoscope 2022, kindly hosted by the Ministry of Communications and Digitalisation, with the support of the Ghana-India Kofi Annan Centre of Excellence in ICT, will take place in Accra, from 7 to 9 December at the National Communications Authority of Ghana, with online participation provided too.

The concept of extended reality (XR) has captured imaginations worldwide. It has inspired blockbuster films envisioning a metaverse of enthralling virtual worlds. To technologists, the promise of truly immersive experiences equates to key problems to solve in interoperability and user-perceived quality. These come down to the ability of different worlds, physical or virtual, to interact seamlessly and offer experiences that can convince and captivate us. Focusing on the technical developments that are contributing to a new and more immersive future, the research selected for presentation also sheds light on the standards that are needed to aid in this digital transformation and provide a safe, human-centred future. The various sessions highlighted different perspectives on future networks; design and implementation of augmented reality systems; building, deploying, and managing QoE in XR communications; Metaverse interoperability in hyper-connected and hyper-personalized virtual environments; and challenges for XR and Holographic-Type Communication. A survey of XR Standards was also presented.

This year's conference included a local University Exhibit that offered the opportunity to local students and professors to present research on a variety of topics related to the broad field of ICT, and an invited special session on "The Metaverse and the Future of Education - Frameworks, features, potential applications, challenges and opportunities". The video demonstration track provided an online demonstration on how to make extended reality safe and secure for teenagers.

I would like to express my great appreciation to the Kaleidoscope community and the larger ITU Academia membership for their enduring support to this series of conferences. With over 160 academic and research institutes now members of ITU, the Kaleidoscope series is certain to continue growing in strength.

My sincerest thanks go to our technical co-sponsors, the Institute of Electrical and Electronics Engineers (IEEE), and the IEEE Communications Society (IEEE ComSoc). I would also like to thank our academic partners and longstanding ITU members, Waseda University, the Institute of Image Electronics Engineers of Japan (I.I.E.E.J.), the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, the Chair of Communication and Distributed Systems at RWTH Aachen University, the European Academy for Standardization (EURAS), the University of the Basque Country, the Liverpool John Moores University, and the University of Limerick.

I would especially like to thank the General Chairman of Kaleidoscope 2022, Collins Yeboah-Afari, Director-General of the Ghana-India Kofi Annan Centre of Excellence in ICT (AITI-KACE), Accra, Ghana and the members of the Host Committee: Fred Yeboah and Yaw Okraku-Yirenkyi from AITI-KACE, and Kwame Baah-Acheamfuor from the Ministry of Communications and Digitalisation, Accra, Ghana; the members of the Kaleidoscope 2022 Technical Programme Committee (TPC) and the members of the Steering Committee: Christoph Dosch, IRT GmbH; Kai Jacobs, RWTH Aachen University; Mitsuji Matsumoto, Professor Emeritus Waseda University; Roberto Minerva, Télécom SudParis; Gyu Myuong Lee, Liverpool John Moores University; Eva Ibarrola, University of the Basque Country; Tiziana Margaria, University of Limerick and particularly Mostafa Hashem Sherif, USA who also serves as the TPC Chair.

Dr Chaesub Lee
Director
ITU Telecommunication Standardization Bureau

General Chairman's message

Dr Collins Yeboah-Afari
Director-General
Ghana-India Kofi Annan Centre of Excellence in ICT
(AITI-KACE), Accra, Ghana

The focus of the ITU Kaleidoscope 2022 (K-2022) academic conference is on how technical standards can help boost quality of experience and interoperability in extended realities, ensuring that the new immersive experiences benefit humanity at large.

In a world where digital transformation continues to expand, immersive technologies that merge physical and virtual worlds are becoming more popular for their potential to improve our quality of life, explore new social and cultural dimensions, and unlock new business opportunities. The promise of the metaverse, a term that recently gained attention, is to allow an even greater overlap of our digital and physical lives. Despite remarkable technological advances, current XR applications are a largely individual and local experience. To deliver a widespread adoption of XR type of services and applications and achieve the vision of a metaverse, communication networks have a key role to play.

K-2022 is held from 7 to 9 December at the National Communications Authority, Accra, Ghana, with the support of the Ghana-India Kofi Annan Centre of Excellence in ICT, which I lead as Director-General.

The conference programme includes nine original academic papers retained from a total of 14 submissions, after a double-blind peer-review process. The video demonstration track, which was introduced in Kaleidoscope two years ago, features one demo.

The authors of the best papers presented at the conference will share a prize fund of CH 6 000. In addition, young authors of up to 30 years of age presenting accepted papers will receive Young Author Recognition certificates.

I would like to thank the Technical Programme Committee (TPC) Chair, Mostafa Hashem Sherif, and all reviewers, whether members of the TPC or anonymous reviewers, for assisting the Steering Committee (SC) in making the selection; the members of the SC that were essential also in preparing four keynote sessions and in inviting one paper dealing with the conference's theme, and the video demonstration track. My special thanks also go to the Host Committee members, Fred Yeboah and Yaw Okraku-Yirenkyi (AITI-KACE), Kwame Baah-Acheamfuor (Ministry of Communications and Digitalisation, Accra, Ghana) that also helped organise the local University Exhibit, and to all local staff who supported the logistics and the local organization of the conference. Special mention must be made of the Honourable Minister for Communications & Digitalisation, Ursula Owusu-Ekuful and the Director-General of the National Communications Authority, Joe Anokye. Their commitment and valuable support to the successful execution of the conference is deeply appreciated.

The K-2022 conference proceedings (as well as those of the previous conferences) can be downloaded free of charge from http://itu-kaleidoscope.org.

As customary with the Kaleidoscope conferences, per agreement with the IEEE, all papers accepted and presented at the conference will be included in the IEEE *Xplore* Digital Library. In addition, selected papers will be considered for publication in a special-feature section of the *IEEE Communications Standards Magazine*.

Finally, I would like to thank the ITU team led by Alessia Magliarditi and including Emer Windsor, Erica Campilongo and Martin Adolph, as well as Gent Bajrami and Ilia Londo from the ITU-T Operations and Planning Department, for their unwavering support.

Dr Collins Yeboah-Afari
General Chairman

TABLE OF CONTENTS

Session 2: Augmented reality systems: design and implementation

Session 3: Services in future network

Video Demonstration Track

COMMITTEES

Steering Committee

- Christoph Dosch, ITU-R Study Group 6 Vice-Chairman; ARD, Germany
- Eva Ibarrola, University of the Basque Country, Spain
- Kai Jakobs, RWTH Aachen University, Germany
- Gyu Myoung Lee, Liverpool John Moores University, United Kingdom
- Tiziana Margaria, University of Limerick, Ireland
- Mitsuji Matsumoto, Waseda University Emeritus Professor, Japan
- Roberto Minerva, Télécom SudParis, France
- Mostafa Hashem Sherif, Consultant, United States

Host Committee

- General Chairman: Collins Yeboah-Afari, Director-General, Ghana-India Kofi Annan Centre of Excellence in ICT (AITI-KACE), Accra, Ghana
- Fred Yeboah, Ghana-India Kofi Annan Centre of Excellence in ICT (AITI-KACE), Accra, Ghana
- Yaw Okraku-Yirenkyi, Ghana-India Kofi Annan Centre of Excellence in ICT (AITI-KACE), Accra, Ghana
- Kwame Baah-Acheamfuor, Ministry of Communications and Digitalisation, Accra, Ghana

Secretariat

- Alessia Magliarditi, ITU Kaleidoscope Coordinator
- Emer Windsor, Executive Assistant
- Erica Campilongo, Collaborator
- Martin Adolph, Technical Adviser
- Pascal Borde, Promotional support

Technical Programme Committee

- Mostafa Hashem Sherif (Consultant, USA)

- Olasupo Ajayi (University of the Western Cape, South Africa and University of Lagos, Nigeria)
- Antoine Bagula (University of the Western Cape, South Africa)
- Thomas Basikolo (International Telecommunication Union, Switzerland)
- Daniel Batista (University of São Paulo, Brazil)
- Paolo Bellavista (University of Bologna, Italy)
- Rodrigo Campos Bortoletto (Instituto Federal de São Paulo, Brazil)
- V. Michael Bove (Consultant, USA)
- Alexander Brem (University of Stuttgart, Germany)
- Giovanni Cicceri (University of Messina, Italy)
- Antonio Corradi (University of Bologna, Italy)
- Udhaya Kumar Dayaan (Trane Technologies, USA)
- Luca Foschini (University of Bologna, Italy)
- Joan Garcia-Haro (Technical University of Cartagena, Spain)
- Kai Jakobs (RWTH Aachen University, Germany)
- Ved Kafle (National Institute of Information and Communication Technology, Japan)
- Dimitar R. Kolev (National Institute of Information and Communications Technology (NICT), Japan)
- Katarzyna Kosek-Szott (AGH University of Science and Technology, Poland)
- Dhananjay Kumar (Anna University, India)
- Gyu Myoung Lee (Liverpool John Moores University, UK)
- Christophe Maudoux (CNAM, France)
- Kazuhide Nakajima (NTT Corporation, Japan)
- Vishnu Ram Ov (Consultant, India)
- Tien Dat Pham (National Institute of Information and Communications Technology, Japan)
- Antonio Puliafito (University of Messina, Italy)
- Domenico Rotondi (FINCONS SpA, Italy)
- Stefano Secci (Cnam, France)
- DongBack Seo (ChungBuk National University, Korea (Rep. of))
- Valerio Torti (LUISS University of Rome, Italy)
- Christos Tselios (University of Patras, Greece)
- Taavi Valdlo (Estonian IT Standardization Technical Committee, Estonia)

KEYNOTE SUMMARIES

METAVERSE INTEROPERABILITY WITH COMPOSABILITY IN HYPER-CONNECTED AND HYPER-PERSONALIZED VIRTUAL ENVIRONMENTS

Junseong Bang, ETRI

A metaverse platform is a simulated environment in which human users can interact with computer-generated objects and other users. Each metaverse platform operates multiple virtual worlds according to the purpose of the service. In order to connect the virtual worlds of heterogeneous metaverse platforms and allow users to freely navigate the metaverse, it is necessary to consider interoperability.

In the metaverse, users can be not only players, but also creators, developers, and operators. Users influence the entire platform beyond the use of the system. Therefore, it is necessary to look at not only interoperability in terms of systems, but also in terms of data, services, and the economy. For a deeper understanding of the metaverse, consider the service environment by dividing it into hyper-connected and hyper-personalized virtual environments. This helps to guide potential discussions about interoperability as well as planning for services. Considering services in hyper-connected and hyper-personalized virtual environments, as well as interoperability in terms of systems, data, services, and economy will be discussed.

In order to guarantee ethics and user rights without interfering with the free activities of users in the digital society called the metaverse, a discussion on programmable ethics in relation to interoperability is also included.

Metaverse: Challenges for Extended Reality and Holographic-Type Communication in the Next Decade

Ian F. Akyildiz
ITU Journal on Future and Evolving Technologies (ITU J-FET)
Email: ian.akyildiz@itu.int

I. Introduction

The metaverse is an iteration of the Internet with immersive and hyperreal 3D digital virtual worlds. It can recreate the physical world using technologies such as digital twins, avatars, and holograms. Besides creating digital models of physical environments, the metaverse allows the existence of virtual information, objects, human beings, and environments that do not exist in the physical world. The mixture of real and virtual content in the metaverse can significantly improve users' capabilities for work, entertainment, education, healthcare, and manufacturing [1].

Development of the metaverse includes three major areas: (1) metaverse content/service design, (2) metaverse core and edge network design, and (3) fully connected networks for metaverse devices in the physical world.

First, the metaverse virtual content and services emphasize social connections and interactions with the physical world. The design of virtual content and services requires digital computing, modeling, and rendering technologies.

Second, the core and edge networks are needed to deliver metaverse content and services to end users. Although existing computer networks can provide basic metaverse services, the newly-proposed compute- and data-intensive networks [2] and the New Internet Protocol (New IP) [3] are efficient in processing metaverse data packets.

Last, metaverse devices, sensors, actuators, and displays must be fully connected considering different levels of data rates, latency, and reliability. This keynote will cover all three areas with special emphasis on wireless communication and networking in the second and third areas.

5G and beyond wireless systems have enabled the creation and development of the metaverse. However, their performance cannot support the high Quality-of-Experience (QoE) metaverse. The most popular tools to get access to the metaverse are the extended reality (XR) Head-Mounted Displays (HMDs) [4]. However, some users experience frustration, eye strain, sickness, and prolonged use is prohibitive due to their large weight and low resolution of the HMDs [5]. Although working and living in the metaverse provides various choices and support that are not available in reality, the experience of using existing HMDs is not convenient and even creates anxiety for some users, according to a recent study [5].

The metaverse has become one of the driving applications for 6G and beyond wireless systems since 5G and beyond wireless systems cannot fully address wireless communication and networking issues. It is anticipated that the 6G and beyond wireless systems will provide high-data-rate, low-latency, high-reliability, and fully-connected data communications for metaverse devices, which can enable users to comfortably enter and use metaverse applications with high QoEs. In this keynote, we will focus on wireless communication and networking challenges for metaverse accessing technologies, namely, the XR technologies, including Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR), and Holographic-Type Communications (HTC). The main contribution of this keynote is two-fold. First, we introduce fundamental technologies of XR and HTC and identify the technical gaps to provide high QoE access to the metaverse. Second, we find the potential solutions in 6G and beyond wireless systems that fill these gaps and discuss future research challenges.

II. Metaverse and Challenges for Extended Reality and Holographic-Type Communication

Wireless communication systems are used in the physical world to provide users with access to the metaverse and connect machines (sensors, robots, actuators, etc.) with metaverse computing servers. Wireless systems in the metaverse mainly have the following two types of communication.

- Human and Metaverse Communication (HMC). The communication between human beings and the metaverse uses specialized devices, including HTC displays and sensors, XR HMDs, and legacy computers and smart devices. These devices connect the user in the physical world with the content in the virtual world and they have sensing units that can track users' motion and gather input data which is used in the metaverse for computing. The computing results are sent back to users' devices for display and actuation.
- Machine and Metaverse Communication (MMC). The metaverse utilizes active and passive sensors to monitor physical objects, machines, and environments. The aggregated data can be used to develop hyperreal virtual models in the metaverse. Also, the metaverse uses actuators to interact with and change the physical world. The communication between the metaverse and sensors and actuators can use 6G and beyond solutions that support massive IoT where a large number of devices can be connected within a small area.

Note that the HMC and MMC are bidirectional, which reflects the interactions between the physical world and the virtual world. Human and human communications and machine and machine communications existed before the era of the metaverse. They can use existing communication technologies without using any metaverse systems. When a metaverse user communicates with another user, the communication becomes HMC since they have to go through the metaverse systems. Similarly, when two machines communicate with each other in the metaverse, their communication belongs to the MMC

This keynote focuses on XR and HTC, which belong to HMC technologies. Computers and smart devices can be used to access metaverse content. However, the user experience is not immersive, i.e., users need to look at screens with a limited size. As the content and activities in the metaverse increase, simply using computers and smart devices as gateways is not enough, and more immersive spatial computing technologies, such as XR which includes AR, MR, and VR, are preferred [4]. XR uses HMDs which integrate displays, inputs, and various sensors together. HTC is another way to get access to the virtual worlds which can use HMDs and holographic displays [6]. It can display truly immersive 3D objects as holograms using light field displays, which can be directly observed by users with naked eyes. Without HMDs, the experience of living and working in the virtual worlds becomes very similar to real life.

XR is a spatial computing technology that enables users to observe and interact with virtual worlds. XR technologies are categorized based on Milgram and Kishino's Reality-Virtuality Continuum [7], where AR only has simple virtual content and the majority of the environment in AR is real, while VR has fully virtual content, and the user's observation using VR devices is not related to the user's surrounding real environment. MR comprises mixed real and virtual content, and the virtual content can be interactive and dynamic. Although MR and AR have overlapped, MR is considered a broader concept than AR with more interactive and richer virtual content.

HTC is expected to provide truly immersive experiences for users when getting access to the virtual worlds. HTC is characterized by the delivery of holograms and other multi-sensory media (mulsemedia) through wireless and wired networks [6]. A hologram is a recording of a light field that preserves the original depth and parallax of real 3D objects. Hologram videos could provide six-Degrees-of-Freedom (6DoF) immersive viewing experiences to users, that is, users can freely move forward/backward (surging), up/down (heaving), or left/right (swaying) to select the favorite viewing angle of a 3D scene [8]. Furthermore, HTC research aims to stimulate all the senses of human perception, including sight, hearing, touch, smell, and taste, to provide a truly immersive experience to users via mulsemedia. Mulsemedia is the media that stimulates three or more senses, which extends traditional multimedia that usually supports only two senses of sight and hearing. HTC systems may collect mulsemedia data using various sensors at the source and then regenerate the environment using various actuators, displays, and speakers at the destination. Users can select the senses depending on the available hardware and privacy and security issues. A generic HTC system consists of the source, the destination, and the networks. The source uses various sensors to capture holographic content, synchronizes multi-sensory data, encodes the holographic data, and follows HTC networking protocols to send data packets. The HTC networks deliver source data with guaranteed performance in bandwidth, latency, reliability, etc. which are defined by the HTC-enabled use cases. The destination receives and renders data for display, utilizes various actuators to regenerate the environment at the source, performs synchronization among multi-sensory media, and provides useful feedback to the source and the network if necessary. The design of an entire HTC system requires knowledge from several technical areas, such as sensors and actuators, computer vision, data compression, and communication and networking. The rest of this section highlights current technologies that are most relevant to communications.

This keynote will provide more details about XR devices, XR communication and networking, hologram presentation and display, hologram compression, hologram network streaming protocols, and mulsemedia, which are essential technologies for the high QoE metaverse. After that, we will discuss the wireless communication challenges and potential solutions for XR and HTC. Since both of them require high reliability, high data rates, low latency, and full connectivity to support high QoE metaverse applications, we introduce generic solutions and also consider their unique characteristics. The general research directions to meet the required performances for wireless XR and HTC include the following aspects.

- Reliable high-data-rate low-latency wireless communications are required to allow users to move freely without constraints. Cables that are tethered to some of the existing XR HMDs limit mobility.
- Accurate high-fidelity wireless sensing technologies can be used to monitor and predict users' motion and render content adaptively. Existing XR HMDs use the integrated head, hand, and eye tracking sensors. The ultimate goal is to remove HMDs and users can get into the metaverse without wearing any devices. Wireless sensing of the motion of the head, hand, body, and eyes can efficiently free the users.
- New Internet protocols and wireless networking technologies are needed to provide new services and new capabilities to deliver metaverse content with different QoEs.
- Computing and content caching need to be moved from the core cloud towards the network edge, in order to reduce the end-to-end latency and support highly interactive applications.

This keynote will provide details about research challenges and solutions for communication, sensing, networking, caching, computation, and metaverse user experiences.

III. Conclusion

This keynote will introduce metaverse accessing technologies including the extended reality (XR) and Holographic-Type Communications (HTC). We classify different wireless communications in the metaverse and point out the unique requirements for XR and HTC. Then, we discuss research challenges and potential solutions in 6G and beyond wireless systems. This keynote aims to stimulate future research around XR and HTC systems.

References

[1] F. Tang, X. Chen, M. Zhao, and N. Kato, "The roadmap of communication and networking in 6g for the metaverse," *IEEE Wireless Communications*, 2022.

[2] Y. Cai, J. Llorca, A. M. Tulino, and A. F. Molisch, "Joint compute-caching-communication control for online data-intensive service delivery," *arXiv preprint arXiv:2205.01944*, 2022.

[3] R. Li, K. Makhijani, and L. Dong, "New IP: A data packet framework to evolve the internet," in *2020 IEEE 21st International Conference on High Performance Switching and Routing (HPSR)*. IEEE, 2020, pp. 1–8.

[4] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for Extended Reality (XR)," *ITU Journal on Future and Evolving Technologies*, vol. 3, no. 1, pp. 1–15, 2022.

[5] V. B. S. Kalamkar, N. Nouri, E. Ofek, M. Pahud, J. J. Dudley, J. Hu, P. O. Kristensson, M. Weerasinghe, K. Č. Pucihar, M. Kljun *et al.*, "Quantifying the Effects of Working in VR for One Week," *arXiv preprint arXiv:2206.03189*, 2022.

[6] I. F. Akyildiz and H. Guo, "Holographic-type Communication: A New Challenge for The Next Decade," *ITU Journal on Future and Evolving Technologies*.

[7] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[8] Z. Liu, Q. Li, X. Chen, C. Wu, S. Ishihara, J. Li, and Y. Ji, "Point cloud video streaming: Challenges and solutions," *IEEE Network*, vol. 35, no. 5, pp. 202–209, 2021.

MINOHEALTH.AI: A CLINICAL EVALUATION OF DEEP LEARNING SYSTEMS FOR THE DIAGNOSIS OF PLEURAL EFFUSION AND CARDIOMEGALY IN GHANA, VIETNAM AND THE UNITED STATES OF AMERICA

Darlington Akogo

minoHealth AI Labs, karaAgro AI, Runmila AI Institute, Ghana

A rapid and accurate diagnosis of medical conditions like cardiomegaly and pleural effusion is of the utmost importance to reduce mortality and medical costs, and artificial intelligence has shown promise in diagnosing medical conditions. We evaluated how well Artificial Intelligence (AI) systems, developed by minoHealth AI Labs, perform at diagnosing cardiomegaly and pleural effusion, using chest x-rays from Ghana, Vietnam and the USA, and how well AI systems perform when compared with radiologists working in Ghana. The evaluation dataset used in this study contained 100 images randomly selected from three datasets. The deep learning models were further tested on a larger Ghanaian dataset containing 561 samples. Two AI systems were then evaluated on the evaluation dataset, whilst we also gave the same chest x-ray images within the evaluation dataset to four radiologists, with 5 - 20 years' experience, to give their independent diagnoses.

For cardiomegaly, minoHealth.ai systems scored an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.9 and 0.97 while the AUC-ROC of individual radiologists ranged from 0.77 to 0.87. For pleural effusion, the minoHealth.ai systems scored 0.97 and 0.91, whereas individual radiologists scored between 0.75 and 0.86. On both conditions, the best performing AI model outperforms the best performing radiologist by about 10%. These models will be of great use in regions, such as sub-Saharan Africa, where there are few radiologists. They can potentially be used to augment the effort of radiologists to improve the diagnosis and treatment of chest conditions.

KEYNOTE PAPER

EXPLORING THE REALVERSE: BUILDING, DEPLOYING, AND MANAGING QOE IN XR COMMUNICATIONS

Pablo Pérez – eXtended Reality Lab, Nokia

ABSTRACT

Immersive communication based on extended Reality (XR) is foreseen to be the next step in human communication. For that challenge, we propose the "Realverse": an approach where physically distant people can interact as if they were in the same physical space. This can be achieved by implementing a distributed reality solution, including four fundamental elements: remote presence, embodied interaction, visual communication, and shared immersion. In this paper we describe our recent research and approach to this topic, as well as its implications in terms of implementation, network architecture, and quality of experience. We also propose some insights for future standardization efforts.

Keywords - 5G, distributed reality, extended reality, quality of experience

1. INTRODUCTION

In recent years there has been a strong increase in interest in Virtual, Augmented and extended Reality (VR/AR/XR) applications, catalyzed in part by the company Meta's commitment to developing new applications and experiences around the concept of the metaverse. This interest is also fueled by the wide availability of moderately priced VR and AR devices, such as Meta Quest, HTC Vive, Hololens or MagicLeap. This wave of technology, whose beginning we can date in the development of the Oculus Rift at the end of 2012, was also one of the leitmotifs that supported the commitment to 5G networks, in anticipation of the connectivity needs (especially bitrate) that would be needed.

The particularity of the wave of interest around the metaverse is the application of XR technologies fundamentally for immersive communication, defined as "exchanging natural social signals with remote people, as in face-to-face meetings, and/or experiencing remote locations (and having the people there experience you) in ways that suspend disbelief in *being there*" [1]. Immersive communication has already had a strong presence in the research community for the last decade, studied from the perspective of telepresence [2], Social VR [3], or within the framework of Computer-Supported Cooperative Work (CSCW) [4].

Despite the fact that there are as many versions of the metaverse as there are people talking about it, most of the visions that are proposed are usually based on the interaction of avatars, virtual or realistic, inside a highly personalized virtual space. In this sense, the interaction between people happens *in* the metaverse, and is strongly (and very actively) mediated by the system. Our approach to immersive communications, on the other hand, is based on the same idea from which the telephone and the video call were born: the interaction between people *through* technology, which must be as transparent as possible, such that two physically distant people can communicate and interact exactly as if they were together in the same place (as if one was visiting the other). We call this vision "Realverse".

Implementing the realverse translates into capturing and transmitting, in real time, the environment of one or more people, who are remote, and mixing them consistently with the environment of the local person, so as to generate a single shared reality, composed of parts of the realities of each of the participants in the conversation. This composition of realities based on video captured and transmitted in real time is called Distributed Reality (DR) [5].

The study of the realverse (or DR) is interesting from two points of view. In the first place because, in the same way that voice and video calls have established themselves as two of the main communications applications today, the "XR call" will be the natural evolution of them, once that the necessary technology reaches a sufficient level of accessibility. And, on the other hand, because DR is a variant of XR focused on communications between systems, which therefore makes intensive use of communications networks. The analysis of the realverse requirements will allow us to design and dimension the 5G and beyond-5G (B5G) networks necessary to support future XR-based services.

In this paper we will describe the fundamental building blocks of a realverse implementation, as well as the different solutions that exist, or that we have researched, for each of them. We will also describe the implications that it has for the network, and which are the main network constraints regarding the Quality of Experience (QoE). Finally, we will discuss the opportunities that it opens for standardization.

2. BUILDING BLOCKS OF THE REALVERSE

The key psychological feature which allows the existence of metaverses, *realverses*, and any other XR paradigm, is the sense of presence. From the early experiments with virtual reality systems in the 1980s and 1990s, it was discovered that the technology created in the users the illusion of "being there", within the virtual environment. This perceptual illusion is called "sense of presence" or simply "presence", defined as "the subjective experience of being in one place or environment, even when one is physically situated in another" [6]. The sense of presence elicited by XR environments is closely tied to the psychological process of consciousness; it can be considered "consciousness within a restricted domain" [7]: You feel, think and behave as if the place were real, even though you know it is not. This property is key in the success of immersive technologies, since consciousness is a very primitive cognitive function, not associated with any cortical region [8]. In other words, the perceptual illusion of "being there" created by XR technology really addresses the deepest layers of the human brain, thus making it universally applicable to any kind of person.

This perceptual illusion of presence involves different factors [9]. Biocca *et al.* identified three key components [10]: spatial presence, the illusion of being in a different place; social presence, the illusion of being and interacting with other people; and self-presence, the illusion of having one's body integrated in the virtual experience. Creating the realverse means addressing those components by transmitting or generating this sense of presence. Technologicaly, it requires implementing the four perceptual building blocks shown in Figure 1 [11]:

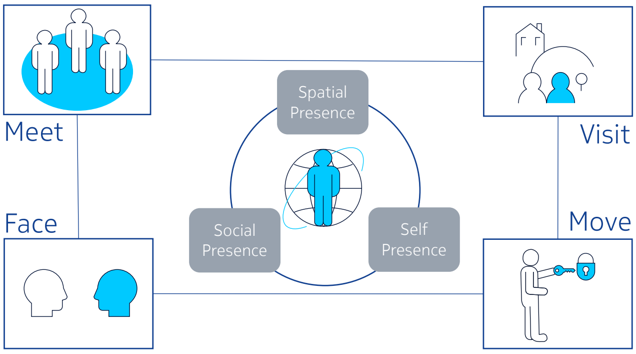


Figure 1 – Building blocks (or fundamental elements) of XR communication systems: face, visit, meet, and move [11].

1. **Face** is the property of the system to transmit in real time a visual representation of the other person, e.g. through a video-conferencing system. This element enables **visual communication**. Seeing the other person is key to transmitting non-verbal communication cues, including showing objects of the personal space.

2. **Visit** is the property of the system to transmit in real time a visual representation of the *surroundings* of the other person. This enables **remote presence**: the sense of "being there", in the physical environment of the remote person and being able to operate and discuss about it.

3. **Meet** is the property of the system to represent the other person in the same (virtual or physical) space as the user. This enables **shared immersion**: being immersed in the same (virtual or physical) environment and interacting with the same (virtual or physical) objects.

4. **Move** is the property of the system itself to represent the user within it and enable its **embodied interaction**. It means that the actions of the users are represented within the system and allow the user to interact with it.

As we showed in [11], all XR communication systems include one or several of these building blocks, and the combination of blocks they address has a strong impact on the technical architecture they implement. In the next subsection we will describe how we are addressing each of them, with the target of building a distributed reality system which is able to provide all of them simultaneously.

2.1 Visit: remote presence

To address the problem of immersive remote presence we have developed *The Owl*, a prototype 360-degree video-based telepresence system. 360-degree videos are video recordings where a view in every direction is recorded at the same time, shot using an omnidirectional camera or a collection of cameras. Our prototype, shown in Figure 2, consists of a commercial omnidirectional video camera, a control system on a Raspberry Pi 4, a backend in the cloud, and a client for Meta Quest developed in the 3-D version of the game engine Unity. The system allows video transmission in real time, in equirectangular projection[1], with 4K resolution (8 Mbps) and conversational delay (< 500 ms), as well as audio in both directions. The approach is similar to other prototypes in the state of the art, such as *ExLeap* [13] or *Show Me Around* [14].

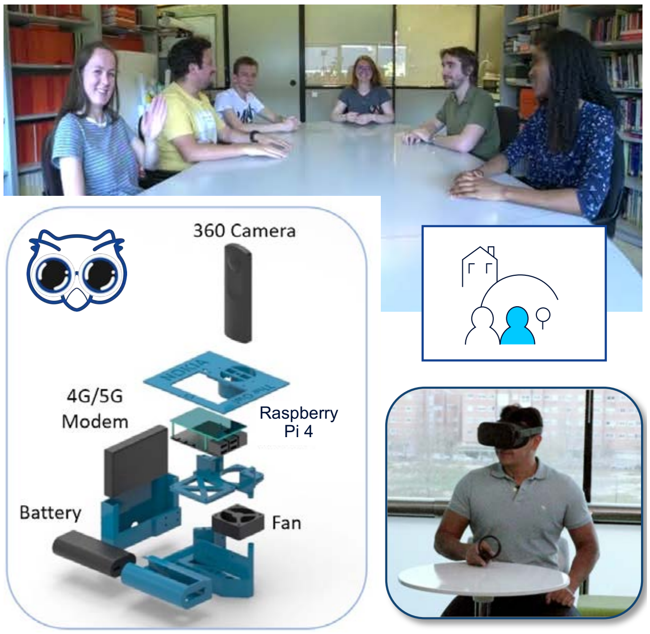


Figure 2 – *The Owl*: an immersive communication system prototype (bottom left), which allows a person wearing a head-mounted display (bottom right) to feel present in a remote location and interact with the people there (top).

The Owl has been field-tested in different scenarios, and has allowed us to evaluate 360 degree video technology for real-time communications in use cases such as education [15] or hybrid conferences, with face-to-face and virtual attendees [16]. For better assessment of its QoE, a novel methodology

[1] The equirectangular projection maps the longitude and latitude of the sphere videos to the horizontal and vertical coordinates of the rectangular video [12].

has been developed, to simultaneously measure the effect of the system on socio-emotional parameters such as presence or empathy, and on technical aspects such as visual quality [17].

Although 360 video technology has obvious limitations, since it only provides three Degrees of Freedom (DoF) in movement, it is enough to elicit a high degree of spatial and social presence [17]. This sense of presence is increased if the user feels part of the scene, that is, if the remote participants address him or her. In a study with elderly people with neuronal degradation, we have also seen how this high sense of presence is maintained even in the face of processes of cognitive degradation, severe dependency, or symptoms of depression [18].

2.2 Move: embodied interaction

In commercial VR solutions, the representation of the person within the immersive environment (*avatar*) is usually implemented through virtual hands. Each virtual hand mimics the movement of the real hand, either when it interacts with a VR game controller, or by using hand detection and tracking algorithms from the cameras integrated in the Head-Mounted Display (HMD). As an alternative, we propose the use of *video-based avatars*[2], where a camera integrated in the HMD is used to capture the video egocentrically (i.e. from the point of view of the user), and the silhouette of the hand or the body itself is detected and segmented to integrate it into the immersive scene [19], as shown in Figure 3.

Figure 3 – Video-based avatars. *Left*: A person wearing an HMD with an attached egocentric camera, so that her body is captured from the camera, segmented, and integrated within a virtual environment. *Right*: The virtual environment as perceived by the HMD user, with she can see her own body.

To validate our proposal, we have developed a prototype in which we integrate an HMD with a stereoscopic camera, properly calibrated, so that the image captured by the camera is displayed in the correct position within the immersive environment. This image, in turn, is sent to a server at the edge, where the image segmentation algorithm runs. The final result is sent back to the HMD and is projected onto the virtual scene. The system works with photon-to-photon latency less than 100 milliseconds, when using transmission over WiFi without collision with other users [20].

This architecture allows us to deploy different algorithms for the processing of the scene, without increasing the computational load in the HMD. Thus, for the segmentation of the person's silhouette, both simple color-based algorithms and algorithms based on semantic segmentation by deep learning [21] have been used. It is also possible to include in the scene some objects from the local environment, with which the user can interact, such as keyboards [22], pens to take notes [17], or tools to perform physical tasks [23] .

The use of video-based avatars provides very high levels of spatial presence and self-presence, significantly improving on solutions based on VR controllers [22], and even those based on hand tracking [20]. It should also be borne in mind that the implementation of video-based avatars is technically more complex, and there is still room for improvement both in execution time and in segmentation accuracy. Additionally, the segmentation of elements of the local scene allows interaction with physical objects. This interaction can be used, for example, in training programs that require some manual work [23].

2.3 Face: visual communication

The combination of the two elements shown so far, **Visit** and **Move**, allows a user to feel present in a remote environment, to interact with the people who are there, as well as with the objects in their environment. The next step is to add other potential remote users to the scene, who will be represented by their avatars [16]. In the distributed reality context, these avatars should be real-time representations of other users, typically captured from an array of 3D depth-sensing cameras[3] and rendered as a photo-realistic avatar.

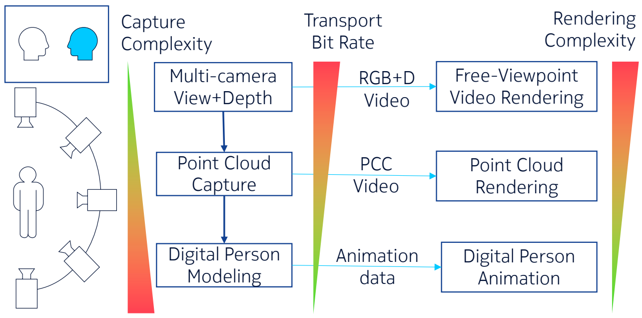


Figure 4 – Alternatives for the implementation of real-time capture and representation of avatars.

There are different approaches to capturing and transmitting these types of avatars. Each of them has different requirements in terms of the computational complexity of the capture and rendering process, as well as the bit rate of

2 We use *avatar* to describe the representation of the user within the immersive environment, regardless of whether this representation is generated as Computer-Graphic Imagery (CGI) animation, or it is a video capture of the user inserted in this scene. We use the term *video-based* avatar for the latter case.

3 3D depth-sensing cameras, or view+depth cameras, capture, for each color pixel, an estimate of the distance of that pixel to the camera (i.e. its *depth*).

data that need to be transmitted in real time (Figure 4):

- Free-Viewpoint Video (FVV) avatars [24]. An FVV system generates synthetic views of a 3D scene from a virtual viewpoint chosen by the user by combining the video information from several real reference cameras, which capture the scene synchronously from different viewpoints. In this solution, the cameras transmit color and depth information to the remote end, where the projection of the avatar in the virtual scene from the user's point of view is generated in real time. This simplifies image capture, but requires high bandwidth and processing power at the remote end. This approach has been recently standardized as MPEG Immersive Video (MIV) in ISO/IEC 23090-12 [25].

- Point cloud avatars [26]. A point cloud is a set of individual 3D points, each of which, in addition to having a 3D position, may also contain a number of other attributes, particularly color. In capture, the color and depth images obtained by the cameras are used to form a point cloud, which is compressed and transmitted using Point Cloud Compression (PCC) technology such as MPEG-I V-PCC (ISO/IEC 23090-5) [27]. On reception the point cloud is decoded and rendered within the immersive scene.

- Digital person model avatars [28]. In this case, a very detailed graphical representation of the captured person is generated in the capture, based on some model such as SMPL [29]. This description is transmitted and decoded at the remote end, where a previously modeled user avatar is animated. This option is the one that requires less transmission bandwidth and reception process.

2.4 Meet: shared immersion

The last step to compose a complete distributed reality experience is to integrate the different components in the scene so that the result is consistent, and the different users have the sensation of sharing the same space (Figure 5) .

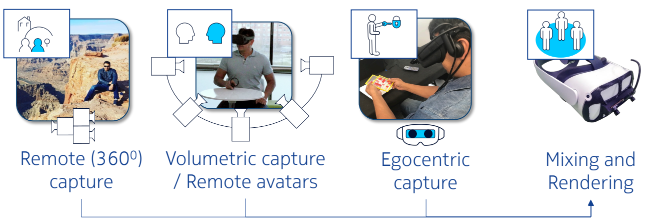


Figure 5 – Representation of the different video flows (visit, face, move) that need to be merged to create a consistent immersive shared experience (meet).

From an implementation point of view, the solution is known. It is enough to simply compose the virtual scene from its different elements (the remote capture, the egocentric capture and the avatars of the remote users) and render it on the HMD using a graphics engine such as Unity3D. In fact, when the shared scene is purely virtual, the insertion of the different elements is trivial, since the 3D geometry is shared by all users.
However, when it comes to mixing captured sources (video streams), the problem is much more complex, since there does not have to be a correspondence between the physical spaces that surround the different people. In this sense, the solutions proposed in the state of the art are always partial: represent the users separated by a window [30] or use digital twins of the physical environment of the remote user in the environment of the local user [31]. Further research is needed to be able to create a distributed reality scene with enough QoE to actually provide the user with the sense of teleportation needed to fulfill the vision of the realverse.

3. IMPLEMENTATION AND QUALITY OF EXPERIENCE

The implementation of a distributed reality solution requires, as has been seen, capturing, transmitting, processing and composing a set of 2D and 3D video streams in real time, and rendering the result on an HMD. Therefore, it is necessary to have an adequate communications and processing infrastructure for this. In addition, DR is presented as an intensive and demanding technology in terms of bandwidth, latency and processing resource requirements. Therefore, it is necessary to know adequately the trade-offs between the use of these resources and the obtained QoE.

3.1 The realverse over the 5G Network

Figure 6 represents our reference architecture to deploy the realverse over a 5G network, and its relationship to the building blocks described in the previous section. To be mobile, the HMD needs to connect wirelessly to the network, integrating the 5G User Equipment (UE) function. We assume, in principle, that the composition and rendering of the final image is done on the device itself, since the latency requirements of a remote rendering at 90 frames per second are challenging to achieve, in a scalable manner, for any mobile network in the short term. However, part of the processing associated to the rendering of the experience, such as the segmentation of the egocentric image or the detection of objects on it, can be derived to a system at the edge of the network (Multi-access Edge Computing, MEC).

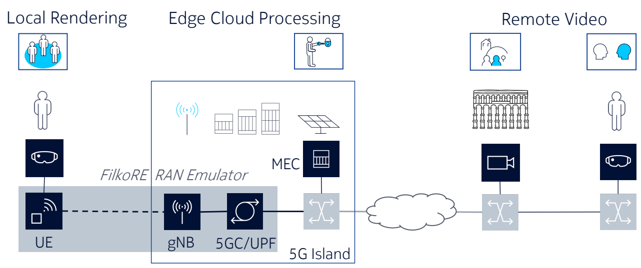


Figure 6 – Reference archiecture to implement the realverse over a 5G/B5G network.

Streams representing remote users and locations, on the other hand, must be transmitted in real time from the location where they occur. Although it is not represented in the figure, these flows may require additional processing (for example, creating a pointcloud representation of the avatar from the set of view+depth cameras) which could also be carried out at the edge of the network, close to the capture. Specifically relevant for the optimization of XR comunication is the viewport-dependant processing of the remote video flows (either 360 or pointcloud) in the edge: transmitting only the part of the scene which is actually being seen by the HMD user, thus saving bandwidth on the downlink channel.

Offloading the algorithms to the edge cloud, e. g. by using MEC, is necessary in order to guarantee sufficient processing capacity to execute them, for example if we are talking about semantic segmentation neural networks [21]. However, the requirements that this segmentation be done in real time and with a sufficiently high frame rate, imply that the network must support uplink traffic peaks in the order of gigabits per second, and have Round-Trip Times (RTTs) up to a MEC servcer of a few milliseconds [32]. When testing such demanding algorithms with currently deployed 5G networks, even those with the highest capacity operating in the millimeter band, it is frequent that the algorithms need to work on reduced frame resolution or frequency to fit within the network capacity [24]. In order to study in detail the interaction of networks with XR systems, we have developed a full-stack emulator of the 5G access network (FikoRE), with which we can test configurations not yet available on the market [33] .

Emulating the network with FikoRE, we have been able to test the operation of the segmentation algorithms described in the Section 2.2 for different configurations of a 5G and B5G networks [34]. In order to carry out the segmentation with sufficiently low latency and, therefore, to be able to deploy a DR service with sufficient QoE, it is necessary to use a radio access network in millimeter band, with at least 400 MHz of bandwidth and a symmetrical Time-Division Duplex (TDD) configuration for uplink and downlink.

3.2 Towards a quality of experience model

As mentioned above, XR communications systems, and the realverse is no exception, operate under constraints of bandwidth, latency, and computing power (or, equivalently, energy). It is therefore necessary to know the relationship between these restrictions and their impact on the quality of experience, in order to dimension, operate and monitor the network effectively.

QoE assessment in XR communications is a complex task. Recommendation ITU-T P.1320, recently published, advises on aspects of importance for QoE assessment of telemeetings with extended reality elements [35]. The goal is to define the human, context, and system factors that affect the choice of the QoE assessment procedure and metrics when XR communication systems are under evaluation. Among the System Influencing Factors (SIFs) for QoE, the Recommendation addresses three categories: the representation of the user and the world, the effect of rendering, and the restrictions of the communication network, which loosely match the three main system blocks described in Figure 6: local, edge, and remote.

Recommendations such as P.1320 are relevant, but still insufficient. The ultimate goal of QoE modeling is to have statistical models that allow the proper design of the network. This is a technically complicated challenge, due to the complexity and heterogeneity of the flows involved. Currently, the ITU-T has drafted Recommendations with parametric models (opinion models) for the most common telecommunications services: voice (G.107), video call (G.1070), IP television (G.1071), or online gaming (G.1072). In all cases, these are quite complex models, designed and developed with a multitude of parameters, and which, however, model much simpler communications systems. Therefore, it is necessary to approach simpler models as a first approximation, which capture the main interactions involved in the XR service, and which evaluate the order of magnitude of the relationship between the network restriction (for example, bandwidth) and the QoE. For this, we rely on previous versions of this exercise, for use cases of tele-operated driving [36] or virtual reality [37]. As a starting point, ITU-T has recently published the technical report GSTR-5GQoE, which describes the most relevant factors to perform this analysis in several use cases involving real-time video transmission over 5G networks [38]. The methodology applied in this technical report can be used to identify the most relevant QoE requirements of the service, using them to build a simplified parametric model.

Figure 7 shows a reference model for our approach. The main restrictions to which the system is subjected are bandwidth, latency and energy. This fundamentally affects two elements of QoE: visual quality and end-to-end latency. The visual quality is directly affected by the compression level, related to the throughput available to transmit; although the processing capacity (energy) or the execution time of the algorithms will also influence it. In the same way, the latency will be determined by the round-trip time of the network, to which are added other factors that affect the transmission, such as the relationship between the coding bitrate and the network throughput. This analysis process will be executed in parallel for the different flows involved, giving rise to a degradation factor I for each of them. The final process consists of merging the different contributions I to obtain a rating factor R, which can be translated into an expected value of the Mean Opinion Score (MOS) [39].

3.3 Throughput requirements

The throughput requirement for a video stream (B), regardless of the format, is simply the number of pixels (or image points) per second that need to be transmitted (P), multiplied by the average number of bits used to transmit each pixel (K). For a given scene, capture and compression technology, the rate K represents the degree of compression achieved. Since most video coding techniques use lossy compression (due to some quantification process), for a given B, increasing K (and therefore increasing $B = P \times K$) results in an improvement of the perceived quality.

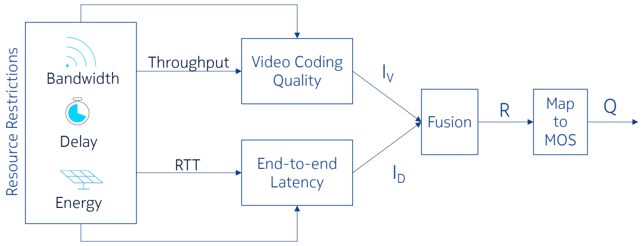


Figure 7 – Simplified QoE model for the realverse. Restrictions on bandwidth, delay and energy would affect visual quality and latency. Statistic quality models for them should be provided and fused into a global QoE model.

In conventional 2D video, this relationship is normally modeled as exponential, following the IQX hypothesis [40]:

$$I_V = 1 - e^{-K_0 \frac{B}{P}} \tag{1}$$

where K_0 represents the compression efficiency, B is the bit rate of the coded stream, and P is the number of pixels per second. This K_0 value captures the dependency on the codec efficiency (including energy considerations) and the spatio-temporal content complexity [41].

One important property is that, if we use 360-degree video in equirectangular projection, the video can be analyzed using the same tools as 2D video [42]. Since, in most immersive communication scenarios, the camera is typically set on a fixed position, and the scene does not usually have intense motion, the spatio-temporal complexity of the resulting content is moderately low, resulting also in relatively low bit rates, less than 10 Mbps for a 4K video in equirectangular projection [43].

Pointcloud transmission scenarios are less mature, and therefore the ranges of total number of points in the representation, its refresh rate (and therefore the total P), as well as the compression efficiency for realistic DR scenarios, are still under research. Common ranges for typical representations of 800 thousand to one million points range from 5 to 100 Mbps [27]. Bit rate can also be reduced using adapting schemes, where the part of the scene where the remote user is looking at is transmitted with higher bit rate than the rest of the scene.

3.4 Latency requirements

The effect of end-to-end delay in QoE has been modeled for conversational and interactive applications, and it is typically characterized by a function with three steps: a first threshold where delay is not important, a fast and linear decay, and a longer tail. The mathematical form of such function may be piecewise linear [44], logistic [45], log-logistic [37], or algebraic [36].

For illustration purposes, we select the latter, which comes from Recommendation ITU-T G.107:

$$I_D = \begin{cases} 1 - \frac{1}{2}\left\{\left(1 + x^6\right)^{\frac{1}{6}} - 3\left(1 + (x/3)^6\right)^{\frac{1}{6}} + 2\right\}, & \text{if } x \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

$$x = \log_2\left(\frac{T}{T_m}\right) \tag{3}$$

Where T is the “interaction lag" (the application-level end-to-end delay), and T_m is a the only model parameter, which can be considered as the threshold where latency starts to be noticeable. A property of this function is that $I_D(T = 4T_m) = 0.5$. Other models may have a second parameter controlling the decay.

The model to be applied will strongly depend on the type of flow where the delay happens. For voice conversation T_m is established around 100 ms, and it can be even higher if there is visual feedback from the other side. For the performance of adaptive compression schemes, users tend to be more tolerable, and the value can be increased to 200-500 ms [46]. Tasks which require quick interactivity and response, such as driving, may use lower values (e.g. 30 ms [36]), even though trained operators can quickly adapt to perform under much worse conditions [47].

In general, we can say that the detailed understanding of the effect of latency in QoE has still room for research before good parametric models are developed that can safely apply to a large range of XR communication scenarios.

4. OPPORTUNITIES FOR STANDARDIZATION

The development of the realverse and, in general, of communications based on extended reality, is still in its infancy. In fact, although existing technology has already shown that it is possible to create communication experiences between people in immersive environments, there is still some way to go before its application is massive. In this scenario, several opportunities open up to standardize key elements of the development of the metaverse in any of its interpretations. We will now talk about two of them.

On the one hand, it would be desirable to address the interoperability of the various systems from the outset, so that it would be possible to build communication solutions that were not fully captured by (and vertically integrated into) a particular platform. Since text messaging and video calling services are mainly provided by major social media platforms and hyperscalers, a new opportunity arises for the new generation of communications to be installed again on open and interoperable systems, as happened with telephony. In this context, our realverse proposal is particularly relevant, as it focuses more on peer-to-peer communications than on integrating users into a specific platform (or metaverse). Of course, the need to implement part of the processing capacity at the edge means that it is not only necessary to standardize communication protocols, but also Virtual Network Functions (VNFs) capable of offering specific services.

On the other hand, there is a need to develop new standards for evaluating and monitoring the quality of experience. To the aforementioned Recommendation ITU-T P.1320, new subjective evaluation methodologies must be added, as well as objective metrics for monitoring and dimensioning the network. ITU-T has already started addressing this problem, with the analysis of QoE influencing factors for VR (G.1035 [48]) and AR (G.1036 [49]), as well as with a methodology to assess the quality of immersive video (P.919 [50]). However, the subjective evaluation of immersive communications systems is a significantly more complex task than simply assessing the visual quality offered by the HMD.

First, the ability of XR technology to elicit a sense of presence, even with much poorer visual quality than that available in, say, high-definition video on a screen, makes it not enough to measure subjective quality-of-the-pixel. It is also necessary to measure socio-emotional aspects such as the feeling of presence or the ability to improve empathy in communications [17]. In order to carry out experiments that allow measuring differences in these aspects, more sophisticated methodologies are needed, which include interaction between several people and carrying out different communication tasks [11]. The challenge ahead is being able to develop a Recommendation that is precise enough to define a set of repeatable tasks and measurements across multiple labs, while also being broad and complex enough to cover many types of systems and measure various perceptual and socioemotional properties. The Immersive Media Group (IMG) of the Video Quality Experts Group (VQEG) is currently starting a project to design and validate an assessment methodology that can address this challenge for a wide range of immersive communication systems.

Finally, with regard to objective quality models, the range of pending work is wide. Today there are no sufficiently consolidated metrics that allow us to accurately model, measure or monitor the objective quality of an immersive communication system. No doubt new Recommendations will emerge over the years that address this problem. In the meantime, there is only recourse to simple solutions like the one outlined in Section 3.2.

5. CONCLUSIONS

Throughout this paper we have explored the idea of the realverse and its implications. First of all, we have proposed the realverse as the vision that several people in distant places can meet each other and interact as if they were visiting the same physical place. The concept of the realverse is closely linked to its practical implementation, distributed reality: combining audiovisual content captured in different places, in such a way that it gives rise to a consistent immersive experience, allowing people to interact naturally.

Throughout the text we have also proposed a way to implement a first version of the realverse from its fundamental building blocks. Our previous work on immersive telepresence with 360 video and on video-based avatars shows that this technology provides a very high level of presence, even though it still has technical limitations. We have also seen that these components must be completed by including video-based avatars also for remote users and, above all, by generating a shared virtual space that is geometrically consistent with the physical space in which the different participants move.

Subsequently, we have proposed and analyzed a network architecture to deploy distributed reality services in a 5G and B5G network. We have described the basic quality models that exist in the state of the art, and we have proposed a way to be able to have coarse parametric models that allow improving the dimensioning of the network.

Finally, we have analyzed the main opportunities for standardization that arise from this technology: on the one hand, in the design of an interoperable system (using telephony as an example) and, on the other, in the development of subjective methodologies and objective metrics to assess the quality of experience.

ACKNOWLEDGEMENT

The author would like to acknowledge the whole Nokia eXtended Realty Lab, as well and the Image Processing Group (GTI) at Universidad Politécnica de Madrid, since this keynote paper shows the work done jointly with researchers of both teams. Special thanks to Marta Orduna, Carlos Cortés, Diego González-Morín, and Ester González-Sosa, from whom I have learnt so much in the last few years.

This work has received funding from the Spanish Administration Agency *Centro para el Desarrollo Tecnológico Industrial* (CDTI) under project IDI-20210977 (IMMINENCE) as an International Technical Cooperation R&D Project.

REFERENCES

[1] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974–990, 2012.

[2] I. Rae, G. Venolia, J. C. Tang, and D. Molnar, "A framework for understanding and designing telepresence," in *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pp. 1552–1566, 2015.

[3] J. McVeigh-Schultz, A. Kolesnichenko, and K. Isbister, "Shaping pro-social interaction in VR: an emerging design framework," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[4] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst, "Revisiting collaboration through mixed reality: The evolution of groupware," *International Journal of Human-Computer Studies*, vol. 131, pp. 81–98, 2019.

[5] A. Villegas, P. Pérez, and E. González-Sosa, "Towards a distributed reality: a multi-video approach to XR," in *Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems*, pp. 34–36, 2019.

[6] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[7] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 332–339, 2005.

[8] S. B. Snider, J. Hsu, R. R. Darby, D. Cooke, D. Fischer, A. L. Cohen, J. H. Grafman, and M. D. Fox, "Cortical lesions causing loss of consciousness are anticorrelated with the dorsal brainstem," *Human brain mapping*, vol. 41, no. 6, pp. 1520–1531, 2020.

[9] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton, "A survey of presence and related concepts," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–39, 2017.

[10] F. Biocca, "The cyborg's dilemma: Progressive embodiment in virtual environments," *Journal of computer-mediated communication*, vol. 3, no. 2, p. JCMC324, 1997.

[11] P. Pérez, E. Gonzalez-Sosa, J. Gutiérrez, and N. García, "Emerging immersive communication systems: Overview, taxonomy, and good practises for qoe assessment," *Frontiers in Signal Processing*, 2022.

[12] W.-C. Lo, C.-L. Fan, J. Lee, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "360 video viewing dataset in head-mounted virtual reality," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp. 211–216, 2017.

[13] A. Izumihara, D. Uriu, A. Hiyama, and M. Inami, "ExLeap: Minimal and highly available telepresence system creating leaping experience," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1321–1322, 2019.

[14] A. Nassani, L. Zhang, H. Bai, and M. Billinghurst, "ShowMeAround: Giving virtual tours using live 360 video," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2021.

[15] R. Kachach, M. Orduna, J. Rodríguez, P. Pérez, Á. Villegas, J. Cabrera, and N. García, "Immersive telepresence in remote education," in *Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE'21)*, pp. 21–24, 2021.

[16] R. Kachach, S. Morcuende, D. Gonzalez-Morin, P. Perez, E. Gonzalez-Sosa, F. Pereira, and A. Villegas, "The Owl: Immersive telepresence communication for hybrid conferences," in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 451–452, 2021.

[17] M. Orduna, P. Pérez, J. Gutiérrez, and N. García, "Methodology to assess quality, presence, empathy, attitude, and attention in 360-degree videos for immersive communications," *IEEE Transactions on Affective Computing*, pp. 1–1, 2022.

[18] P. Perez, E. Vallejo, M. Revuelta, M. V. Redondo Vega, E. Guervós Sánchez, and J. Ruiz, "Immersive music therapy for elderly patients," in *ACM International Conference on Interactive Media Experiences*, pp. 47–52, 2022.

[19] P. Perez, E. Gonzalez-Sosa, R. Kachach, J. Ruiz, I. Benito, F. Pereira, and A. Villegas, "Immersive gastronomic experience with distributed reality," in *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–6, 2019.

[20] D. Gonzalez-Morin, E. Gonzalez-Sosa, P. Perez-Garcia, and A. Villegas, "Bringing real body as self-avatar into mixed reality: A gamified volcano experience," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 794–795, IEEE, 2022.

[21] E. Gonzalez-Sosa, A. Gajic, D. Gonzalez-Morin, G. Robledo, P. Perez, and A. Villegas, "Real time egocentric segmentation for video-self avatar in mixed reality," 2022.

[22] A. Villegas, P. Perez, R. Kachach, F. Pereira, and E. Gonzalez-Sosa, "Realistic training in VR using physical manipulation," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 109–118, 2020.

[23] C. Cortés, M. Rubio, P. Pérez, B. Sánchez, and N. García, "Qoe study of natural interaction in extended reality environment for immersive training," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 363–368, IEEE, 2022.

[24] P. Pérez, D. Corregidor, E. Garrido, I. Benito, E. González-Sosa, J. Cabrera, D. Berjón, C. Díaz, F. Morán, N. García, *et al.*, "Live free-viewpoint video in immersive media production over 5g networks," *IEEE Transactions on Broadcasting*, 2022.

[25] V. K. M. Vadakital, A. Dziembowski, G. Lafruit, F. Thudor, G. Lee, and P. Rondao Alface, "The MPEG Immersive Video Standard - Current Status and Future Outlook," *IEEE Multimedia*, 2022.

[26] M. Montagud, J. Li, G. Cernigliaro, A. El Ali, S. Fernández, and P. Cesar, "Towards socialvr: evaluating a novel technology for watching videos together," *Virtual Reality*, pp. 1–21, 2022.

[27] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, *et al.*, "Emerging mpeg standards for

point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[28] A. C. S. Genay, A. Lécuyer, and M. Hachet, "Being an avatar" for real": a survey on virtual embodiment in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[29] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," *ACM transactions on graphics (TOG)*, vol. 34, no. 6, pp. 1–16, 2015.

[30] J. Lawrence, D. B. Goldman, S. Achar, G. M. Blascovich, J. G. Desloge, T. Fortes, E. M. Gomez, S. Häberling, H. Hoppe, A. Huibers, C. Knaus, B. Kuschak, R. Martin-Brualla, H. Nover, A. I. Russell, S. M. Seitz, and K. Tong, "Project Starline: A high-fidelity telepresence system," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, vol. 40(6), 2021.

[31] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-Me: An adaptive avatar for mixed reality remote collaboration," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[32] D. G. Morín, P. Pérez, and A. G. Armada, "Toward the distributed implementation of immersive augmented reality architectures on 5g networks," *IEEE Communications Magazine*, vol. 60, no. 2, pp. 46–52, 2022.

[33] D. G. Morin, M. Morales, P. Pérez, A. G. A. A. Villegas, *et al.*, "Fikore: 5g and beyond ran emulator for application level experimentation and prototyping," *arXiv preprint arXiv:2204.04290*, 2022.

[34] D. González Morín, M. J. López Morales, P. Pérez, and A. Villegas, "Tcp-based distributed offloading architecture for the future of untethered immersive experiences in wireless networks," in *ACM International Conference on Interactive Media Experiences*, pp. 121–132, 2022.

[35] ITU-T, "Qoe assessment of extended reality (xr) meetings," *Recommendation P.1320*, 2022.

[36] P. Pérez, J. Ruiz, I. Benito, and R. López, "A parametric quality model to evaluate the performance of tele-operated driving services over 5g networks," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12287–12303, 2022.

[37] B. Krogfoss, J. Duran, P. Perez, and J. Bouwen, "Quantifying the value of 5g and edge cloud on qoe for ar/vr," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–4, IEEE, 2020.

[38] ITU-T, "Quality of experience (QoE) requirements for real-time multimedia services over 5G networks," *Technical Report GSTR-5GQoE*, June 2021.

[39] T. Hoßfeld, P. E. Heegaard, M. Varela, and S. Möller, "Qoe beyond the mos: an in-depth look at qoe via better metrics and their relation to mos," *Quality and User Experience*, vol. 1, no. 1, pp. 1–23, 2016.

[40] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[41] L. Janowski, P. Romaniak, and Z. Papir, "Content driven QoE assessment for video frame rate and frame resolution reduction," *Multimedia tools and applications*, vol. 61, no. 3, pp. 769–786, 2012.

[42] M. Orduna, C. Díaz, L. Muñoz, P. Pérez, I. Benito, and N. García, "Video multimethod assessment fusion (vmaf) on 360vr contents," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 1, pp. 22–31, 2019.

[43] J. Gutiérrez, P. Pérez, M. Orduna, A. Singla, C. Cortés, P. Mazumdar, I. Viola, K. Brunnstrom, F. Battisti, N. Cieplinska, D. Juszka, L. Janowski, M. I. Leszczuk, A. Adeyemi-Ejeye, Y. Hu, Z. Chen, G. Van Wallendael, P. Lambert, C. Diaz, J. Hedlund, O. Hamsis, S. Fremerey, F. Hofmeyer, A. Raake, P. Cesar, M. Carli, and N. García, "Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T Rec. P.919," *IEEE Transactions on Multimedia*, pp. 1–14, 2021.

[44] S. Wang and S. Dey, "Modeling and characterizing user experience in a cloud server based mobile gaming approach," in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*, pp. 1–7, IEEE, 2009.

[45] S. Moller, S. Schmidt, and S. Zadtootaghaj, "New itu-t standards for gaming qoe evaluation and management," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2018.

[46] C. Cortés, P. Pérez, J. Gutiérrez, and N. García, "Influence of video delay on quality, presence, and sickness in viewport adaptive immersive streaming," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–4, 2020.

[47] K. Brunnström, E. Dima, T. Qureshi, M. Johanson, M. Andersson, and M. Sjöström, "Latency impact on quality of experience in a virtual reality simulator for remote control of machines," *Signal Processing: Image Communication*, vol. 89, p. 116005, Nov. 2020.

[48] ITU-T, "Influencing factors on quality of experience for virtual reality services ," *Recommendation G.1035*, Nov. 2021.

[49] ITU-T, "Quality of Experience (QoE) Influencing Factors for Augmented Reality Services," *Recommendation G.1036*, 2022.

[50] ITU-T, "Subjective test methodologies for 360 video on head-mounted displays," *Recommendation P.919*, Oct. 2020.

INVITED PAPER

A SURVEY OF EXTENDED REALITY (XR) STANDARDS

Gillian Makamara and Martin Adolph

International Telecommunication Union (ITU)

NOTE – The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the International Telecommunication Union or its membership.

ABSTRACT

This paper surveys completed and ongoing work of Standards Developing Organizations (SDOs), fora and consortia dealing with technical aspects of extended Reality (XR). In line with the XR definition introduced by [1], specifications considered here address environments containing real or virtual components or a combination thereof, i.e. standards dealing with, inter alia, Augmented Reality (AR), Virtual Reality (VR), and metaverse. While aiming to be comprehensive, this survey can only provide a snapshot of the standards landscape taken at the time of submission of the manuscript.

Keywords – Augmented reality (AR), extended reality (XR), metaverse, SDO, specification, standard, standardization, survey, virtual reality (VR)

1. INTRODUCTION

The past two years have highlighted, perhaps more than ever, the key role of information and communication technologies (ICTs) in managing disruptions brought about by global challenges, such as the COVID-19 pandemic.

As people have shifted behavior, reduced their face-to-face interactions and avoided gatherings, events, travel and public transportation, access to broadband networks helped those connected to move some of their physical interactions online, to the virtual space.

Work From Home or "WFH" emerged as perhaps the most prominent outcome of this shift, leaving office spaces empty. The download figures of online meeting applications such as Microsoft Teams or Zoom have skyrocketed. In no time, physical distancing and confinement changed the way we teach and learn, work out, attend academic conferences, and engage in trade shows or standards meetings, usually comprising real and/or virtual elements.

Coined long before the pandemic, and not limited to virtual background pictures and (augmented reality) face filters in video conferences, the term extended reality, or XR for short, saw a strong growth in interest and attention over the past two years. In one of its most recent Recommendations [1], ITU defines XR as "an environment containing real or virtual components or a combination thereof, where the variable X serves as a placeholder for any form of new environment" e.g. Augmented Reality (AR), Mixed Reality (MR), or Virtual Reality (VR).

Examples for immersive technologies and experiences covered by the XR umbrella term have existed in the past, whether in the form of the Pokémon Go mobile AR gaming app, the Oculus VR and HoloLens MR headsets [2]. The announcement [3] of the rebranding of Facebook to Meta in October 2021 popularized the closely related term metaverse, a "hybrid of today's online social experiences, sometimes expanded into three dimensions or projected into the physical world." We contend that both the pandemic and the rebranding have served as catalysts to bringing these technologies and their capabilities and application scenarios to the attention of a broader audience.

Growing interest in and adoption of XR go hand in hand with calls for interoperability, to avoid market fragmentation, create a level playing field which provides access to new markets, and through economies of scale, which can help reduce costs for all: manufacturers, service providers, and consumers.

In this paper, we introduce the organizations responding to these calls for XR standards and categorize and summarize their respective deliverables.

The work is based on a review of related literature, including [4, 5], study of the documentation of the standardization groups covered, discussions with individuals involved in the activities described, and our own personal experience in the ITU.

While aiming to be comprehensive, we note that this survey can only provide a snapshot of the standards landscape taken at the time of submission of the manuscript in September 2022.

2. ORGANIZATIONS ACTIVE IN XR STANDARDIZATION

In this section, we introduce, in alphabetical order, the organizations currently involved in, or associated with, XR-related standardization. We briefly describe their respective scope and membership composition. Where such information is available, we highlight opportunities for academia to engage in the XR standards making activities described.

Table 1 provides a high-level summary of the organizations covered in this section, including the respective working group(s) dealing with XR.

2.1 Standards Developing Organizations (SDOs)

2.1.1 3rd Generation Partnership Project (3GPP)

The 3rd Generation Partnership Project (3GPP) unites seven regional and national telecommunications SDOs (the Organizational Partners) to develop reports and specifications of cellular telecommunication technologies, including radio access, core network and service capabilities, which provide a complete system description for mobile telecommunications.

Work on XR-related specifications is currently undertaken by two working groups of the Services & Systems Aspects Technical Specification Group (TSG SA), and one working group of the Radio Access Networks Technical Specification Group (TSG RAN).

SA WG4 addresses topics including: XR-based services and traffic characteristics; glass-type AR; VR conferencing; immersive voice and audio services and associated extension for headset interface tests of User Equipment (UE); immersive audio and video quality. SA WG2 discusses aspects of immersive media architecture. RAN WG1 explores XR application performance for 5G New Radio (NR). The working groups meet regularly and come together for their quarterly TSG Plenary meeting, where their work is presented for information, discussion, and approval.

2.1.2 European Telecommunications Standards Institute (ETSI)

The European Telecommunications Standards Institute (ETSI) is a recognized regional standards body dealing with telecommunications, broadcasting and other electronic communication networks and services. ETSI's Industry Specification Group (ISG) Augmented Reality Framework (ARF) aims to specify relevant components and interfaces required to ensure interoperability of AR components, systems, and services.

ETSI membership comprises private sector organizations, governmental organizations, user associations, universities, and public research bodies. Participation in the ISG is open to members and non-members; for the latter, a fee is charged for participation in ISG meetings.

2.1.3 Institute of Electrical and Electronics Engineers Standards Association (IEEE SA)

An operating unit within IEEE, the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA) develops global standards in a broad range of industries, including power and energy, consumer technology and consumer electronics, biomedical and health care, information technology, telecommunications, automotive and transportation. IEEE SA work on XR-related topics is spread across several committees, working groups and projects, including:

- IEEE VR/AR Standards Committee (CTS/VRARSC),
- IEEE 1857 Working Group (Audio Video Coding),
- IEEE P2048 Working Group (VR/AR),
- IEEE P2843 Working Group (Measuring Accessibility Experience and Compliance),
- IEEE P2874 Working Group (Spatial Web protocol, architecture and governance), in collaboration with the Spatial Web Foundation,
- IEEE 2888 Working Group (Interfacing cyber and physical world),
- IEEE 3079 Working Group (Human factors for immersive content),
- IEEE P7030 Working Group (Global XR ethics).

The standards development process is open to members and non-members, alike, with privileges for IEEE SA members in the balloting.

2.1.4 International Organization for Standardization / International Electrotechnical Commission, Joint Technical Committee, Subcommittee 29 (ISO/IEC JTC 1, SC29)

Joint Technical Committee (JTC 1) of International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), is a consensus-based, voluntary international standards group for information technology. Subcommittee 29 (SC29) "Coding of audio, picture, multimedia and hypermedia information" comprises both, the Joint Photographic Experts Group (JPEG) and the Moving Picture Experts Group (MPEG).

JPEG Pleno aims to provide a standard framework for representing new imaging modalities, such as texture-plus-depth, light field, point cloud, and holographic imaging. MPEG has released collections of standards to digitally represent immersive media (MPEG-I) and specifications aimed at facilitating interoperability between virtual worlds (MPEG-V).

2.1.5 International Telecommunication Union (ITU)

The work of the International Telecommunication Union's Radiocommunication Sector (ITU-R) and Standardization Sector (ITU-T) is conducted through contribution-driven Study Groups.

Immersive video conferencing, MR and VR were among the application scenarios considered by Working Party 5D (WP5D) of ITU-R Study Group 5 (responsible for terrestrial services) in its foundational work leading to the IMT-2020 (5G) specifications. Situated between enhanced mobile broadband and ultra-reliable and low latency communications, AR features in the "5G triangle" [6], depicting usage scenarios of IMT for 2020 and beyond. WP5D's preliminary draft work on technology trends of terrestrial IMT systems towards 2030 and beyond identifies XR-enabled interactive immersive experiences among the use cases driving innovation in mobile network technologies. ITU-R Study Group 6 (responsible for broadcasting service) explores aspects of immersive audio and immersive video in a broadcasting context.

Most of the 11 ITU-T Study Groups have activities and outputs related to XR, including Study Group 9 (transmission and distribution of immersive content over cable networks), Study Group 11 (testing procedures for AR applications), Study Group 12 (immersive media quality of experience (QoE)), Study Group 13 (trusted networks for immersive media), Study Group 16 (immersive live experience systems and services), and Study Group 20 (using AR and VR to monitor and control IoT devices). In January 2022, Study Group 16 established a correspondence group for metaverse standardization (CG-Metaverse) to assess standardization needs in the context of its mandate.

The resulting voluntary international standards, called ITU-R and ITU-T Recommendations respectively, are developed by, and based on the consensus of the organization's membership, comprising industry, academia, and, reflecting its role as a United Nations specialized agency, 193 states from around the globe.

2.1.6 World Wide Web Consortium (W3C)

W3C's Immersive Web Working Group is aiming to bring XR content, applications, and services to the web by developing standardized APIs to interact with XR devices and sensors in web browsers. W3C is hosted jointly by four academic institutions and has membership options for universities.

2.2 Consortia, fora, and trade associations

2.2.1 Consumer Technology Association (CTA)

The Consumer Technology Association (CTA) is a trade association representing the U.S. consumer technology industry. It is perhaps best known as the owner and producer of the annual Consumer Electronic Show (CES). The CTA's XR Working Group supports the growth of companies developing technologies and services in the AR, MR and VR segments.

A separate AR/VR Committee develops standards, recommended practices, and technical reports related to AR, VR and related technologies. Participation in the committee is open to CTA members and non-members including academia who pay an annual participation fee.

2.2.2 Immersive Digital Experiences Alliance (IDEA)

The Immersive Digital Experiences Alliance (IDEA) is an industry alliance working towards developing royalty-free technical specifications that define interoperable interfaces and exchange formats to support the end-to-end conveyance of immersive volumetric and/or light field media.

2.2.3 Khronos Group

The Khronos Group is an industry consortium creating interoperability standards for 3D graphics, AR, and VR, among other areas. The Khronos membership also includes some 20 universities.

2.2.4 Metaverse Standards Forum

Initiated by the Khronos Group in June 2022, the Metaverse Standards Forum intends to bring together companies and standards organizations to foster alignment on requirements and priorities for metaverse interoperability standards and accelerate their development and deployment. Rather than developing standards itself, the forum's goal is to coordinate requirements and support for existing standards relevant to the metaverse, and to provide a venue for discussion and coordination between standards organizations and companies building metaverse-related products [7]. Membership is currently at no charge and open for for-profit and non-profit organizations, including universities.

2.2.5 Streaming Video Technology Alliance (SVTA)

The Streaming Video Technology Alliance (SVTA) is an industry group spanning the streaming video value chain. It produces documentation on best practices, specifications, and requirements on identified key challenges of video streaming. Topics discussed include caching, measurement / QoE, and VR / 360-degree video. While not a standards body itself, the alliance submits its specifications to the appropriate standards body for ratification.

2.2.6 Virtual Reality Industry Forum (VRIF)

The principal purpose of the VR Industry Forum is to further the widespread availability of high-quality audiovisual VR experiences, for the benefit of consumers. Like SVTA, the VRIF does not develop standards, but relies on, and liaises with, SDOs such as MPEG for the development of standards in support of VR services and devices. Membership in the VRIF is open to academic institutions.

2.3 Research networks

2.3.1 Qualinet

Qualinet is a network of European researchers with the main scientific objective of developing methodologies for subjective and objective quality metrics addressing trends in multimedia communication systems. Qualinet outputs include research papers, position papers (white papers) and databases containing audiovisual and subjective datasets. Findings of the research network have been contributed to international standardization efforts on immersive media experiences, notably to ITU-T Study Group 12.

3. XR STANDARDS

An analysis of published outputs from the organizations introduced above reveals two main themes being explored in XR standardization work: (1) building capabilities for XR interoperability by establishing common understanding (standard terminology), identifying key system and user requirements (design guidelines and system standards) and developing compatible interfaces and data formats for XR services and applications; and (2) defining XR user experience requirements addressing accessibility and quality aspects. In this section, we introduce a selection of existing technical standards and specifications, grouped by theme, relevant to the creation, delivery, and deployment of XR experiences.

Table 2 includes an overview of the standards and specifications covered in this paper.

3.1 Interoperability

3.1.1 Terminology, use cases

Terminology is often the first element to be standardized for any new technology as it provides the foundation for interoperability. Current work on standard terminology for XR tends to feature more within standards offering guidance on other topics under the XR umbrella, as in the case of [1], rather than standalone lexicons such as CTA-2069-A.

Although varying definitions are adopted within these and other standards, terminology seems to converge towards a common level of understanding. For example, adopted definitions for AR in a sample of standards [1, 8, 9] converge towards a uniform interpretation of AR as a superimposition of virtual objects to the "real"/physical world. ITU-T P.1320 as well as CTA-2069-A and CTA-2085 standards altogether outline key terms for XR and could potentially be used in combination as a "comprehensive XR glossary". Ongoing work under IEEE P2048, which aims to define terminology for immersive video, audio, and VR/AR devices, could also complement this list once complete.

XR is being applied extensively e.g. in entertainment and gaming, healthcare, e-commerce, education, etc. and an analysis of its application in these areas provides guidance to standards developers in understanding user needs, identifying requirements needed to meet user expectations and developing a baseline for interoperability requirements. XR use cases have been described across a range of standards-related literature (both normative and informative publications) including SVTA's extended reality brief, VRIF's guidelines on live VR services, multiple 3GPP technical reports including TR 26.862 (immersive teleconferencing), TR 26.928 (XR services in 5G), TR 26.918 (VR use cases) and TR 26.998 (AR use cases over 5G), an ETSI report (ETSI GR ARF 002) on AR industrial use cases and specifications (ETSI GS MEC 002 and GS MEC-IEG 004) on XR over mobile edge computing, as well as a few ITU Recommendations including ITU-T G.1036 (AR use cases) and ITU-T H.430.3 (use cases on Immersive Live Experiences (ILEs)).

In addition to these *broader* use case descriptions, some standards offer guidance on deploying specific XR use cases such as IEEE 1589 (AR-assisted learning), ITU-T F.740.2 (AR artwork) and ITU-T J.301 (AR-enabled smart television services).

ITU-T Technical Report GSTR-5GQoE on QoE requirements for real-time multimedia services over 5G networks discusses MR offloading implementation scenarios, architecture and derives relevant KPIs [10].

3.1.2 Design guidelines, system standards, and APIs

Baseline requirements for implementing XR are provided in use case publications (as outlined in Section 3.1.1), e.g. ITU-T H.430.3 which covers not only use cases but also key requirements and architectural considerations for implementing ILE.

Complementary to this are the other parts of the ITU-T H.430.x series which offer additional guidance on implementing ILE including high-level functional requirements (H.430.1) and architecture (H.430.2), signalling aspects (H.430.4) and a reference model for ILE presentation (H.430.5). ITU-T J.302 offers detailed system requirements for AR-enabled smart TV services to complement the use case descriptions in ITU-T J.301. Additional video and system parameters supporting immersive media in TV broadcasting services are specified in ITU-R BT.2123-0.

ETSI specifies a functional reference architecture (ETSI GS ARF 003) and detailed interoperability requirements (ETSI GS ARF 004-x series) for AR components, systems, and services in addition to the requirements described in ETSI GR ARF 002.

VRIF outlines interoperability guidelines and best practices for VR content production and distribution for VR implementers in VRIF Guidelines 2.3 using MPEG standards described in the following section as a basis.

Standard models and interfaces are also required to generate virtual components and facilitate data exchange on XR applications. The catalog of published work on this subject features a variety of Application Programming Interface (API) definitions and requirements for representing virtual 3D scenes.

The Khronos group's XR API suite provides developers with tools to create interoperable XR experiences (OpenXR) and generate 3D scenes/assets (3DCommerce and gITF), and high-performance low-latency 3D graphics (Vulkan, OpenGL and WebGL).

From the ISO/IEC suite, ISO/IEC 23090-8 defines APIs for immersive media processing and the extensible 3D (X3D) standards provide a format and architecture for the representation of 3D scenes/objects (ISO/IEC 19775 and 19776 series). Specific formats for XR are also defined in ISO/IEC 18038 to 18040, 18520 and 23488.

Although not published yet, ongoing work in W3C is set to provide developers with APIs to build immersive applications that can be accessed on the web under the WebXR series.

3.1.3 Data presentation and formats

The suite of available standards on data formats for immersive media has largely been outputs from MPEG and JPEG and comprises of ISO/IEC 23000-13 and the ISO/IEC 21794, 23005 and 23090 series. The completed seven-part ISO/IEC 23005 (MPEG-V) set describes an architecture and outlines requirements for interoperability between virtual and real worlds. ISO/IEC 23000-13 blends some elements from this series to specify AR data formats.

The five-part ISO/IEC 21794 (JPEG Pleno) set complements the MPEG-V series by providing a framework and coding tools to facilitate the representation of new image modalities required to support immersive environments. Currently published work in this series specifies the general framework (Part 1) and coding tools for light field modalities (Part 2). The series also offers guidance on conformance testing (Part 3) and implementing reference software (Part 4) for the series. The final piece in the series (Part 5), currently under development, is exploring a similar set of coding tools for holographic imaging.

Completed work in the ISO/IEC 23090 (MPEG-I) series specifies a set of data formats and coding tools for immersive media. This work is composed of formats, conformance testing and reference software implementation for omnidirectional media (parts 2 and 17), improved features for versatile video coding (Part 3) and data compression and carriage for point clouds (parts 5 and 10). Ongoing activities on the ISO/IEC 23090 series are looking into specifying coding formats for immersive audio/video [11] and volumetric, point cloud and versatile video types.

Outside the ISO/IEC standards suite, ITU-R BT.2133-0 describes elements required for the transport of immersive audio-visual content using the omnidirectional media format (specified in ISO/IEC 23090-2) in IP-based broadcasting systems.

IEEE 1857.9 specifies an efficient coding toolset for immersive visual content and current work in progress under the IEEE P2048 series is exploring data formats for immersive audio and video. The IDEA's Immersive Technologies Media Format (ITMF) specification provides a reference for data exchange formats supporting end-to-end delivery of immersive media.

3.2 User experience

3.2.1 Accessibility and human factors

XR technologies impose very high requirements for usability. This requirement is even greater for users with specific needs, particularly those with sensory impairments or limited mobility (where movement is a component of user interaction). [12] provides more insight on accessibility requirements within immersive media and standardization opportunities in this context.

To support the delivery of optimally accessible XR experiences, the W3C's working group note on 'XR accessibility user requirements' and ETSI EG 202 848 outline a set of accessibility guidelines containing accessible user design principles and architectural considerations for XR system designers and product developers. CTA 2095 also offers some considerations for XR users with limited mobility.

Complementary to this work, the IEEE's accessibility and digital inclusion working group are currently developing testing criteria to measure XR accessibility in IEEE P2843.

3.2.2 Quality aspects

Assessing quality of experience is a crucial factor in evaluating overall user experience. Activities of the ITU, IEEE and 3GPP have yielded a series of technical specifications and standards on conducting QoE evaluation of XR.

Current publications provide guidance on understanding factors influencing user-perceived experience for VR (ITU-T G.1035) and AR (ITU-T G.1036), outline a set of quality metrics (ISO/IEC 23090-6 and 3GPP TR 26.929) and KPIs (3GPP TR 26.998, ITU-T GSTR-5GQoE) for immersive media and describe assessment methodologies for immersive audio (3GPP TS 26.259 and TS 26.260), 360º video (ITU-T P.919), XR telemeetings (ITU-T P.1320), and perceptual quality of 3D visual content (IEEE 3333.1 series).

IEEE 3079 provides a different perspective to QoE of XR. It focuses on improving VR user experience by providing guidelines on reducing simulator sickness which is an undesirable effect experienced by some VR users.

Other existing publications on XR-related quality offer guidance on network performance aspects (3GPP's 5G capacity study for XR (TR 38.838)), QoS assurance considerations (ITU-T Y.3109), cable network specific requirements to deliver VR services (ITU-T J.1631), and testing procedures for AR applications (ITU-T Q.4066).

4. CONCLUSIONS

In the recent past, immersive technologies and experiences covered by the extensive reality umbrella term have experienced a rapid growth in interest and adoption. This interest is reflected in the activities of standards developing organizations, industry fora and consortia, and other initiatives, aiming to provide guidance in this domain, foster interoperability, and facilitate even greater XR adoption.

In this paper, we have attempted to provide a snapshot of these growing number of activities and their sponsoring organizations, broadly classifying the work in two categories: (1) building capabilities for XR interoperability, and (2) defining XR user experience requirements.

Looking ahead, we anticipate that the number of organizations involved, and their XR-related initiatives will continue to increase, likely addressing technical (e.g. security, privacy) and non-technical areas (e.g. ethics) outside these two categories.

Regular surveys of the XR standards landscape can increase awareness and inform coordination and collaboration efforts, benefitting manufacturers, service providers and consumers.

REFERENCES

[1] Recommendation ITU-T P.1320, "Quality of experience (QoE) assessment of extended reality (XR) meetings," 2022.

[2] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for extended reality (XR)," ITU J-FET, vol. 3 no. 2, pp. 273-287, 2022.

[3] Introducing Meta: A Social Technology Company. https://about.fb.com/news/2021/10/facebook-company-is-now-meta/, 28 October 2021.

[4] J. Skowronek, A. Raake, G. H. Berndtsson, O. S. Rummukainen, P. Usai, S. N. B. Gunkel, M. Johanson, E. A. P. Habets, L. Malfait, D. Lindero and A. Toet, "Quality of Experience in Telemeetings and Videoconferencing: A Comprehensive Survey." IEEE Access, vol. 10 no. 10, pp. 63885-63931, 2022.

[5] A. Perkis, C. Timmerer, S. Baraković, J. B. Husić, S. Bech, S. Bosse, J. Botev, K. Brunnström, L. Cruz, K. D. Moor, A. de Polo Saibanti, W. Durnez, S. Egger-Lampl, U. Engelke, T. H. Falk, J. Gutiérrez, A. Hameed, A. Hines, T. Kojic, D. Kukolj, E. Liotou, D. Milovanovic, S. Möller, N. Murray, B. Naderi, M. Pereira, S. Perry, A. Pinheiro, A. Pinilla, A. Raake, S. R. Agrawal, U. Reiter, R. Rodrigues, R. Schatz, P. Schelkens, S. Schmidt, S. S. Sabet, A. Singla, L. Skorin-Kapov, M. Suznjevic, S. Uhrig, S. Vlahović, J.-N. Voigt-Antons and S. Zadtootaghaj, "QUALINET white paper on definitions of immersive media experience (IMEx)." European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online), May 25, 2020.

[6] Recommendation ITU-R M.2083-0, "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," 2015.

[7] M. Koziol, "The Metaverse Needs Standards, Too." IEEE Spectrum, 31 August 2022. https://spectrum.ieee.org/metaverse-standards-forum

[8] ISO/IEC 18039, "Information technology – Computer graphics, image processing and environmental data representation – Mixed and augmented reality (MAR) reference model," 2019.

[9] IEEE 1589, "IEEE Standard for Augmented Reality Learning Experience Models," 2020.

[10] Technical Report ITU-T GSTR-5GQoE, "Quality of experience (QoE) requirements for real-time multimedia services over 5G networks," 2022.

[11] V. K. M. Vadakital, A. Dziembowski, G. Lafruit, F. Thudor, G. Lee, and P. R. Alface, "The MPEG Immersive Video Standard — Current Status and Future Outlook." IEEE MultiMedia, vol. 29, no. 3, pp. 101-111, 2022.

[12] M. Montagud, P. Orero, and S. Fernández, "Immersive media and accessibility: hand in hand to the future," ITU Journal: ICT Discoveries, vol. 3 no. 1, 2020.

Table 1 – Organizations active in XR standardization covered in this paper

Organization	Membership information	Type	Working Group(s)/Committees with activities related to XR
3GPP	https://webapp.etsi.org/3gppmembership/QueryForm.asp	SDO	TSG RAN WG 1 (Radio Layer 1 (Physical Layer))
			TSG SA WG 2 (Architecture)
			TSG SA WG 4 (Multimedia Codecs, Systems and Services)
CTA	https://members.cta.tech/cta-member-directory	Trade association	AR/VR Committee
			XR WG
ETSI	https://www.etsi.org/membership	SDO	ISG ARF (Augmented Reality Framework)
IDEA	https://www.immersivealliance.org/members/	Industry consortium	Live Action WG
IEEE SA	https://standards.ieee.org/about/memberorgs/full-member-listing/	SDO	VR/AR Standards Committee (CTS/VRARSC)
			1857 WG (Audio Video Coding)
			P2048 WG (VR/AR)
			P2843 WG (Measuring Accessibility Experience and Compliance)
			P2874 WG (Spatial Web protocol, architecture and governance)
			2888 WG (Interfacing cyber and physical world)
			3079 WG (Human factors for immersive content)
			P7030 WG (Global XR ethics)
ISO/IEC JTC1	https://www.iso.org/committee/45020.html?view=participation	SDO	SC24 (Computer graphics, image processing and environmental data representation)
			SC29 (Coding of audio, picture, multimedia and hypermedia information)
ITU-R	https://www.itu.int/hub/membership/our-members/directory/?myitu-members-directory=true&request=search-companies&sector=ITU-R	SDO	SG5 (Terrestrial services)
			SG6 (Broadcasting service)
ITU-T	https://www.itu.int/hub/membership/our-members/directory/?myitu-members-directory=true&request=search-companies&sector=ITU-T	SDO	SG9 (Broadband cable & TV)
			SG11 (Protocols, testing & combating counterfeiting)
			SG12 (Performance, QoS & QoE)
			SG13 (Future networks)
			SG16 (Multimedia & digital technologies)
			SG20 (IoT, smart cities & communities)
			CG-Metaverse
Khronos Group	https://www.khronos.org/members/list	Consortium	3D Commerce
			glTF
			OpenGL
			OpenVX
			OpenXR
			Vulkan
			WebGL

Organization	Membership information	Type	Working Group(s)/Committees with activities related to XR
Metaverse Standards Forum	https://metaverse-standards.org/members/	Forum	
Qualinet	http://www.qualinet.eu/members/partner-institutions/	Research network	Task Force 7 (TF7): Immersive Media Experiences (IMEx)
SVTA	https://www.svta.org/svta-members/	Industry consortium	VR/360-Degree Video Study Group
VRIF	https://www.vr-if.org/members/	Forum	Guidelines WG
			Requirements WG
W3C	https://www.w3.org/Consortium/Member/List	SDO	Accessible Platform Architectures WG
			Immersive Web WG

Table 2 – XR standards and specifications covered in this paper

Organization	Working Group/Committee	Specification #	Specification title	Specification publication date (latest)
3GPP	TSG RAN WG 1	TR 38.838	Study on XR (Extended Reality) evaluations for NR	
	TSG SA WG 4	TR 26.862	Immersive Teleconferencing and Telepresence for Remote Terminals (ITT4RT) Use Cases, Requirements and Potential Solutions	
		TR 26.918	Virtual Reality (VR) media services over 3GPP	2022-05
		TR 26.928	Extended Reality (XR) in 5G	2022-05
		TR 26.929	QoE parameters and metrics relevant to the Virtual Reality (VR) user experience	2022-05
		TR 26.998	Support of 5G glass-type Augmented Reality / Mixed Reality (AR/MR) devices	2022-04
		TS 26.259	Subjective test methodologies for the evaluation of immersive audio systems	2022-04
		TS 26.260	Objective test methodologies for the evaluation of immersive audio systems	2022-05
CTA		CTA-2069-A	Definitions and Characteristics of Augmented and Virtual Reality Technologies	2020-11
		CTA-2085	Definitions and Characteristics for VR Video and VR Images	2019-11
		CTA-2095	Best Practices for Limited Mobility in XR	2021-05
ETSI	ISG ARF	GR ARF 002	Augmented Reality Framework (ARF); Industrial use cases for AR applications and services	2019-07
		GS ARF 003	Augmented Reality Framework (ARF); AR framework architecture	2020-03
		GS ARF 004-x series	Augmented Reality Framework (ARF); Interoperability Requirements for AR components, systems and services	2021
	ISG MEC	GS MEC 002	Multi-access Edge Computing (MEC); Phase 2: Use Cases and Requirements	2022-01
		GS MEC-IEG 004	Mobile-Edge Computing (MEC); Service Scenarios	2015-11
	TC HF	EG 202 848	Human Factors; Inclusive eServices for all: Optimizing the accessibility and the use of upcoming user-interaction technologies	2011-02
IDEA		ITMF Spec	Immersive Technologies Media Format specification suite	2021-12
IEEE SA	1589 WG	1589	Standard for Augmented Reality Learning Experience Model	2020-04
	1857 WG	1857.9	Standard for Immersive Visual Content Coding	2022-03

Organization	Working Group/Committee	Specification #	Specification title	Specification publication date (latest)
	P2048 WG	P2048 series	Standard for Virtual Reality and Augmented Reality	
	P2843 WG	P2843	Standard for Measuring Accessibility Experience and Compliance	
	P3079 WG	3079	HMD based VR Sickness Reducing Technology	2021-04
	P3333.1 WG	P3333.1 series	Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices based on Human Factors	
		IEEE 3333.1.1	Standard for Quality of Experience (QoE) and Visual-Comfort Assessments of Three-Dimensional (3D) Contents Based on Psychophysical Studies	2015-07
		IEEE 3333.1.2	Standard for the Perceptual Quality Assessment of Three-Dimensional (3D) and Ultra-High-Definition (UHD) Contents	2017-12
		IEEE 3333.1.3	Standard for the Deep Learning-Based Assessment of Visual Experience Based on Human Factors	2022-05
ISO/IEC	JTC1/SC24	18038	Information technology — Computer graphics, image processing and environmental representation — Sensor representation in mixed and augmented reality	2020-04
		18039	Information technology — Computer graphics, image processing and environmental data representation — Mixed and augmented reality (MAR) reference model	2019-02
		18040	Information technology — Computer graphics, image processing and environmental data representation — Live actor and entity representation in mixed and augmented reality (MAR)	2019-05
		18520	Information technology — Computer graphics, image processing and environmental data representation — Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR)	2019-01
		19775 series	Information technology — Computer graphics, image processing and environmental data representation — Extensible 3D (X3D)	
		19776 series	Information technology — Computer graphics, image processing and environmental data representation — Extensible 3D (X3D) encodings	
		23488	Information technology — Computer graphics, image processing and environment data representation — Object/environmental representation for image-based rendering in virtual/mixed and augmented reality (VR/MAR)	2022-05
	JTC1/SC29	21794 series	Information technology — Plenoptic image coding system (JPEG Pleno)	
		23000-13	Information technology - Multimedia application format (MPEG-A) — Part 13: Augmented reality application format	2017-11
		23005 series	Information technology — Media context and control	
		23090 series	Information technology — Coded representation of immersive media	
ITU-R	SG5	M.2083-0	IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond	2015-09
	SG6	BT.2123	Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting	2019-01
		BT.2133	Transport of advanced immersive audio visual content in IP-based broadcasting systems	2019-10

Organization	Working Group/Committee	Specification #	Specification title	Specification publication date (latest)
ITU-T	SG9	J.301	Requirements for augmented reality smart television systems	2014-10
		J.302	System specifications of augmented reality smart television service	2016-10
		J.1631	Functional requirements of E2E network platforms to enhance the delivery of cloud-VR services over integrated broadband cable networks	2021-11
	SG11	Q.4066	Testing procedures of augmented reality applications	2020-09
	SG12	G.1035	Influencing factors on quality of experience for virtual reality services	2021-11
		G.1036	Quality of experience influencing factors for augmented reality services	2022-07
		GSTR-5GQoE	Quality of experience (QoE) requirements for real-time multimedia services over 5G networks	2022-06
		P.919	Subjective test methodologies for 360º video on head-mounted displays	2020-10
		P.1320	Quality of experience assessment of extended reality meetings	2022-07
	SG13	Y.3109	Quality of service assurance-related requirements and framework for virtual reality delivery using mobile edge computing supported by IMT-2020	2021-04
	SG16	F.740.2	Requirements and reference framework for digital representation of cultural relics and artworks using augmented reality	2021-06
		H.430.1	Requirements for immersive live experience (ILE) services	2018-08
		H.430.2	Architectural framework for immersive live experience (ILE) services	2018-08
		H.430.3	Service scenario of immersive live experience (ILE)	2018-08
		H.430.4	Service configuration, media transport protocols, signalling information of MPEG media transport for immersive live experience (ILE) systems	2019-11
		H.430.5	Reference models for immersive live experience (ILE) presentation environments	2020-08
Khronos Group	3D Commerce WG		Real-time Asset Creation Guidelines	2020-10
	3D Formats WG		glTF Specification	2021-10
	OpenXR WG		OpenXR API Specification	2019-07
SVTA	VR/360-Degree Video Study Group		eXtended Reality Brief	2022-03
VRIF	Guidelines WG	VRIF 2.3	VR Industry Forum (VRIF) Guidelines 2.3	2021-01
W3C	APA WG	NOTE-xaur	XR accessibility user requirements	2021-08

SESSION 1

SOME PERSPECTIVES ON FUTURE NETWORK

INTEGRATED NETWORK CONTROL ARCHITECTURE FOR TERRESTRIAL AND NON-TERRESTRIAL NETWORK CONVERGENCE IN BEYOND 5G SYSTEMS

Ved P. Kafle; Mariko Sekiguchi; Hitoshi Asaeda; Hiroaki Harai

National Institute of Information and Communications Technology, Japan

ABSTRACT

To seamlessly cover urban and rural areas, mountains and deserts, as well as sea and air, with high quality ubiquitous communication services, the convergence of Terrestrial Networks (TNs) and Non-Terrestrial Networks (NTNs) such as satellites is essential in beyond 5G communication systems. This paper presents the scenarios and architectures of integrated and individual network control systems for TN and NTN convergence in the control plane. The integrated network control system enables end-to-end service design, orchestration, performance monitoring, closed-loop control, and automation. It promotes interoperability in the control plane of TN and NTN domains composed of virtualization-supporting infrastructures and possibly managed by different organizations. This paper is related to ITU-T Study Group 13's activities of standardizing fixed, mobile and satellite convergence technologies.

Keywords— Extended reality, integrated network control, interoperability, non-terrestrial networks, quality of experience, terrestrial and non-terrestrial network convergence,

1. INTRODUCTION

Fifth Generation (5G) mobile networks, which are currently being deployed worldwide, have the capability of connecting a massive number of devices and offering ultra-reliable, low latency, mobile broadband services over sharable, virtualized communication and computational infrastructure [1]. They offer user-customized, high-quality services required for emerging applications based on novel technologies of High Definition (HD) videos, Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR). However, 5G mobile network services are largely confined to highly populated urban areas and cannot cover 100% of the globe because of economic and geographical difficulties. The deployment of 5G cellular services in remote and rural areas with a very low population density is not lucrative as it would result in unnecessarily higher costs per user. Moreover, it is difficult to deploy 5G terrestrial radio infrastructure in deserts and mountainous regions. Therefore, to seamlessly cover urban and rural areas, mountains and deserts, as well as sea and air, the convergence of Terrestrial Networks (TNs), such as cellular and data networks, and Non-Terrestrial Networks (NTNs) such as satellites is essential in 5G and Beyond 5G (B5G) systems [2].

TN and NTN convergence is required not only for covering rural, remote, deserts, mountains, sea, and airspace by a high-quality ubiquitous communication service but also keeping the service intact when natural disasters (e.g. earthquake, tsunami, forest fire and floods) hit the earth and damage TN infrastructures. Thus, TN and NTN convergence would enhance the robustness of communication systems and make them dependable for speedy public rescue and disaster risk reduction.

In this paper, we present a high-level architecture framework of integrated network control systems for the realization of TN and NTN convergence. Various segments of a TN such as wireless local area networks, cellular mobile networks, and data networks are already interconnected by standard technologies. However, NTN components such as satellites are still working as isolated communication systems. Satellites stationed at different altitudes such as Geostationary Earth Orbit (GEO), Medium Earth Orbit (MEO) and Low Earth Orbit (LEO), together with recently deployed Unmanned Aerial Vehicles (UAV) such as drones and High-Altitude Platform Systems (HAPS) are going to be key components of NTN systems.

Research and development of novel technologies enabling TNs and NTNs convergence has started recently in various countries. However, there is no standard technology developed yet for the convergence and interoperability of TNs and NTNs. Therefore, various Standards Development Organizations (SDOs) such as the International Telecommunication Union (ITU), 3rd Generation Partnership Project (3GPP), and European Telecommunications Standards Institute (ETSI) have recently started studying the various issues of TN and NTN convergence.

The proposed integrated network control architecture is expected to contribute to the convergence of TNs and NTNs in the control plane. It enables the design, orchestration, monitoring, and control of end-to-end network service across TN and NTN domains or segments, which may be composed of equipment produced by different manufacturers and operated by different organizations. It empowers network service providers to offer high quality services and seamless mobility from a TN to NTN and vice versa.

The remainder of this paper is organized as follows. Section 2 discusses related research, development, and standardization activities. Section 3 presents scenarios of TN and NTN convergence in the control plane. Section 4 describes the functional architecture of integrated and individual network control systems for TN and NTN convergence. Section 5 concludes the paper by listing future work items.

2. RELATED WORK

Non-terrestrial networks, composed of satellites, HAPS and drones, will have the capability of offering communication services on a large scale because of their wide coverage around the world and robust multilink transmission capability. High-Throughput Satellites (HTS) are being developed for the extension of an urban 5G system's small cell coverage to rural and remote areas. Although satellite communications in the past used to be for specific purposes such as television broadcasting or international telephone trunk lines, recent deployments of very high throughput GEO satellites and LEO satellite constellations have extended their usage to general purpose telephony and data communication services [3].

Worldwide efforts of research, development, and standardization of TN and NTN convergence technologies have been progressing rapidly. The European Commission (EC) has recently funded several projects to investigate TN and NTN integration issues and develop suitable solutions. The project named "Satellite and terrestrial networks for 5G (SaT5G)" [2] has investigated the extension of Network Function Virtualization (NFV) and Software-Defined Networking (SDN) capabilities to satellite network configuration and management. It has proposed technology for seamlessly integrating NTN into the 5G Management And Orchestration (MANO) system, while addressing the issues of multi-links, heterogeneous transport, control and user planes harmonization, security, caching, and multicast. Similarly, another project named "5G agile and flexible integration of satellite and cellular (5G- ALLSTAR)" has investigated the issues of multiple access when combining cellular and satellite access technologies for reliable, ubiquitous broadband services [4]. More specifically, it has explored a set of technologies such as new radio-based satellite access, multi-connectivity, and spectrum sharing between cellular and satellite systems to provide global connectivity and support mission critical applications. Similarly, the project named VITAL (Virtualized hybrid satellite-terrestrial systems for resilient and flexible future networks) has studied the federated resource management of hybrid satellite-terrestrial networks by applying NFV and SDN in satellite networks [5]. It has studied three scenarios of virtualization and sharing of satellite communication resources, 5G backhauling services, and federated satellite-terrestrial access services.

The European Space Agency (ESA) funded project "Demonstrator for satellite-terrestrial integration in the 5G context (SATis5)" has aimed to build a large-scale, live, proof-of-concept testbed of 5G integrated satellite-terrestrial networks [6]. The testbed is used to demonstrate the applicability of satellite technology to realize the 5G use cases of enhanced Mobile Broadband (eMBB) and massive Machine-Type Communication (mMTC) services.

Research projects related to TN and NTN convergence are also progressing in Japan. In the Beyond 5G/6G whitepaper published by the National Institute of Information and Communications Technology (NICT) recently [8], TN and NTN convergence has been considered a key technological innovation for serving the future society of 2030s and beyond. TN and NTN convergence has been considered indispensable for the realization of three novel use-case scenarios of the future society, namely, cybernetics avatar society, city on the moon, and transcending space and time. Similarly, the Space ICT Promotion Initiative Forum (SPIF) [9] has been promoting the multifaceted development of space-related communication technologies and applications by establishing a close collaboration between industry, academia, and government. It is channeling Japan's collective efforts on sharing information on the latest technological and application development trends, identifying cooperation/competition areas, and examining various strategies.

SDOs are also actively progressing the standardization of TN and NTN convergence technologies. In ITU, the scenarios of satellite access integration with terrestrial networks were discussed by using the term *ManyNet* under the activities of Focus Group on Technologies for Network 2030 (FG-NET2030) [10]. Recently, ITU-T Study Group 13 has started the study of requirements, capabilities, interfaces, service continuity, and procedures for fixed, mobile and satellite convergence in networks beyond IMT-2020 [11].

3GPP has published several related technical reports. 3GPP TR 22.822 presents a study on using satellite access in 5G systems [12]. It describes twelve use cases by categorizing them into three groups (i.e. service continuity, ubiquity, and scalability) and drawing several requirements from each use case. 3GPP TR 28.802 discusses the management aspect of satellites in 5G [13]. It identifies the issues associated with service and network management of satellite integrated 5G and presents a reference architecture for management and the associated solutions. 3GPP TR 23.737 presents a study on architecture aspects for satellite access in 5G [14]. It has identified the impact areas and solutions (including procedures) for resolving radio access and core network related issues. Similarly, ETSI TR 103 611 presents the various integration scenarios of NTN in 5G systems and related architecture options [15].

Several academic research papers published recently have also discussed the diverse aspects of TN and NTN convergence. A comprehensive survey of the state-of-the-art of integrated satellite-terrestrial networks toward 6G has been presented in [16]. It describes various kinds of simple integration architectures at the physical component levels and their typical applications, characteristics, and challenges. Similarly, [17] has investigated the problem of inter-domain routing between the ground segments of satellite networks and traditional terrestrial Internet Service Provider (ISP) networks.

This work complements the above related work by proposing an umbrella architecture of integrated network control system for TN and NTN convergence in the control plane. It provides an architectural framework in which various functions can be integrated to make it useful for the design and orchestration of high-quality, reliable end-to-end communication services, and effectively control the communication and computational resources of both TN and NTN segments on the basis of end-to-end service performance.

3. SCENARIO OF TN AND NTN CONVERGENCE

In this section, we discuss the scenarios of TN and NTN convergence and their integrated network control system.

3.1 Scenario of integrated network control

Figure 1 shows an image of terrestrial and non-terrestrial networks collectively monitored and controlled from the integrated network control system in the control plane. The targeted integrated network control system would enable monitoring the resource utilization and performance metrics of various network segments and controlling of their resources dynamically to provide high-quality, end-to-end communication services. It will enable ultra eMBB, ultra mMTC, and Ultra-Reliable Low Latency Communication (URLLC) services available anywhere and at any time. It will also enable the provisioning of emerging applications such AR/VR, Immersive Living (IL), automated driving, and other smart control systems by connecting billions of IoT devices and robots to the communication system.

Each network segment or network domain (e.g. B5G cellular network, data network, GEO/MEO/LEO satellites, and HAPS) is controlled in the control plane by its own control system. The control plane obtains monitoring data (i.e. control data) from the underlying network elements through network control internal interfaces. The control data is collected, stored, updated, and shared by the control data service function. The control data service function provides control data to the network control functions. The network control functions process control data to assess the current states of network elements and their performances, predict their future states, and formulate control decisions so that the network can meet service performance requirements while optimally utilizing available computational (i.e. CPU, GPU, and memory) and storage resources as well as communication resources such as

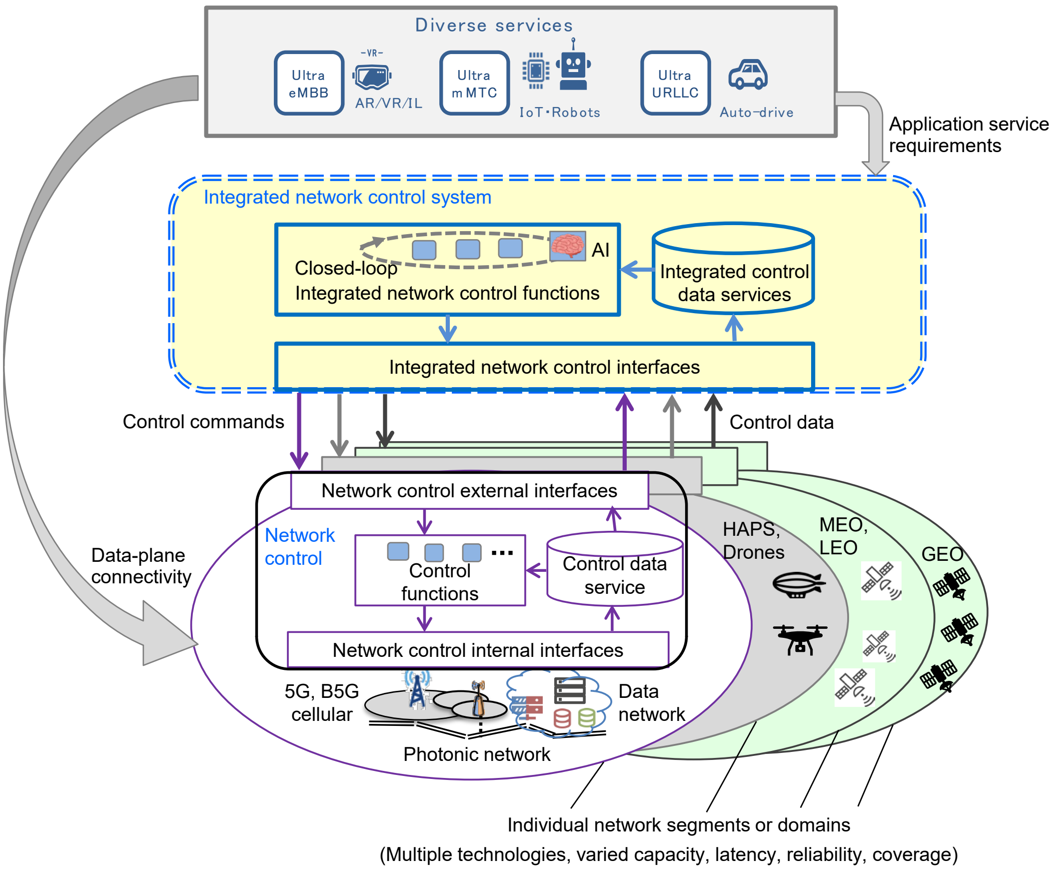
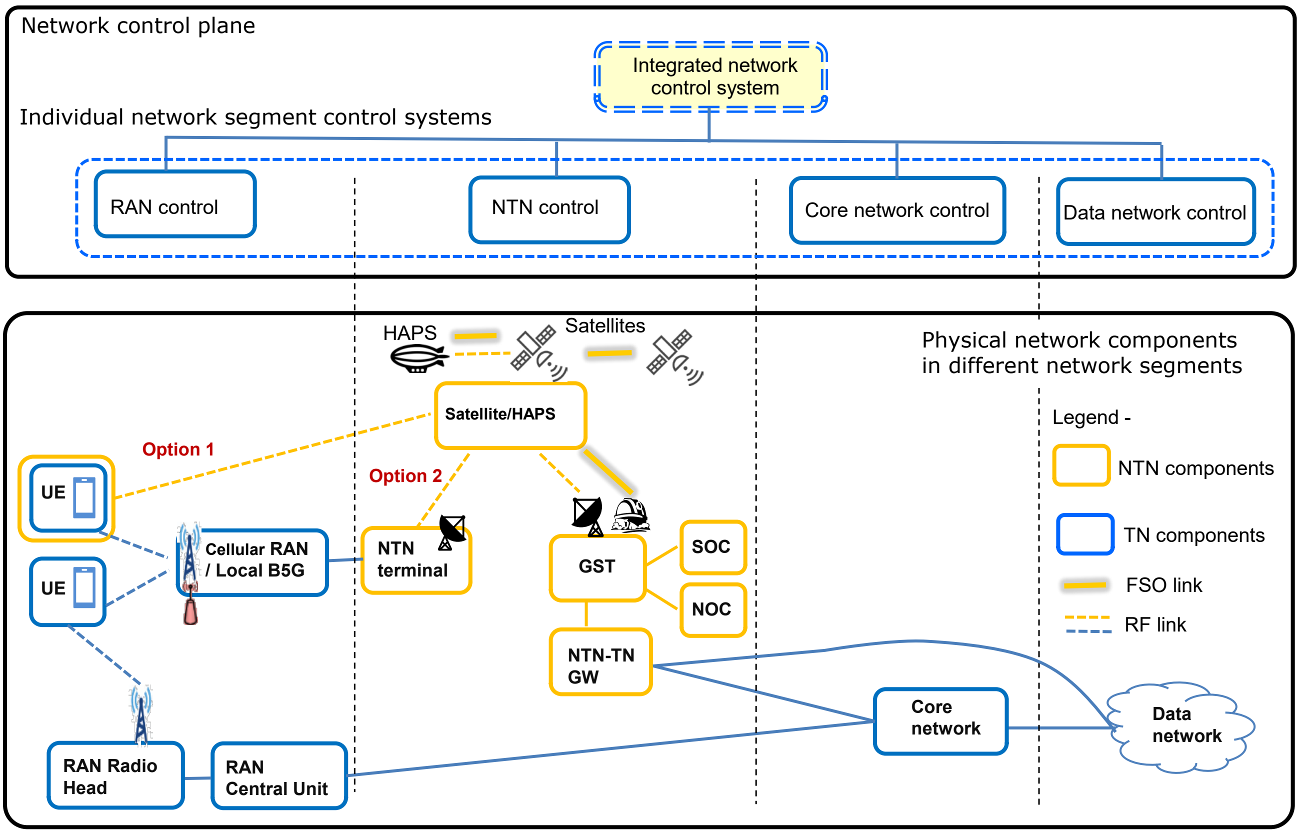


Figure 1 - Scenario of TN and NTN convergence through integrated network control system



HAPS: High-altitude platform system, SOC: Satellite operation center, NOC: Network operation center, GST: Ground station, GW: Gateway, RAN: Radio access network, UE: User equipment, FSO: Free-space optical, RF: Radio frequency

Figure 2 – Physical network and control plane components of TN and NTN convergence

Radio Frequencies (RFs) and Free Space Optical (FSO) communication wavelengths. The control decisions are given to underlying network elements via network control internal interfaces. The individual network segment's control plane also possesses network control external interfaces, which provide control data to and obtain control commands from the integrated network control system.

The integrated network control system is mainly composed of three sets of functional components: integrated network control interfaces, integrated control data service, and integrated network control functions. The integrated network control interfaces possess the capabilities to collect control data from the heterogeneous types of individual network segments and to provide network control commands to them. They can collect monitoring data from an individual network segment's control plane either on an on-demand basis (i.e. by sending a request and receiving the data in response) or on a subscription basis (i.e. by subscribing the data of a specific event or topic by using a publish-subscribe paradigm and regularly obtaining the relevant data). Control data is managed (i.e. collected, stored, updated, processed, and shared) by the integrated control data service function. It provides control data to the integrated network control functions. The integrated network control functions process control data using closed loop operations of advanced data processing techniques such as artificial intelligence and machine learning models and assess the end-to-end network's status in terms of resource availability, utilization, and performance. They predict the future states of networks and resource demands of the service. They then make appropriate control decisions (e.g. network reconfiguration or adjustment of resources) while fulfilling the requirements of keeping the performance at a desired level, optimally utilizing available resources, and maintaining the operational overheads low.

3.2 Physical network and control plane components

Figure 2 illustrates the components of various network segments and their control planes.

Physical networks are data networks, cellular core networks, NTN, and access networks. Data networks are composed of the Internet, cloud computing infrastructure, data storage, and application servers. The cellular core network is composed of various functions for user authentication, mobility management, resource management, session management, etc. of cellular networks as defined and specified in 3GPP standards. Similarly, NTN contains satellites and HAPS in space and Satellite Operation Center (SOC), Network Operation Center (NOC), Ground Stations (GSTs), Non-Terrestrial Network and Terrestrial Network Gateways (NTN-TN GWs), and NTN terminals on the earth. The RF or FSO links are used to communicate between the GST and satellites/HAPS or among satellites and HAPS.

Similarly, NTN-TN GWs can be connected to the cellular core network or directly to data networks if cellular core functions are not required by the radio access network connected to NTN terminals.

The access network can be a fixed access network (e.g. fiber-to-the-home) to connect fixed terminals (e.g. computers and smart devices in home or office) or a radio access network (e.g. cellular WAN consisting of RAN central units and radio heads) to connect mobile user equipment (e.g. smartphones). The fixed access networks (not shown in the figure) are usually connected to data networks, while RAN networks are connected to the cellular core network through high-speed, multi-core optical fiber networks [7] so that seamless mobility of mobile users from a TN to an NTN can be supported. Similarly, the cellular RAN including local B5G/6G access networks and various types of wireless access points installed on the ground, airplanes or ship can be connected to NTN though an NTN terminal such as a Very Small Aperture Terminal (VSAT) acting as a gateway between the terrestrial RAN and NTN. 5G New Radio (NR) UEs also have capabilities of connecting to satellites directly.

In the control plane, each network segment is managed by its own network control system. Although different kinds of NTN networks (e.g. GEO satellites, LEO satellites, and HAPS) are shown collectively managed by a single NTN control system in Figure 2, each NTN segment with airspace components stationed at a given altitude may be controlled and managed by a distinct network control system.

4. INTEGRATED NETWORK CONTROL SYSTEM

In this section, we describe the functional architecture of integrated network control system and its components.

4.1 Architecture of integrated network control system

Figure 3 shows the functional architecture of an integrated network control system. This figure extends the network control plane shown in Figure 1. The major components of integrated network control architecture are: (1) Integrated network control interfaces, (2) Integrated control data service, (3) End-to-end (E2E) network status analysis, (4) Application requirements and user intent analysis, (5) E2E resource allocation, management, and optimization, and (6) Integrated control and orchestration functions.

The integrated network control interfaces communicate with the network control external interfaces of individual network segments (e.g. radio access, NTN, cellular core, and data networks) to collect control data from them and send control commands to them. The interfaces also contain functions to collect control data in different granularity of time scales and details as required. There should be provisions to adjust the granularity and frequency of control data generation and transmission so that the control system overhead will not overwhelm the integrated network control system performance. The control interfaces can be based on flexible and extendable standard protocols (e.g. RESTful API), which can be easily adapted to efficiently communicate with individual network control systems.

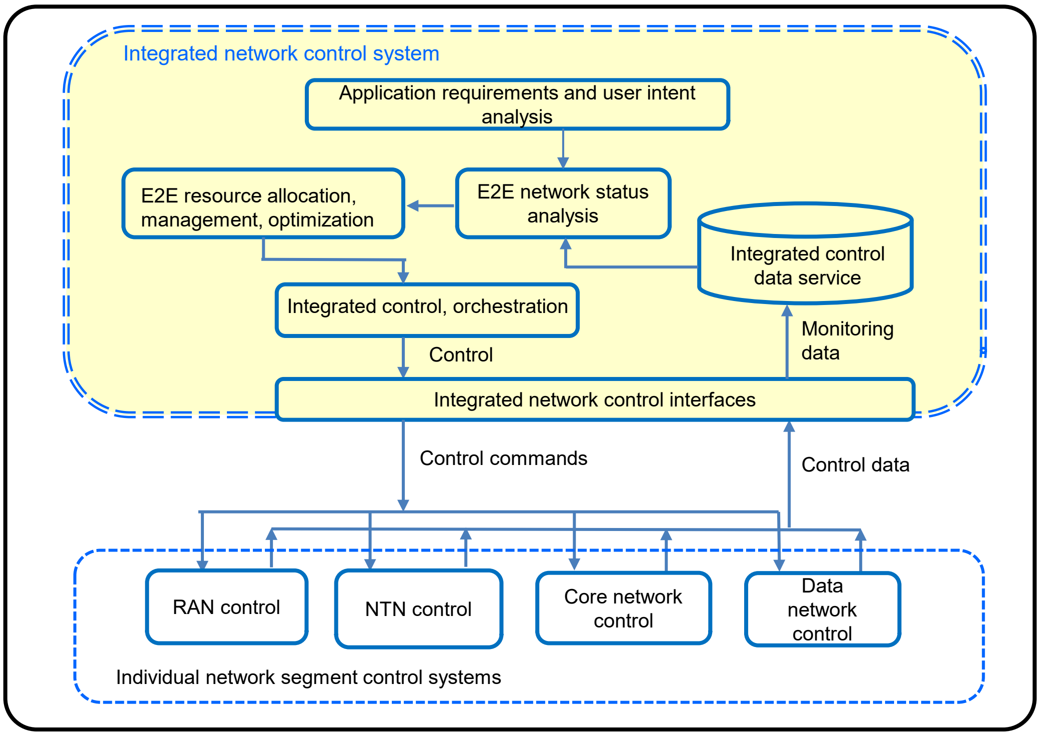


Figure 3 - Functional architecture of integrated network control system

The collected control data from various network segments are managed by the integrated control data service functions. The data service functions are responsible for checking the data consistency and providing control data in dynamically adjustable granularity to the E2E network status analysis function through on-demand and streaming data services.

The E2E network state analysis function processes integrated control data to assess the status of end-to-end network services by analyzing the status of resource utilization and performances of all involved network segments. It also uses the input given by the application requirements and user intent analysis component to judge the network service quality with respect to service level agreements. The application requirements and user intent analysis component includes necessary functionalities to interpret the user and service provider expectations expressed in abstract policies (called intents) and convert them into the control system-understandable parameters that can be mapped into well-defined service configuration templates.

The E2E resource allocation, management and optimization functions formulate optimal control decisions on the basis of the current and predicted future states of network segments involved in an end-to-end communication service, together with the prediction of available resources and their demands. Different kinds of network segments can have different kinds of resources on demand. For example, data networks may demand more computational (e.g. virtual machines, CPU, GPU, memory) and storage resources, while the NTN may demand more RF bandwidths for feeder and service links or replacement of RF links by FSO links due to changes in weather conditions and user traffic demands. Different combinations of resource allocation strategies can be considered as feasible candidate solutions and evaluated their fitness by using mathematical optimization models and/or machine learning models. The optimal resource allocation and control decisions are provided to the integrated control and orchestration function.

The integrated control and orchestration function executes the control decision by generating a list (or configuration file) of appropriate resource control commands for each network segment involved in the end-to-end network service and sending the command lists to all network segments through integrated network control interfaces.

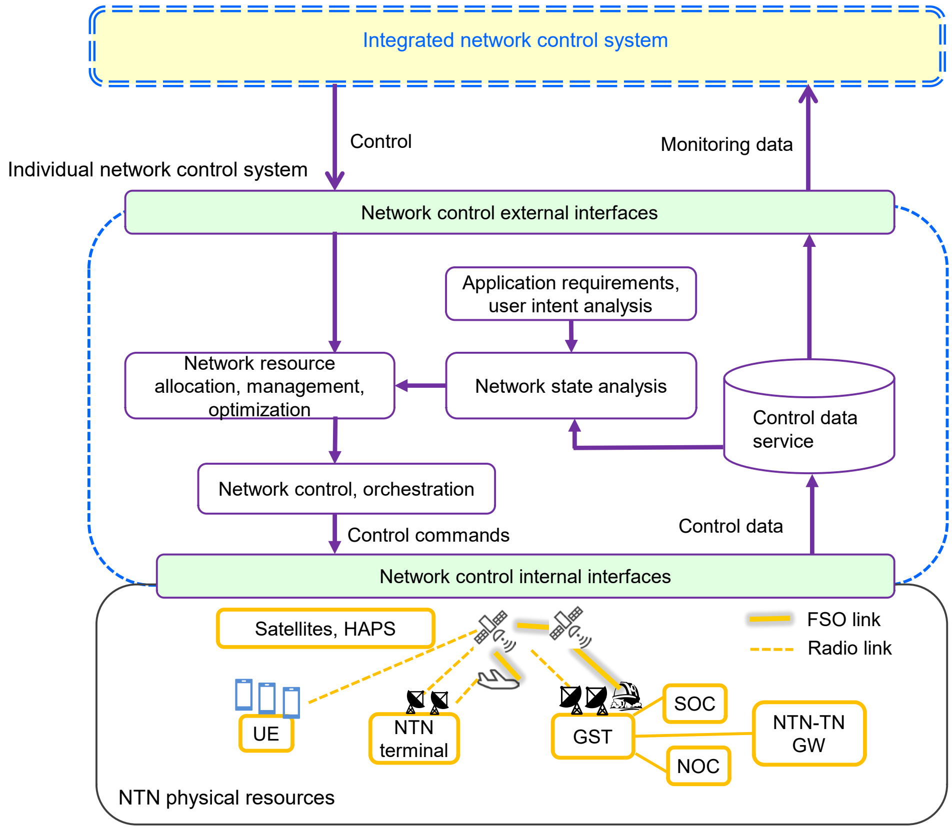


Figure 4 - Functional architecture of individual network segment

4.2 Control system architecture of individual network segment (assuming NTN)

Figure 4 shows the functional architecture of an individual network segment's control system, assuming the NTN physical components as the controlled elements. The individual network control system's functional architecture components are similar to the functional components of integrated network control architecture. The major difference is that the scope of each function is not E2E network service, but within the given network segment.

The network control internal interfaces are used for receiving monitoring/control data from the physical or functional network elements in different granularity and time scale. Both the types and values of control data may differ in different kinds of network segments. Standard frameworks and methods for control data collection in terrestrial networks have been defined by various SDOs such as ITU, IETF, 3GPP, and ETSI. For example, control data collection and transmission functions of terrestrial 5G networks are well defined in 3GPP technical specifications on architecture enhancement to support network data analytics, such as 3GPP TS 23.288 [19]. They can collect control data based on subscription to events occurring in network functional elements, such as the User Plane Function (UPF), Session Management Function (SMF), and Access and Mobility management Function (AMF), and from local data repositories collocated with individual functional elements. Similarly, for NFV and SDN-based virtualized data networks, the control data collection methods and interfaces have been specified in detail in ETSI NFV standards [20]. However, NTN network segments still lack standard frameworks of network control internal interfaces for collecting control data and issuing control commands from remotely. The concept and approaches used in TN data collection can be gradually introduced into NTN segments as the new generation of NTN space and ground components also support NFV and SDN.

NTN control data can be roughly divided into the following four groups: (1) preset system parameters, (2) service/user requirements, (3) system monitoring parameters, and (4) NTN link control parameters. The preset system parameters such as satellite orbit coordinates and number of satellites (in a constellation), coordinates and number of GSTs, number of communication interfaces, their types (e.g. physically/logically connected, FSO or RF links, transparent or regenerative, and antenna patterns), beam coverage size and numbers, fixed or movable beam types, total bandwidth, and total power are static parameters. Similarly, service/user requirements such as service type (e.g. voice, video, browsing, and sensors data), and QoS requirements (e.g. minimum traffic rate, tolerable call drop or packet loss rate, maximum latency, and security levels) are also static parameters. System monitoring parameters are dynamic parameters such as the number of active user terminals, resource (RF and FSO) utilization, weather conditions, Doppler shift, and performance (e.g. throughput, latency, jitter, packet loss and call drop rates). NTN link control parameters such as beam RF bands or FSO wavelengths and coding rate (with or without encryption) are dynamic parameters.

In current NTN systems, this control data is available in different components. For example, satellite coordinates, satellite conditions, and inter-satellite link quality can be obtained from the SOC. GST availability, weather conditions, feeder and service link quality, and GST interface measurements (e.g. frequencies, Doppler shift, modulation methods and transmission speeds) can be obtained from the GST. Similarly, service/user requirements, service types and their QoS requirements (e.g. traffic rate, tolerable call drop, maximum latency, and security levels), and link performances for payload transmission (e.g. throughput, latency, jitter, and packet loss rates) can be available from the NOC. Therefore, the SOC, GST, and NOC should be able to provide telemetry data to the NTN control system and receive telecommands from there through standard network control internal interfaces.

The control data service function contains data repository and related functions for maintaining and providing control data to both the internal network state analysis function and external integrated network control system through the network control external interfaces. As in the integrated control system, the network state analysis function of the individual network control system can apply machine learning and other data mining techniques to process the control data, together with service requirements and user intent parameters, and assess the network status (e.g. resource utilization, performances, anomaly detection, fault and failure prediction). The control data is provided to the integrated network control system based on subscriptions, whose formats are negotiated in advance.

As stated earlier, the integrated network control system processes data obtained from TN and NTN network segments to assess the E2E network service performance and make network control decisions. The network control decisions are fed to the individual network control system's network resource control, management and optimization function. The network resource allocation, management, and optimization function also receives the network state analysis results from the local network state analysis function. It uses both the E2E network control decision made by the integrated network control system and local network state to make an appropriate decision to control its network segment. The decisions are given to the network control and orchestration function.

The network control and orchestration functions generate control commands and send them to the related controlled elements. For example, the control commands in NTN can be for increasing bandwidth allocation to a beam, changing beam projection angles, selection of a different GST, switching RF links to FSO links, rerouting of traffic in satellite constellation or satellite-HAPS networks, and so on. The difference between the global network control and orchestration functions of the integrated network control system and local individual network segment control system is that the global network control commands can be generated in an abstract format while the local control commands are generated in the form of a given network segment technology-specific syntax. For example, the NTN control and orchestration function should be able to generate telecommands that can be directly executed by the NOC and SOC, without requiring any transcoding.

4.3 Features

The architecture presented in this section possesses the following features, which play crucial roles in addressing several challenges of TN and NTN convergence, mainly promoting the existence of multiple NTN operators and empowering richer end-to-end service provisioning than an individual network operator can offer by using only its own resources.

a. End-to-end network control and monitoring

In order to offer highly reliable and dependable ubiquitous communication services, while optimally utilizing available computational and communication resources, end-to-end network service control and management is essential. This architecture integrates different network segments in the control plane and enables their smooth convergence in the data plane for offering high-quality communication services.

b. End-to-end network resource sharing

Network resource sharing has been essential for providing cost-effective communication services. Thanks to NFV and SDN technologies and standardized common interfaces, communication and computational resource sharing in virtualized data networks and cloud computing infrastructures has become a common practice recently. In the near future, NTN will also have to provide similar types of resource sharing capabilities so that satellite operators can share their NTN resources (e.g. ground stations and satellite interfaces) with many TN operators and application service providers through standard control interfaces. This architecture is aimed at the development and standardization of such resource sharing interfaces and functions.

c. Unified representation of resources

Different network segments or domains have different kinds of resources in their focus for control and management. TN virtualized data networks are in focus of the efficient management of computational and storage resources while NTN are in focus of the efficient management of RF link and GST resources. These resources and their capabilities should be represented by unified and standard expressions so that they can be unambiguously identified and interpreted by the integrated network control system.

d. Technology-agnostic control operation

For TN and NTN seamless convergence, technology-agnostic control operations are essential. This architecture provides a framework in which various TN and NTN domains can be plugged in through standard interfaces. The availability of technology-agnostic control systems and open interfaces will promote free competitions among equipment manufacturers and vendors and attract new startups to participate in innovative services creation and marketing.

e. Promoting network control automation

Network control and management automation is essential to deal with the complexity of TN and NTN convergence. This architecture framework supports the development of data-oriented, closed-loop control mechanisms, leading to the automation of individual and integrated network control functions.

5. CONCLUSION AND FUTURE WORK

This paper presented the scenarios and architectures of integrated and individual network control systems for TN and NTN convergence. The integrated control system architecture can accommodate various control functions for the design and orchestration of end-to-end network services, their performance monitoring, and resource control of both TNs and NTNs. The content of this paper is related with ITU-T Study Group 13, which is standardizing technologies for the fixed, mobile and satellite convergence. It is also related with activities of other SDOs such as 3GPP and ETSI.

In future work, we will extend this work with the detail design of each functional component of the integrated network control system architecture by making it compliant with ETSI Zero-touch network and Service Management (ZSM) [18] and Open Source NFV Management and Orchestration (MANO) [20]. We will develop an experimental system for its validation and performance evaluation. We will also bring this research outcome to ITU for standardization.

Acknowledgement

The authors are grateful to their colleagues Yuma Abe, Kentaro Ishizu, Amane Miura, and Morio Toyoshima for actively supporting this work.

REFERENCES

[1] V. P. Kafle, P. Martinez-Julia, and T. Miyazawa, "Automation of 5G network slice control functions with machine learning," *IEEE Communications Standards Magazine*, vol. 3, no. 3, pp. 54-62, 2019.

[2] K. Liolis et al., "Use cases and scenarios of 5G integrated satellite-terrestrial networks for enhanced mobile broadband: The SaT5G approach," *International Journal of Satellite Communications and Networking*, vol. 21, Issue 4, pp. 1-22, Nov. 2017.

[3] "Transcending connectivity," *SES White Paper*, Sep. 2017.

[4] 5G ALL-STAR – "5G AgiLe and fLexible integration of SaTellite And cellular," *EC Horizon 2020 Project*, https://cordis.europa.eu/project/id/815323, visited on 15th June 2022.

[5] VITAL – "Virtualized hybrid satellite-terrestrial systems for resilient and flexible future networks," *EC Horizon 2020 Project*, https://cordis.europa.eu/project/id/644843, visited on 15th June 2022.

[6] "Satellite is 5G, SATis5 – Demonstrator for Satellite-Terrestrial Integration in the 5G Context," *Whitepaper, The European Space Agency*, Nov. 2020.

[7] B. J. Puttnam, G. F. Rademacher, and R. S. Luis, "Space-division multiplexing for optical fiber communications," *OSA Journal Optica*, vol. 6, no. 9, pp. 1186-1203, Sep. 2021.

[8] "Beyond 5G / 6G White Paper," English version 0.9, National Institute of Information and Communications Technology, Apr. 2022. https://beyond5g.nict.go.jp/en/, visited 15th June 2022.

[9] Space ICT Promotion Initiative Forum (SPIF), Japan. https://spif.nict.go.jp/, visited on 15th June 2022.

[10] "Network 2030 Architecture Framework," Technical Specifications, FG-NET2030 – Focus Group on Technologies for Network 2030, ITU-T, June 2020.

[11] ITU-T Study Group 13 – Future Networks and Emerging Network Technologies, https://www.itu.int/ITU-T/go/sg13, visited on 15th June 2022.

[12] 3GPP TR 22.822 (V1.0.0), "Study on using Satellite Access in 5G Stage 1 (Release 16)," May 2018.

[13] 3GPP TR 28.808 (V17.0.0), "Study on management and orchestration aspects of integrated satellite components in a 5G network (Release 17)," Mar. 2021.

[14] 3GPP TR 23.737 (V17.2.0), "Study on architecture aspects for using satellite access in 5G (Release 17) ," Mar. 2021.

[15] ETSI TR 103 611 (V1.1.1), "Seamless integration of satellite and/or HAPS (High Altitude Platform Station) systems into 5G and related architecture options," *Satellite Earth Stations and Systems (SES)*, June 2020.

[16] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications and challenges," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 437-461, 2022.

[17] G. Guiliari et al., "Internet backbone in space," *ACM SIGCOMM Computer Communications Review*, vol. 50, no. 1, pp 25-37, 2020.

[18] ETSI GS ZSM 002 (V1.1.1), "Zero-touch network and Service Management (ZSM), Reference Architecture," Aug. 2019.

[19] 3GPP TS 23.288 (V17.5.0), "Architecture enhancements for 5G System (5GS) to support network data analytics services (Release 17)," June 2022.

[20] ETSI Network Functions Virtualisation (NFV), https://www.etsi.org/technologies/nfv, visited on 15th June 2022.

TOWARDS COMPUTING AND NETWORK CONVERGENCE: QOE-ORIENTED SERVICE ANYCAST BASED ON SRV6

Zicheng Wang; Yetong Wang; Xian Gao; Shuai Wu; Wei Lin

Inspur Communications Technology Co., Ltd., China

ABSTRACT

The new mobile Internet services represented by extended Reality (XR) have brought new challenges to computing and networks. A new generation of Information and Communication Technology (ICT) infrastructure with both computing and communication capabilities is being promoted. Traditional load balancing technologies cannot sense the connection between users and services. It is also difficult to support large-scale distributed computing service instances, so the service experience provided can be poor.

In this paper a service anycast system based on a Segment Routing over IPv6 (SRv6) data plane is proposed, which can provide a Quality of Experience (QoE)-oriented load balancing capability for the network according to the obtained status information of network and computing service instances. Besides, an IP address identification mechanism is also proposed to help the control plane handle traffic engineering policies, and to support efficient service resource discovery, seamless mobility, and service continuity.

Keywords – Anycast, Computing and Network Convergence (CNC), load balance, SRv6

1. INTRODUCTION

With the global promotion and deployment of 5G networks, the large bandwidth and low latency wireless connectivity provided has brought new opportunities for the development of interactive multimedia applications [1]. New immersive and interactive applications represented by XR not only need to collect and process feedback data from various sensors in real time to evaluate the information including user's position, speed, perspective, but also need to quickly render highly realistic images with high resolution based on this information [2]. Therefore, the rapid rise of XR-type applications not only puts the network's service assurance capability to the test while also posing new challenges to the computing performance of terminals that are highly constrained in terms of power consumption and volume. It is a feasible solution to use edge computing to offload real-time workloads on terminals with extremely low service processing latency, while some non-real-time computing tasks can be forwarded to the central cloud for completion, i.e., a collaborative cloud-edge-terminal system [3].

As Network Functions Virtualization (NFV) is adopted by more and more equipment vendors, various network equipment including base stations will also have the ability to provide edge computing services. In addition, due to the transition to cloud-native for all types of application servers, a complete application will likely be divided into multiple decoupled micro-services, which implement business logic through mutual calls between Application Programming Interfaces (APIs) [4], and XR-type applications are no exception. These micro-services will be distributed on different computing nodes, including edge computing nodes and central cloud computing nodes, etc. In the foreseeable future, computing infrastructures will be distributed in large numbers over a huge spatial scale [5], XR-type application service instances can be deployed at all or part of the micro-service components in a distributed manner, and user terminals often access XR-type applications on the move. This leads to a high degree of time variability in the service instance status and the link between the user and the service instance. The previous DNS-based load balancing mechanisms [6] are difficult to fully utilize the decentralized deployment of computing resources to provide good QoE to users because they do not sense the network status between users and service instances. However, the load balancing solution based on the service status list maintained by the client requires terminal devices to collect service node status information and measure network status information in real time. This introduces a large resource overhead for low-power devices, such as Internet of Things (IoT) devices. Since the cloud-edge-terminal collaboration process of various applications, including XR, will be abstracted from the network's perspective as the mutual invocation of multiple micro-service components between user terminals and micro-service clusters and across micro-service clusters, which provides a potential solution for processing node and link selection among a large number of service nodes and guarantees service experience quality, that is, load balancing through the control plane of the network.

In this paper an overlay service anycast system based on the SRv6 protocol is proposed, which utilizes the underlay routing and traffic engineering functions provided by the

SRv6 protocol to deliver the overlay anycast packets to the appropriate computing service gateway. By establishing a network address identification mechanism, the readability and manageability of the control plane are improved when configuring the forwarding policy and mobility policy of each computing service. In addition, the control plane will also obtain the information of computing service instance and network status to support for the load balancing decision of anycast service. In summary, our contributions are as follows:

1) Identification mechanism of network address: The globally unique anycast IP address assigned to the computing service instance will identify its service ID information, so that the control plane can maintain the mapping relationship between computing service and anycast IP address. The computing service gateway will be assigned multiple Segment Identifiers (SIDs), and each SID will also identify its corresponding computing service information. The SID directly reflects which computing services are deployed behind the computing service gateway, which enables the control plane to discover the computing services, and forward the computing service traffic to the appropriate computing service gateway. The user identity information is identified in the IP address of the terminal to facilitate the control plane to provide users with seamless service mobility and continuity.

2) Enhanced control plane: To support the load balancing decision among computing service instances, the status of each computing service instance is obtained through the interface between the control plane and the computing infrastructure controller.

3) QoE-oriented load balancing decision: The control plane can select one or more computing service instances for user terminals to provide high-quality services based on the computing and network KPI requirements of the computing service, the available area limitation of the computing service, the status information of each computing service instance, and the current connection status between user terminal and each available computing service instance.

The remainder of this paper will be organized as follows. In Section 2, the load balancing requirements of the computing and network convergence for ubiquitous computing services is summarized and in Section 3, some existing related work is reviewed and a potential solution is elaborated in Section 4. The content of Section 5 is analyzing the relationship between our research content and existing standards. Section 6 is a conclusion of this paper.

2. REQUIREMENTS

Based on the background introduction above, and with reference to [7] and [8], the requirements of computing and network convergence infrastructure platform used to support large-scale distributed applications and services such as XR and Vehicle to X (V2X) are summarized.

2.1 Service resource discovery

Since computing services deployed on highly distributed heterogeneous infrastructures such as edge computing nodes and cloud computing nodes are often dynamically instantiated and de-instantiated according to indicators such as the number of users, occupancy cost, energy consumption, and estimated service quality, service load balancing through the network control plane assumes the network control plane can efficiently and dynamically discover the currently stand-by service instances.

2.2 QoE-oriented service instance selection

For users, the QoE of some computing services should be fully guaranteed, and the traditional best-effort policy is unacceptable. For example, high latency and packet loss will likely cause dizziness and nausea to users using XR applications [9]. Therefore, the network control plane should be able to understand the demand information of computing services for network quality as well as node status, and filter and route user service requests to service instances with better quality of experience.

2.3 Seamless mobility and service continuity

Since the network and node status is highly time-varying, the service instances selected by the network control plane for users will also change frequently. In addition, due to the movement of the user, the gateway attached may also change. Since frequent service interruptions are less acceptable than QoE degradation, the network control plane should ensure that users do not switch to other service instances when they use session-based computing services.

3. RELATED WORKS

3.1 Edge computing support in 3GPP system

5G New Radio (NR) provides support for enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and enhanced Machine-Type Communication (eMTC) services, and exposes various network capabilities to provide service quality assurance support for edge computing applications. Uplink Classifier (UL CL) and IPv6 multi-homing mechanisms can easily help edge computing application traffic offload to specific edge computing nodes [10]. A 3GPP system also helps User Equipment (UE) discover edge computing services by introducing an edge computing enablement layer on UEs and edge computing servers [11], which utilizes the capabilities exposed by the core network to provide mobility management and service continuity for edge computing services. However, it only provides well-established enabling tools for 3GPP systems and edge computing applications and has not yet provided support for heterogeneous networks and cloud-edge-terminal collaboration scenarios. Besides, edge computing applications need to be built based on edge enabling layers, which also brings a lot of extra work.

3.2 Named data networking

A Named Data Network (NDN) is different from IP networks by the mechanism of addressing a specific node by IP address for the purpose of acquiring specific data or functions. Consumers pull the corresponding data or functions by sending interest packets to the network, and thus they can naturally adapt to scenarios where multiple computing nodes provide the same service and provide seamless mobility [12]. In addition, the Content Store (CS) in the NDN architecture can also provide reuse of computational results. However, it is difficult to implement in a practical deployment environment since NDN requires a lot of modifications to the existing IP network and lacks a push mechanism.

3.3 Anycast proxy

IETF [13] defines the IPv6 addressing architecture, which covers the definition and usage of IPv6 anycast addresses. However, because IP anycast is difficult to support session-based stateful upper-layer services, and the rapidly changing anycast members make the routing table unstable, the anycast protocol is only widely used in single-round-trip service scenarios such as DNS. To address this problem, some proxy-based anycast solutions have been proposed [14]. Through the tunnel connection established between the client proxy and the server proxy, anycast communication is enabled on the overlay, while the communication is still using unicast for the underlay network. By configuring the proxy, a suitable server can be chosen for the client to provide services, and this process is transparent. However, a large number of IP tunnels need to be maintained between proxies, which brings great challenges to the performance of proxy devices.

3.4 Dyncast

Paper [15] proposes a dynamic anycast architecture (CFN Dyncast), which is similar to the idea of the anycast proxy. But the difference is that CFN Dyncast uses the modified BGP protocol to inform the computing service status of each routing node to provide dynamic load balancing among the anycast member nodes and significantly reduce the control plane overhead. However, since session affinity is implemented by the aging mechanism, it is difficult to support seamless mobility across routing nodes.

4. SOLUTION DESIGN

4.1 Overview

In this section, the reason why the SRv6 protocol is used for the user plane in the following proposed solution design will be illustrated as well as how it enables service resource discovery, QoE-oriented service instance selection, and seamless mobility.

4.2 Why use SRv6

SRv6 is a segment routing technology implemented using the IPv6 protocol, which is designed based on the concept of source routing. Therefore, packets can decide their routing paths at the ingress, which is used to support traffic engineering in IPv6 networks [16]. The prerequisite of performing load balancing for multiple computing service instances sharing the same anycast address through the network is that the service request packets sent by the user terminal with the anycast address as the destination address through the user plane of the network can be forwarded to a suitable computing service instance. As a result, underlay traffic engineering is the basis for anycast of overlay services and the flexibility provided by SRv6 and the good affinity with IP networks are the reasons why the SRv6 protocol is chosen as the underlay traffic engineering protocol.

4.3 Architecture design

This paper draws on the design ideas of the anycast proxy and Dyncast, and improves it with the SRv6 protocol to support service resource discovery, QoE-oriented service instance selection and seamless mobility. The following is the design of the proposed architecture in Figure 1:

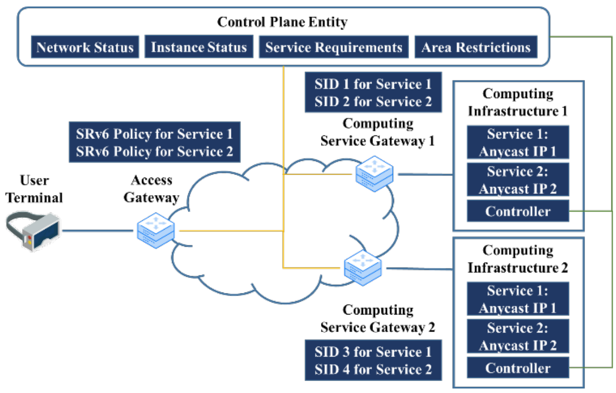


Figure 1 – Architecture design

1) Access gateway: In the overlay service anycast system, the access gateway acts as the gateway of the user terminal, driven by the SRv6 protocol, and communicates with the computing service gateway through a unicast address. It functions similarly to the proxy of the user terminal and performs forwarding according to the SRv6 policy issued by the control plane by identifying the user's computing service upstream traffic with an anycast destination IP address and encapsulating the corresponding Segment Routing Header (SRH). The SRH contains an SID list, which is used to identify the hop-by-hop node in the forwarding process of the data packet. In the forwarding process of computing service traffic, the SID list eventually transmits the data packet to the computing service gateway specified by the control plane. In actual deployment, it can be a home router, User Plane Function (UPF), Broadband Remote Access Server (BRAS), etc., and the access gateway can also function as a computing service gateway.

2) Computing service gateway: In the overlay service anycast system, the computing service gateway is used as the gateway of the computing service instance, driven by the SRv6 protocol, and communicates with the access gateway through a unicast address. Its role is similar to the proxy of computing service instances and the operator of the computing service instances will configure multiple SIDs for the computing service gateway. The SID is a special IPv6 address used in the SRv6 protocol to identify a forwarding action or function on a routing node device. On the computing service gateway, each SID corresponds to a computing service instance deployed behind the computing gateway. When a packet with a destination address of a certain SID is received, the computing service gateway will strip the SRH and forward the inner data packet to the back-end computing service instance. In actual deployment, it can be the Point of Presence (POP) device of the cloud data center and the gateway device of the edge computing node, and the computing service gateway can also have the function of an access gateway.

3) Control plane: In the overlay service anycast system, the control plane can dynamically discover the configured SID on the computing service gateway and provide SRv6 policies and corresponding SRHs for the access gateway. The control plane can also monitor the network quality and status between each access gateway and computing service gateway in real time and obtain the status information of each computing service instance through the interface with each computing infrastructure controller. The control plane also maintains the computing and network KPI requirements of each computing service. Therefore, each access gateway can be configured with an SRv6 policy that leads to one or a group of computing service gateways that meet the requirements through the control plane, so as to access the computing service under the premise of ensuring the service quality of experience.

4) Computing service instance: In the overlay service anycast system, the computing service instance acts as the server of each computing service. All computing instances that provide the same computing service share the same anycast IP address globally, so there is one and only one computing service instance behind a computing service gateway. When multiple computing service instances have to be deployed behind the same computing service gateway, the anycast IP address should be shared through other local L3 load balancers.

4.4 Service resource discovery based on IP address identification

The IPv6 protocol provides a vast address space, which makes it possible to carry more information in IP addresses. By using the computing service identification information carried in the IP address, it can help the control plane to make decisions on the routing paths of specific users and specific computing services, as in Figure 2.

In this paper, the IP prefix part of the anycast IP of the computing service instance will be globally reserved to prevent it from being assigned to other unicast interfaces. The interface ID part of the anycast IP address of the computing service instance will be determined by the unique ID of the computing service provided by the computing service instance, which can be managed and generated through a certain computing service registration mechanism.

The SID of the computing service gateway corresponds to a computing service instance behind the computing service gateway in this paper, so in order to help the control plane discover the computing service and make policy decisions, the unique ID of the computing service is added to the SID in the function field, while the locator field remains allocated by the network. In this way, the control plane can discover the back-end ready service instances through the SIDs enabled on the computing service gateway in order to manage and configure SRv6 policies.

Anycast IP Address of Service Instance	
IP Prefix	Interface ID
Reserved Globally	Service ID
SID of Computing Service Gateway	
Locator	Function
Allocated By ISP	Service ID
User Terminal IP Address	
IP Prefix	Interface ID
Allocated By ISP	Client ID

Figure 2 – IP address identification design

4.5 Load balancing strategy

Once the control plane obtains the computing services supported by each computing service gateway, it can make load balancing decisions by configuring the SRv6 policies for the access gateway. Unlike other traditional load balancing solutions, the control plane, as a part of the network infrastructure, is able to know the connection status and quality between each access gateway and computing service gateway, which provides the possibility for QoE optimization. Since the control plane itself does not maintain the status information of each computing service instance, this information must be obtained from the computing infrastructure controller. Typical computing service instance status information may include load information and performance information.

In addition, in order to reduce the number of SRv6 policies active on the access gateway, the control plane should also obtain the service area restrictions of each computing service instance and network KPI requirements of each computing service instance. The SRv6 policies corresponding to the service instance that do not meet the service area restrictions and KPI requirements will not be configured to the access gateway,

As an example, shown in Figure 3, a user terminal requests access to computing service 1. Computing service 1 is deployed on hundreds of computing infrastructures globally, but after being filtered by service area restrictions, only three computing infrastructures in the vicinity can provide the service to the user's access gateway. The three computing infrastructures are respectively connected to three computing service gateways. By monitoring the network status, the control plane learns that the bandwidth and delay from the access gateway to the three computing service gateways are 50Mbps@8ms, 70Mbps@12ms, and 80Mbps@20ms, respectively. At the same time, the control plane obtains the load status of the three computing service instances from the three computing infrastructure controllers, which are 85%, 40%, and 30%, and the performance indicators are 0.6 0.8 1.0 respectively. According to the computing service requirements for computing and network KPIs, the delay is less than 15ms and the load is less than 70%, therefore the control plane selects the computing service instance on computing infrastructure 2 for the user terminal and configures the SRv6 policy of computing service 1 for the access gateway, which contains the SRH leading to SID2.

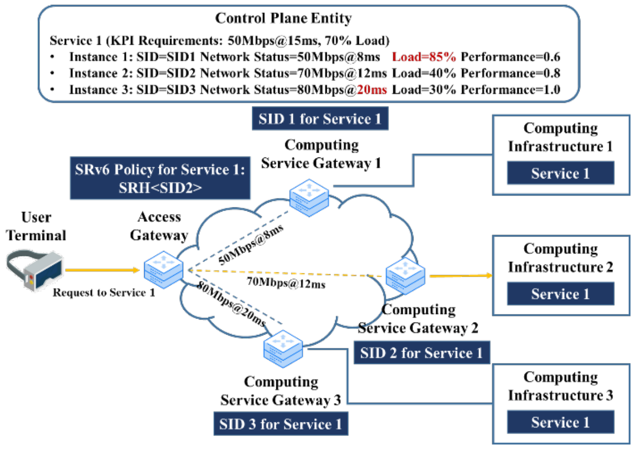


Figure 3 – Load balancing decision for Service 1

As shown in Figure 4, an SRv6 policy can be composed of multiple candidate paths. A candidate path is selected by its priority parameter and composed of multiple SID lists with different weights. This provides a variety of different strategies for load balancing of computing services. For edge computing services that require extremely high reliability, the access gateway can directly and autonomously adjust the values of priority parameters. For stateless services, the weighting mechanism can also be used to achieve load sharing among multiple computing service instances to maximize throughput.

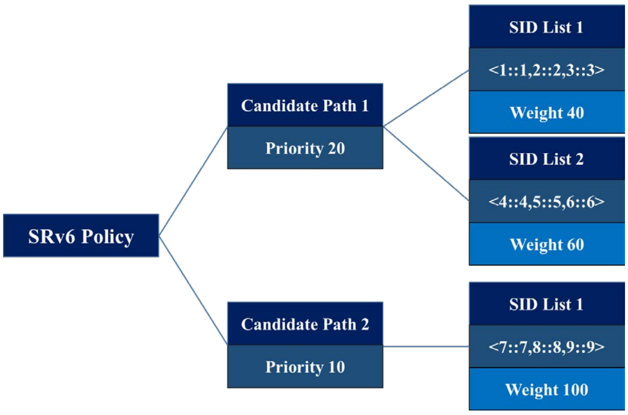


Figure 4 – SRv6 policy

The SRv6 protocol provides good support for the separation of the user plane and the control plane, and one access gateway can be controlled by more than one control plane entity, so the deployment method of the control plane has extremely high flexibility.

4.6 Supporting seamless mobility and service continuity

Due to the highly time-varying nature of the network and node status, the SRv6 policy configured on the access gateway will also change frequently, which may lead to interruption of the ongoing stateful services of the user terminal. Therefore, a service continuity mechanism for stateful services is necessary in the overlay service anycast system. Since the control plane controls the SRv6 policy of the access gateway, the user terminal can initiate registration with the control plane when stateful services are required. Subsequently, the control plane will assign a high-priority SRv6 policy specific to the user terminal to the access gateway in order to ensure that the computing service instance will not be changed during the session, as in Figure 5. It should be noted that the current SRv6 protocol has not provided a filtering mechanism for source address, so it has to enhance the function of the access gateway.

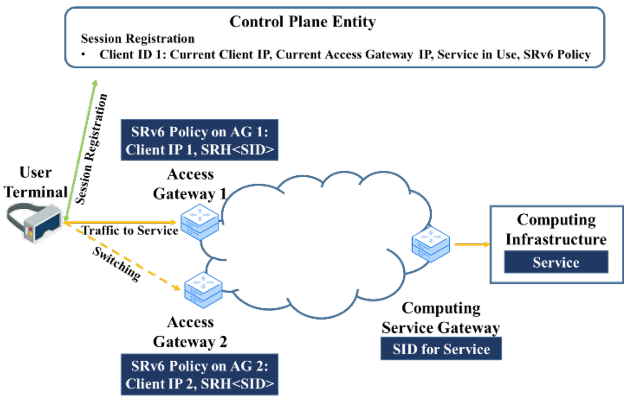


Figure 5 – Session registration procedure

In this paper, an IP address can be used to identify a computing service as well as a user terminal. Similar to the

SID of the computing service gateway, the IP address of the user terminal may also include a globally unique client ID. When a user switches to a new access gateway, such as 5G UE switching between edge UPFs, the IP prefix of the user terminal may also change. However, by introducing the client ID into the IP address of the user terminal, the control plane can still allocate the same computing service instance to the terminal and assign a separate SRv6 policy with the destination of the previous computing service gateway to the new access gateway to achieve seamless mobility which is transparent to the user terminal.

4.7 Establishment of testbed

In order to evaluate the performance of the system, a testbed is being established. Some Software-Defined Wide Area Network (SD-WAN) Customer Premises Equipment (CPEs) are being adapted to the SRv6 protocol for functioning as access gateways and computing service gateways. An SD-WAN controller is being modified to function as the control plane, while the interfaces to the computing infrastructure controllers are deployed. The assessment and analysis will be released in future work.

5. RELATIONSHIP TO STANDARDS

The new generation of services represented by XR brings new challenges not only for network performance but also for computing performance. Joint optimization of computing performance and network performance is a key issue for next-generation computing and network convergence infrastructure, and relevant standardization work has also been started.

Recommendation ITU-T Y.2501, which has been approved in ITU-T SG13, defines the computing power network architecture from the perspective of resource allocation optimization by network. In this paper, the overlay service anycast system proposed works on both the CPN resource layer and CPN control layer, and performs computing resource allocation towards users by traffic engineering. The work item Y.CAN-reqs proposed a networking solution to support cloud computing including distributed cloud, combine computing information with network information and user's service requirements, and finally guarantee the service requirements both on computing and the network, which is the QoE-oriented load balance mentioned in this paper.

Furthermore, there are also some CNC-related work items under study, and this paper provides a preliminary technical solution for these proposals.

6. CONCLUSIONS

In this paper the challenges of supporting XR-type business applications to the existing infrastructure and the technical requirements for load balancing of computing services through the network control plane are summarized. To provide a better solution, existing related work is reviewed and its deficiencies are assessed. Finally, an overlay service anycast system is proposed based on SRv6 to support service resource discovery, QoE-oriented service instance selection, seamless mobility and service continuity. In the future, we will evaluate its performance in the lab.

REFERENCES

[1] Sumit Patil, Rahul Kumar, "Accelerating Extended Reality Vision With 5G Networks," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA).

[2] Hans-Jürgen Zepernick, "Toward Immersive Mobile Multimedia: From Mobile Video to Mobile Extended Reality," 2018 Third International Conference on Informatics and Computing (ICIC).

[3] Anousheh Gholami, John S. Baras, "Collaborative Cloud-Edge-Local Computation Offloading for Multi-Component Applications"

[4] Kaustabha Ray, Ansuman Banerjee, Nanjangud C. Narendra, "Proactive Microservice Placement and Migration for Mobile Edge Computing," 2020 IEEE/ACM Symposium on Edge Computing (SEC).

[5] Xiongyan Tang, Chang Cao, Youxiang Wang, Shuai Zhang, Ying Liu, Mingxuan Li, and Tao He, "Computing power network: The architecture of convergence of computing and networking towards 6G requirement," China Communications, vol. 18, pp. 175-185, 2021.

[6] "Extension Mechanisms for DNS," IETF RFC 6891, 2013.

[7] Shuo Wang, Xing Zhang, Yan Zhang, Lin Wang, Juwo Yang, and Wenbo Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," IEEE Access, vol. 5, pp. 6757-6779, 2017.

[8] Nam Ky Giang, Rodger Lea, Michael Blackstock, and Victor C.M. Leung, "Fog at the Edge: Experiences Building an Edge Computing Platform," 2018 IEEE International Conference on Edge Computing (EDGE).

[9] M. S. Elbamby, C. Perfecto, M. Bennis and K. Doppler, "Toward low-latency and ultra-reliable virtual reality," IEEE Netw., vol. 32, pp. 78-84, Mar./Apr. 2018.

[10] "System architecture for the 5G System (5GS)," 3GPP TS 23.501.

[11] "Architecture for enabling Edge Applications," 3GPP TS 23.558.

[12] Abderrahmen Mtibaa, Reza Tourani, Satyajayant Misra, Jeff Burke, and Lixia Zhang, "Towards Edge Computing over Named Data Networking," 2018 IEEE International Conference on Edge Computing (EDGE).

[13] "IP Version 6 Addressing Architecture," IETF RFC 4291, 2006.

[14] Tim Stevens, Marc De Leenheer, Chris Develder, Filip De Turck, Bart Dhoedt, and Piet Demeester, "ASTAS: Architecture for scalable and transparent anycast services," Journal of Communications and Networks, vol. 9, pp. 457-465, 2007.

[15] Bing Liu, Jianwei Mao, Ling Xu, Ruizhao Hu, and Xia Chen, "CFN-dyncast: Load Balancing the Edges via the Network," 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW).

[16] "Segment Routing Architecture," IETF RFC 8402, 2018.

TOWARDS A MORE FLEXIBLE NETWORKING LANDSCAPE

David Lou[1]; Marinos Charalambides[2]

[1]zhe.lou@huawei.com Huawei Technologies Duesseldorf GmbH

[2]marinos.charalambides@gmail.com Independent Researcher

ABSTRACT

Technological advancements leading to 5th generation networks mainly focused on improving coverage and performance in terms of bandwidth and latency. While these will likely remain aspects of continuous improvement, along with issues on reliability and security, this paper argues that flexibility is a key property that 6G networks should exhibit in order to overcome important limitations of the current networking landscape and fulfill emerging user needs and application requirements. We identify key areas that can contribute towards more flexibility, present existing efforts on relevant technologies and discuss the research challenges that need to be addressed in order to reach the desired level of flexibility.

Keywords – 6G networks, flexible Internet, network intelligence

1. INTRODUCTION

It is evident that network infrastructures are already hard to operate and manage [28], while emerging trends such as the introduction of computational capabilities at network edges bring additional complexities. In addition, current networking practices, architectures and technologies have contributed in making the networking landscape rather rigid. As a result, demanding services in terms of performance and reliability are difficult to deploy, the operational expenditure remains high, and the integration of new technologies sometimes requires nasty hacks that are costly in many ways.

While the capacity and latency of networks continue to improve, we believe that flexibility should be a key characteristic of future networks, which deserves attention from the research community [27]. Making the network more flexible will enable adaptive behavior based on emerging events, programmatic control of resources, fine-grained resource management, and will extend the performance boundaries. Together, these can simplify the deployment and operation of new services, cut service deployment times drastically, reduce operational costs, and achieve more efficient use of resources.

To this end, the paper focuses on four key areas that can contribute significantly towards increasing the degree of flexibility for the benefit of network operators, service providers and users: (i) network configuration and adaptability, Section 2; (ii) cloud-native networking, Section 3; (iii) network softwarization, Section 4; and (iv) network addressing, Section 5. For each of these areas we present the state of play and identify the gaps, while in Section 6 we discuss the main challenges and research opportunities.

2. NETWORK CONFIGURATION AND ADAPTABILITY

Current practices in the telecommunications domain rely on long-lived configurations, e.g. in the order of days, which do not cater for changes in the operating conditions or in the performance requirements of running services. More flexible solutions that allow shorter configuration cycles and adapt the network behavior accordingly are of interest to network operators but also to other vertical industries operating demanding services. The main obstacles in realizing dynamic reconfiguration functionality in telecommunication networks are the centralized management platforms that are in place today and the not-so-straightforward way by which network devices can be collectively programmed.

2.1 Network automation

Although centralization simplifies the implementation of network management systems, it renders the automation of management tasks rather difficult. Due to the large geographic footprint of operator networks, centralized solutions introduce significant lag and overhead when collecting state information. In addition, algorithms that optimize for the whole network tend to be time consuming. As such, autonomic networking principles envisioned by many operators and vendors, in which automated decision-making was based on the analysis of continuously collected monitoring data, were never applied to large telco networks, in contrast to data center networks that are characterized by much smaller network diameters.

To enable adaptive behavior, and thus realize self-driving networks, management and control planes require decentralization so that relevant functions can execute very close to the infrastructure (if not in it) and base their decisions on local knowledge. Such a setting would allow the timely detection of events that adversely affect the service performance, and the swift computation and enforcement of appropriate remedy actions, for example rerouting through an alternative path or selecting another server hosting an application instance. In addition, to warrant the performance of demanding services delivered over large network infrastructures, real-time telemetry is needed so that an accurate view of the resources can be constantly maintained. The vast amount of collected information can overwhelm network management systems, hence efficient mechanisms are needed that can dynamically self-tune and achieve the right trade-off between accuracy and overhead [22], [23].

2.2 High-level programmability

Network programmability has always been a challenging issue with solutions traditionally involving complex low-level scripts and multiple vendor-specific interfaces, thus making desired changes in the behavior of network devices a daunting task. Programmability has been a topic of discussion since the late 90s with proposals on open APIs [14] and active networks [3], [24]. While the idea of programming control information in headers that defines the treatment of packets in the network initially seemed promising, it was eventually dismissed on the grounds of significant security risks.

In an effort to reduce the network administration complexity, the policy-based management technology [21] was extensively researched around the same time, in which policies govern the network behavior offering a certain degree of programmability. These are technology-independent rules that enhance the functionality of network devices by introducing interpreted logic that can be dynamically changed without modifying the underlying implementation. Alongside the core PBM technologies, techniques to refine high-level goals to concrete configuration parameters had been developed at the time [2], [16]. These, however, were not fully automated and were also based on application-specific policy refinement patterns, which limited their applicability to specific domains, e.g. quality of service and security. Another issue concerned the possibility of configuration inconsistencies as a result of policy conflicts [5].

More recently, the decoupling of the control plane from the forwarding hardware in the Software-Defined Networking (SDN) paradigm [12] empowers operators with a highly flexible approach to control the behavior of the network, since control functions/policies (realized in soft form) can be easily added and removed. While the southbound interface of the SDN architecture, i.e. OpenFlow, has undergone several iterations and allows us to define the treatment of traffic in switches, the northbound interface that supports the interaction between the controller and external applications, as well as the east-west bound interface for decentralized controller settings, have not reached adequate maturity levels. In a parallel effort, the P4 language [4] addresses a more general problem than OpenFlow and has been designed to program the behavior of the forwarding plane irrespective of the supported network protocol and the type of controller.

Although P4 has gained traction in the networking community, the language is fairly low level and therefore somewhat complex even for simple packet processing operations. We believe that simpler approaches are needed, which do not necessarily require expert knowledge on the underlying networking technologies. The abstraction level at which the resources are programmed should be raised. Instead of composing code that details ’how’ a goal is achieved, directives are used to define ’what’ needs to be achieved. Alongside a declarative approach for expressing the desired network behavior, a mechanism is needed that automatically decomposes high-level goals to specific configuration settings, thus allowing the network to flexibly adapt to changes in business objectives, administrative goals, and application requirements. Research on intent-based networking [6], [25] has taken some initial steps in this direction, but a concrete language and decomposition solutions have not been made available yet.

3. CLOUD NATIVE NETWORKING

Cloud computing has brought enormous advantages in terms of elasticity, scalability, and automation in computational and storage infrastructure. However, large cloud providers tend to be based around a handful of large data centers strategically positioned at central locations on the Internet. Users are, therefore, often located hundreds or thousands of kilometers away from the computational nodes hosting the services they are accessing. This implies network and processing latencies in the order of hundreds up to thousands of milliseconds for many users in addition to the inefficiencies and cost implications of transferring data over large distances and long network paths in the case of data-intensive applications. This also has additional non-technical disadvantages such as data being stored and processed in regions under different jurisdictions to those where the users are located. In terms of security and privacy, the concentration of clouds in a few central locations turns them into easier targets for (cyber and physical) attacks and facilitates surveillance taps into the fiber optic links interconnecting data centers.

The above issues drove research in academia and industry to investigate distributed deployments of a vast array of computational resources in the vicinity of both the users and data sources [18], [20]. Based on the flexibility concerning, for example, the locations at which applications execute and the selection of nodes from where user requests are served, the edge compute paradigm envisions an environment where services can be accessed with a much lower latency, while at the same time reducing the traffic footprint on the network and enabling better resilience to failures and attacks.

The edge computing concept was initially envisioned as part of 5G deployments [9], [10] for drastically reducing the latency of cloud-based applications. But it did not prove to be the case as there are still major obstacles to overcome before it becomes a reality. There needs to be seamless integration between communication and computation technologies, which have been traditionally evolving on their own parallel paths. Carefully crafted management functionality can provide the right glue and can allow the two to be operated and optimized in a unified manner.

To meet the performance requirements of future demanding services, telecommunication network and service management practices need to become more dynamic and flexible in nature and match the capabilities of those in cloud domain. Going beyond the traditional telecommunication virtualization technologies and platforms, a cloud-native approach is needed, which hides the underlying heterogeneous infrastructure by operating on smaller containers, such as those provided by Docker. This will provide a level of abstraction above the compute infrastructure allowing applications to be formed from one or more components executing within lightweight containers. The management logic for coordinating the container-hosting locations, the allocation of user requests among the distributed set of application execution points, and the decisions concerning the usage of resources will need to follow a multidimensional optimization framework.

4. NETWORK SOFTWARIZATION

Although hardware implementations in the network are in general very fast and have predictable behavior due to exhaustive testing, they are rigid in terms of location and available resources, and their update/maintenance tends to be complicated. To overcome these limitations, the research community has been investigating software-based solutions which can achieve greater agility and cost effectiveness.

The two main technologies that drive the so-called "softwarization" of telecommunication networks are Network Functions Virtualization (NFV) [7] and SDN [12]. The former allows us to replace network equipment, such as load balancers and firewalls, with software that executes on commodity servers, while the control data plane separation in SDN moves control functions outside network devices and into dedicated controller entities. SDN largely simplifies network management tasks and allows advanced network intelligence to be flexibly added to controllers without needing to upgrade network devices. NFV, on the other hand, allows network functions to be created/migrated according to the needs and for updates to be performed remotely, while the use of virtualization technologies enables the dynamic scaling of resources allocated to network functions.

The significant advantages such technologies offer have been driving the trend towards software-based solutions. The fact, however, that multiple technologies are needed to manage physical and virtual resources, alongside the imminent integration of computational capability in the network, can complicate the interaction between administrators and the infrastructure. Hence, there is a need for a unified platform, e.g. in the form of a network operating system, through which the interaction can be simplified in order to flexibly program the operation of resources and services.

5. NETWORK ADDRESSING

While the Internet Protocol (IP) has been adopted as the de facto layer-3 protocol for Internet-based communications, it has inherited several limitations of its original design as scalability and applicability issues were not properly considered at the time. Although IPv6 solves the problem of address depletion, it still inherits a number of the IPv4 problems concerning, for example, the fixed number of address bits and the fixed semantics associated with the addresses. These limitations have been demonstrated by the emergence of several networking scenarios, which require significantly more flexibility in network addressing, both in terms of length and semantics.

5.1 Low power IoT networks

The low device complexity of IoT devices defines the properties of communication technologies developed for this domain, which use short (e.g. 16-bit) addresses in order to reduce the header size, communication overheads and memory requirements. IPv4 32-bit addresses are already long, resulting in expensive operations in multi-hop routing scenarios, while IPv6 128 bit-addresses worsen the problem and would not even fit within the maximum transmission unit of some IoT protocols [19]. Header compression techniques can potentially decrease the IPv6 overhead and fragmentation of packets would allow coping with the bulky protocol. These, however, involve power-hungry operations inappropriate for resource-constrained IoT devices and hence require the use of gateways.

Due to its inherent rigidity in supporting only the single addressing semantic of topological location, IP protocol is not able to meet the requirements of a wide range of IoT application scenarios in which addresses can take various forms to identify, for example, communication endpoints, physical objects, data types, and locations within a geographic area [1]. In addition, support for multiple semantics would be lost in a scenario where address translation is needed when connecting multiple IoT networks over the Internet. Flexibility in the address length, which can also be seen as elasticity in the address space, caters to an increasing number of specialized network deployments and would allow IoT networks to use the length that best fits their scale and constraints. In addition, flexible addressing can enable seamless communication between nodes, bypassing the need for expensive address mappings.

5.2 Highly dynamic network topologies

The semantically rigid nature of IP addresses, which refers only to the topological location of a network interface, poses a significant challenge to networks with highly dynamic topologies. The constantly changing data plane connectivity of satellite networks [8] constitutes a representative example, where the use of IP protocol is rather restrictive since the relation between network nodes can change as well as that between the nodes and the user endpoint. Another example is that of vehicular networks, where the communication can be over direct links between vehicles and roadside infrastructure but where it can also involve multi-hop connections between vehicles [11]. As in the case of satellite networks, the dynamics of communicating nodes result in fluid topologies, which require maintaining and updating topological IP addresses on a frequent basis.

The address structure could instead be flexible enough to support multiple semantics that apply to environments with highly dynamic topologies. In addition to reducing the complexity of such networks, a flexible addressing scheme can allow for richer policies that enhance packet treatment in terms of routing performance and security. A representative example is the approach in [26], which uses semantic addresses to represent virtual switches at fixed space locations (based on geo-coordinates) being traversed by satellites, and which can result in reduced service disruptions caused by satellite handover events.

Table 1 – SoTA and challenges summary

Area	State-of-the-Art	Research Challenges
Network configuration and adaptability	Long-lived configurations, centralization, OpenFlow, P4	Automation: decentralized management, light-weight telemetry High-level programmability: intent decomposition, configuration consistency
Cloud-native networking	Communication and computation treated separately	Efficient resource management algorithms (joint optimization), multi-provider resource federation
Network softwarization	SDN, NFV	Network operating system, abstractions, APIs, common functionalities
Network addressing	Fixed address length and semantics	Elastic addressing scheme, semantically-enhanced addresses and routing mechanisms

6. CHALLENGES AND RESEARCH OPPORTUNITIES

This section discusses the main research challenges associated with the four areas described in the previous parts of the paper that can allow for more flexibility in the network. These are summarized in Table 1.

6.1 Network configuration and adaptability

Self-driving capabilities in large network infrastructures, with quick detection and reaction to network events, can be achieved by the decentralization of network intelligence. This, however, will require a communication protocol (and relevant east-west interfaces) to synchronize the local views of the distributed decision-making entities and to coordinate their actions so as to avoid inconsistent configurations (e.g. the introduction of loops). At the same time, the reduced computational capability of management nodes, as a result of decentralization, will make it difficult to cope with collecting and processing the vast amount of data generated in the network. Hence, a key challenge concerns the development of lightweight telemetry mechanisms that execute close to the data source and can strike the right balance between accuracy and overhead. Such a mechanism could use, for example, machine learning-based classification methods for automatically reducing the dimensionality of data by efficiently discarding non-useful information.

Concerning the programmability aspect, intent-based networking has been gaining traction in the last few years evidenced by efforts in standardization bodies [6], [15], the emergence of relevant workshops, and the inclusion of the topic in international conferences. Besides the need of a common intent specification language, key challenges include a generalized intent decomposition mechanism, the incorporation of feedback to ensure the continuous enforcement of intent, the automated selection of the most appropriate action(s) to execute (given multiple options) for achieving a specific objective, and tools to ensure configuration consistency.

6.2 Cloud-native networking

The embedding of computation in the network poses several challenges which mainly stem from the requirement that resource management algorithms should no longer treat different resource types in isolation but instead optimize them jointly. In conjunction with the need of real-time reconfigurations to support demanding applications executing in the network, such algorithms should be designed with efficiency as a prime objective. In addition, the resource scarcity in edge computing environments naturally forces the use of container-based technologies, which operate on a finer granularity compared to conventional virtualization technologies and can achieve better usage. The management logic for coordinating the container-hosting locations, the allocation of user requests among the distributed set of application execution points, and the decisions concerning the usage of resources will need to follow a multidimensional optimization framework.

In edge compute scenarios it is likely that multiple providers will need to collaborate to form a large cloud infrastructure both in terms of resources and geographical footprint, so as to support a wide range of services and large customer bases. To this end, another challenge not only concerns the design of models and architectures for effective federation of resources across providers, but also approaches that warrant the confinement of data within the boundaries defined by specific laws and regulations, e.g. GDPR.

6.3 Network softwarization

Although some controller implementations in the network domain, e.g. ONOS [17], have been coined operating systems, their functionality is limited to that of control and management. The equivalent can be said about Kubernetes [13], a container orchestration platform, for the compute domain. The design and development of a fully-fledged Operating System (OS) for networks is the main challenge in this area. Key issues to address include the definition of objects and abstractions which the OS will handle and that can be unambiguously referenced (e.g. links, paths, topologies, slices, containers, network and management functions, etc.), but also the design of APIs through which services can interact with those objects to specify, for example, resources and requirements. The infrastructure programmability could be realized with a high-level intent engine that can use primitives on the attributes of the same objects and abstractions.

Another challenge concerns the derivation of functionalities and facilities that will be common for all services, so that developers do not need to re-implement these, i.e. for reusability. While device drivers responsible for interacting with network devices for low-level configurations are important, a network OS can utilize existing controller and container management implementations as the bottom end to enforce desired configurations.

6.4 Network addressing

The penetration of IoT scenarios requires power-efficient operations that suit the characteristics of such networks. A key challenge in this area is the design of a new elastic addressing scheme and associated mechanisms that can allow the use of short addresses while data is communicated within the IoT network (or other specialized deployments) and longer ones when being routed over the public Internet. Another challenge concerns the design of a new structure for network addresses and associated identifiers that can allow for a rich set of semantics to be embedded in the packet headers, going well beyond the traditional topological locations. While doing so, it is important to avoid adding significant overhead, if any, and to develop routing mechanisms that can efficiently process the new semantics.

7. SUMMARY

In this paper we stress the need of enhancing future networks with more flexibility given the key benefits in doing so, and we presented four areas which we believe can have a significant impact towards achieving this objective. Based on our analysis, important research challenges need to be addressed in order to improve the degree of flexibility, which present interesting opportunities for academia and industry.

In subsequent work we plan to delve deeper into each of these areas and develop specific solutions.

REFERENCES

[1] Alliance for Internet of Things Innovation, “Identifiers in internet of things, February 2018. Available at: https://euagenda.eu/upload/publications/identifiers-in-internet-of-thingsiot.pdf

[2] A.K. Bandara, et al., “Policy refinement for IP differentiated services quality of service management, IEEE Transactions on Network and Service Management (TNSM), vol. 2, no. 2, 2006.

[3] S. Bhattacharjee, K.L. Calvert, E.W. Zegura, “An architecture for active networking,” In proceedings of International Conference on High Performance Networking, Boston, USA, April 1997.

[4] P. Bosshart et al., “P4: Programming protocol-independent packet processors,” ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 3, pp.87-95, 2014.

[5] M. Charalambides et al., “Policy conflict analysis for DiffServ quality of service management,” IEEE Transactions on Network and Service Management (TNSM), vol. 6, no. 1, pp. 15-30, March 2009.

[6] A. Clemm, et al., “Intent-based networking - Concepts and definitions,” IRTF Draft, September 2020.

[7] ETSI NFV ISG, “ETSI Network Functions Virtualisation (NFV) Industry Standards (ISG) Group Draft Specifications,” http://docbox.etsi.org/ISG/NFV/Open, December 2014, Accessed on January 31, 2021.

[8] L. Han, R. Li, “Problems and requirements of satellite constellation for internet,” Internet Draft, July 2021.

[9] N. Hassan, K.L.A. Yau, C. Wu, “Edge computing in 5G: A review,” IEEE Access, vol. 7, pp.127276-127289, August 2019.

[10] Y.C. Hu, M. Patel, D. Sabella, N. Sprecher, V. Young, “Mobile edge computing - A key technology towards 5G,” ETSI white paper, no. 11, pp.1-16, September 2015.

[11] J. Jeong, “IPv6 wireless access in vehicular environments (IPWAVE): Problem statement and use cases,” Internet Draft, September 2021.

[12] H. Kim, N. Feamster, "Improving network management with software defined networking," IEEE Communications Magazine, vol. 51, no. 2, pp. 114119, February 2013.

[13] Cloud native computing foundation, "Kubernetes: Production-Grade Container Orchestration," https://kubernetes.io/, Accessed on 6 February 2022.

[14] A. Lazar, "Programming telecommunication networks," IEEE Network, vol. 11, no. 5, pp. 8-18, 1997.

[15] C. Li, et al., "Intent Classification," IRTF Draft, October 2018.

[16] J.R. Loyola, et al., "A methodological approach toward the refinement problem in policy-based management systems," IEEE Communications, Topics in Network and Service Management, vol. 44, no. 10, October 2006.

[17] Open networking foundation, "ONOS: open network operating system," https://opennetworking.org/onos/, Accessed on 6 February 2022.

[18] M. Satyanarayanan, "The emergence of edge computing," IEEE Computer, vol. 50, no. 11, pp.30-39, January 2017.

[19] P. Sethi, S.R. Sarangi, "Internet-of-Things: Architectures, protocols and applications," Journal of Electrical and Computer Engineering, January 2017.

[20] W. Shi, J. Cao, Q. Zhang, Y. Li, L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp.637-646, 2016.

[21] M.S. Sloman, "Policy driven management for distributed systems," Journal of Network and Systems Management (JNSM), vol. 2, no. 4, pp. 333-360, December 1994.

[22] G. Tangari, M. Charalambides, D. Tuncer, G. Pavlou, "Self-adaptive decentralized monitoring in software-defined networks," IEEE Transactions on Network and Service Management (TNSM), vol. 15, no. 4, pp. 1277-1291, 2018.

[23] G. Tangari, M. Charalambides, D. Tuncer, G. Pavlou, "Accuracy-aware adaptive traffic monitoring for software dataplanes," IEEE/ACM Transactions on Networking (ToN), vol. 28, No. 3, pp. 986-1001, June 2020.

[24] D.L. Tennenhouse, D.J. Wetherall, "Towards an active network architecture," Multimedia Computing and Networking, vol. 2667, pp. 2-16, January 1996.

[25] D. Tuncer, M. Charalambides, G. Tangari, G. Pavlou, "A northbound interface for software-based networks," In proceedings of 14th IFIP/IEEE International Conference on Network and Service Management (CNSM), Rome, Italy, November 2018.

[26] G. Zheng, et al., "Virtual data-plane addressing for SDN-based space and terrestrial network integration," In proceedings of 22nd IEEE International Conference on High Performance Switching and Routing (HPSR), 2021.

[27] W. Kellerer, et al., "How to Measure Network Flexibility? A Proposal for Evaluating Softwarized Networks", IEEE Communications Magazine, vol. 56, iss. 10, October 2018.

[28] M. Kuehlewind, et al., "Challenges in Network Management of Encrypted Traffic", https://arxiv.org/ftp/arxiv/papers/1810/1810.09272.pdf, white paper by the EU-H2020 MAMI project.

SESSION 2

AUGMENTED REALITY SYSTEMS: DESIGN AND IMPLEMENTATION

A FRAMEWORK FOR THE DESIGN, IMPLEMENTATION AND EVALUATION OF A MULTI-VARIANT AUGMENTED REALITY APPLICATION

Sophie Westfahl[1]; Dany Meyer-Renner[1]; Antoine Bagula[2]

[1]University of Applied Sciences Neu-Ulm, Germany

[2]University of the Western Cape, South Africa

ABSTRACT

Augmented Reality (AR) is one of the key technologies of the fourth Industrial Revolution (4IR) and plays an increasingly important role in many companies. However, while the demand for new AR applications is rapidly increasing, fundamental best practices and frameworks for the industrial AR sector are still scarce or in their infancy stage. This paper addresses this gap by proposing a framework for the design and efficient implementation of AR applications with multiple models and variants. The proposed framework is built around: i) a development process that describes the different steps for the design of a model-based AR application and its implementation with Unity and Vuforia model targets; and ii) a multilayer orchestration model that describes the different interactions between a user and a server layer. The proposed framework is successfully implemented, and its performance analyzed using both quantitative and qualitative evaluation based on the Brooke's System Usability Scale.

Keywords - Complexity of variants, industrial augmented reality, model-based tracking, unity, Vuforia

1. INTRODUCTION

Immersive technologies play an important role in the fourth Industrial Revolution (4IR). While Virtual Reality (VR) simulates a virtual world, Augmented Reality (AR) creates a highly visual, context-sensitive and interactive environment that extends the "real" physical world with virtual elements. The world is becoming increasingly complex and technologically focused, with most companies aiming to integrate trending technologies into their products. Among these technologies, AR is gaining momentum, especially in the industrial sector due to its innovative and competitive way of presenting information and enhancing the customer's experience. Moreover, as many companies are exposed to AR for the first time, fundamental approaches, best practices and frameworks for its development are tools which are increasingly needed by the industry. This need, for example, is illustrated by the automotive sector where AR is increasingly being used in both the manufacturing and maintenance operations. However, these tools are still scarce or in their infancy stage. This paper's aim is to fill this gap by proposing a framework for the design and implementation of AR applications with multiple models and variants.

The main contributions of this paper are threefold. Firstly, we propose an AR development process that describes the key steps needed for the design of a model-based AR application and its implementation with Unity [1] and Vuforia [2]. Secondly, a multilayer orchestration model that describes the different interactions between a user plane and a server plane of a multi-variant AR ecosystem is proposed. Lastly, the evaluation of an AR implementation is presented as a use case showcasing how a real car's main features and functionalities can be visualized with virtual animations and textual annotations to the main parts as augmented instructions. The remainder of this paper is as follows. Section 2 reports on related work while Section 3 presents the methodology used, and Section 4 proposes a generic approach for the design and implementation of a model-based AR prototype with Unity and Vuforia. Section 5 focuses on the design and implementation for a complexity of variants. The performance analysis of the proposed framework including a qualitative and quantitative evaluation is presented in Section 6. Section 7 presents a few lessons learned from the research work and Section 8 concludes the paper with a short discussion and subsequent summary.

2. RELATED WORK

The use of AR in the industrial sector is steadily increasing and deployed more and more in service, manufacturing, sales and marketing, design, operation and training. However, while plenty of developed use cases have been described by different authors [3], [4], [5], most of the AR literature refers to specific applications. Only a few exceptions provide general and fundamental approaches for development, but there is no approach for an implementation for multi-variant use cases. Linowes and Babilinski give an introduction to AR and present the development with Unity, Vuforia and other frameworks. Furthermore, they introduce different recognition targeting methods and major concepts for AR development in different areas [6]. Liull et al. provides a step-by-step approach for developing an image-based AR application with Unity and Vuforia [7]. Hameed et al. introduce the development with Vuforia object targets and tested the tracking quality in different circumstances [8].

These approaches can be used as guidelines for developing a model-based AR application with Unity and Vuforia, but are progressively outdated with technology upgrades. Furthermore, these approaches usually refer to only one model and are often neither practical nor realistic since industrial use cases often have to consider several models [9]. To the best of our knowledge, while most of the existing literature refer to one product, there is not yet any scientific approach proposed for the design and implementation for multi-variant products.

This research gap raises the issue of how a basic model-based AR prototype can be developed and implemented for products with numerous variants.

3. METHODOLOGY

The Design Science Research (DSR) method according to Oesterle et al. [10] was used as the methodology for the development of the proposed framework and its use case. Therefore, in an iterative process of analysis, design, implementation and evaluation, three artifacts were created. To simplify the development, at first a simple 3D model was used instead of the actual product. Therefore, the first and second artifact are abstractly illustrated by a 3D printed train model in toy size. Only the third artifact was developed for the real car.

The first artifact displays virtual information about a locomotive and is a best practice for developing a basic AR application with Unity and Vuforia. The second artifact deals with the complexity of variant and is also illustrated by a locomotive but with three additional wagons which represent the different variants. The last artifact deals with the real car and combines the findings from the first two artifacts. The application shows the AR information about multiple car variants. Each artifact was evaluated. In this paper, only the performance evaluation of the second variation artifact and the usability evaluation of the car application is examined more closely.

4. DEVELOPMENT OF AN AR APPLICATION

This section presents the development process of an AR application which includes the design phase and the implementation phase.

4.1 Design

To be able to expand the physical world, it must first be captured. This usually takes place via video stream which is recorded by a calibrated camera (visual tracking). Virtual objects can only be correctly integrated into reality if the AR system knows where the virtual object should appear and what spatial relationship the viewer (or camera) has to this position [11]. After determining the position, location and orientation of the camera, the target position is recognized. Targets can be e.g. horizontal surfaces, artificial markers or existing objects in the environment, such as 2D surfaces or 3D models. The latter is also referred to as model-based tracking and is based on edges, textures and optical flow.

In the following, particular attention will be paid to visual, model-based tracking with mobile devices. After target detection, the virtual content is incorporated into reality in a correct perspective by an optical overlay of the video stream. Virtual content can be e.g. 2D images or video, 2D or 3D text, 3D models as well as sound or others. As an output, the user can see virtual elements in correspondence to the real environment.

The development process can be described with six major steps (Figure 1) which can be traversed in an iterative process. Subsequently, a basic understanding of the individual steps should be conveyed. Best practices and implementation are described in Section 4.2.

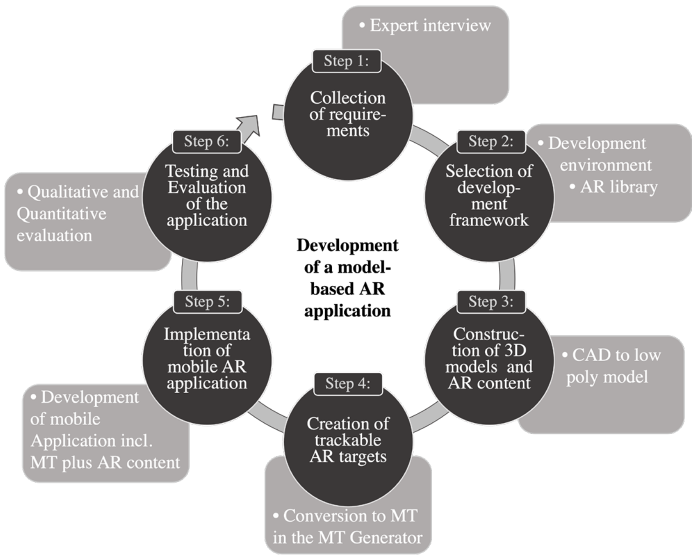


Figure 1 – Overview of the development process of an AR application

Step 1: Firstly, the functional and non-functional requirements towards the application must be collected.

Step 2: Selecting the right development environment is decisive for a successful implementation and can save a lot of resources. ARCore (Google) and ARKit (Apple) are well known but both are platform specific. There are also other platforms to be considered such as Wikitude, 8[th] Wall or Vuforia. Berger et al. examined several AR tools and compared the target tracking abilities [12]. Saar analyzed six frameworks regarding tracking and rendering together with Unity [13]. However, it is hard to find valid comparison literature because the technology is frequently further developed. In addition, the choice of the AR tool depends very much on the use case. Since many AR tools only provide the library for tracking and rendering, a separate development environment is needed. Here, the two tools Unity and Unreal usually stand out. According to Salama, both engines are not very different in their main functionalities. If a project requires graphically intense performance, Unreal Engine seems to be the better option. For small companies looking to target multiple platforms and monetize effectively, Unity is more suitable [14].

Step 3: The most important step of the development process is the collection of data. To augment a three-dimensional object, a virtual model of this object is needed. In some cases, a CAD model is already available, otherwise it can be created by 3D scanning it. One should ensure, that the object has enough edges or clear textures for the tracking software to recognize. Furthermore, 3D models are shown using polygons. Due to performance limitations of the end device, the number of polygons per model is limited.

Step 4: Some AR tools require a conversion from a 3D model to a trackable AR model. This procedure depends on the chosen platform.

Step 5: In the development environment, the 3D model and the virtual content must be connected. This may be done via easy drag and drop or via scripts. Also, additional behavior can be determined, or user interfaces can be created here. This step includes all programming-related steps.

Step 6: After finishing with development, the application can be tested and executed to the end device. The output is either a native application or a project that can be included in an already existing application. Some platforms also offer WebAR, which can be accessed by a QR code for example, and no mobile application is needed. Especially the tracking quality should be tested.

4.2 Implementation

The implementation is described according to the above-mentioned development process. For simplicity, it is not yet implemented for the car but a 3D printed locomotive.

Step 1: Our requirements towards the application were collected in an expert interview with the stakeholder. For the first artifact, the requirements are to become familiar with the technology and implement a first "Hello World" application which displays random AR content on the locomotive. The application should then be executed for Android and iOS.

Step 2: For this project, we used Vuforia and Unity for development. While Vuforia provides the AR library and all incoming functions regarding AR, Unity is a multifunctional cross-platform game engine for the development environment. As of now, Vuforia has a free basic plan for up to 20 Model Targets (MTs) and supports all common platforms. Additionally, a detailed documentation and developer forum is available. Unity's personal plan is free, and they offer a free learning path with video tutorials and live sessions. Both platforms are intuitive to use and especially Unity as beginner friendly as there are many tutorials and forum articles. Vuforia can easily be imported to a Unity project after creating an account for both platforms. After import, just a free license key must be entered in the Vuforia Engine settings.

Step 3: Before the actual development of the AR application, the most important step is the creation of 3D models for the AR target and virtual content. For the application to track a real object, the exact 3D model of it is needed. The model can deviate by up to 20% from the real object according to Vuforia. Additionally, these 3D models must not exceed 20 parts and 400,000 polygons because of the end device performance [15]. This is particularly a problem when the 3D data comes from CAD data. In addition, a smaller number of polygons automatically leads to a smaller data size of the models. The number of polygons for the virtual content is not explicitly limited, but also here the application performance is much better if the models are smaller. As a practical hint, it can be said that it is best or advisable to already set up the position of the 3D model to the coordinates (0|0|0), as these can no longer be adjusted, and it makes subsequent editing in Unity much easier. It is also important to distinguish that in most CAD programs the z-vector points upwards while in Unity it is the y-vector.

Step 4: For the real object to be tracked, its 3D model must be converted to an MT, which is Vuforia's term for 3D targets. Therefore, the 3D model needs to be imported to the Vuforia Model Target Generator, a separate program that converts an existing 3D model into a Vuforia Engine dataset. Several file formats can be used for import, but experience has shown that *.fbx* is the best, because it is most common and can also be used in Unity later. In the Model Target Generator, scale and texture might be adjusted. Next, the complexity of the model is shown inclusive of the number of polygons. If required, the model can be automatically simplified here. After that, it must be specified how much the object is moving. One can choose the motion hint between static (no motion of object), adaptive (limited motion of object) and dynamic. Using static motion will significantly improve tracking quality and save power during tracking. Lastly, a guide view can be selected. A guide view is a preview image of the real object which indicated to the user the distance and angle for fastest tracking. When the AR system has successfully tracked the real object, the guide view disappears. Especially for inexperienced users, it is a good assistance for tracking. Finally, the MT can be created, and a dataset is generated.

Step 5: After creating the MT dataset, the next step is the actual programming of the application. A simple augmentation of an object can be done without writing a single line of code. In the Unity project, import the MT dataset, delete the camera and replace it with an AR camera. In the hierarchy an MT object can be added, and the imported dataset is automatically selected and appears in the scene view. From here, any AR content can be created. Our MT is a locomotive, and the AR content are red wheels and a 3D text. The AR content must be made as a child object of the locomotive in the hierarchy. For demonstration, either a webcam can be connected and the project be played in the editor, or the project is executed on a mobile device.

Executing the project for mobile devices is done within the build settings. Therefore, a camera usage description must be added in the player settings. For Android, the scripting backend needs to be changed from from ”Mono Scripting” to ”IL2CPP” and both ”ARMv7” and ”ARM64” architectures need to be activated in addition due to missing ARCore support for 32-bit apps on 64-bit devices from Google. For iOS, an *.xcodeproj* file is created which can be distributed to a device via XCode. For Android, an *.apk* file is created which can be directly transferred to the device.

Step 6: The images in Figure 2 show the result on an Android-device with a 12 MP camera. When the application is opened, the guide view appears and the locomotive can be tracked. After successful tracking, the guide view disappears, the wheels are highlighted in red, and a 3D text is displayed. For this artifact we only tested the tracking quality which was good.

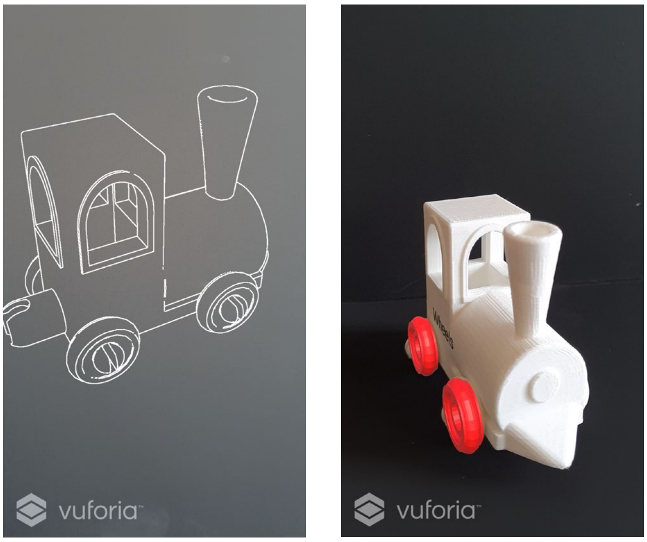


Figure 2 – Outcome of the application with a locomotive MT

5. FRAMEWORK FOR HANDLING THE COMPLEXITY OF VARIANTS

The design of a framework for handling the complexity of variants is presented in this section, followed by its basic implementation. The section concludes with a more efficient implementation of the variants.

5.1 Design

A multi-variant product is a product with multiple variants which can arise from different situations. For our augmented car, we started to augment one single car model but, in practice, the number of models should not be limited to one. This is because the car can have several model series or different customizations, such as an upgrade on the sunroof, the rear trunk or other car features. An AR application should work for all the variants. This led to a multi-variant application with a number of variants which can amount to several hundred MTs because each single variant is, in principle, a different 3D model that needs to be augmented. The storage of each MT locally on the end device is a challenging issue which can result in a huge application size and poor performance. Moreover, in many situations, a multi-variant product is developed while only one variant is usually needed by the user. The idea behind our framework is to have all variants stored remotely and at usage, when the user enters the variant number in the application, only the matching variant is downloaded. To visualize the different variants, a previously used locomotive is extended with three wagons representing other variants. However, as explained above, when using the application, a user must select a variant, such as wagon 1 for example, and only the 3D model associated to the variant is downloaded to the application.

For our use case, we used a server for remote storage, but any other location and computing infrastructure can be used. The remote storing of the variants consists of storing on a remote server or suitable computing infrastructure the MT with AR content and its corresponding dataset. Both are uploaded to the server and later downloaded on demand, during runtime, to the application. To upload the MT with AR content to the server, it must be converted into a suitable file format. For this, Unity offers the so-called AssetBundles for downloading non-code Assets at runtime. The uploading of the MT dataset to the server requires checking closely its structure. An MT dataset contains two kinds of files: i) dataset information files imported in the Unity Editor folder and ii) a .xml and .dat file containing the 3D model, which are imported to the Streaming Assets folder in Unity. Files in the Editor folder are only for development in Unity and not available in builds at runtime. The Streaming Assets folder contains files that are placed in the normal file system of the target device and can be accessed via a path name.

If all files are remotely stored on the server, they must somehow be downloaded from the server after a request from the application. The mapping of the files can simply be done by the file names. AssetBundles can be directly downloaded to the application, or they can be loaded as an addressable asset. Here the asset is assigned an address via which it can be loaded. Furthermore, there is another approach called cloud content delivery which is used for variants that are frequently updated with their previous versions getting outdated. Using cloud content delivery, the assets are directly stored in a cloud database and the current version is marked. This service is only free for the first 50 GB of bandwidth every month.

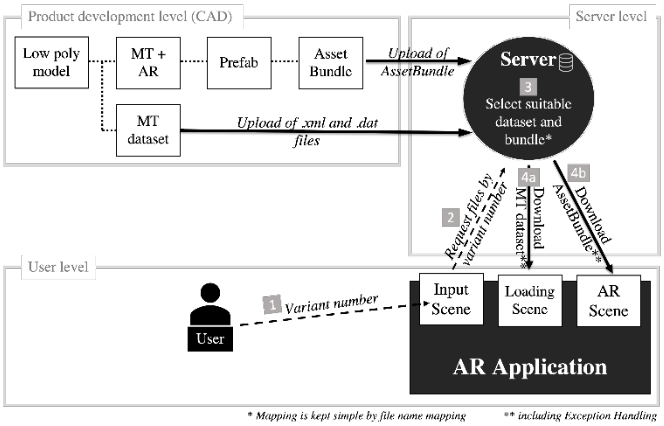


Figure 3 – AR orchestration model

Figure 3 illustrates an AR orchestration model for the implementation with AssetBundles. The orchestration model reveals three levels. On the user level the variant is selected, and the application is executed. The server level stores all MT dataset files and AssetBundles and is responsible for the mapping, which is simply done by file names. The product development level is where the CAD files are made for AR targets.

It is also possible to only store the MT with AR content remotely, and the dataset remains in the application. This makes sense when the model is always the same, but the AR content changes frequently. Similarly, for a smaller application size with larger data size, only the MT with AR content can remain in the application and the dataset is loaded dynamically.

5.2 Implementation

The structure of the application is divided into *Input scene*, *Loading scene* and an *AR scene*. In the Input scene, the variant is selected while in the Loading scene, the corresponding dataset is downloaded from a server, and in the AR scene, the AssetBundle is loaded and afterwards detected, tracked and augmented.

All models are made into MTs and provided with AR content according to the generic approach described above. Afterwards all scenes are created. To upload the MTs with AR content to the server, they have to be made to AssetBundles. Since AssetBundles are not included in Unity by default, an extra C#-script must be added to the editor folder. Each MT first needs to be made to a Prefab, which is a Unity format and acts like a template. Then it is converted to an AssetBundle. Two separate bundles for Android and iOS are automatically created, because they differ for each platform. All AssetBundles are stored in the project folder and can be uploaded to a server from there. For demonstration purposes, a domain was bought, and a Secure File Transfer Protocol (SFTP) access created. A server and client software can be used to connect to the server and execute file transfers. After the upload, all files can be deleted in the project or remain in there to update them later.

In addition to the MT and AR content, the MT dataset must also be uploaded. The Streaming Assets files .xml and .dat contain the 3D model to be tracked and the Guide View. Both files are also uploaded to the server. Afterwards, they can be deleted from the project in the Vuforia folder inside the Streaming Assets. The other files can remain in the project for direct test of the application in the Unity Editor or they can also be removed. All files can now be accessed by the URL e.g. *url.com/filename*.

After successfully uploading all files, the structure for downloading them to the application on runtime must be created. In the input scene, a unique serial number is entered by the user. This indicates which variant must be downloaded.

While the dataset files need to be downloaded, before Vuforia is initialized (e.g. before the AR scene), all MTs with AR content are directly loaded by the AR scenes. The dataset is stored locally on the device and the AssetBundles are just temporarily available and are unloaded when exiting the scene. The dataset files are loaded using *UnityWebRequest*. If those files already exist on the device, they are overwritten. During development, it turned out that during runtime, files cannot be downloaded to the Streaming Assets folder at iOS. Therefore, the dataset files are stored with the persistent data path, which points to the user library folder, in which data can be stored that wants to be kept between runs.

To load the AssetBundles at runtime, an empty MT object needs to be created in every AR scene, which later gets assigned by the downloaded AssetBundle.

5.3 Modularization of variants

For some use cases, many variants just slightly differ from each other. Car variant 1, for example, has four doors, a central locking system and a radio. On the other hand, Car variant 2 also has 4 doors and a radio but manual locking. If both variants need to be augmented, most of the AR content will be the same and it will be redundant to create the AR content for the radio and the four doors twice. To spin this further, if for example the AR content for the radio changes, all variants need to be updated manually and in the worst case, this affects a large number of models. For this reason, we have considered disassembling all variants into their individual components. In that way, we do not have the whole car anymore, but all its sub-models such as the radio, doors, central locking system and manual locking. The variants (different car models) are no longer loaded as a whole, but the individual components (radio sub-model, etc.) are loaded on demand.

We have expanded the above-described framework for variants by the components and again abstractly implemented this with the train models. Technically, no new knowledge is needed, but the architecture of the application slightly changes. A train variant consists of three components (e.g. locomotive, wagon 1, wagon 2 and wagon 3). When the user enters the application, the train variant number is entered. A database or comparable system stores the information on which train number contains which components. In the

Loading scene, all component datasets are then downloaded. Theoretically, all components can then be tracked at once. But currently, with Vuforia, only one MT can be tracked at the same time. That is why we created a separate AR scene for each component and the scenes are tracked successively.

If the user selects train number 001, the train consists of the locomotive, wagon 1 and wagon2. All three datasets (but not the dataset of wagon 3) are downloaded from the server. In the first AR scene, the locomotive AssetBundle is loaded and augmented. In the second AR scene, wagon 2 is augmented and the AssetBundle is loaded.

It is assumed that the locomotive is part of five train variants and the AR content for it needs to be updated. Instead of changing five variants, only the one component is changed. Since all components are remotely stored, the application does not even need to be updated with what is better for the user.

6. EVALUATION

This section presents the evaluation of all prototypes as envisaged in design science research. The evaluation of the various prototypes with multiple components and the final prototype are described below.

6.1 Performance

To evaluate the performance major actions were marked in scripts and printed to the console with a time stamp. These actions include:

- User activities including scene changes
- Successful download of files
- Application requests to the server
- Instantiating the empty MT object with the AssetBundle
- Target tracking status
- Unload time of an AssetBundle

The application was executed on an iOS device and the performance of the XCode console output was evaluated. The app started on an iPhone X and a speed test indicated a download rate of 100MB/s. We used a simple virtual server that is reachable for everyone. The used dataset files had an average size of 83 KB.

In Table 1, the significantly reduced XCode console output can be seen. Line 1 shows that the app has started and Vuforia is initialized. Then, the start scene is opened and the variant with its components is selected.

The *Loading scene* is entered in second 30.09 and all dataset files (3 .dat and 3 .xml files) are downloaded to the device within 0.69 sec. The dataset of Wagon 3 was not downloaded, because it was not selected in the *Input scene*.

After the successful download, the first AR scene to track the first component is automatically opened. The AssetBundle is requested from the server and downloaded. The correct dataset is automatically created out of the earlier downloaded .xml and .dat file. It took 0.66 sec., from entering the *AR scene* to an instantiated MT (line16). Line 17 indicates the presentation of the guide view and line 18 a successful tracking of the locomotive. From showing the guide view until successful tracking, almost a second passed. This value heavily depends on the user and in what angle the device is held. Successful tracking of Wagon 1 (lines 27 and 28) took more than 4 sec. When the next scene is opened, the MT resets tracking (line 19) and the AssetBundle is unloaded (line 20).

The performance evaluation is given by time but cannot be classified and compared to other frameworks or applications because there is no reference. For the given use case it was important to get results that are acceptable for the user and less than one second to download all datasets and about 2 seconds to load and track the MT is in an acceptable range.

Table 1 – Performance evaluation: XCode console output

line	***XCode* Console output**
1	14:35:14.39 Vuforia Initialized
2	14:35:15.46 Entering Start scene
3	14:35:29.56 Components selected.
4	14:35:30.09 Entering Loading Scene
5	File saved at: */var/mobile/Containers/Data/Application/Documents/ Vuforia/Lok.xml*
6	File saved at: . . . */Vuforia/Lok.dat*
7	File saved at: . . . */Vuforia/Wagon1.xml*
8	File saved at: . . . */Vuforia/Wagon1.dat*
9	File saved at: . . . */Vuforia/Wagon2.xml*
10	File saved at: . . . */Vuforia/Wagon2.dat*
11	14:35:30.77 Entering AR Scene 1
12	Requesting bundle at *http://www.westfahlsophie.com/Lok-IOS*
13	Creating dataset: Vuforia/Lok.xml/Lok
14	14:35:30.83 Target: - EMPTY - NO_POSE - NOT_OBSERVED
15	Loaded: ModelTarget (4)(Clone)
16	14:35:31.43 Target: Lok NO_POSE - NOT_OBSERVED
17	14:35:31.48 Target: Lok NO_POSE -RECOMM_GUIDANCE
18	14:35:32.41 Target: Lok TRACKED - NORMAL
19	14:35:34.52 Target: Lok NO_POSE - NOT_OBSERVED
20	UnloadTime: 7.174000 ms
21	14:35:34.49 Entering AR Scene 2
22	14:35:34.53 Target: - EMPTY - NO_POSE - NOT_OBSERVED
23	Requesting bundle at *http://www.westfahlsophie.com/Wagon1-IOS*
24	Creating dataset: Vuforia/Wagon1.xml/Wagon1
25	Loaded: ModelTarget (1)(Clone)
26	14:35:34.89 Target: Wagon1 NO_POSE - NOT_OBSERVED
27	14:35:34.95 Target: Wagon1 NO_POSE -RECOMM_GUIDANCE
28	14:35:38.97 Target: Wagon1 TRACKED - NORMAL

6.2 Usability

Brooke notes that usability as such cannot be clearly defined and is always dependent on a context and a user group. To still be able to quantify and evaluate usability, he presented the System Usability Scale (SUS) in 1996 [16]. It includes ten statements about the user-friendliness of an app, which test subjects can agree or disagree with. One vote is given for each statement on a Likert scale ranging from one (strongly disagree) to five (strongly agree) [16]. Brooke's proposed questions were slightly customized and are shown below:

1. I think that I would use the AR application frequently to get information about my car.
2. I found the AR application unnecessarily complex.
3. I thought the AR application was easy to use.
4. I think that I would need the help of a technical person to use or understand the AR app.
5. I found the various functions of the AR application were well integrated.
6. I thought that the AR application still has many bugs and doesn't work properly.
7. I would imagine that most people would learn to use the AR application very quickly.
8. I found the AR application very cumbersome to use.
9. I felt very confident using the AR application.
10. I needed to learn a lot of things before I could use the AR application.

The result of the SUS is a value in a range from 0 to 100. The questions are structured in such a way that in the best case, the surveyed alternately ticks the rightmost (odd-numbered questions) and then the leftmost (even-numbered questions). For each question, a score between zero and four is calculated. The sum of all scores is multiplied with 2.5 to get the final SUS score. The following rules apply to the calculation of the questions:

- Odd-numbered questions: *Score= arithmetic mean-1*
- Even-numbered questions: *Score= 5-arithmetic mean*

The usability of the final prototype was evaluated with Brooke's SUS. The application uses AR to display context-information about multiple car models. We conducted the experiment with 11 test subjects, who were asked to test the prototype and answer ten questions with predefined answers via an online survey. The participants do not represent the existing diversity of age, gender and user behavior, because at this stage of preliminary work it was not necessary to provide representative results. Nevertheless, it is worth mentioning that no testers had previous experience with an AR instruction application.

We calculated the SUS score of our application with the results from the survey according to the above explained method and obtained an SUS-value of 90. To classify this value, a grading scale presented by Lewis and Sauro was used, to assign the SUS score a grade and a normalized percentage. They determined an average value of 68 and assigned it the grade "C" [17]. Comparing the calculated SUS result to the grading scale, it is rated with the best grade (A+) and therefore belongs in the percentage range of the best 96-100%.

7. DISCUSSION AND LESSONS LEARNED

This section discusses the approach of the framework and presents some key lessons learned during the research work.

7.1 Discussion

The proposed framework describes the implementation for model-based use cases with multiple variants. In contrast to other previous referred approaches, this framework provides state of the art information about the implementation with Vuforia MTs and the handling of a variety of models. The usage of AssetBundles is only one possible solution, but it is most suitable for the given use case.

The evaluation regarding usability is not representative at this stage of work, because the number of participants is not sufficient, whereby the application is still in an infantile state. Nevertheless, a first impression on usability may be contained and that downloading models during runtime has no negative influence on it. The performance evaluation is given by time, whereas the values strongly depend on Internet speed, end device and model size. The results cannot be referenced or compared to other frameworks but can be classified subjectively in relation to the requirements of the use case.

7.2 Lessons learned

During the development process we faced multiple challenges and learned a few lessons.

- We started developing without any previous experience in AR and only basic programming skills. Especially for novices, it is important to have fundamental blueprints and best practices. There is a lot of literature about specific use cases, but there is a scarcity in fundamental and generic approaches for design, implementation and evaluation of model-based AR applications.

- When it comes to practice, generic approaches come to their limits due to multi-variant products. In the automotive industry, offering customer-specific variations has been standard since the early 1990s. These must be implemented as efficiently as mass production which also includes an efficient creation of instruction material.

- A car always has different series and customizations, and each of them is another variant of the original car with another instruction. Each variant must be implemented and included in the instructional application which leads to a large amount of data and inefficient performance. Such a handling of the complexity of variants in an AR context has not been adequately researched yet.

8. CONCLUSION

The contribution of this paper is the development of a framework for the implementation of a mobile AR application with model-based tracking for multi-variant products. We started by introducing a fundamental approach for designing and implementing model-based AR applications with one model. This approach was expanded with a framework for the design and implementation for numerous variants due to the fact that most of the industrial use cases involve more than one model. Some of the key features of the proposed multi-variant framework include:

- It was designed to store the MT with the AR content and its dataset remotely e.g. on a server. This can enable endlessly models to be implemented with a small application size and good application performance. As implemented, during runtime, the application requests only the demanded dataset and MT, which is stored as an AssetBundle. Both are then downloaded.

- It does not only work for products with multiple variants but can also be used for applications with a lot of different products. Because all models are stored remotely, updating is easier and user-friendlier because no application update is required for every model update. In a last step, we presented a best practice for extensive models. Many products have multiple variants but they only differentiate in single parts. Therefore, it makes sense to disassemble those products into its single components and augment each component. This allows a more efficient creation of AR models and simplifies updating single parts.

- It is not limited to model-based targets. It can be implemented with other targets as well and even with VR assets because it is a general approach for loading assets dynamically on demand during runtime. It is successfully implemented in a real industrial use case and has proven itself.

- Furthermore, it was evaluated quantitatively and qualitatively based on its usability. For the performance evaluation, the scripts were commented and the XCode console output analyzed. The results are not comparable but it can be said that from a user-perspective, they are acceptable. The usability was evaluated with Brooke's System Usability Scale and achieved a SUS-value of 90 which corresponds to an A+.

REFERENCES

[1] Unity Real-Time Development Platform | 2D, 3D VR & AR Engine: https://unity.com/

[2] Vuforia Developer Portal: https://developer.vuforia. com/

[3] J. Murthy, R.R. Dsouza, and A.R. Lavanya, "Visualize the 3D Virtual Model Through Augmented Reality (AR) Using Mobile Platforms," in *Recent Advances in Civil Engineering*, Springer, p. 837-863, 2022.

[4] A. Adam, Y. Yusof, K. Kamran, A.Z.A. Kadir,

[5] T.H. Sam, S. Nazrein, and D.A. Memon, "The Implementation of a Novel Augmented Reality (AR) Based Mobile Application on Open Platform Controller for CNC Machining Operation," in Recent Trends in Mechatronics Towards Industry 4.0, Springer, pp. 227-234, 2022.

[6] C.O. Reddy, "Data visualization using marker-based Augmented Reality techniques," Journal of Positive School Psychology, pp. 3850-3856, 2022.

[7] J. Linowes, and K. Babilinski, "Augmented reality for developers: Build practical augmented reality applications with unity, ARCore, ARKit, and Vuforia," Packt Publishing Ltd, 2017.

[8] X. Liull, Y.H. Sohn, and D.W. Park, "Application development with augmented reality technique using Unity 3D and Vuforia," International Journal of Applied Engineering Research, vol. 13 no. 21, pp. 15068-15071, 2018.

[9] Q. Hameed, H.A. Hussein, M.A. Ahmed, and M.B. Omar, "Development of Augmented Reality-based object recognition mobile application with Vuforia," Journal of Algebraic Statistics, vol. 13 no. 2, pp. 2039-2046, 2022.

[10] M. Hüllenkremer, "Erfolgreiche Unternehmen arbeiten mit Produktkonfiguratoren/Companies boost success with product configurators," Industrie Management, Publisher, Location, vol. 19 no. 1, pp. 37-40, 2003.

[11] H. Oesterle, J. Becker, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E. Sinz, "Memorandum zur gestaltungsorientierten Wirtschaftsinformatik, Zeitschrift für betriebswirtschaftliche Forschung, vol. 6 no. 62, pp. 664–672, 2010.

[12] M. Tönnis, "Augmented Reality: Einblicke in die erweiterte Realität," Springer, 2010.

[13] C. Berger, and M. Gerke, "Comparison of Selected Augmented Reality Frameworks for Integration in Geographic Citizen Science Projects," The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 43, pp. 223-230, 2022.

[14] O. Saar, "Survey of Existing Augmented Reality Frameworks," 2021.

[15] R. Salama, and M. Elsayed, "A live comparison between Unity and Unreal game engines," Global Journal of Information Technology: Emerging Technologies, vol. 11 no. 1, pp. 1-7, 2021.

[16] PTC Inc., Model Target Generator User Guide: https://library.vuforia.com/objects/model-target-generator-user-guide

[17] J. Brooke, "SUS: A quick and dirty usability scale," Usability Eval. Ind., vol. 189, 1995.

[18] J. Lewis, and J. James, "Item Benchmarks for the System Usability Scale," vol. 13, pp. 158-167, 2018.

ENHANCING USER EXPERIENCE IN PEDESTRIAN NAVIGATION BASED ON AUGMENTED REALITY AND LANDMARK RECOGNITION

Dhananjay Kumar[1]*; Shreayaas Iyer*[1]*; Easwar Raja*[1]*; Ragul Kumar*[1]*; Ved P. Kafle*[2]

[1]Department of Information Technology, Anna University, MIT Campus, Chennai, India

[2]National Institute of Information and Communications Technology, Tokyo, Japan

ABSTRACT

Pedestrian navigation using traditional mapping systems is constrained by the inherent limitations of existing digital online mapping services. The major challenges include complete reliance on GPS for user localization and inferior user experience caused by lack of information about the surroundings, especially in unknown environments. In this paper, we design and develop a marker-less augmented reality-based pedestrian navigation system which can handle navigation even in the absence of GPS as well as improve user experience by providing a novel landmark recognition feature, which allows users to identify nearby buildings or streets during navigation. To mitigate the absence of a GPS signal, a user localization method utilizing a step count-based distance estimator is proposed. The performance comparison with existing state of the art techniques and devices shows locational accuracy of 2.5 meters on average and a step count detection accuracy increase of nearly 0.5% with a latency of 70 milliseconds in an urban environment. The proposed solution is intended to be used as a mobile application on smartphones and has a potential to contribute to the smart city-related standardization activities of ITU-T Study Group 16.

Keywords – Augmented reality, landmark recognition, pedestrian navigation, step count estimation

INTRODUCTION

Mobile Augmented Reality (MAR) is rapidly growing with reports indicating a compound annual growth rate of nearly 26% by the year 2025 [1]. It is believed that in the year 2021, nearly 810 million people had access to MAR and projected growth indicates this number to rise up to 1.73 billion people in 2024 [2]. The applications development in MAR need to support the real-life use cases adapted using augmented reality providing new avenues for day-to-day activities. For example, in tourism it allows travelers to broaden their perception of their physical environment by providing virtual information about various available activities [3]. In the field of navigation, MAR faces a few challenges in terms of user interface mainly due to the size and computational power variance of various smartphones [4]. In a location-aware system, one of the objectives is to maintain a constant user interface across all user locations. Locational awareness in mobile devices can be brought about by using sensors such as a Global Positioning System (GPS), accelerometer and magnetometer which allow any MAR application to pinpoint the location of the user. This location awareness allows a MAR system to assist navigation, especially for pedestrians (as shown in Figure 1). Usage of augmented reality in navigation enhances it by rendering virtual signs or information stamps on the user's screen. Traditional navigation systems using a two-dimensional online mapping service (e.g., Google Maps, Bing Maps), although they provide state of art directional services, are constrained by their inherent limitations. Popular digital map services such as Google Street View do not provide sufficient information required for the detection of various points of interests during navigation. The visual data in Street View may be outdated. Furthermore, Google Maps does not provide walking directions in offline mode [5]. The existing popular MAR navigational systems (e.g., Google Maps Live View) do not support AR-based navigation in areas where Google Street View is not available [6]. There is a need to incorporate landmark detection in MAR-based pedestrian navigation in order to improve user experience in real-time.

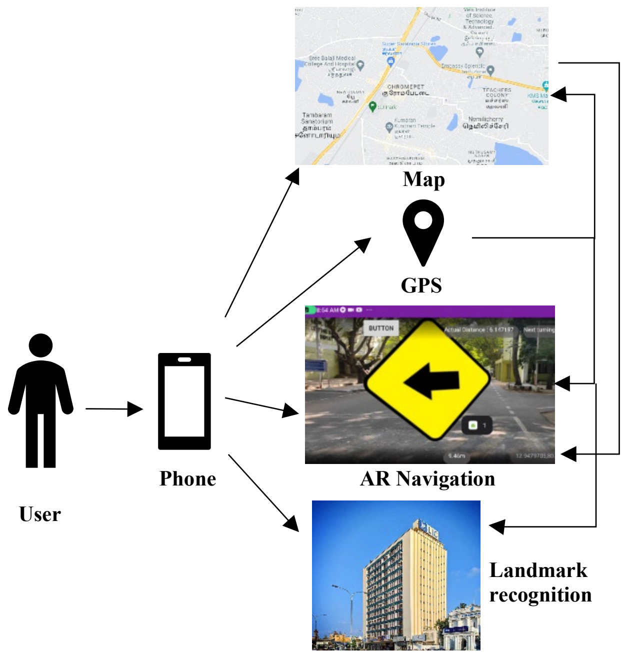


Figure 1 – Schematics of pedestrian navigation

The absence of GPS signals in public places such as underground shopping malls, subway stations, and basements of multistorey buildings constitutes a requirement of a new user localization mechanism in digital online maps to support walking directions. Counting the footsteps of the user is useful in estimating the distance covered by them between successive GPS signals. Real-time step counting algorithms are useful for indoor pedestrian navigation by creating a probabilistic map which stores geographic data primitives [7]. Accurate step counts can be estimated by using a high pass filter on the accelerometer readings. Another key parameter while estimating distance from step counts is to take into account the step length of the user. However, step length changes dynamically while the user walks and this step length needs to be detected at every point [8]. One of the shortcomings of existing step counters for Android devices is the latency in generating the step count. According to the official Android documentation [9], the built-in step counter requires at least 10 seconds to display accurate step count. In our proposed work, we aim to reduce this step count latency by leveraging a temporal filtering technique coupled with a threshold-based magnitude filter.

Recommendation ITU-T Y.5462 [10] discusses spatio-temporal information services for smart cities requiring navigational and positioning systems for citizens to determine a person's location as well as provide a digital mapping system displaying instructions for moving from one place to another. The MAR-based solution for pedestrian navigation needs to be considered for the enhancement of user experience.

In this paper, we present a marker-less Mobile Augmented Reality-based Pedestrian Navigation System (MAR–PNS), which provides a visual solution for the enhancement of user experience by incorporating landmark detection in pathways, and improves the accuracy of a navigational system in the absence of GPS signals. The system uses an AR framework to render the turning points and landmarks on demand by the mobile user. It allows landmark recognition by relying on the user's current location and direction in which they're pointing their phone. The system mitigates the absence of GPS in scenarios such as underground subways by integrating a step count-based user localization method. The MAR–PNS model complies with Recommendation ITU-T Q.4066 "Testing procedures of augmented reality applications" [11] and can be a solution for smart city navigational systems, while meeting the requirements of Recommendation ITU-T Y.4562 "Functions and metadata of spatiotemporal information service for smart cities" [10].

The remainder of the paper is organized as follows. Section 2 provides the architectural details of the proposed system and Section 3 describes the algorithms developed during the course of our work. Section 4 describes the implementation environment and evaluates our system in terms of locational accuracy and other metrics. Section 5 provides concluding remarks and offers scope for future work.

2. PROPOSED SYSTEM MODEL

The architecture of the proposed MAR–PNS consists of three modules namely the AR Navigation (ARN) module, Step count-based Distance Estimation (SDE) module, and the Landmark Recognition (LR) module (Figure 2). The AR Navigation module involves getting the optimum route between source and destination on digital map, details about the pedestrian user (such as estimated length of distance between legs), turning point coordinates, and generating / rendering AR turning point indicator objects according to the proposed algorithm. The SDE module helps in localizing the user's position with respect to the turning point in the absence of GPS. There are two major inputs from the user i.e., the destination address and the live feed captured by their mobile camera. The destination point-based route information provided by the digital map service is integrated with the ARN module along with the LR to enhance the pedestrian navigation in real time. The LR module estimates the location of the landmark which the user's camera is pointed to and displays relevant information about the landmark in the MAR device. The AR objects generated by the ARN and LR modules are anchored to their geo-positions, localized with respect to the user's camera view and rendered on the user's screen.

2.1 AR Navigation (ARN) module

The ARN module is designed to display walking direction in an AR environment based on turning point markers. It uses a digital mapping service to obtain an optimal route from source to destination. The last known location of the user is obtained from GPS and the user's position is localized by polling the user's position either by using GPS or the step count-based distance estimation module. Whenever the user is near a turning point, the turning point direction (AR object) is generated (right or left direction depending on the turning point) and rendered on the user's screen.

2.2 Step count-based Distance Estimation (SDE) module

This module mitigates the unreliability or unavailability of GPS during navigation. It uses the step count of the user and performs arbitrary calculations to translate the number of steps to the distance covered by the user using parameters such as the user's height and waking duration. The trade-off decision in minimizing latency was taken based on the fact that in pedestrian navigation a few missteps do not affect the overall working of the application and an error threshold can be maintained. This design consideration is also taken keeping in mind that even intermittent GPS signals can help in localization of the user with the distance estimation module. The other design requirement includes the overall physique of the user which can affect the foot stride.

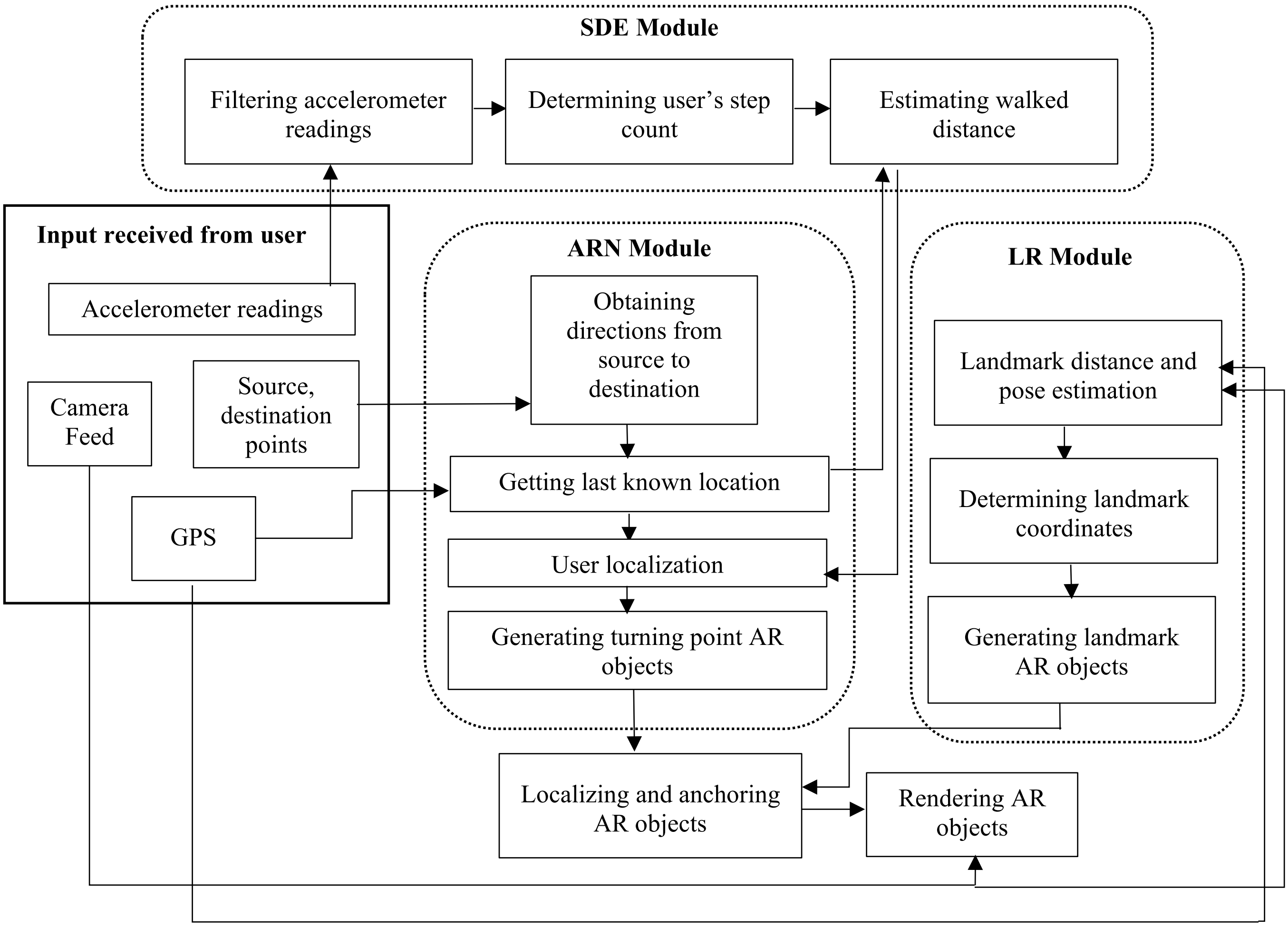


Figure 2 – Proposed system architecture

The function of the SDE module involves filtering the real-time accelerometer readings obtained from the user's phone using the proposed algorithm to determine the step count and, in turn, estimate the distance walked by the user.

2.3 Landmark Recognition (LR) module

This module aims to help the user identify landmarks such as buildings or roads during the course of navigation. It allows the user to point their camera at some building and displays the name of the building or the landmark using AR objects rendered on the user's handheld device screen. The main advantage of the LR module lies in its simplicity and ease of use without using too much of processing power. It uses user-specific data such as their height to calculate the distance of the user from the landmark they are pointing to and uses the phone's built-in sensors to approximate the pose of the phone with respect to the True North. The LR module can display details about the landmark to the user by estimating its latitude and longitude and generating AR objects.

3. ALGORITHM DEVELOPMENT

The three major modules in the proposed MAR–PNS system require independent designing and development of algorithms. The ARN and SDE modules are designed to work in tandem in areas where a GPS signal is not available. However, the LR module uses GPS to recognize landmarks.

3.1 Algorithm of ARN module

The ARN module is used to render the AR turning point objects whenever the user is near a particular turning point. This is achieved by localizing the user's position with respect to the turning point either by using GPS or by using the step count-based user localization (SDE module) in the absence of GPS. The flowchart of the ARN algorithm is depicted in Figure 3. The online directional map service finds an optimal route from source to destination. Since the turning points obtained from the route appear one after the other, a first-in first-out pattern is used to process them. The user's position is polled at every point during navigation to find whether the user is at a particular distance away from the turning point. Using the user's location, a decision is made to determine whether the turning point has been crossed by the user before or not. If not crossed, the turning point is localized and

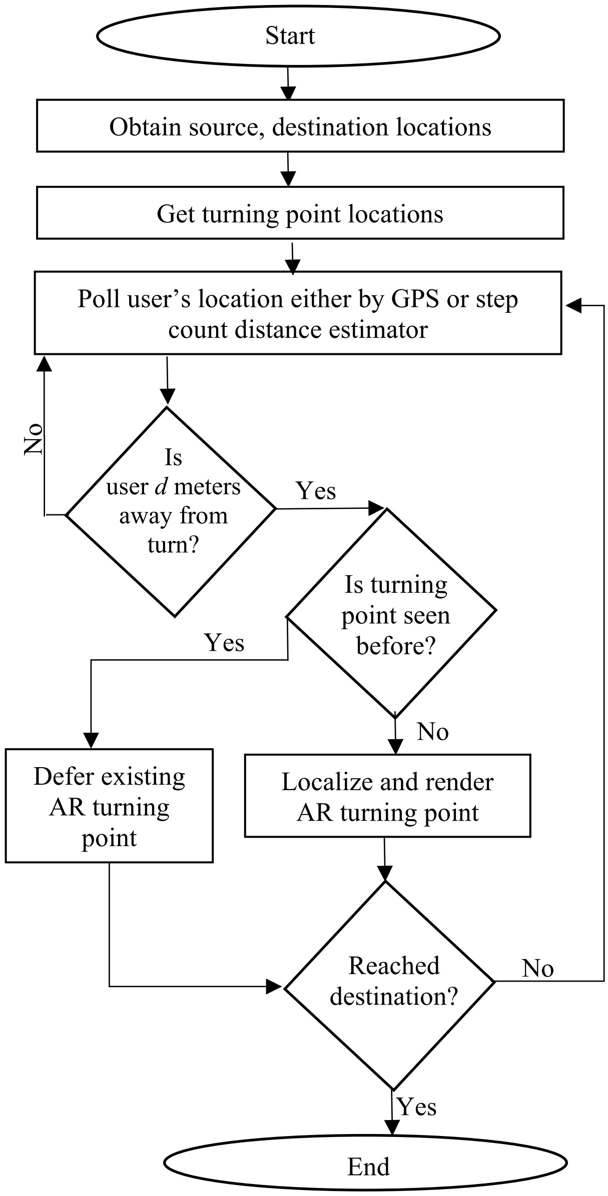


Figure 3 – Flowchart of ARN algorithm

rendered on the user's screen. Otherwise, the current turning point AR object is removed, and the process is continued until the user reaches the next direction point, and finally to the destination. The step count-based distance estimation algorithm (Figure 4) is used to determine the steps walked by the user by filtering their phone's accelerometer data and estimating the distance covered by the user as a product of the calculated footsteps and the length of the foot stride. The algorithm involves finding the magnitude threshold of accelerations. The readings exceeding the threshold are temporally filtered to ensure that each value of potential footstep is at least x milliseconds after the previous footstep. This temporal filtering ensures reduction in false positives which could be caused by noisiness of the phone's sensor. Before filtering the accelerometer data, it is normalized since raw accelerometer data is obtained in three axes namely X, Y and Z. After each foot-step detection, the cumulative count is multiplied by the foot-stride length of the user to obtain distance covered by the user. The foot-stride length of the user is obtained by performing initial calibration tests which involve the user walking a certain distance. This calibration test is described by the gait calibration algorithm (Algorithm 1). The gait calibration algorithm is performed in the beginning when the user launches the navigation module for the first time.

3.2 Algorithm of LR module

The landmark recognition algorithm (Algorithm 2) used in the LR module aims to assist the user in discovering landmarks such as buildings or other objects whose location/identity is unknown to them. The working of this algorithm involves using the user's height, the current location of the user and the bearing of the user's phone (i.e.) azimuth. The user has to point their phone at the base of the building/object to obtain the details of the landmark. The method uses a trigonometric tangent function to obtain the distance of the user from the base of the landmark given the user's phone's height (approximated from the user's height) and the angle of inclination of the user's phone (obtained from sensors) [12]. Using the approximated distance, current location of the user and the azimuth, the location of the landmark is estimated using the inverse haversine formula [13]. Upon getting the location of the landmark, its details are obtained using the online digital mapping service and displayed to the user using the AR renderer.

Algorithm 1 - Gait calibration algorithm

Input – User's current location, Phone's GPS
Output – User's foot-stride length
1. Initialize start location as user's current location 2. Increment step count of the user using the footstep counting method 3. Update the user's current location using GPS 4. If the user doesn't stop, go to Step 2 5. Find the distance between the starting location and the current location using the Haversine formula [14] 6. $Foot\_stride\_length = distance / step\_count$ 7. Return the foot-stride length

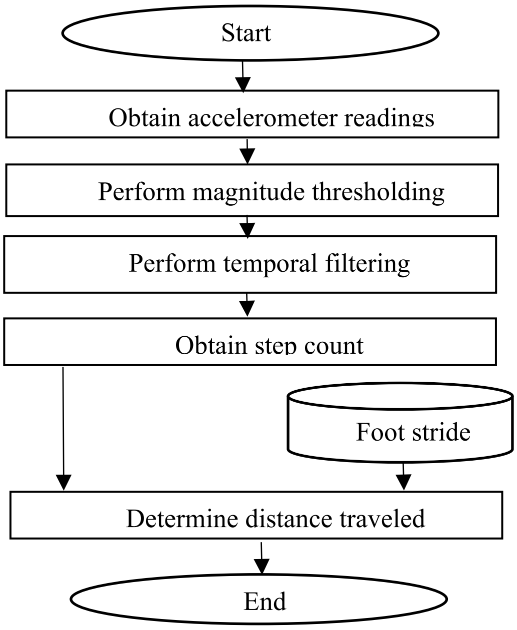


Figure 4 – Flowchart of SDE algorithm

Algorithm 2 - Landmark recognition algorithm

Input – User's height, User's current location, Phone's azimuth
Output – Location of pointed landmark
1. Initialize accelerometer, magnetometer sensors 2. Obtain the landmark view from the camera upon user action 3. Calculate the angle of inclination (tilt) of the phone using the accelerometer and magnetometer sensors [12] 4. Find the distance of the base of the landmark from the user by applying the tangent function to the height and angle of inclination 5. Using the distance, azimuth, and the current location, find the latitude and longitude of the landmark using the inverse Haversine formula [13]. 6. Return the landmark location

4. IMPLEMENTATION AND RESULTS

The MAR-PNS was developed using Android Studio (Arctic Fox 2020.3.1) [15], Google Cloud Platform [16] and Beyond AR Framework [17]. Google Directions API [18] and Google Places API [19] are used for getting directions and details of locations. Each module was tested on Android smartphones in an urban environment and the results were validated against the physical locations and distances in real time.

The ARN and LR modules were implemented using three activities namely the *Map activity*, *AR activity* and *landmark activity* in Android Studio. Map activity displays the user's current location on a map and allows the user to select a destination point to where they need AR directions. The AR activity displays the AR turning point of objects during navigation. The landmark activity allows the user to point their phone at an unknown landmark and displays its details in AR. The SDE module was implemented using the accelerometer reading of the user to estimate their position in the absence of GPS. The AR activity and the SDE activity are interconnected to implement the hybrid approach of user localization. Some screenshots of the implemented Android application are illustrated in figures 5 and 6. Figure 5 displays the working of the ARN module and Figure 6 illustrates the LR module.

4.1 ARN and LR modules

The two main parameters for testing AR-based navigational applications are locational accuracy and latency of object generation and latency. The first parameter is mainly crucial for the LR module. The testing for the landmark recognition module involved validation of the location received from Google Maps and the location calculated by the landmark recognition algorithm. The findings of our experiments are shown in Table 1.

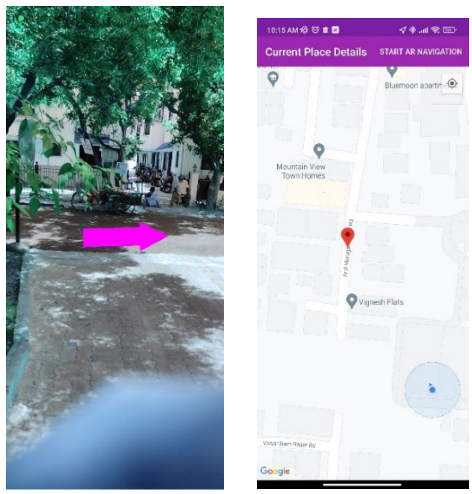


Figure 5 – ARN module screenshots

The second parameter being the latency of object appearance on the user's screen depends on the following three factors: the framework used for rendering AR objects, complexity of the algorithm used for localization of object position, and the device specific functionalities such as processor speed. The major contribution of latency lies in the framework chosen for the application. Experimentally it is found that the latency of AR object appearance is in the range of 60ms to 80ms with an average of 74 ms.

Table 1 – Locational accuracy and latency

Actual location	Estimated location	Distance deviation (m)	Latency (ms)
12.949037, 80.140572	12.949024, 80.140599	3.3	72
12.948213, 80.139994	12.948214, 80.140013	2.1	77
12.948849, 80.140914	12.948865, 80.140930	2.5	82
12.949498, 80.139833	12.949510, 80.139847	2	71
12.950599, 80.140618	12.9506038, 80.1406339	1.8	68
	Average	2.34	74

4.2 SDE module

In testing the step count module, we experimented the step count-based localization application by performing tests for three scenarios, namely slow walking, fast walking and running. A threshold for acceleration was chosen (11 to 13) based on a trial-and-error method for each experiment and the user had to slow-walk, fast-walk and run for 30 steps five times for each threshold. The average steps for each test case are displayed in Table 2.

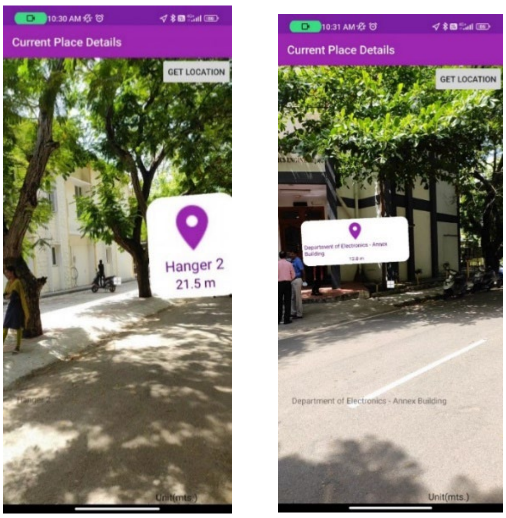


Figure 6 – LR module screenshots

The value with the least step count error in all three test cases was chosen as the magnitude threshold for the SDE algorithm. The error rates are visualized in the form of a graph shown in Figure 7. It is observed that, setting the threshold value to 11.96 yields the maximum accuracy and the model is compatible for all three modes of pedestrian travel. The error rate comparison of our experimental findings with two smart watches (Realme Dizo Watch 2, and MI Band 3) is presented in Table 3.

Table 2 – Determination of threshold

Threshold	Slow Walking (30)	Fast Walking (30)	Running (30)
11.45	20.4	28.4	27.6
11.7	25.6	28.2	27.8
11.96	29.4	29	28
12.34	24	28.8	27.8
12.59	20.6	26	27

As observed in Table 3, the step count estimator in the proposed MAR-PNS model is well suited when the user is walking as the error rate is 0.5% less than of the smart watches. However, it is observed that it shows 2% more error when the user is running. This model thus allows us to estimate the relative position of the user with respect to the turning point as soon as possible in the absence of GPS and smoothens the navigation process. Using the results obtained experimentally, the threshold value was set to 11.96 in the SDE algorithm.

Figure 7 – Error rate step counts for various thresholds

Table 3 – Step count validation

Model	Navigational Mode	Average Estimated steps (out of 30)	Error %
Realme Dizo Watch 2	Walking	29.1	3%
	Running	28.5	5%
MI Band 3	Walking	29	3.34%
	Running	28.8	4%
MAR-PNS System	Walking	29.2	2.67%
	Running	28	6.67%

The distance estimation sub-module was analyzed with Root Mean Squared Error (RMSE) as a metric to validate the distance covered by the user. The experiments were carried out locally, where the user with test application walked a stretch of road whose distance was already known. The RMSE for the distance deviation was observed and is listed in Table 4.

Table 4 – Distance estimation experiments

Actual Distance (m)	Calculated Distance (m)	Squared error (m)
25	23.4	2.56
30	27.6	5.76
52	53.2	1.44
70	69.3	0.49
100	99.4	0.36
	RMSE	1.45

5. CONCLUSION AND FUTURE WORK

The proposed MAR–PNS system provides means for mitigating the drawbacks of traditional pedestrian navigation systems and existing MAR-based navigational systems. The major contribution of this work is its novel methodology for discovering nearby landmarks whose locations are unknown by just pointing the phone’s camera at those landmarks. This improves the user experience as well as opens new junctures for landmark recognition without using the computationally intensive traditional methods which rely on image or video processing. Another contribution of this work is cutting down on the reliance of the navigational systems on GPS and providing a local way of estimating the user’s position by using their step count.

The step count-based user localization module requires the user to hold the phone in their hand while walking. This can be improved in future by applying an additional filter which mitigates the effect of noise when the user is keeping the phone in other positions such as their pockets or shaking the phone while walking. Integrating this module with a stationary device such as a smart watch can further improve its accuracy.

The ARN module can be further enhanced by adding a text to speech translation feature, which could be used to speak out the directions to users, and thereby improving the overall accessibility of the application.

The incorporation of computer vision for landmark recognition and comparing the performance with MAR–PNS and GPS-based identification could be another future area of work.

REFERENCES

[1] T. Alsop, "Mobile augmented reality (AR) market revenue worldwide from 2019 to 2025", Statista, 2021. https://www.statista.com/statistics/897587/world-augmented-reality-market-value/

[2] T. Alsop, "Number of mobile augmented reality (AR) active user devices worldwide from 2019 to 2024", Statista, 2021. https://www.statista.com/statistics/1098630/global-mobile-augmented-reality-ar-users/

[3] E. Cranmer, M. tom Dieck, and P. Fountoulaki, "Exploring the value of augmented reality for tourism", Tourism Management Perspectives, vol. 35, July 2020, doi: 10.1016/j.tmp.2020.100672.

[4] A. Dirin and T.H. Laine, User Experience in Mobile Augmented Reality: Emotions, Challenges, Opportunities and Best Practices. Computers 2018, doi: 10.3390/computers7020033.

[5] Google Maps offline navigation. https://support.google.com/maps/answer/6291838?hl=en&co=GENIE.Platform%3DAndroid

[6] Google Live View. https://support.google.com/maps/answer/9332056?hl=en&co=GENIE.Platform%3DAndroid/

[7] M. Garcia Puyol, D. Bobkov, P. Robertson, and T. Jost, "Pedestrian Simultaneous Localization and Mapping in Multistory Buildings Using Inertial Sensors," in IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 4, pp. 1714-1727, August 2014, doi: 10.1109/TITS.2014.2303115.

[8] A. Abadleh, E. Al-Hawari, E. Alkafaween, and H. Al-Sawalqah, "Step detection algorithm for accurate distance estimation using dynamic step length," 2017 18th IEEE International Conference on Mobile Data Management (MDM), 2017, pp. 324-327, doi: 10.1109/MDM.2017.52.

[9] Android Motion Sensor documentation. https://developer.android.com/guide/topics/sensors/sensors_motion

[10] Recommendation ITU-T Y.4562 "Functions and metadata of spatiotemporal information service for smart cities," December 2020.

[11] Recommendation ITU-T Q.4066 "Testing procedures of augmented reality applications," September 2020.

[12] Android position sensor documentation. https://developer.android.com/guide/topics/sensors/sensors_position

[13] Calculate distance, bearing and more between Latitude/Longitude. https://www.movable-type.co.uk/scripts/latlong.html

[14] J. Inman, Navigation and nautical astronomy for the use of the British seamen. London: F & J Rivington, 1847.

[15] Android Studio Arctic Fox. https://android-developers.googleblog.com/2021/07/android-studio-arctic-fox-202031-stable.html

[16] Google Cloud Platform. https://cloud.google.com/

[17] Beyond AR Framework. https://github.com/BeyondAR/beyondar

[18] Google Directions API. https://developers.google.com/maps/documentation/directions/overview

[19] Google Places API. https://developers.google.com/maps/documentation/places/web-service/overview

THE KNOWLEDGE GRAPH AS THE INTEROPERABILITY FOUNDATION FOR AN AUGMENTED REALITY APPLICATION: THE CASE AT THE DUTCH LAND REGISTRY

Alexandra Rowland[1]; Erwin J.A Folmer[1]; Tony Baving[2]

[1]University of Twente & Kadaster, The Netherlands

[2]Kadaster, The Netherlands

ABSTRACT

The concept of the knowledge graph supports insight to a given context through the provision of standards-based mechanisms for accessing open and interoperable data. In doing so, the graph uses the power of the web to integrate data from distributed data sources and make this data available to end users in a transparent, flexible and application-independent manner, either by simply displaying data in the browser based on a dereference unique resource identifier or in an application built using the knowledge graph as the source. With the latter approach, the knowledge graph remains independent of the applications making use of it as a data source, where the connection between the graph and the application is achieved through interfaces which are completely based on open standards, most commonly through the use of a SPARQL endpoint. Indeed, chatbot applications often make use of the knowledge graph in this way but this paper aims to present the potential for Augmented Reality (AR) applications to be similarly built using knowledge graphs. By presenting this potential, this paper will argue that AR applications exemplify the potential opportunities that fully open, interoperable and standards-based approaches to data publication, such as the development of knowledge graphs, have and, therefore, will become key drivers within the organization in the investment of the further development of the concept of the knowledge graph in the future and improved accessibility of data for end users.

Keywords – augmented reality, geospatial information, knowledge graph, interoperability, open standards

1. INTRODUCTION

In a recent update to the approach to linked geospatial data publication at the Dutch Land Registry and Mapping Agency [1], the Knowledge Graph (KG) is the central architectural concept supporting interoperability between distributed data sources and end-user applications allowing users to interact with the authoritative data published by the organization. The approach makes use of open standards to publish the key registers maintained by the organization as linked data, to combine the key registers themselves under a standardized schema and to make the resulting in a completely standards-based KG available to developers and end users. Within this architecture, the KG is designed to remain an application-independent data source with a standardized data model and applications can connect to the data source using a range of fully standardized endpoint options as an interface where each standardized interface option is made available for a specific target user group.

An example of an application which can make use of the KG through a standardized interface is an Augmented Reality (AR) application developed for Kadaster[1] users which provides said users with building information about a building which is scanned and selected using their smartphone cameras. In practice, this application performs two simple SPARQL queries on the Kadaster Knowledge Graph (KKG), the results of which are available to the application through a SPARQL endpoint. The information that is made available through the application has provenance in various key registers and external datasets which are then combined in the KG and, through the simple SPARQL queries, made available to the application in a flexible and developer-friendly manner. The interoperability between this application in particular, and, as will be argued, any application in general, is achieved through the use of open standards both in the delivery of the KG, the interface between the data source and application and within the application itself.

The successful development and implementation of this application for Kadaster using the architectural approach previously defined provides insight into the potential advantages and disadvantages of using this approach, both from a technical and organizational perspective. Indeed, the approach supports interoperability between the data source and the application in a very simple manner, one that is easily

1 https://www.kadaster.nl/ .

understood and maintained and offers the possibility of adding more data options over time in a relatively flexible manner. From the organizational perspective, the KG improves the governance of the data contained within. From the perspective of the end user, the application offers a low-threshold, user-friendly manner in which to access, interact and provide feedback on the geospatial data published by Kadaster. Overall, this paper will argue that applications such as the AR application exemplify the opportunities that such an architectural approach holds for a decoupled, flexible ecosystem and how such an application will act as key drivers for further investment in implementing this approach more widely; particularly within the context of governmental organization.

2. KNOWLEDGE GRAPHS FOR DATA INTEROPERABILITY

The Knowledge Graph (KG), a data representation model in which data is stored as a graph comprised of nodes and edges mapped to ontologies to form a semantic network, was introduced as a means of connecting and integrating data from different sources. Indeed, a knowledge graph or semantic network can be defined as representing 'a network of real-world entities – i.e. objects, events, situations or concepts – and illustrates the relationship between them' [2]. Although the concept of the knowledge graphs is not recent, the term was first popularized by Google in 2012 and has since formed part of technological solutions from various multinational corporations and as part of search engines such as Google and Yahoo. During implementation, a knowledge graph makes use of various data management models including the traditional database model, a graph model and a knowledge base model wherein the formal semantics (from various domains) are defined [3].

In general, the knowledge graph makes use of the Resource Description Framework (RDF) structure for data representation and, as such, there are three main components which make up said graph, namely nodes, edges and labels. Objects, places or persons, for example, can be represented using a node and the relationships between these nodes are represented by an edge. This node-edge-node structure, or the subject-predicate-object structure, is known as a triple, a data representation structure which underpins the publication of data as linked data. An example of this triple structure in the context of geospatial data is provided in Figure 1.

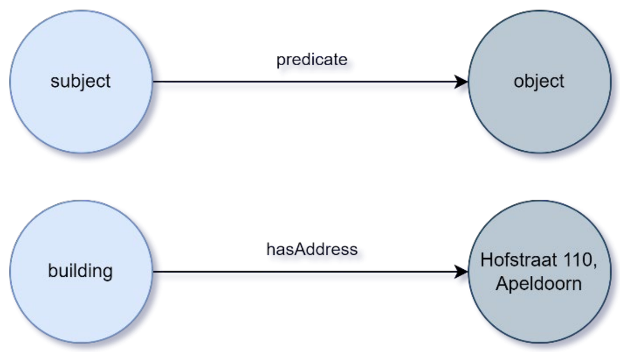


Figure 1 – Triple structure (subject-predicate-object) using a geospatial example

The semantic web, in facilitating the creation of linked, machine-readable data on the web, makes use of RDF, the associated schema (RDFS) and a range of Web Ontology Languages (OWLs) as the core web technologies which structure data on the web. These web technologies formalize the semantics of data from various domains using defined ontologies and capture these semantics in triples using Unique Resource Identifiers (URIs) for nodes and relations [4, 5], enabling location-independent cross-machine readability and interaction [6]. A single triple can be connected iteratively with other triples, forming an expanded graph data model known as the knowledge graph. These connected triples can be semantically enriched and placed in context through the application of (domain) ontologies and connection to other knowledge bases using semantic web technologies and standards.

The interoperability of data made available on the web, such as data made available as part of a knowledge graph, is supported by the efforts of the World Wide Web Consortium (W3C), the main standardization organization for the worldwide web. This consortium publishes and maintains a variety of open standards, including common linked data standards such as OWL, RDF, SPARQL, PROV and SKOS, which aim to foster compatibility and agreement in the publication of data on the web. Where reusability of open standards is high, there is a shared understanding of the meaning of data published using these standards and, therefore, interoperability and potential for reuse of datasets available on the web is higher.

The following section discusses the architecture used in the creation and ongoing development of the Kadaster Knowledge Graph (KKG). The use of open standards published by the W3C in the modeling of the KKG is a central element of this architecture. The standards used in this solution include OWL-based ontologies, RDF and RDFS for data representation as well as elements of the SKOS and PROV vocabularies for knowledge organization and provenance links between source datasets and the KKG. Naturally, the resulting linked data is also made available through a SPARQL endpoint, another example of a standards-based implementation within the solution architecture.

3. BUILDING THE KADASTER KNOWLEDGE GRAPH

Kadaster, in its role as the national mapping and land registry, publishes and maintains several key geospatial registers. Contained in these datasets is administrative and spatial data concerning property ownership and real estate rights; both on a national and international level. In carrying out this role, the organization provides Dutch society legal certainty with regards to all administrative and spatial data on property and rights involved. Recently, Kadaster has also actively developed an approach for publishing and maintaining several key registers and geospatial datasets as linked (open) data, a standards-based approach to linked data publication which has recently been updated [1] and which is central to the initial and ongoing development of the Kadaster Knowledge Graph (KKG). The SPARQL endpoints for all key registers, the KKG itself and information about any updates to datasets can be found and used in Kadaster's triple store environment (https://data.labs.kadaster.nl/).

The KKG is composed of several key registers including the Key Register for Addresses and Buildings (BAG), the Key Register for Large Scale Topography (BGT) and the Key Register for Topography (BRT). Each key register and the dataset is first made available as a siloed linked data source through an Extract, Transform and Load (ETL) process based on several steps extensively discussed by [1] and as illustrated in Figure 2.

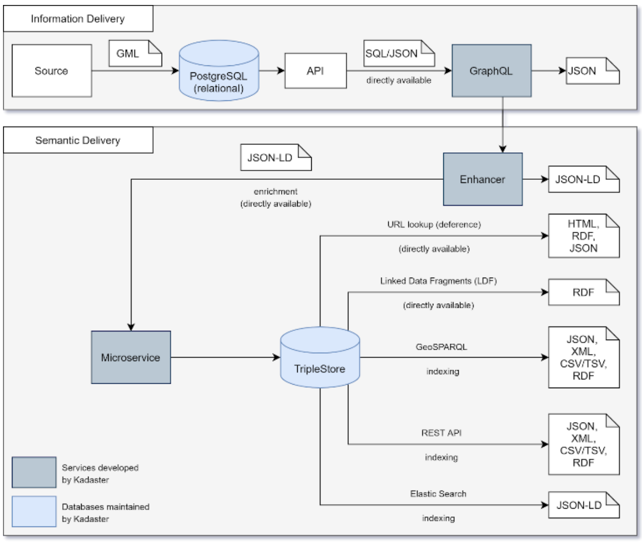


Figure 2 – Architecture supporting the ETL process which produces linked data [6]

To briefly summarize this process, each key register or dataset is loaded from a relational database source to the PostgreSQL instance following a Geography Markup Language (GML) indexing step [7] and is then made accessible through a GraphQL endpoint. An internally developed microservice, denoted in the figure above as the Enhancer, is then used to query the data and return the results in JSON-LD serialization format. This microservice also publishes the resulting data to the triple store, an instance of TriplyDB (https://triplydb.com/), which in turn makes the data available in a number of serialization formats based on the instantiation of services (e.g. a SPARQL service) as shown above. A preceding step to the loading of data into the triple store is a SHACL validation step which ensures any loaded data complies with a defined data model and to ensure that data has not been partially lost or fundamentally changed over the course of the transformation. Indeed, the process is automated as much as possible, in order to avoid any human errors in the transformation.

For each dataset, the linked data model used for publication is made available in linked data 'as-close-to-the-source' as possible. This approach is taken for a number of reasons. Firstly, data models for key registers are often defined in Dutch law, supporting the legal certainty of the registered using a given data model. Keeping the linked data model as close to the source supports, to a certain extent, the extension of this legal certainty to the linked data from of the register. Beyond this step, no other legislation requirements were included in the scope of linked data publication at Kadaster. Additionally, the linked data publication of a siloed key register is then also recognizable for domain experts and ensures that the linked data is directly usable for these users.

When combining these datasets for the construction of the KKG, a central data model is used which simplifies some of the complexity seen in the siloed datasets with the goal of making the KKG more suited for a wider range of user groups, including domain experts and developers as well as interested citizens, researchers and industry experts. In the initial development of the KKG, the data model used for publication was based on the schema.org specification[2]. Current versions of the KKG use the same architecture defined above for publishing linked data in the KKG but are now using the first version of the Samenhangende Object Registratie (SOR)[3]; a data model in development by Geonovum (https://www.geonovum.nl/), a Dutch government foundation which supports the standardization and management of geospatial information in the Netherlands. This data model, which in English is called the Connected Object Registration, is being developed to improve the way the base registers and other related datasets are connected based on, for example, the KKG.

As illustrated in Figure 3 (below), the KKG is realized through the creation of a layer on top of the key registers as linked data and is created by performing mappings between the data models of the source datasets and the SOR data model. As extensively discussed by [1], the implementation

2 https://schema.org/ .

3 https://www.geobasisregistraties.nl/basisregistraties/doorontwikkeling-in-samenhang/objectenregistratie

of these mappings are done through LD views[4], each of which transform the data from the key register source dataset to linked data conforming to the SOR model based on predefined SPARQL construct queries. These LD views, in using SPARQL construct functionality, take a part of a data model from a key register and map this to an associated part of the SOR model. This process is a key part of the architecture outlined by the authors as this serves to preserve the provenance and traceability between the source data and that available in the KKG.

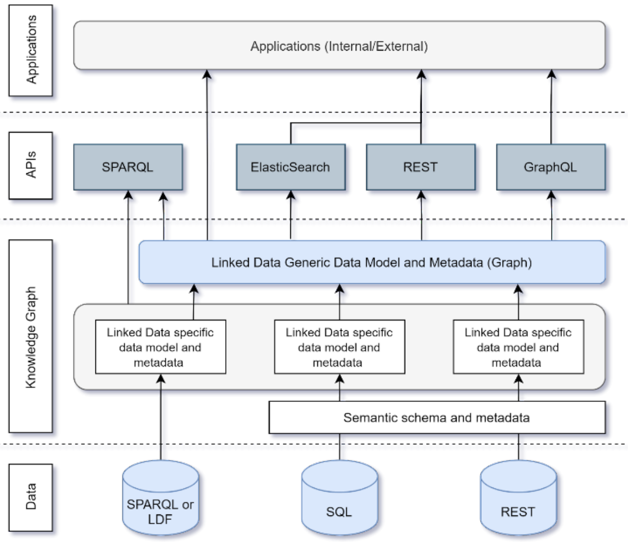


Figure 3 – Architecture for the implementation of the Kadaster Knowledge Graph [6]

The KKG is now available for use in the Kadaster triple store environment and contains approximately 680 million triples. The data in the KKG is updated on a quarterly basis in line with the quarterly updates of the linked data publication of the various key registers and an updated notification is posted in the Kadaster triple store on completion. As highlighted in Figure 2, it is the intention that the KKG is used directly by applications which access the data various service options (e.g. SPARQL, REST or GraphQL) and, indeed, it is in this manner that the Augmented Reality (AR) application described in the following sections makes use of Kadaster data in providing functionality to end users.

4. THE AUGMENTED REALITY APPLICATION

Within Kadaster, the KKG is increasingly being used as a data source for various applications, particularly where the applications require the availability of data from various key registers and/or other data sources. Making use of the KKG, therefore, allows this data to be queried and used in the application without the need to perform additional transformation or integration steps in order for the data to be used in an integrated manner within the given application.

One such application, the most recent of which to be developed using the KKG as the data source, is the Augmented Reality (AR) application. As illustrated in Figure 4, the application allows the user to scan their environment for buildings within a given location and then choose a specific building to retrieve information about that building on their screen. As such, the application can be categorized as a visualization application for geospatial data and is an example of an application which supports users with using geospatial data from the KKG without needing any knowledge of linked data or querying languages, supporting Kadaster's ambition for making geoinformation available for everyone. The application itself is querying information from the KKG which is then rendered by the application and displayed as shown in Figure 4.

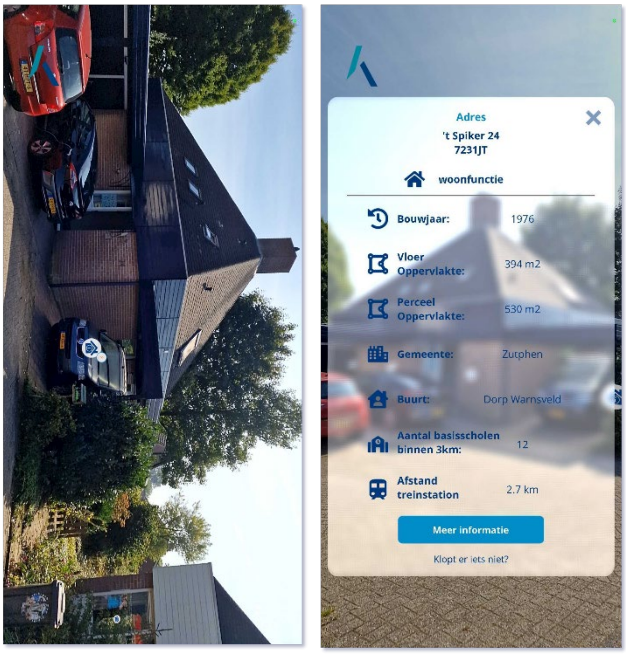


Figure 4 – Screenshots displaying house icon and building information from the augmented reality application built on the Kadaster Knowledge Graph

What is not obvious to the user when the information is being displayed is that the information has provenance in several distributed data sources, both those maintained by Kadaster as well as those provided by other organizations such as the Bureau of Statistics in the Netherlands (Dutch acronym: CBS). Indeed, the address of the building (adres), the building year (bouwjaar) and the floor size of the building (oppervlakte) are all originally sourced from the Key Register for Addresses and Buildings (BAG). The neighborhood (buurt) and the number of schools within 3 km (aantaal basisscholen binnen 3 km) are originally published in the CBS districts and neighborhood dataset. These, based

4 https://labs.kadaster.nl/demonstrators/architectuur-selfservice/LDViews/ .

on the architectures presented in the previous sections, are now all published in the KKG and made available to the AR application based on an endpoint.

For the application to provide users with the information seen in the figure above, the KKG endpoint needs to be queried by three different queries and that data made available through an API to be integrated into the application itself. Both queries are SPARQL queries being performed on the SPARQL endpoint for the KKG and the resulting APIs for each query are simply integrated as API variables in the application. In line with the system architecture illustrated in Figure 3, SPARQL APIs are used as the interface between the KKG as the application-independent data source and the AR application itself. As previously noted, SPARQL, an open standard, is the most widely used query language for linked data and allows users to query data flexibly and make use of the returned data in an application given that the user has knowledge about the structure of the data model. The linked data available through a SPARQL API can be returned in a range of formats including JSON-LD, Turtle[5] and N-Quads[6]. These formats are native linked data serialization formats, each with a different structure.

The first query, which is available here, has a single input query parameter in the form of a point geometry. This geometry is defined by the application itself in fetching the latitude and longitude of the user's location, combining these to form the point geometry string and using this string as input for the first SPARQL query. The query returns the building identifier and associated polygon for each building within a 100 meter radius of the defined point (i.e. the user's location) defined by the application. The results of this query are visualized only as a house icon above a building (see Figure 4) where the icon is centered on a building based on the central point of the polygon returned during this query. This first query is introduced to support better performance in the application. Indeed, querying building information within a radius requires a large amount of computational power and this query returns a minimal amount of information which then allows successive queries to be performed on a subset of the data, improving overall performance of the application.

Once the house icon has been visualized in the application, the user is able to tap on the house icon to indicate that they are interested in seeing more information about that particular building. In selecting a given building, the building identifier associated with that icon is then used as the input parameter for the second query, which is available here. The KKG data model makes a distinction between buildings and building units (in Dutch: verblijfsobjecten), where the attributes such as building year are associated with a whole building and usage function and floor size are associated with the building unit. In the application, this distinction allows the users to click on the house icon and then see a list of all building units contained within the building in order to choose which unit in particular they would like to see information on. This functionality in the application is illustrated in Figure 5 and is supported by the second query which takes the building identifier from the first query as input and returns all information about building units contained within that building, including house numbers, letters and additions as well as the postcode and street name associated with that address. Naturally, this functionality is most useful when multiple units exist within a building but is still available in the application when only a single unit is present. The selection made in this step is used as input for the following step.

Figure 5 – Screenshots displaying building units from the augmented reality application built on the Kadaster Knowledge Graph

The final query, which is available here, makes use of the identifier associated with the building unit selected by the user and then displays all information associated with both the whole building such as the building year and statistical information, as well as the information only associated with a given unit, including floor size and usage function. An example of this information is visualized and presented to the end user and can be seen in Figure 4.

As more information becomes available in the KKG from new data sources, the same endpoints can be used to display more detailed information about a building in the AR application. The only changes that would need to be made are in the application itself to transform and render the information for display on the screen in a similar manner as above. From the user perspective, the AR application serves

5 https://www.w3.org/TR/turtle/

6 https://www.w3.org/TR/n-quads/ .

one of Kadaster's main organizational aims, namely, to provide everyone (from developers and data engineers to students and interested citizens) with the ability to use and interact with the geospatial data they maintain and publish. An application in this form, specifically, provides a low-threshold, gamification-type experience for the user, who only needs access to a smartphone and network connection; arguably improving the accessibility and overall experience of interacting with Kadaster's data and the KKG.

Currently, the application is made available for download[7] and use within the Netherlands based on the fact that this is the spatial extent of the data published in the KKG. The usability of the application has been tested internally as part of the finalization of this innovation project and users are able to provide feedback about the data or application through the application itself. No formal user experience research has been carried out to date but is seen as something that should be considered as a next step in order to improve the functionality and data made available through the application.

5. DISCUSSION

What the previous sections aim to highlight is the novelty, success, and value in implementing an application architecture where data storage and application are distributed entities. The development and implementation of the AR application as described in the previous section allows for reflection on both the practical advantages and disadvantages of the overall system architecture proposed by [1] (see Figure 3).

5.1 A standards-based approach

Firstly, SPARQL, as a standardized SQL-based query language, is relatively easy to understand and use, given a basic understanding of the SQL language. Data can also be made available through an API in a format that is recognizable to developers unfamiliar with linked data, making the data easy to use in a range of applications. The only disadvantage in this respect is the small learning curve required to be able to write appropriate, performant SPARQL queries for integration in the application.

Secondly, the data required by the AR application is available through three separate SPARQL APIs by design. This is the key component of interoperability between the KG and application because the APIs used as interfaces are open standards-based interfaces. These APIs are then simply integrated into the application as API variables and the application needs only to render the data for visualization. The advantage of this approach is that data is available through a single data source and using the same schema making it easy to simply display. Previously, because the data would have been returned from a range of sources using a range of schemas, it would have been necessary to appropriately transform and integrate the data and then render it for display. This, of course, complicates the development of the application because the application requires data transformation and management functionality rather than simply visualization functionality. Additionally, the use of SPARQL APIs also means that any addition of more information in the future simply requires an updated query and/or an updated API variable in the application code as well as some modification of the way in which the data is rendered. This, therefore, makes the maintenance and extension of the application much easier. A key disadvantage in this respect is that if the underlying schema for the source is changed, the queries will automatically break and, therefore, the application will no longer work as intended.

Finally, and as a consequence of the previously mentioned advantages, this approach also means that the need for copying the data for transformation and/or management purposes is negated. As such, data remains at the source and is always as up to date as possible in the application, again reducing the complexity in the development and ongoing maintenance of the application. The only barrier to this advantage in the approach is the trust that the application developer has in the fact that the source has an acceptable uptime and is as performant as required for the application functionality. A similar trust needs to be had in the fact that the underlying schema used for the data remains stable so that continuous updates to the query and/or rendering do not need to be done to maintain the application. The authority and dependability of the source, therefore, need to be ensured.

5.2 Application-independent data publication

From an organizational perspective, the development of the AR application at Kadaster also highlights several additional issues worth considering when implementing this approach. Firstly, it can be argued that the implementation of this approach is only appropriate when data is not being combined in a triple store for a single purpose. Indeed, if data from various, distributed sources is being copied (as a subset), transformed into linked data and combined to form a type of Knowledge Graph that only serves one use case or application because it, for example, only contains data relevant to that use case, this approach is no longer appropriate. Doing so would, in essence, be comparable to more traditional approaches to data storage where data is combined and used for a specific purpose. The intention of the approach, and what this paper hoped to highlight, was that data from the source could be used in various application contexts without the need to perform any further integration or transformation to the data itself. Depending on the structure of the organization, it would only be necessary to have a single KG which various applications use as a data source where appropriate.

7 https://www.vr-innovations.nl/ar-kadaster-knowledge-graph-download .

It is, of course, possible that a single application might want to use data from various organizations such as, for example, including more detailed information from the Chamber of Commerce about a specific building in the AR application. Where this is required by a given application, the system architecture proposed by [1] comfortably supports this where the queries used by the application are changed to include federation. The use of federated queries supports all the advantages previously mentioned while also allowing for more data views to be added to the application without any need to change the application architecture. From an organizational perspective, the only requirements needed to still take advantage of this approach are that one KG per relevant organization is available rather than multiple APIs per dataset and that there is a single SPARQL endpoint per KG on which to perform federated queries. Being able to perform federated queries has the advantage of making use of various datasets at the source, having the data in your application as up to date as possible and having few ongoing maintenance requirements over time.

5.3 The provision of 'geoinformation for everyone'

Looking specifically from the perspective of the targeted end user, the requirements of which are perhaps not explicitly captured in the architecture outline or not explicitly discussed by [1], the AR application itself exemplifies the type of application which can make use of a KG to provide users with easier access to (geospatial) data for a given purpose. Indeed, without such an application, the non-expert user might not have the skills to access the information contained within the KG and, comparatively to the use of chatbots, this application, whilst offering developers and organizations all the advantages mentioned above, also supports users with a low-threshold and perhaps even gamified means of using Kadaster's data. For Kadaster, this could mean increased use of geospatial data by a user group which previously might have been excluded.

6. CONCLUSIONS

Kadaster, the Dutch National Land Registry and Mapping Agency, has recently started publishing Dutch key registers as linked data and integrating these in Kadaster Knowledge Graph (KKG), a fully standards-based knowledge graph. As part of this approach, an architecture was proposed where the KKG is defined as an application-independent data source and various API formats were used to interface between the data source and the application in question. One such application, an augmented reality application allowing users to display information about a building on their smartphone screens, successfully implemented this architecture.

The development and implementation of the AR application on the KKG provided an illustrative test case for the standards-based system architecture proposed by [1] as illustrated in Figure 3. This test case highlighted various advantages of using this approach, arguably the most important of which is the ability to store the data independently of the application while still allowing the developer to have complete control over the content and format of the data it receives from the source. In practice, it was clear that this independence between data storage and application challenges the traditional architecture and requirements set that accompany innovative applications such as the AR application but that over time the advantages of this approach become clear to developers. While offering many advantages to the developer, the end user themselves also benefit from the ability to develop and implement a wider range of applications because it improves the accessibility of authoritative, actual data to end users without the skills to access data through APIs. This application, in particular, provides users with a gamified way of interacting with Kadaster's data, hopefully acting as an additional driver for the ongoing development of these types of applications in the future.

From an organizational perspective and in widening the scope to include multiple applications using the same data, it is hoped that this paper highlights the advantages of decoupling the data source from the applications built on top of them. Indeed, the true proof of the architecture used by the AR application can only be seen when additional applications with a wider range of functionality based on the KG are using only SPARQL as the interface; a process that has already been initiated to a large degree of success within the context of Kadaster.

For Kadaster, the successful implementation of the AR application using the KKG as outlined by the architecture approach defined in previous work supports a vision for the future with regards to an ecosystem of (Dutch) government datasets [8]. This vision includes the creation of organizational KGs using standardized models which can be queried using a single endpoint. These organizational KGs can be federated in SPARQL queries and the results integrated into innovative applications, such as the AR application, serving a range of use cases benefiting citizens. Indeed, this architecture negates the need for these large governmental datasets to be copied and transformed to serve the needs of applications and supports the ability of organizations themselves to control and standardize how their data is integrated. By doing so, the data being used in the applications remains actual and assurance is provided to the end user as to the trustworthiness of the source. Overall, this ecosystem, as exemplified by the implementation of the AR application, supports governments in playing an important and increasingly central role in supplying authoritative data that can be directly consumed by businesses to build innovative applications benefiting citizens of their country and beyond.

REFERENCES

[1] A. Rowland, E. Folmer, W. Beek, R. Wenneker, "Interoperability and Integration: An Updated Approach to Linked Data Publication at the Dutch Land Registry". ISPRS International Journal of Geo-Information. 2022.

[2] IBM Cloud Education, "Knowledge Graph", 12 April 2021. Available online: https://www.ibm.com/cloud/learn/knowledge-graph#:~:text=A%20knowledge%20graph%2C%20also%20known,the%20term%20knowledge%20%E2%80%9Cgraph.%E2%80%9D (accessed on 12 April 2022).

[3] Ontotext. "What Is a Knowledge Graph?" Available online: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/ (accessed on 12 April 2022).

[4] J. Tekli, A.A Rjeily, R. Chbeir, G. Tekli, P. Houngue, K. Yetongnon, M.A Abebe, "Semantic to Intelligent Web Era. In proceedings of the 5th International Conference on Management of Emergent Digital Ecosystems, Luxembourg. Association for Computing Machinery, New York, NY, USA. Vol 20. 2013.

[5] S. Guus, Y. Raimond, "RDF 1.1 Primer". W3C Working Group Note. 2014. Available online: http://www.w3.org/TR/rdf11-primer/ (accessed on 20 April 2022).

[6] C. Bizer, R. Cyganiak, T. Heath, "How to Publish Linked Data on the Web". 2007. Available online: http://wbsg.informatik.uni-mannheim.de/bizer/HowtoPublishLinkedData.htm (accessed on 02 May 2022).

[7] L. Van den Brink, P. Janssen, W. Quak, J. Stoter, "Linking Spatial Data: Automated Conversion of Geo-Information Models and GML data to RDF". International Journal of Spatial Data Infrastructure. Vol. 9, p.p. 59-85. 2014.

[8] E. Folmer, S. Ronzhin, J. Van Hillegersberg, W. Beek, R. Lemmens, "Business Rationale for Linked Data Governments: A Case Study at the Netherlands' Kadaster Data Platform", IEEE, Access 8, 70822-70835. 2020.

SESSION 3

SERVICES IN FUTURE NETWORKS

RESEARCH ON ASSET ADMINISTRATION SHELL STANDARD SYSTEM ARCHITECTURE

Quanbo Lu[1]; Xinqi Shen[2]; Mei Li[1]*

[1]School of Information Engineering, China University of Geosciences Beijing, Beijing, China
[2]China Academy of Information and Communications Technology, Beijing, China
*Corresponding author: maggieli@cugb.edu.cn

ABSTRACT

Asset Administration Shell (AAS) is an important enabling technology to implement digital twins for Industry 4.0(I4.0). It establishes cross-company interoperability. However, in the process of AAS application, there is a lack of AAS standard references. Aiming at the problem of lacking AAS standardization, this paper analyzes the requirement of AAS standards from three perspectives: AAS concept, the implementation of AAS key technologies, and AAS applications. As a basis for research to establish the three dimensions of the standardized architecture, AAS standard architecture is formed from the integration of architecture under different perspectives. This provides guidance for the study and formulation of standards related to AAS standardization.

Keywords –Asset administration shell, common technology, key technology, standards

1. INTRODUCTION

AAS is an important technology to depict, simulate, optimize, and visualize the physical world in the information world. It provides an effective way to realize the global industrial and social development trends such as digital transformation, intelligence, services, and green sustainability. At present, AAS technology is widely concerned and studied by industry and academia [1-3]. According to statistics, 134 researchers from 22 countries, including the United States, South Korea, and Italy, and more than 103 institutions have carried out AAS theory and application research. Relevant research results have been published. At the same time, ABB, Bosch, Amazon, FESTO, and other world-famous companies carried out the implementation and application of AAS in related fields [4-6]. For example, in the manufacturing field, Siemens in Germany have studied the application of AAS technology to the whole life cycle process of a product (e.g. product design, production, manufacturing, operation, services, and recycling) [7].

However, in the research and application of AAS, the following problems have been found: (1) The standard lack of AAS-related terms, system architecture, and application results in users' different understanding of AAS from different technical requirements. (2) The standard lack of AAS-related models, data, connection, and services has led to problems, such as difficult integration, poor consistency, low compatibility, and difficult interoperability between models. (3) The standard lack of AAS tools and platforms results in users' confusion about the use of AAS.

The above problems have seriously hindered the further development and application of AAS. There is an urgent requirement for guidance and reference of relevant standards. At present, there is no release of AAS standards at home and abroad. Current AAS-connected international standard development initiatives and their working documents are as follows: Working group IEC/TC65/WG24 AAS for industrial applications is formulating IEC 63278-1 ED1 AAS for industrial applications-Part 1: AAS structure, IEC 63278-2 ED1 AAS for industrial applications-Part 2: information meta model, and IEC 63278-3 ED1 AAS for industrial applications-Part 3: security provisions for AASs. Subcommittee ISO/IEC JTC 1/SC 41 Internet of Things and digital twin is formulating ISO/IEC AWI 30172 Digital Twin-Use cases and ISO/IEC AWI 30173 Digital Twin-Concepts and terminology. The standardization council I4.0 pursues the goal of initiating digital production standards in Germany and coordinating them nationally and internationally. Industrial digital twin association is the one-stop shop for the digital twin, an alliance of active creators working together to make the digital twin practical for industry through open technologies. OMG-s the Digital Twin Consortium drives the awareness, adoption, interoperability, and development of digital twin technology.

Therefore, combined with the previous AAS research and application exploration, this paper establishes AAS standard system architecture from six aspects: AAS basic common standards, key technology standards, tool/platform standards, evaluation standards, security standards, and industry application standards. It provides reference for AAS standard researchers.

2. AAS STANDARD REQUIREMENTS ANALYSIS

2.1 AAS concept

People from different fields, different needs, and different levels have a different understanding of AAS. This will lead to difficult communication in AAS research. Therefore, there is an urgent need for basic common standards (e.g. the standard of AAS-related terms and system architecture) to help strengthen the understanding of AAS. It promotes the AAS concept.

2.2 The research and implementation of key technology

From physical entity, information entity, connection, information data, and services, this paper gives a reference for the implementation of AAS technology. Due to the lack of reference to AAS-related models, data, connection, and services, the compatibility and interoperability of AASs developed by different development teams are poor. It results in problems, such as difficult integration and consistency among models. The specific aspects are as follow:

(1) The physical entity is the foundation of AAS model. Due to the lack of standard guidance on the real-time state perception of physical entities, different developers do not know what the physical entities need to perceive, map, and reverse control. It results in inconsistent physical entity interfaces, unclear technology levels, and an unclear function boundary.

(2) The information entity is the true, objective, and complete mapping of a physical entity in the digital space. It is the carrier of information data. The information entities developed by different people have poor consistency in the description level and low format compatibility. Therefore, their availability is low.

(3) Information data is the core driving force of AAS. It provides accurate and comprehensive information sources for the fusion of information entities and physical entities. Due to the lack of standard guidance on the representation, classification, preprocessing, storage, use, and testing of information data, data classification is not unified and data formats are incompatible. It results in poor universality, interoperability, and data fusion.

(4) The connection realizes the interconnection between physical entities, information entities, information data, and services. Due to the lack of standard guidance on its connection mode, information transmission, interaction mechanism, and test method, it is difficult to be compatible with input and output.

(5) Service is the medium for the efficient performance of AAS functions. Due to the lack of standard guidance on its description method, development requirements, operation management, and test evaluation, the services developed by different personnel have poor compatibility and low interoperability.

2.3 The AAS implementation

With the development of cloud computing, Internet of Things, big data, and artificial intelligence, AAS applications have gradually expanded to various industries, especially in the manufacturing industry [3]. However, in the development of the AAS implementation and application, the following problems must be solved:

(1) Before implementing AAS, we should consider whether to apply AAS in combination with their needs and conditions. For example, they must consider industry applicability and an input-output ratio. Therefore, AAS functional and technical requirements guide us to conduct an applicability assessment and decision analysis.

(2) Once we make a decision to use AAS, the next problem is how to implement AAS, such as what software and hardware conditions it needs, what tools and platforms it relies on, and what functions it needs. Therefore, it needs the implementation requirements, tools, and platforms to guide AAS.

(3) After AAS implementation, we need to evaluate the comprehensive benefits brought by the use of AAS and the comprehensive performance of the AAS system (such as accuracy, security, stability, availability, and ease of use). It provides iterative optimization and a decision-making basis for the application in the next stage. Therefore, standards of AAS evaluation, safety, and management provide reference and guidance for the evaluation and safe use of AAS.

3. AAS STANDARD SYSTEM FRAMEWORK

According to the above demand analysis on the AAS standard system, we take into account the rationality, integrity, systematization, and availability of the standard system. The AAS standard system framework is designed as shown in Figure 1. It gives standard guidance from six aspects: basic common standards, key technology standards, tool/platform standards, evaluation standards, security standards, and industry application standards.

(1) AAS basic common standards include terminology standards, reference architecture standards and applicable standards. They focus on the concept definition, reference framework, applicable conditions, and requirements of AAS and provide support for the whole standard system.

(2) AAS key technology standards include physical entity standards, information entity standards, information data standards, connection and integration standards, and service standards, which are used to guide the research and implementation of AAS key technologies, ensure the effectiveness of key technologies in the AAS implementation, and break the technical barriers of collaborative development and module interchangeability.

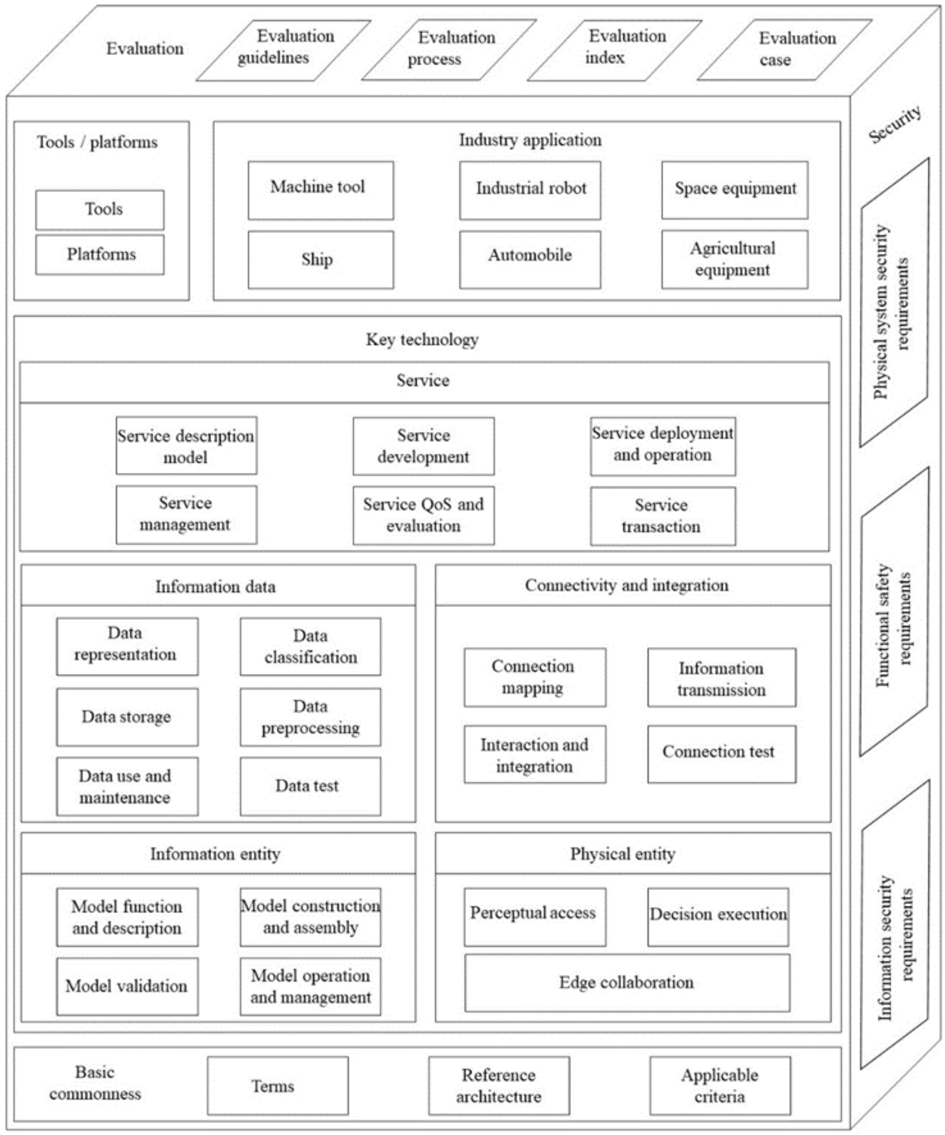


Figure 1 – The AAS standard system framework

(3) An AAS tool/platform standard includes two parts: tool standard and platform standard. They are used to standardize the technical requirements of tool/platform, such as function, performance, development, and integration

(4) AAS evaluation standards include evaluation guidelines, evaluation process standards, evaluation index standards, and evaluation use case standards, which are used to standardize the test requirements and evaluation methods of an AAS system.

(5) AAS security standards include three parts: physical system security requirements, functional security requirements, and information security requirements, which are used to standardize the technical requirements for the safe operation of personnel and the safe storage, management, and use of various information in the AAS system.

(6) An AAS industry application standard takes into account the technical differences of an AAS application in different fields and different scenarios. On the basis of basic common standards, key technical standards, tool/platform standards, evaluation standards, and safety standards, it standardizes the application of AAS in specific industries such as machine tool/platform standards, tool, industrial robot, space equipment, ship, automobile, and agricultural equipment.

4. AAS STANDARD ARCHITECTURE

The AAS standard system consists of various substandard systems of AAS. The AAS standard system structure is shown in Figure 2, including six parts: basic common standards, key technology standards, tool/platform standards, evaluation standards, safety standards, and industrial application standards.

5. AAS BASIC COMMON STANDARDS ARCHITECTURE

AAS basic common standards mainly regulate the basic and general standards of AAS. The relevant standards and main contents are shown in Figure 3, including the following aspects:

(1) We need to define AAS related concepts and correspond-ing abbreviations help users understand AAS concepts, and provide support for the formulation of other standards. AAS related terms include AAS main concept definitions and key technologies, such as AAS, information data, digital thread, AAS submodel, AAS system, physical entity, information entity, connection, and digital service.

(2) AAS can be divided into unit level AAS, system level AAS, and complex system level according to the function and structure of physical entities [8-10]. The reference architecture standard specifies the layering rules of the above three levels, the AAS architecture and the reference architectures of each part to help users clarify the AAS layering method, architecture, and the relationship between each part. Reference architecture standards include AAS layered criteria, AAS overall reference architecture, unit level AAS reference architecture, system level AAS reference architecture, complex system level AAS reference architecture, AAS model overall architecture, AAS physical entity, AAS information entity, information data, connection and interaction, AAS application/service platform, AAS security reference architecture, and AAS evaluation reference architecture [11].

(3) The applicable standards regulate the applicability requirements of AAS and help users decide whether to apply AAS, including functional requirements, performance requirements, safety requirements, reliability requirements, maintenance requirements, and consistency requirements.

6. AAS KEY TECHNICAL STANDARDS ARCHITECTURE

6.1 Physical entity standards framework

Physical entity standards mainly regulate the perceived access, decision execution, and edge cooperation of physical entities. The relevant standards and main contents includes the following aspects:

(1) The perceptual access standard specifies the relevant technical requirements for perceptual access of physical entities such as people, machines, objects, and environment in the AAS system. It ensures the normalization and compatibility of information acquisition and uploading of physical entities, including the perception requirements of static attribute and dynamic attribute.

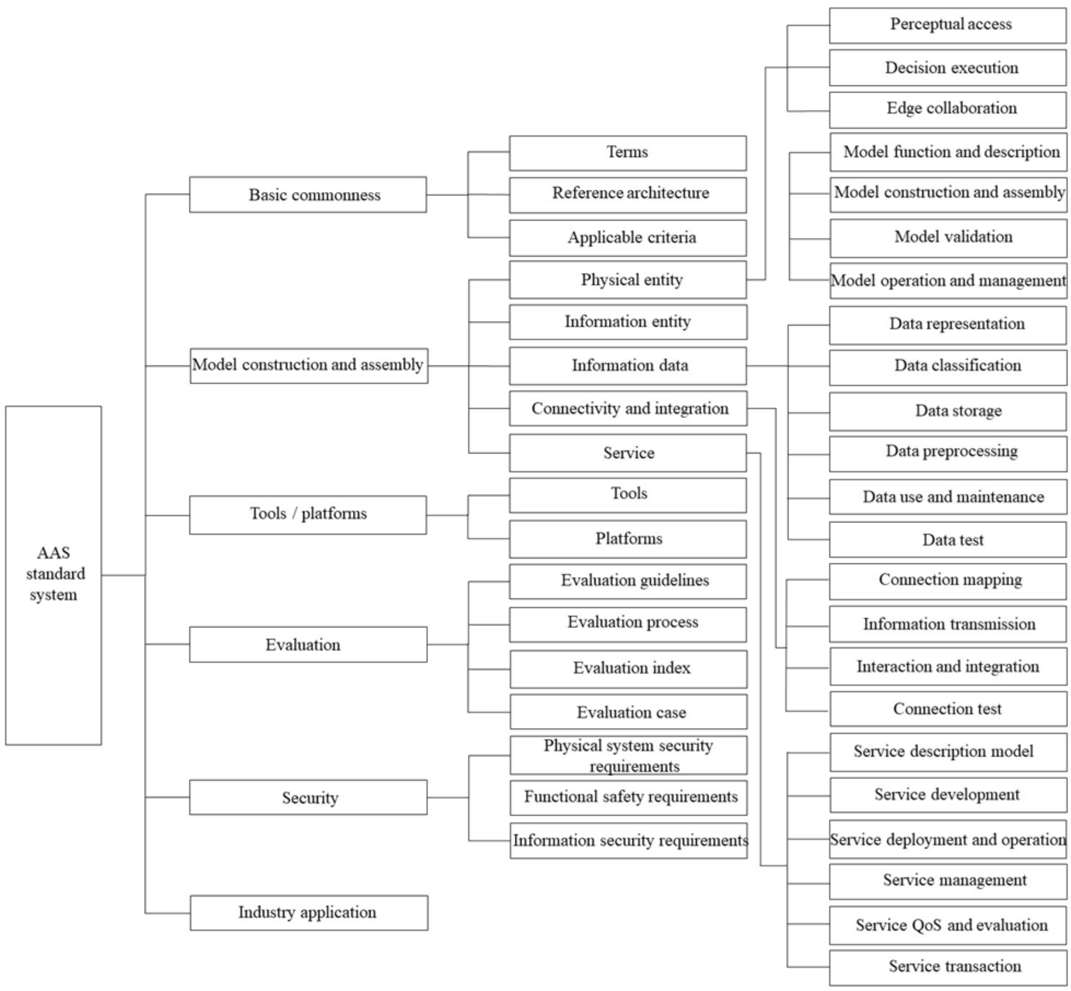


Figure 2 – The AAS standard architecture

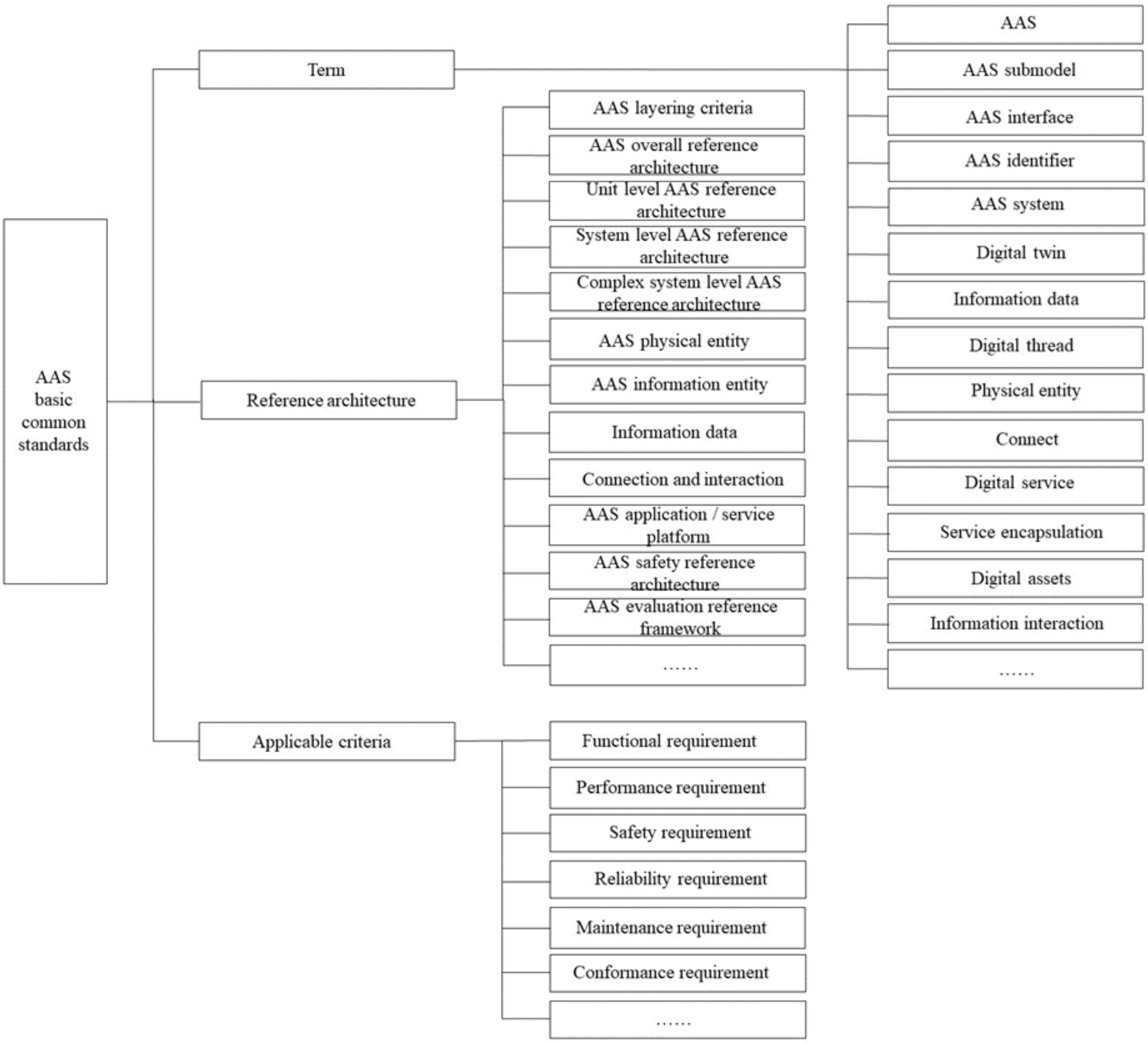


Figure 3 – AAS basic common standards framework

(2) The decision execution standard specifies the relevant technical requirements for the decision execution of physical entities, including control instruction access requirements, time sensitive requirements, and execution effectiveness requirements.

(3) The edge end collaboration standard specifies the technical requirements of physical entities, including the deployment architecture, function boundary division criteria, computing interface, computing performance, and network performance of the edge end.

6.2 Information entity standards framework

The information entity standard mainly regulates the model function, model description, model construction, model assembly, model verification, model operation, and model management. The relevant standards and main contents include the following aspects:

(1) The model function and description standard specify the technical requirements of multi-dimensional models [12], including the physical function, behavior function, rule function, dimension function, scale function, capability description, performance description, modeling language, model encapsulation, model input, and output description.

(2) The model construction and assembly standard specify the technical requirements related to the construction and assembly of multi-dimensional models, including the modeling rules, modeling environment, modeling process, model assembly rules, model assembly interface, model assembly process, and model extension.

(3) The model verification standard specifies the technical requirements related to the hierarchical verification of individual models and assembly models, including the verification rules, verification environment, verification process, function verification, performance verification, consistency verification, and compatibility verification.

(4) Model operation and management standards standardize the technical requirements related to model operation and management, including the model operation environment, model operation configuration, real-time monitoring of model operation, model simplification/lightweight principle, and model update.

6.3 Information data standards framework

Data is fundamental to driving the operation of AAS system. The information data standard is mainly used to standardize the representation, classification, storage, pretreatment, maintenance, and testing of information data. The relevant standards and main content include the following aspects:

(1) The data representation standard specifies the technical requirements of AAS data, including information data representation criteria, data index, data structure, data temporal relationship, and data spatial relationship.

(2) The data classification standard specifies the information data categories, including classification criteria, historical data, and live data. The historical data and live data include physical entities, information entities, service generated status data, control data, derived data, knowledge data, simulation data, and management data.

(3) The data storage standard specifies relevant technical requirements for information data storage, including distributed storage, local storage, storage media, and data access.

(4) After the data preprocessing standard obtains the original data of physical entities, information entities and services. It needs to preprocess them. The relevant technical requirements include data cleaning, data reduction, data transformation, data specification, data association, and data integration.

(5) The data use and maintenance standard specifies the relevant technical requirements in the use of information data, including the use environment, data fusion, data visualization, data optimization, data loading, data sharing, and data maintenance.

(6) The data testing standard specifies the technical requirements related to data testing, mainly including the data testing process, data standardization test, data integrity test, data accuracy test, data compatibility test, and data usability test.

6.4 Connection and integration standards framework

The connection and integration standards mainly regulate the data connection and integration of AAS physical entities, information entities, services, and databases. The relevant standards and main contents include the following aspects:

(1) The connection mapping standard specifies the connection mapping of physical entities, information entities, and services, including connection mode, mapping model, and mapping dictionary.

(2) The information transmission standard specifies the technical requirements for data transmission among physical entities, information entities, and services, including transmission protocol, real-time transmission, transmission reliability, and transmission security.

(3) The interaction and integration standard specifies the technical requirements for the interaction and integration between physical entities, information entities, and services.

(4) The connection test standard specifies the technical requirements related to the connection test between physical entities, information entities, and services, including the connection compatibility test, connection reliability test, connection time sensitivity test, interactive function and performance test, and system integration test.

6.5 Service standards framework

Service standards mainly regulate the service description model, service development, service deployment, service operation, service management, service QoS, evaluation, and service transaction. Relevant standards and main contents include the following aspects:

(1) The service description model standard specifies the relevant technical requirements for the AAS service description model, including modeling rules, modeling environment, modeling language, and modeling process.

(2) The service development standard specifies relevant technical requirements for service development, including the function description, design specification, development environment, development process, service encapsulation, and service template library.

(3) The service deployment and operation standard specifies the technical requirements related to AAS service deployment and operation, including deployment requirements, operation environment, service instantiation, operation monitoring, service collaboration, and service operation fault tolerance.

(4) The service management standard specifies the relevant technical requirements for service life cycle process management, including service search, supply and demand matching, service optimization, service scheduling, service composition, service fault tolerance, and service collaboration.

(5) The service QoS and evaluation standard specifies the technical requirements related to AAS service QoS and evaluation, including time, cost, reputation, satisfaction, and non-functional service quality.

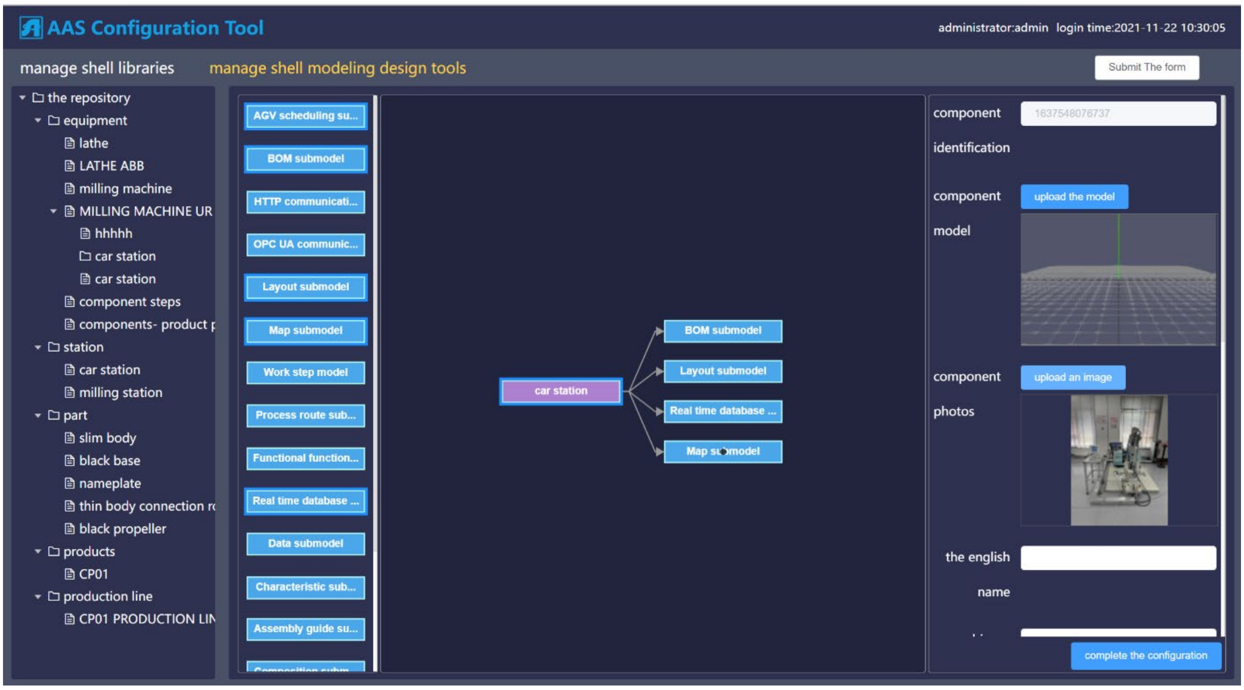


Figure 4 – AAS configuration tool

(6) The service transaction standard specifies the technical requirements related to AAS service transaction, including service transaction platform, service transaction mode, service transaction security, service pricing and billing, and service authorization.

7. AAS TOOL/PLATFORM STANDARDS ARCHITECTURE

The AAS tool/platform standard is used to standardize the tool/platform involved in the AAS implementation. The relevant standards and main content include the following two aspects:

(1) The tool standard specifies the relevant technical requirements for the AAS tool, including tool function, tool performance, tool operating environment, tool secondary development, and tool integration.

(2) The platform standard specifies the relevant technical requirements of the AAS platform, including the platform function, platform performance, platform operating environment, platform use and maintenance, platform interface and integration, and platform security.

8. AAS EVALUATION STANDARDS ARCHITECTURE

The AAS evaluation standard is used to standardize the test requirements and evaluation methods of the AAS system. The relevant standards and main contents include the following aspects:

(1) The evaluation guidelines specify the basic requirements for the testing and evaluation process of the AAS system, including the evaluation purpose, evaluation type, evaluation level, evaluation environment, evaluation tools, and evaluation confidentiality and security.

(2) The evaluation process standard specifies the technical requirements related to the AAS testing and evaluation, including evaluation analysis, evaluation preparation, evaluation method selection, evaluation steps, and evaluation documents.

(3) The evaluation index standard specifies various index requirements involved in the AAS system test and evaluation process, such as stability, reliability, ease of use, security, consistency, maintainability, support ability, portability, expansibility, adaptability, self-evolution, testability, economy, and satisfaction.

(4) The evaluation use case standard specifies relevant technical requirements for test and evaluation use cases of the AAS system, including use case selection, use case verification, and use case archiving.

9. AAS SAFETY STANDARDS ARCHITECTURE

AAS safety standards are used to regulate the safety of the AAS system. Relevant standards and main content include the following aspects:

(1) The safety requirements for physical systems specify the safety requirements for physical systems in the AAS system, including physical system safety, electrical system safety, mechanical system safety, intrinsic safety, and functional safety.

(2) The functional safety requirements specify the technical requirements related to the AAS safety functions, including the safety risk analysis of the information system, the safety function design of the information system, and the safety integrity level assessment of the information system.

(3) The information security requirements specify various information security related technical requirements involved in the AAS system, including information security risk analysis of the information system, information data security, and network security of the twin system.

10. CASE STUDY: AAS CONFIGURATION TOOL

The AAS standard system architecture can help develop the AAS tool/platform. Therefore, according to AAS standard system architecture, we have design an AAS configuration tool. It can describe an entity asset in accordance with the AAS standard specification, and it can be recognized by the AAS operation tool. The AAS configuration tool needs to meet the characteristics of standardization, scalability, compatibility, interoperability, and closed-loop. The AAS configuration tool is shown in Figure 4.

11. CONCLUSION

This paper analyzes and discusses the construction requirements of AAS standards from three dimensions: AAS concept, the implementation of AAS key technologies, and AAS application. This paper explores and establishes a set of AAS standard system architectures, and expounds the standard system from six aspects: basic commonality, key technologies, tools/platforms, evaluation, security, and industry application. It is expected that relevant work can play a role in the research and formulation of AAS standards.

REFERENCES

[1] J. N. Chen and J. H. Zhou, “Revisiting Industry 4.0 with a Case Study,” In 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifaxno, Canada, 2018, pp.1928-1932.

[2] Q. B. Lu, X. Li, G. Y. Li, M. Li, Mei, J. H. Zhou, J. D. Liu, X. Q. Shen, and F. H. Zhu, “A general asset administration shell platform for production line design,” 1st IEEE International Conference on Digital Twins and Parallel Intelligence, Beijing, China, 2021, pp. 192-195.

[3] X. Ye, M. Yu, W. S. Song, and S. H. Hong, “An asset administration shell method for data exchange between manufacturing software applications,” IEEE Access, vol. 9, pp. 144171-144178, 2021.

[4] M. Redeker, S. Volgmann, F. Pethig, and J. Kalhoff, “Towards Data Sovereignty of Asset Administration Shells across Value Added Chains,” ETFA 2020-25th Annual Conference on Emerging Technologies and Factory Automation, Vienna, Austria, 2020, pp. 1151-1154.

[5] B. S. Lv, Y. P. Zhang, F. Yang, and J. P. He, “Edge Asset Management based on Administration Shell in Industrial Cyber-Physical Systems,” IECON 2020 - 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 2020, pp. 3817-3822.

[6] S. Cavalieri and M. G. Salafia, “Asset Administration Shell for PLC Representation Based on IEC 61131–3,” IEEE Access, vol. 8, pp. 142606-142621,2020.

[7] DIN SPEC 91345: Reference Architecture Model Industrie 4.0 (RAMI4.0), DIN Std. DIN SPEC 91345, 2016.

[8] A. Sidorenko, M. Volkmann, W. Motsch, A. Wagner, and M. Ruskowski, “An OPC UA model of the skill execution interaction protocol for the active asset administration shell,” Procedia Manuf., vol. 55, pp. 191-199, 2021.

[9] T. Terzimehić, A. Bayha, and K. Dorofeev, “Function Blocks for the Interaction with the Asset Administration Shell,” 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 1235-1238.

[10] X. Ye and S. H. Hong, “Toward I4.0 components: Insights into and implementation of asset administration shells,” IEEE Ind. Electron. M., vol. 13, no. 1, pp. 13-25, 2019.

[11] K. Xia, C. Sacco, M. Kirkpatrick, C. Saidy, L. Nguyen, A. Kircaliali, and R. Harik, “A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces, and intelligence,” J. Manuf. Syst., vol. 58, pp. 210-230, 2021.

[12] K. T. Park, J. Yang, and S. D. Noh, “VREDI: virtual representation for a digital twin application in a work-center-level asset administration shell,” J. Intell. Manuf., vol. 32, pp. 501-544, 2021.

RESEARCH AND STANDARDIZATION REQUIREMENTS FOR 5G NETWORK PEAK CONTROL TECHNOLOGY IN VIDEO TRANSMISSION

Zhiji Deng[1, 2]; Zhewei Fu[2]; Ming Liu[2]; Xiangyu Qu[2]; Dong Ding[3]; Qi Ye[2]; Weisheng Kong[2]; Fei Wang[2]; Jinyu Zhang[2]; Hui Wang[2]; Jian Lou[3]

[1]Zhejiang Provincial Key Laboratory of Harmonized Application of Vision & Transmission, China

[2]Zhejiang Dahua Technology Co. Ltd., China

[3]China Mobile Communications Corporation, China

ABSTRACT

The 5G network peak control technology in video transmission scenarios is used to solve the problem of violent fluctuations in network peaks caused by the collision of I-frames (key frames) during multichannel video transmission. Based on the research of key technologies such as I-frame collision detection and network peak shift scheduling, through the Multiaccess Edge Computing (MEC) external interface, the network data and video management system is opened to effectively smooth the network peak during multichannel video transmission, and realize the number and bandwidth of 5G terminal access under 5G base stations. The improved utilization rate is conducive to the large-scale promotion of 5G video surveillance scenarios.

Keywords – 5G network, key frame collision, network peak control, video transmission

1. INTRODUCTION

A 5G network is no longer a simple network for people to use, but a network for the Internet of Everything that includes computing, storage, and network [1-2]. Therefore, applications in all walks of life are the top priority of 5G development. The video surveillance service is one of the typical services of 5G applications. In the past, wireless communication technology was rarely used in video surveillance because Wi-Fi network coverage needs to be constructed and deployed, and outdoor scenes are not easy to cover. However, there are public users in the 4G network, which are easy to influence each other and cause network congestion, and the uplink bandwidth is insufficient, which cannot meet the requirements of video surveillance. Larger bandwidths and stable transmission are required. 5G features large bandwidths, low latency, and wide coverage, which enables video surveillance terminals to have access anytime, anywhere. The device is connected to the network when it is powered on, and deployment and debugging are extremely simple. Every time a camera is deployed, the engineering department is no longer required to deploy the network, and multidimensional data fusion and local storage based on the MEC[3], breaking the data island and ensuring data security.

Using a 5G network for video surveillance transmission can solve the problems of video surveillance that cannot be moved and monitored temporarily, and make up for surveillance loopholes. However, in large-scale deployment, when multiple video channels under the 5G network are transmitted concurrently, the network peak value fluctuates violently. If the 5G network is planned according to the peak value of video transmission, in most cases, the average video transmission bandwidth demand is small, and the network utilization rate is low, resulting in a waste of a large amount of communication resources; while planning the 5G network according to the average video transmission bandwidth, there will be intermittent video transmission freezes, which affects user perception, and the number of 5G cameras connected cannot meet expectations.

The White Paper "5G ToB Experience Standards" clearly explains the characteristics of real-time streaming video transmission: the bandwidth requirement when a single video stream I-frame is sent far exceeds the average code rate, the peak bandwidth is multiplied during the collision of multiple real-time streaming video I-frames, and the more network camera devices, the greater the probability of I-frame collision [4]. Real-time streaming video transmission network peak control technology is commonly used in the industry include: 1. Code flow smoothing sending technology, using the idle interval of real-time streaming video, scattering I-frame burst to the P frame time slot to send, smooth off the peak burst flow. This technology will make the I-frame sending time longer and introduce the sending delay; 2. Adjust the encoding strategy distributes I-frames to multiple P frames to smooth and burst. The scheme will affect subsequent frames when some I-frames are lost, resulting in un-decoding for a long time [5]; 3. Reduce the encoding rate and improve the fluency experience when the network is congested [6]. The scheme will cause video clarity to decrease when dealing with emergencies. Combined with the 5G network peak control technology of multi-video stream I-frame bursts, I-frames are delivered to the network in an orderly manner according to a certain time distribution to achieve the effect of equalized traffic distribution. For each device's video stream, the full-speed transmission of the network will not introduce additional delay, which can effectively improve real-time video transmission efficiency of 5G networks.

However, the current 5G technical standards mainly focus on 5G network capabilities, and there is a lack of technical standards that integrate with industry business characteristics such as video surveillance. To better promote 5G in industry services, it is necessary to deeply study the integration of 5G networks and industry services. The development of standardized technologies combined with business characteristics to achieve 5G networks truly lands in industry applications, and network peak control technology for concurrent transmission of multichannel videos under 5G networks is an important one.

2. DISCUSSION

2.1 Network peaks generation analysis in single-channel video transmission scenario

According to the transmission characteristics of single-channel video, through the analysis of video transmission technology, it can be found that in real-time video monitoring transmission, there is a situation of regularly sending I-frames (key frames), and the traffic when sending I-frames will be ordinary P frames[7-9].

According to the video surveillance service test, data analysis results show: taking Frames Per Second (FPS) equal to 25 as an example, a video frame is generated every 40ms, and an I-frame is generated every 2 seconds. The difference between I-frame and the P frame is large, and the data volume ratio can reach more than 10 times. The network transmission data is shown in Figure 1.

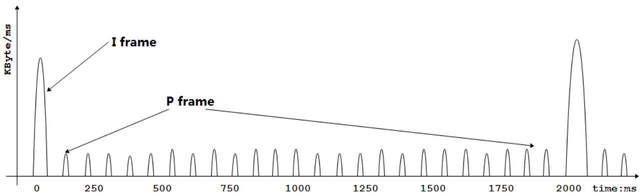


Figure 1 – Network transmission characteristics of real-time streaming video frames

2.2 Network peaks generation analysis in multichannel video concurrent transmission scenario

According to the situation of single-channel video transmission, it can be inferred that when multiple channels of video are transmitted concurrently, I-frames may be sent at the same time (I-frame collision), causing severe fluctuations in network peaks. Since the I-frame sending timing configurations of cameras in the same area are generally the same, when an I-frame collision occurs, it will cause a continuous timing collision and will not disappear automatically.

Based on two front-end video devices testing data, the network data transmission situation when I-frames collide is shown in Figure 2.

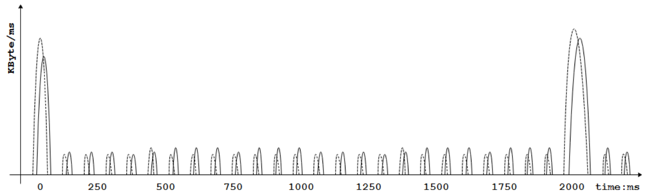


Figure 2 – I-frame collision of two front-end video devices causes the burst of network transmission data to increase

Through the random experiment of 7-channel video concurrent transmission, 7-channel cameras are randomly pulled (I/P frame size ratio is 10:1), 125 operations are recorded, and the probability of collision of 2/3/4/5/6/7 channels is generated is shown in Table 1.

Table 1 – Probability of I-frame collision when 7 devices are concurrently uplinked

Number of collision paths	Occurrence	Total number of detections	Collision probability
None	10	125	8%
2	35	125	28%
3	25	125	20%
4	25	125	20%
5	20	125	16%
6	8	125	6.4%
7	2	125	1.6%

When seven channels of video are transmitted concurrently, the collision probability of I-frame is about 92%. The larger I/P frame ratio, the more Internet protocol Cameras (IPCs) colliding at the same time, resulting in larger data peaks. With the common 5G base station capability of 150M uplink bandwidth, the theoretical number of video transmission paths (calculated according to the 4M code stream of 1080P) is about 36 channels, which is far more than seven channels concurrently, and the probability of its I-frame collision will also be greater.

It is precisely because of the collision of I-frames when multiple video channels are concurrently transmitted that the peak value of the 5G network fluctuates greatly, and the bandwidth resource requirements of the 5G base station are higher. Under the limited bandwidth, it may even cause video freezes, hindering the scale of 5G video surveillance applications.

3. 5G NETWORK PEAK CONTROL TECHNOLOGY

3.1 Technical architecture and workflow

3.1.1 Architecture

The core technology of the 5G network peak control technology in the video transmission scenario is that the video management platform obtains 5G base station network data through the MEC, and then determines whether there is network congestion caused by I-frame collision when multiple video channels are concurrent. When an I-frame collision occurs, the camera's I-frame timing control is used to achieve the effect of staggered traffic, and to achieve 5G network peak control, so as to optimize the video transmission efficiency and improve the utilization of 5G wireless resources. The architecture is shown in Figure 3.

− The technical architecture includes front-end equipment, a 5G access network, MEC and video management platform.
− The front-end video devices are surveillance video capture devices such as cameras, which support the ability to adjust the timing of I-frames.
− The 5G access network consists of wireless network facilities such as 5G base stations.
− In this technical architecture, the MEC connects the 5G wireless network and the video management platform, and provides 5G data forwarding services, I-frame collision detection services, and network peak scheduling services.
− The video management platform provides equipment management services, streaming media services, network peak adjustment policy execution services, etc.

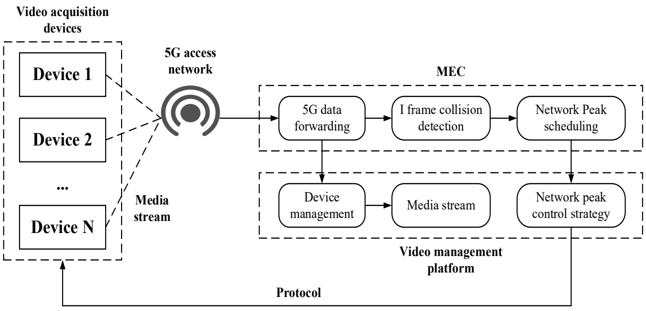


Figure 3 – Diagram of peak control technology of 5G video transmission network

3.1.2 Workflow

The specific workflow of the technical solution is as follows:

Step 1, a group of front-end video equipment is connected to the video management platform through a 5G wireless network and MEC data forwarding. Front-end video equipment realizes device registration and synchronizes the time of front-end equipment through a "device management service".

Step 2, the platform initiates real-time streaming to the camera, and realizes the real-time preview of video surveillance through a "streaming media service".

Step 3, during the real-time video monitoring process, the MEC obtains the parameter information and data flow information of the front-end video equipment from the video management platform and the 5G network for the "I-frame collision detection service" to detect the collision of the equipment's I-frames.

Step 4, the "I-frame collision detection service" performs I-frame collision detection according to whether the I-frame transmission time period of the device (the time period from the start of I-frame transmission to the completion of transmission) coincides with the I-frame transmission time period of other devices. When an I-frame collision is detected, the "Network Peak Scheduling Service" is notified to schedule the I-frame timing.

Step 5, the "Traffic peak scheduling service" according to the I-frame timing distribution of the group of front-end video equipment, arranges the timing of the collision equipment, decides the reasonable I-frame generation timing of the front-end equipment that collides with the I-frame, and manages the I-frame through the MEC and video management platform interface notifying the "Network Peak Adjustment Policy Execution Service".

Step 6, the "Network Peak Adjustment Policy Execution Service" issues I-frame adjustment command to the front-end equipment through the 5G network to adjust I-frame encoding timing of the equipment.

Step 7, the front-end video equipment receives I-frame adjustment command and adjusts the generation timing of I-frame, so as to achieve the time-sharing and off-peak sending of I-frame.

3.2 Key technologies

In the network peak control technology process of the overall 5G video transmission scenario, the key technologies are the MEC-based I-frame collision detection service and network peak shift scheduling service, which have opened up the network data and video management system, and which have made full use of the MEC's open interface. The perception capability of the network is linked with the video application management, which enables better and more accurate rapid policy adjustment based on network conditions.

3.2.1 I-frame collision detection service

The I-frame collision detection service is mainly divided into two steps: I-frame information collection and I-frame collision detection.

I-frame information collection means that the video management platform obtains the information of the front-end equipment and the analysis result of the data stream by the 5G network and synchronizes it to the MEC for I-frame collision detection. The parameter information includes device identification, IP address port, channel number, stream type, I-frame sending start timestamp, I-frame sending completion timestamp, code rate, frame rate, Group of Pictures (GOP) size, etc.

Among them, information such as device identification, IP address port, channel number, and stream type are obtained when the front-end device registers with the video management platform, and the platform synchronizes to the MEC through the data interface with the MEC; and when the video streaming starts, Based on information such as device identification, IP address and port, the MEC notifies the 5G network to parse the corresponding data stream, and records its I-frame sending start timestamp, I-frame sending completion timestamp, bit rate, frame rate, GOP size and other information. The second I-frame triggers I-frame information collection and reports the collected information to the MEC.

I-frame collision detection refers to performing time-sharing hash processing on the collected data. Combined with parameters such as video terminal grouping, number, GOP, frame rate, etc., it is hashed to the corresponding time slot by the method of time-sharing hashing. If there are multiple devices in the same time slot, it is regarded as I-frame collision.

For example, the streaming bit rate is usually configured as FPS=25, GOP=2000ms, and one video frame every 40ms. Hash is performed in units of three video frames (120ms) during hashing, which can fully hash open I-frames and achieve a good hashing effect. When there are many video front-end devices, the hash interval is adjusted according to GOP and FPS, and the minimum interval is 1 frame interval (40ms).

Assume the timing of video frames of three devices, as shown in Figure 4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Device 1	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Device 2	P	P	P	P	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Device 3	P	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P

Figure 4 – Video frame timing distribution of multiple devices

Hash is performed at an interval of 120ms (3 video frames). The corresponding time slots and collisions are shown in Figure 5. I-frames of device 1 and device 3 are hashed to the same time slot 1, which is regarded as I-frame collision.

		1			2			3			4			5			6			7			8			9	
Device 1	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P		
Device 2	P	P	P	P	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P		
Device 3	P	I	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P		

Figure 5 – Video I-frame hash distribution

Since the bandwidth resources of different 5G base stations are not shared, the video streams under different base stations collide with I-frames, which will not affect the peak value of the wireless network under a single base station. Therefore, it is necessary to perform packet detection on front-end equipment. For example, if three devices are under the same 5G base station, they are grouped according to the Cell Global Identity (CGI) of the base station. When I-frame collision detection is performed, the three devices in the group can be detected. Subsequent processing also only focuses on within the group.

3.2.2 *Network peak scheduling service*

Network staggered scheduling refers to calculating the weight of each time slot based on time-sharing hashing after an I-frame collision occurs, adjusting the I-frame timing of the conflicting video device, and forcing its I-frame timing to a time slot with a high idle weight, so as to achieve the collision traffic peak where all front-end video devices are staggered by I-frame in time.

For example, for a group of front-end video devices, device 1 corresponds to timestamp X, hash slot is 1, and the interval is 120ms. I-frame hashing of this group of devices is as follows in Table 2.

Table 2 – I-frame slot hash example

Device	I-frame timestamp	Hash slot
1	X	1
2	X+80ms	1
3	X+500ms	5

It can be seen that the corresponding I-frame time slot hash: time slot 1 has 2 video front-end devices (I-frame collision), time slot 5 has 1 video front-end device, every 16 time slots (time slot 0-time slot 15). It is regarded as a cycle (the front-end video equipment sends I-frames regularly), as shown in Figure 6.

0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Figure 6 – I-frame slot hash example 1

In Figure 6, the minimum value of the time slot is 0, and then I-frame time slot of the conflicting device is adjusted to the time slot with the minimum value of 0. If there are multiple time slots of 0, then the IPC of the collision I-frame will be adjusted to the slot with the largest idle weight.

Weight calculation method:

Hash value: $N_0 \sim N_{15}$ (subject to the number of hashes)

Minimum hash value: $N_{min} = \min(N_0 \sim N_{15})$

The left weight of i-th hash: $L_i =$ the number of consecutive N_{min} on the left

The right weight of i-th hash: $R_i =$ the number of consecutive N_{min} on the right

I-th hash weight:

$$X_i = \begin{cases} 0 & != N_{min} \\ \min(L_i, R_i) + 1 \end{cases}$$

E.g:

The 0th time slot has a minimum value of 0. If there are 11 consecutive minimum values of 0 on the left side, then L_0=11; if there are 0 consecutive minimum values on the right side, then R_0=0. The final 0th slot weight X0=min (L_0, R_0) + 1=1.

The first time slot, the value is not the minimum value of 0, then X_1=0. And so on, the corresponding weight distribution is shown in Figure 7.

Slot Hash Distribution	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Corresponding weight	1	0	1	1	0	1	2	3	4	5	6	6	5	4	3	2

Figure 7 – I-frame slot hash example 2

Therefore, adjust I-frame timing of a device in the conflicting time slot to the time slot with a maximum weight of 6. The adjusted timing hash is shown in Figure 8.

New Hash Distribution	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0

Figure 8 – I-frame slot hash example 3

3.3 Application results

This technology adopts 5G private network access in the park to reduce external network interference. Through multiple network cameras to test the real-time video pull business scene, master the collision distribution of random I-frames with multi-device pull streams, draw the I-frame distribution and traffic on the platform side, and compare the IO renderings of the TCPDUMP packet capture network to ensure the accuracy of the test data. Among them, the camera parameters adopt the mainstream coding configuration of the monitoring industry, as shown in Figure 9.

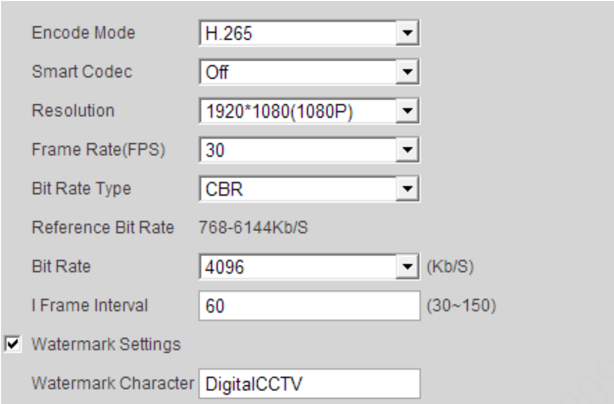


Figure 9 – Network camera video parameter configuration

The application result of this technology in actual scenarios is remarkable. Taking the effect to be achieved before and after the collision of four devices, the flow staggered effect is shown in Figure 9 and Figure 10.

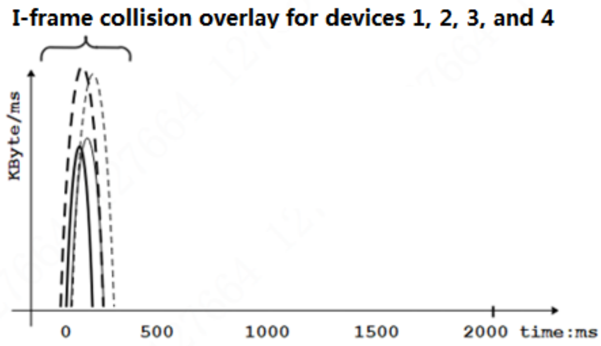


Figure 10 – 4 collision devices before the key frame is staggered

Before peak shift: 4 devices collide with overlapping key frames, and each GOP cycle will overlap, resulting in network shock.

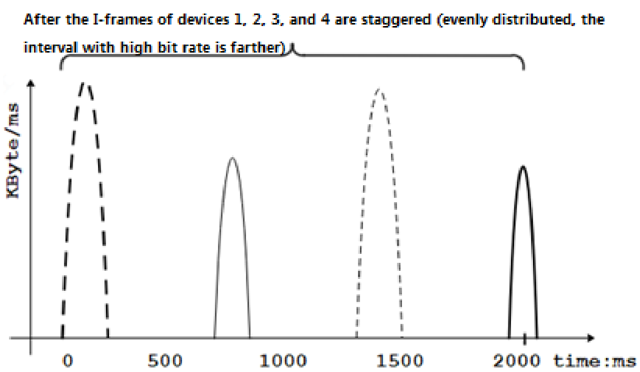


Figure 11 – 4 collision devices after the key frame is staggered

After peak staggering: four device key-frames no longer collide and overlap, and each GOP cycle no longer overlaps, and device key-frames are evenly dispersed.

Video surveillance scenarios have high requirements for real-time video. The real-time playback cache is usually between 100 and 200ms. If the network fluctuation is more than 200ms, there will be visual stun, which requires the network to complete the sending of video frames within 200ms.

Take the common video configuration FPS=25 and GOP=50 as an example, then a video frame is generated every 40ms (FPS=25), and a key frame is generated every 50 frames (GOP=50), and the period is 2000ms. In the 200ms where the key frame of a single device is located, the total number of frames=200ms/40ms=5; the rate=(1 I-frame data + 4 P-frame data)/200ms.

By analogy, in the case of staggered peaks, the 200ms granularity key frames of N devices are evenly dispersed, and the peak rates of staggered peaks are as follows:

Number of I-frames=floor ((N+9)/10)

Number of P frames=N*5-Number of I-frames

Peak rate at peak shift = (number of I-frames * size of I-frames + number of P frames * size of P frames)/200ms

Note: N devices that evenly distribute key frames during peak shifts have a maximum of floor ((N+9)/10) key frames (floor: rounded down) in every 200ms of the period of 2000ms. For example, when 10 devices are evenly distributed, at most floor ((10+9)/10) =1 key frame within 200ms; when 20 devices are evenly distributed, at most floor ((20+9)/10) =2 Key-frames; floor ((21+9)/10) =3 key frames at most when there are 21 units.

In extreme cases, the key frames of N devices are burst in the same 200ms, and the peak rate during full collision is as follows:

Number of I-frames=N

Number of P frames=N*5-Number of I-frames

Full collision peak rate = (number of I-frames * size of I-frames + number of P frames * size of P frames)/200ms

According to the actual measurement, 1080P high-definition real-time video with a bit rate of 4Mbps, the average size of I-frame is 200KB, and the average size of P frame is 20KB. According to the above calculation formula, the peak rate of multiple devices can be calculated as shown in Table 3 (1 Byte = 8 bits).

Table 3 – Peak staggered and full collision peak rate comparison

Number of Device	Peak rate when off-peak (Mbps)	Peak rate at full collision (Mbps)	Staggered peak drop (%)
1	11.2	11.2	0.00%
4	23.2	44.8	48.2%
7	35.2	78.4	55.1%
10	47.2	112.0	57.9%
13	66.4	145.6	54.4%
14	70.4	156.8	55.1%
16	78.4	179.2	56.2%
30	141.6	336.0	57.9%
31	152.8	347.2	56.0%

Taking a common base station with a bandwidth of 150M as an example, it is necessary to transmit video frames in time within 200ms. According to the data in Table 3, the theoretical effect of the number of device access channels is as shown in Table 4 (the network bandwidth planning needs to consider the service guarantee in the most extreme case, that is, full collision condition).

Table 4 – 5G video transmission network peak control technology application effect (1080P)

Stream size	4M
The maximum number of roads that can be accessed in a full collision	13
The maximum number of channels that can be accessed in staggered peaks	30
The increase rate of the number of accessible channels	130%
Full collision bandwidth utilization	34.7%
Staggered bandwidth utilization	80%
Bandwidth utilization improvement rate	130%

Description: Bandwidth utilization= (4M*Number of access channels)/150M

In the same way, the common high-definition video stream is calculated, and after applying the 5G video transmission network peak control technology, theoretically, the number of 5G terminal accesses and bandwidth utilization can be more than doubled. The effect is shown in Table 5.

Table 5 – 5G video transmission network peak control technology application effect

Stream size	4M	8M	12M
The maximum number of roads that can be accessed in a full collision	13	6	4
The maximum number of channels that can be accessed in staggered peaks	30	17	10
The increase rate of the number of accessible channels	130%	183%	150%
Full collision bandwidth utilization	34.7%	32%	32%
Staggered bandwidth utilization	80%	90.7%	80%
Bandwidth utilization improvement rate	130%	183%	150%

In the campus video surveillance environment, the cameras with seven channels of 1080P stream have been measured for more than 100 times. The highest peak rate (7 channels full collision) before the application of the network peak control technology is 73.2Mbps. After the application of the network peak control technology (I-frame without collision), the peak rate is stable at about 37.4Mbps, which is basically close to the theoretical values of 78.4Mbps and 35.2Mbps for seven cameras concurrently. In practical applications, the peak rate drops by 49%.

4. CONCLUSION

With the widespread deployment of 5G networks around the world, 5G applications in various vertical industries have gradually entered the stage of large-scale deployment from pilot demonstrations. In the campus video surveillance scenario, the solution based on 5G networks has the advantages of rapid deployment and temporary deployment and control compared with the traditional wired solution. However, during large-scale deployment, it encounters the problems of large network peak fluctuations and difficult network resource planning. 5G wireless bandwidth, especially uplink bandwidth resources are precious. How to use wireless bandwidth resources more effectively in practical applications is very important in terms of commercial implementation and technical feasibility.

The severe fluctuation of network peaks in 5G video surveillance scenarios is largely caused by the collision of I-frames during multichannel video transmission. Network peak control in 5G video transmission scenarios, based on key technologies such as I-frame collision detection, network peak shift scheduling, etc., through the MEC external interface, the network data and video management system are opened up to form the linkage between 5G network and video management applications, which can effectively smooth multiple channels. When the number of terminal devices is above 7 channels, the network peak value during video transmission increases by 100% percent the number of 5G terminal accesses and bandwidth utilization under 5G base stations and facilitates the large-scale promotion of 5G video surveillance scenarios.

REFERENCES

[1] ZHOU Y Q, TIAN L, LIU L, et al. Fog computing enabled future mobile communication networks: a convergence of communication and computing[J]. IEEE Communications Magazine, 2019, 57(5): 20-27.

[2] BARBAROSSA S, SARDELLITTI S, DI LORENZO P. Communicating While Computing: distributed mobile cloud computing over 5G heterogeneous networks[J]. IEEE Signal Processing Magazine, 2014, 31(6): 45-55.

[3] PATEL M. Mobile-edge computing introductory technical white paper [R]. 2014.

[4] White Paper. 5G ToB Service Experience Standard[R]. 2021.

[5] Guobo Zhang, Xuan Zhang, Chongrong Li. A study of stream smoothing methods for network video surveillance system[J]. J. Huazhong Univ. of Sci. & Tech. (Natural Science Edition). 2016.

[6] Gongyu JIN, Xiaoyong WU, etc. A rate adaptation algorithm for real-time video streaming[J]. JOURNAL OF COMMUNICATION UNIVERSITY OF CHINA (SCIENCE AND TECHNOLOGY). 2022.

[7] CHEN Y, QING J M, LIN B Y, LIU C Q. Research of keyframe extraction technology in massive video data analysis[J]. Electronics World, 2018(18):164-165.

[8] Zhang X L, Li Y, Zhao Y J. Key frame extraction for SLAM based on spatio-temporal slices[J]. Journal of System Simulation, 2018, 30(05):1724-1729.

[9] CAI X H, HUA N, WU D, CHEN W Z. Distributed video analysis decoupling mechanism based on keyframe[J]. Video Engineering, 2015, 39(14):1-4+20.

A COMPARATIVE ANALYSIS OF AUGMENTED REALITY FRAMEWORKS AIMED AT DIVERSE COMPUTING APPLICATIONS

Mfundo A. Maneli[1]; Omowunmi E. Isafiade[1]

[1]Department of Computer Science, Faculty of Natural Science, University of the Western Cape, South Africa

ABSTRACT

Immersive systems such as Augmented Reality (AR) and Virtual Reality (VR) have proven useful in diverse computing domains. However, there is little effort on accuracy measurements within AR applications, which could greatly impact outcomes and decisions in certain domains, such as crime scene investigations, among others. This paper aims to analyze and evaluate two existing prominent AR frameworks, ARCore and ARKit, which support the development of diverse mobile computing applications for immersive systems. This research developed prototype applications and conducted comparison tests of measurement accuracy within the applications. The accuracy was tested using four distance criteria across six different devices, spanning ARCore and ARKit frameworks. A control experiment was used to benchmark the measurement accuracy. Relatively, an instance of the experiment presented ARCore as reliable. Overall, ARKit proved to be more accurate between the two frameworks, with an average accuracy of 99.36% as opposed to 89.42% scored by ARCore. The obtained results can give insight on the choice of framework to consider during AR application development for a specific domain, hence boosting quality of experience.

Keywords – ARCore, ARKit, AR measurements, augmented reality, crime scene investigation

1. INTRODUCTION

Research development in the field of information technology is ever advancing at a rapid pace. Newer and more efficient methods of improving human interactions with their environments are continually devised. These human interactions and environment methods encompass immersive technologies [1]. Augmented Reality (AR) and Virtual Reality (VR) are the two primary types of immersive technology, but other technologies such as Mixed Reality (MR) and extended Reality (XR) also fall under the immersive technologies umbrella. AR is a technology which can superimpose digital perceptual information in the real world [1], as shown in Figure 1. VR, on the other hand, offers a three-Dimensional (3D) computer generated environment, and immersive, and multisensory experience [2], which rely on 3D, stereoscopic, head tracked displays, hand/body tracking and binaural sound, as shown in Figure 2. The earlier methodologies of achieving AR/VR required bulky, computationally-demanding and expensive equipment to operate [3]. As time went by further research and development came into place. Consequently, low powered devices such as smartphones can run AR and power-efficient head mounted displays can facilitate VR [4]. Immersive technologies have been widely adopted in diverse sectors such as education [5], advertising [6], gaming [7], forensic science and crime scene reconstruction [8], among others.

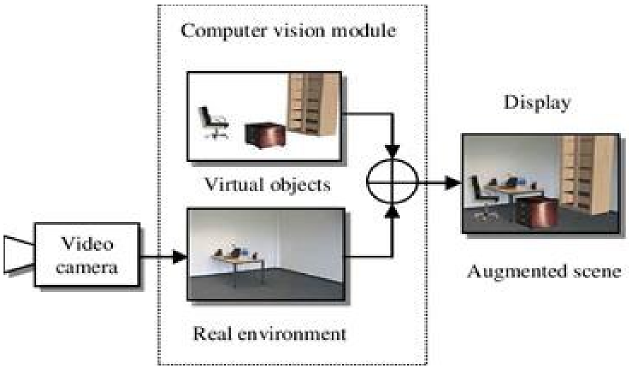


Figure 1 – Augmented reality schematic diagram [9]

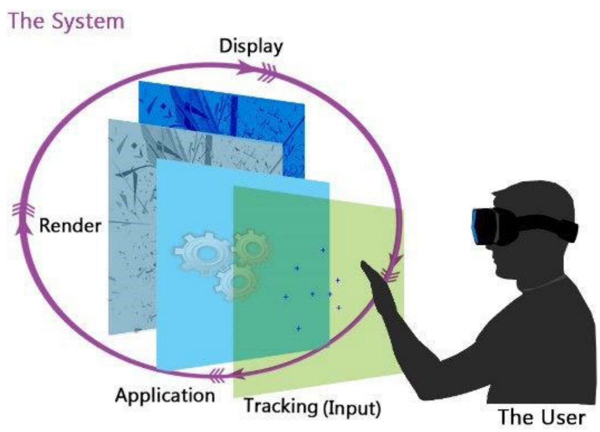


Figure 2 – Virtual reality schematic diagram [2]

Apple and Google are the two mobile technology giants that dominate the mobile operating system sector [10] as shown in Figure 3. Apple provides iOS for its mobile smartphone counterparts, and Google provides the Android OS for an entire range of different smartphone manufacturers. These giants each developed frameworks which cater for augmented reality functionalities such as AR face tracking, mapping, just to mention a few [11], [12]. Google offers ARCore while Apple offers ARKit, both frameworks have the same goal but may not necessarily always achieve the same level of accuracy and reliability in certain applications. It is important to note that the impact of AR accuracy measurements may be graver in certain domains than others. For instance, in crime investigation and forensic science where critical decisions and court verdicts are based on evidence gathered from a crime scene, as opposed to advertisement, where the goal could just be to display the existence and desirable features of an object.

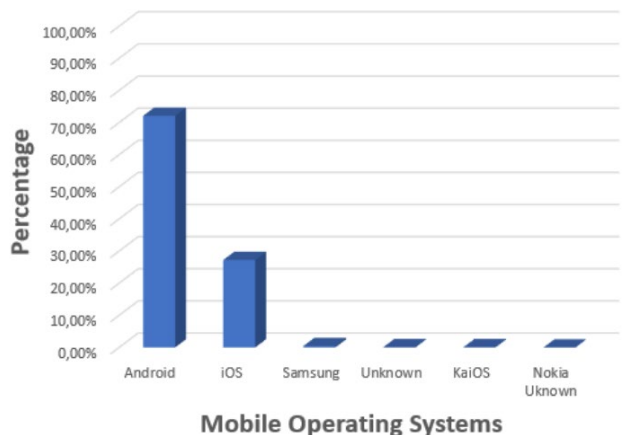


Figure 3 – Mobile operating systems globally [10]

One of the roles played by a crime scene investigator is to capture accurate and reliable crime data at a rapid pace, without the risk of contaminating a scene [13]. Forensic data collection encompasses a wide field such as post-mortem full body documentations [14], blood spatter analysis [15], footwear analysis [16], to mention a few. The use of a Global Positioning System (GPS) for accurate indoor mapping and traditional data capturing methods may not be sufficient for the given tasks. In this paper, we compare the accuracy and reliability between two frameworks which aim to provide accurate augmentation measurements which could be beneficial for forensic investigators. There is limited research that has investigated the comparison between ARKit and ARCore, in particular the accuracy of the AR measurements achieved by these frameworks. This has motivated the need for this paper.

The remainder of this paper is structured as follows: Section 2 presents related research and further justifies the novelty of our research. Section 3 presents the research methodology. Results are presented in Section 4, while Section 5 concludes the paper and provides future recommendations.

2. RELATED WORK

This section discusses some of the previous research efforts that relate to the comparison between the ARKit and ARCore frameworks for diverse computing applications.

P. Nowacki et al., [17] explored the potential of ARCore and ARKit platforms for AR/VR applications. The objective of their study was to evaluate the capabilities of ARKit and ARCore and help in choosing the right framework to speed up prototyping and development of modern AR/VR applications. General performance in terms of CPU/memory usage as well as mapping quality were the major criteria examined. Their work however did not describe trends of other AR/VR frameworks, but rather focused on two main technologies, which are ARKit and ARCore. This paper also did not look at the accuracy derived from AR calculations conducted over ARKit and ARCore.

Z. Oufqir et al., [18] present a study which implements and concretizes the different functionalities available in augmented reality to enrich the real world with additional information. The objective of their study was to evaluate the capabilities of the two libraries ARKit and ARCore and their capability in the development of augmented reality applications.

R. Cervenak et al., [11] in their work, present the possibilities of indoor space mapping and user movement tracking using augmented reality technologies. The objective of their study was to evaluate the possibility of the use of ARKit and ARCore to analyze movement in space without using other navigation technologies. Emphasis is placed on optimizing the algorithms created to track device position in space.

J. Borduas et al., [12] present a study which compares the reliability of four mobile 3D scanning technologies and provides insight and recommendations as to which of these are sufficiently reliable for the customization of respiratory face masks. The objective of the study is to compare the reliability of ARCore: Augmented Faces SDK, ARKit: Face Tracking SDK, ScandyPro app using the raw information of the TrueDepth Structured Light sensor and the 3DSizeMe app using the Structure Sensor by Occipital.

H. Fabrício et al., [19] presents a comparative analysis of augmented reality frameworks aimed at the development of educational applications. The objective of the study is to compare the characteristics of existing frameworks that may allow the development of educational solutions using augmented reality resources, focusing on tools that enable the conception, design, and implementation of mobile applications.

While the aforementioned research efforts have done some work on exploring and comparing ARCore and ARKit frameworks, none of these efforts have been on the accuracy of augmented reality measurements as presented in the current research.

3. DESIGN AND TEST METHODOLOGY

3.1 Test setup

In this research, two developed mobile applications based on ARCore and ARKit are used to determine the accuracy and reliability of augmented reality solutions. Six mobile devices were utilized, which are Samsung S8, S10, S20, A20 and A32 and an Apple iPad Pro 5th generation. The Apple iPad Pro 5th generation represents ARKit while the other devices represent ARCore. The reason for only having one device to represent ARKit is due to the fact that the chosen device is the latest to house a LiDAR scanner and has the best configurations for dynamic ranging which all the other ARCore devices do not possess. The applications were both developed on Unity engine using the ARCore and ARKit XR Plugins versions 4.2.3. The backend code was done using visual studio; a desktop computer was utilized to develop both applications, and finally a tape measure was used to act as a control for the entire experiment. The entire test application layouts are presented in figures 4 and 5. Table 1 outlines the key camera configurations for each device used in this study.

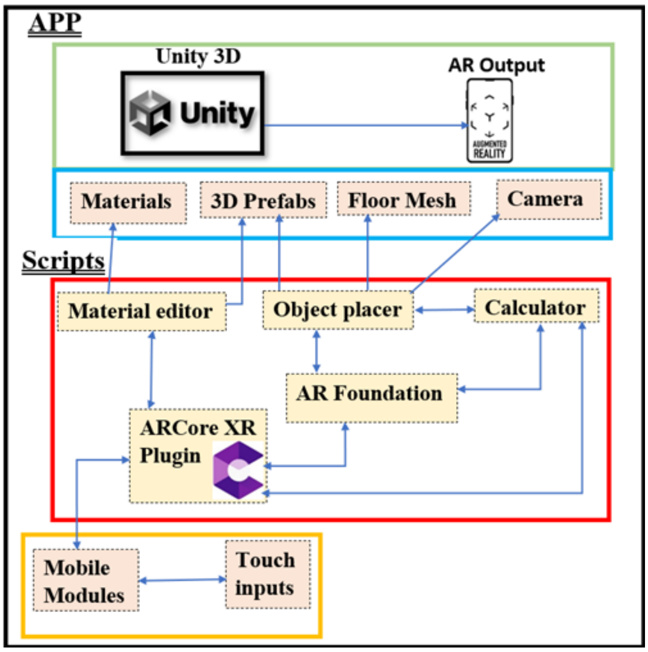


Figure 4 – ARCore application layout

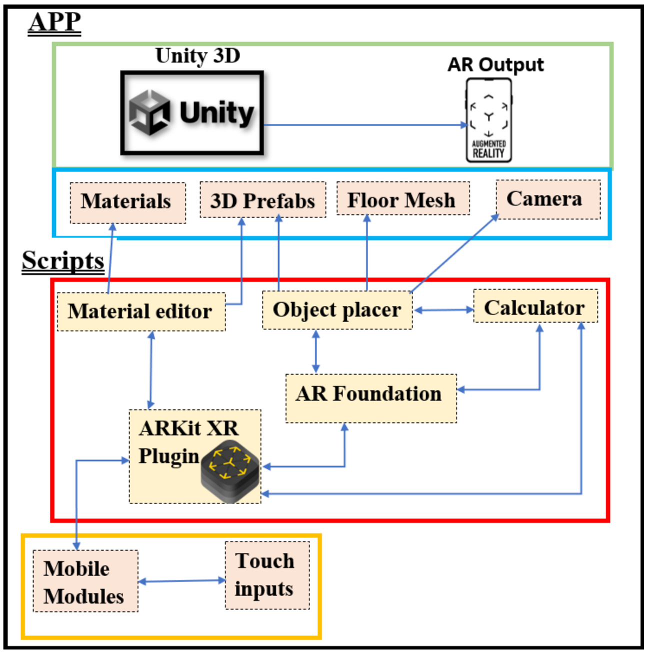


Figure 5 – ARKit application layout

Table 1 – Camera configuration of each device

Device	Camera specifications
Samsung S8	12 MP, f/1.7, 26mm (wide), 1/2.55", 1.4µm, dual pixel PDAF, OIS
Samsung S10	12 MP, f/1.5-2.4, 26mm (wide), 1/2.55", 1.4µm, Dual Pixel PDAF, OIS 12 MP, f/2.4, 52mm (telephoto), 1/3.6", 1.0µm, AF, OIS, 2x optical zoom 16 MP, f/2.2, 12mm (ultrawide), 1/3.1", 1.0µm, Super Steady video
Samsung S20	12 MP, f/1.8, 26mm (wide), 1/1.76", 1.8µm, Dual Pixel PDAF, OIS 64 MP, f/2.0, 29mm (telephoto), 1/1.72", 0.8µm, PDAF, OIS, 1.1x optical zoom, 3x hybrid zoom 12 MP, f/2.2, 13mm, 120° (ultrawide), 1/2.55" 1.4µm, Super Steady video
Samsung A20	13 MP, f/1.9, 28mm (wide), AF 5 MP, f/2.2, 12mm (ultrawide)
Samsung A32	64 MP, f/1.8, 26mm (wide), PDAF 8 MP, f/2.2, 123°, (ultrawide), 1/4.0", 1.12µm 5 MP, f/2.4, (macro) 5 MP, f/2.4, (depth)
Apple iPad 5th Gen	12 MP, f/1.8, (wide), 1/3", 1.22µm, dual pixel PDAF 10 MP, f/2.4, 125° (ultrawide) TOF 3D LiDAR scanner (depth)

3.2 Data collection and evaluation metric

The nature of data required for the experiment is AR measurements. Measurements acquired from the tests are compared to measurements captured with the control tape measure. For consistency, the test environment was controlled such that ambient lighting was kept constant throughout the tests, ensuring adequate lighting for plane detection and AR measurements.

Figure 6 depicts the practical use case of how the 100 cm criteria was derived. A crime scene-related sample scenario was used, where the knife's handle represents point A and the leg of the chair was used as point B (end point). The distance between points A and B was set at 100cm using a tape measure. The same approach was used to set the "10cm", "45cm" and "75cm" distance criteria. The crime scene use case was considered to see how practical AR measurements can assist with measuring crime scene evidence without the risk of contamination by means of physically touching the knife.

The accuracy of the AR frameworks was assessed based on how close they could get to the control criteria. Six test runs were conducted per device per measurement criteria. Four measurement criteria were used, which are "10cm", "45cm", "75cm" and "100cm" at a distance of one meter and two meters across all tests. The average score ($\overline{D}$) of each device per measurement criteria after six test runs was then computed using Equation (1).

$$\overline{D} = \frac{1}{N} \sum_{j=1}^{N} Xj \, . \qquad (1)$$

Where N represents the total number of test runs that was conducted for each device (D) per each distance criterion. Parameter X represents the measured distance between two points (A and B) per time per test for each D.

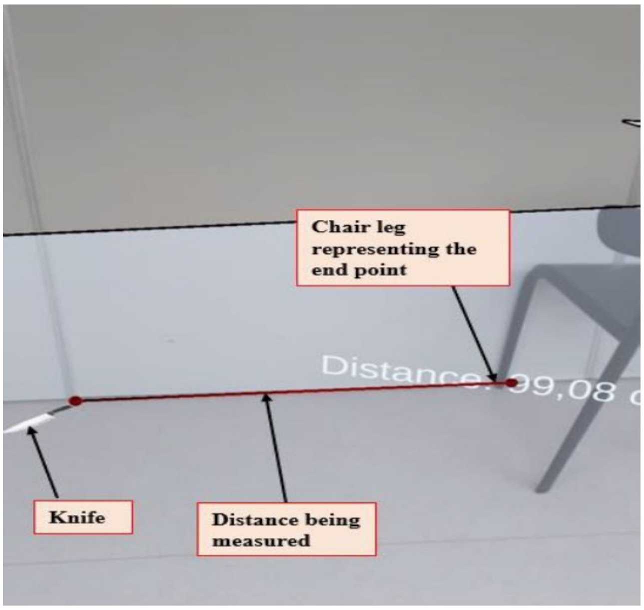


Figure 6 – Crime scene sample scenario with 100 cm AR measurements taken from 1 meter away

4. RESULT AND DISCUSSION

Tables 2 and 3 provide the sample readings of how the ARCore and ARKit devices are performing. AR calculations conducted over ARKit seem to be the most accurate when gauged against other devices running ARCore and the control value. The best and worst results observed for each device are also noted in the tables. The measurements taken with the tape measure (i.e. control) remained constant as expected, and hence marked as N/A in the tables for each distance criterion. The graphs of the readings were then plotted as shown in figures 7, 8, 9, 10, 11, 12, 13 and 14. These graphs reveal the performance trend of the devices spanning ARCore and ARKit when compared to the control measurement.

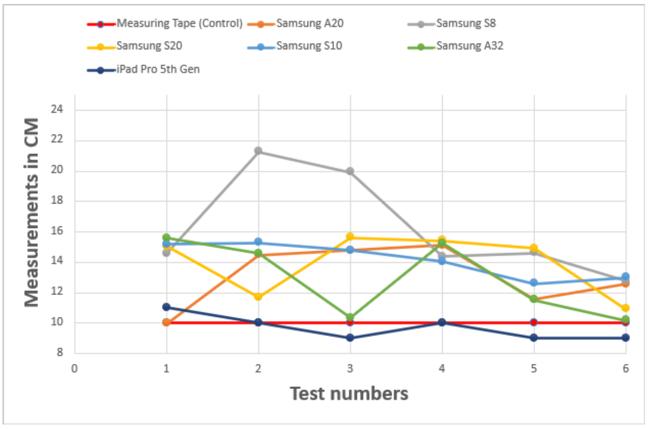


Figure 7 – 10cm AR measurements taken from 1 meter away

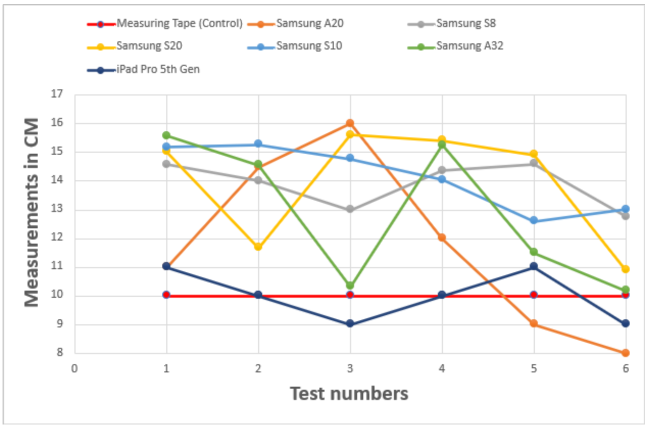


Figure 8 – 10cm AR measurements taken from 2 meters away

While it is hypothesized that the accuracy of the measurements would improve within close proximity and reduce the further away the devices are from the focus area, which is the crime scene scenario, the comparison between 10cm taken at 1 meter and at 2 meters depicts an unexpected trend in which the devices seem to improve in terms of accuracy the further away they are. Only the Samsung S10 and Samsung A32 retain similar scores at the 1-meter mark and 2-meters' mark, as seen in figures 7 and 8.

Table 2 – AR 100cm measurements taken from 1 meter away

Device Name	Result 1 (CM)	Result 2 (CM)	Result 3 (CM)	Result 4 (CM)	Result 5 (CM)	Result 6 (CM)	Average (CM)	Best Result (CM)	Worst Result (CM)
Tape measure	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**100**	**N/A**	**N/A**
Samsung S8	120	110	115	102	105	106	**109.70**	**102**	**120**
Samsung S10	102	98	98	94	106	99	**99.55**	**99**	**106**
Samsung S20	125	111	108	108	109	125	**114.39**	**108**	**125**
Samsung A20	90	99	102	98	99	107	**98.99**	**99**	**90**
Samsung A32	106	100	107	105	108	97	**103.86**	**100**	**108**
Apple iPad 5th Gen	97	98	103	102	98	103	**100**	**98**	**103**

Table 3 – AR 100cm measurements taken from 2 meters away

Device Name	Result 1 (CM)	Result 2 (CM)	Result 3 (CM)	Result 4 (CM)	Result 5 (CM)	Result 6 (CM)	Average (CM)	Best Result (CM)	Worst Result (CM)
Tape measure	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**100**	**N/A**	**N/A**
Samsung S8	**102**	**108**	**104**	**103**	**107**	**99**	**103.83**	**99**	**108**
Samsung S10	**99**	**103**	**104**	**107**	**98**	**102**	**102.17**	**99**	**107**
Samsung S20	**120**	**90**	**106**	**98**	**99**	**110**	**103.83**	**99**	**120**
Samsung A20	**87**	**102**	**86**	**97**	**98**	**107**	**96.17**	**98**	**107**
Samsung A32	**100**	**112**	**104**	**104**	**103**	**97**	**103.40**	**100**	**112**
Apple iPad 5th Gen	**101**	**99**	**102**	**102**	**100**	**98**	**99.83**	**100**	**102**

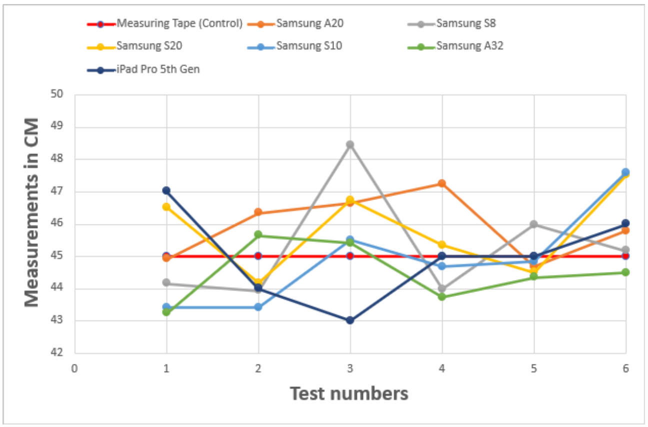


Figure 9 – 45cm AR measurements taken from 1 meter away

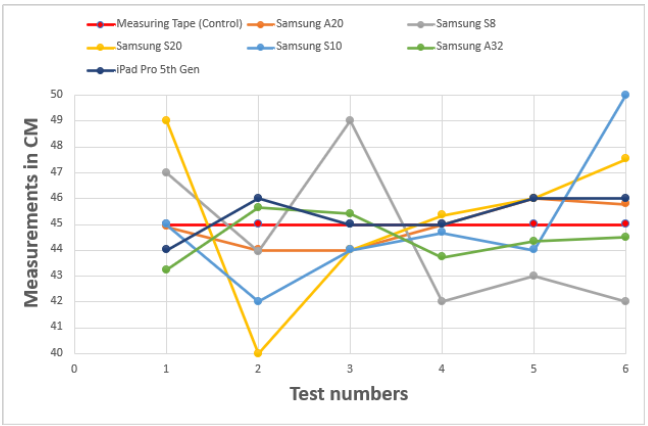


Figure 10 – 45cm AR measurements taken from 2 meters away

The comparison between 45cm taken at 1 meter and at 2 meters depicts a similar trend to the first observation of 10cm taken at 1 and 2 meters away. The majority of devices seem to improve in terms of average accuracy the further away they are. However, the Samsung A20 seems to have better accuracy at 1 meter compared to 2 meters away. We also note that the Samsung A32 retained the same accuracy average at 1 and 2 meters away as seen in figures 9 and 10.

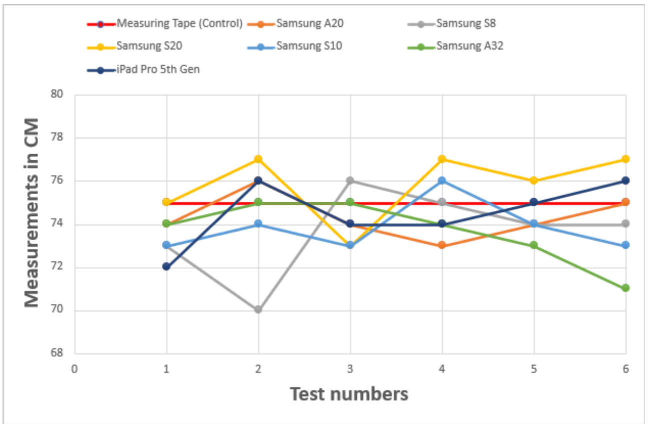


Figure 11 – 75cm AR measurements taken from 1 meter away

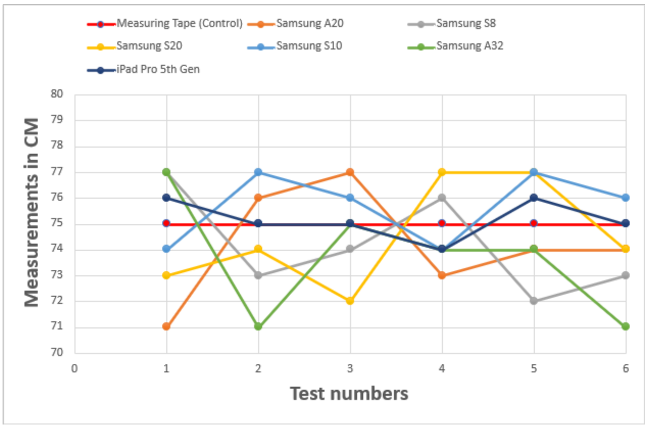


Figure 12 – 75cm AR measurements taken from 2 meters away

The comparison between 75cm taken at 1 meter and at 2 meters depicts a trend whereby the majority of devices seem to improve in terms of accuracy the further away they are. However, the Samsung A20 did not follow this trend and had a higher average accuracy at 1 meter away. The Samsung A32 retains a similar score at the 1-meter mark and 2-meter mark as shown in figures 11 and 12.

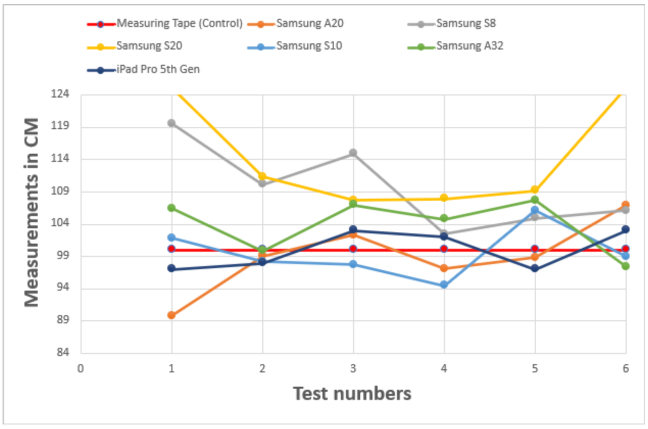


Figure 13 – 100cm AR measurements taken from 1 meter away

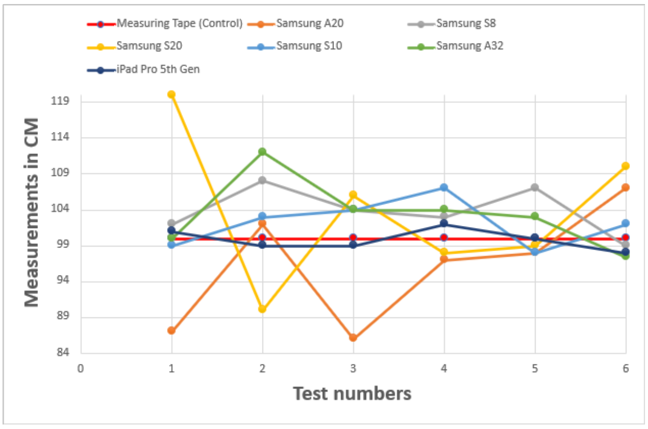


Figure 14 – 100cm AR measurements taken from 2 meters away

The comparison between 100cm measurements taken at 1 meter and 2 meters reveals a trend in which the majority of the devices improved in terms of accuracy the further away they are. The Samsung A20 and Samsung A32 devices decreased in accuracy scores at the 2-meter mark compared to the 1-meter mark as shown in figures 13 and 14.

The average accuracy percentage scores for ARKit are measured by averaging the sum of the six scores per test and comparing them to the control value per control criteria. The same notion was used to calculate for ARCore. Based on all the conducted experiments, comparing the ARKit framework with ARCore framework, the following is noted: In Figure 7, ARKit scored an average accuracy of 96.70% and ARCore obtained an average accuracy score of 59.5%. In Figure 8, ARKit scored an average accuracy score of 100% and ARCore scored an average accuracy of 71.78%. In Figure 9, ARKit scored an average accuracy of 100%, while ARCore scored an average accuracy of 99.38%. In Figure 10, ARKit scored 99.27% and ARCore scored 99.63%. In Figure 11, ARKit scored 99.33% and ARCore scored 99.02%. In Figure 12, ARKit scored 99.78% and ARCore scored 99.24%. In Figure 13, ARKit scored 100% and ARCore scored 94.70%. Lastly, in Figure 14 ARKit scored 99.83 and ARCore 92.12%.

Table 4 – Performance comparison of ARCore and ARKit

Framework	Average accuracy score	Deviation
ARCore	89.42%	10.58%
ARKit	99.36%	0.64%

Table 4 provides a summary of how the two frameworks performed across all devices, spanning ARCore and ARKit in all given tests. These results were obtained by adding all the average percentage values obtained across all tests as shown in figures 7, 8, 9, 10, 11, 12, 13 and 14. Then those total results were divided by 8 to get the average accuracy score shown in Table 4. Table 4 illustrates that ARKit proves to be far superior compared to ARCore regarding accuracy and reliability in AR-related work and diverse computing applications within the context of the criteria used in this study.

5. CONCLUSION AND FUTURE RESEARCH

This work has provided a comparative analysis of two prominent augmented reality frameworks, ARKit and ARCore, aimed at diverse computing applications, using a crime scene-related scenario as a use case. In total, eight tests were conducted with six test runs per test or device to gauge the accuracy and reliability of all devices. Six devices were used and five of those devices ran the application under the ARCore framework, while one device ran the test under the ARKit framework. One device ran the test under ARKit due to the fact that the chosen device is the latest to house a LiDAR scanner and has the best configurations for dynamic ranging, which all the other ARCore devices do not have. Four measurement criteria were used, which are 10cm, 45cm, 75cm and 100cm at a distance of 1 meter and 2 meters across all tests. ARKit proved to be the most accurate and superior between the two frameworks by scoring an average accuracy of 99.36%, as opposed to the 89.42% scored by ARCore. The device running ARKit was the most accurate and reliable in seven out of the eight tests based on the criteria used. Figure 9 illustrates the one occurrence where every device running ARCore performed better than the device running ARKit. When conducting the tests, an unexpected trend occurred whereby most of the devices seemed to improve in terms of accuracy the further away they were from the target being measured. Furthermore, the tests were conducted based on the applications built using the two AR frameworks, and scientifically evaluated and benchmarked to gauge any metric errors or bottleneck capabilities in determining AR measurement outcomes. The results presented in this research can guide future research on the choice of framework to explore for prototyping and development of immersive applications. However, it is also worth noting that the continuous development of mobile technology may impact future performance and choices, as both frameworks may have their strengths and weaknesses.

One thing to note with this study is the limited diversity regarding devices running ARKit. In future work, we plan to use a much larger number of devices capable of running ARKit and offer additional test parameters such as time taken to acquire measurements, system utilization (CPU and RAM), quality mapping and plane detection coverage. These additional tests on multiple ARKit and ARCore devices will ensure a much fairer and broader comparison.

REFERENCES

[1] R. Aggarwal and A. Singhal, Augmented Reality and its effect on our life, in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2019, pp. 510-515. doi: 10.1109/CONFLUENCE.2019.8776989.

[2] H. Kharoub, M. Lataifeh, and N. Ahmed, 3D User Interface Design and Usability for Immersive VR, Applied *Sciences*, vol. 9, p. 4861, Nov. 2019, doi: 10.3390/app9224861.

[3] A. Basu, A brief chronology of Virtual Reality. arXiv, 2019. doi: 10.48550/ARXIV.1911.09605.

[4] R. Schroeder, Virtual reality in the real world: History, applications and projections, *Futures*, vol. 25, no. 9, pp. 963-973, 1993, doi: https://doi.org/10.1016/0016-3287(93)90062-X

[5] P. Chen, X. Liu, W. Cheng, and R. Huang, "A review of using Augmented Reality in Education from 2011 to 2016," in *Innovations in Smart Learning*, 2017, pp. 13-18.

[6] P. A. Rauschnabel, B. J. Babin, M. C. Tom Dieck, N. Krey, and T. Jung, "What is augmented reality marketing? Its definition, complexity, and future," *Journal of Business Research*, vol. 142, pp. 1140-1150, 2022, doi: https://doi.org/10.1016/j.jbusres.2021.12.084

[7] C. Lo, "An acupuncture VR game," 2019, pp. 237-240. doi: 10.1201/9780429058127-49

[8] M. A. Maneli and O. E. Isafiade, "3D Forensic Crime Scene Reconstruction Involving Immersive Technology: A Systematic Literature Review," in *IEEE Access*, vol. 10, pp. 88821-88857, 2022, doi: 10.1109/ACCESS.2022.3199437.

[9] M. Hamidia, N. Zenati, H. Belghit, and A. Bellarbi, *Object Recognition Based on ORB Descriptor for Markerless Augmented Reality*. 2015.

[10] "Mobile Operating System Market Share Worldwide, Statcounter Global Stats." https://gs.statcounter.com/os-market-share/mobile/worldwide (accessed Jul. 14, 2022).

[11] R. Cervenak and P. Masek, "ARKit as indoor positioning system," in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2019, pp. 1-5. doi: 10.1109/ICUMT48472.2019.8970761

[12] J. Borduas, A. Castonguay, P. Laurin, and D. Beland, "Reliability of Mobile 3D Scanning Technologies for the Customization of Respiratory or Face Masks," Oct. 2020. doi: 10.15221/20.34.

[13] J. W. Streefkerk, M. Houben, and J. Dijk, "LNCS 8022 - The ART of CSI: An Augmented Reality Tool (ART) to Annotate Crime Scenes in Forensic Investigation," 2013.

[14] P. Claes, D. Vandermeulen, S. de Greef, G. Willems, J. G. Clement, and P. Suetens, "Computerized craniofacial reconstruction: Conceptual framework and review," *Forensic Science International*, vol. 201, no. 1, pp. 138-145, 2010, doi: https://doi.org/10.1016/j.forsciint.2010.03.008

[15] G. Naveen and Sarvamangala, "A Survey on Techniques Used in Blood Spatter Analysis, *Asian Journal of Engineering and Technology Innovation,* vol. 4, no. 7, 2016.

[16] U. Buck, N. Albertini, S. Naether, and M. J. Thali, "3D docmentation of footwear impressions and tyre tracks in snow with high resolution optical surface scanning," *Forensic Science International*, vol. 171, no. 2-3, pp. 157-164, Sep. 2007, doi: 10.1016/j.forsciint.2006.11.001

[17] P. Nowacki and M. Woda, "Capabilities of ARCore and ARKit Platforms for AR/VR Applications," 2020, pp. 358-370. doi: 10.1007/978-3-030-19501-4_36.

[18] Z. Oufqir, A. el Abderrahmani, and K. Satori, "ARKit and ARCore in serve to augmented reality," *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp. 1-7, 2020.

[19] H. Fabrício, G. Renan and T. Liane, "A Comparative Analysis of Augmented Reality Frameworks Aimed at the Development of Educational Applications," *Creative Education* 2017, doi: 10.4236/ce.2017.89101

VIDEO DEMONSTRATION TRACK

MAKING EXTENDED REALITY SAFE AND SECURE FOR TEENAGERS WITH PARGUARD

Agasthya Gangavarapu, Researcher, Safety4XR.org

1. INTRODUCTION

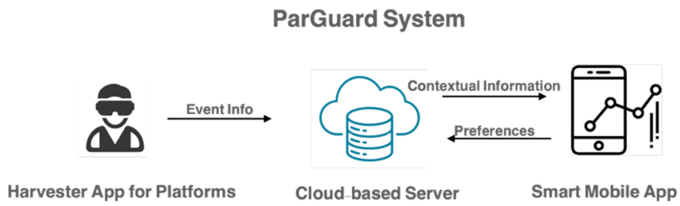
Extended Reality (XR) is expected to grow exponentially with significant investments from big tech companies such as Meta and the introduction of related and supporting new technologies such as web3, blockchain, etc. Like other technologies, most of the discussion and analysis is focused on the positive benefits of XR technologies and systems at the inception stage. If history was any guide, then the negative impacts of XR technologies with ultra-personalization capabilities and advanced immersive technologies are likely to be much worse, especially for teenagers. Based on the current direction of regulations and laws related to the use of technologies, including social media and XR, the current expectation is that parents 'regulate' the usage of the technologies and nudge their teenage children away from the ill effects of XR technologies [1]. This is an unreasonable expectation and an onerous burden placed on the parents given personalized algorithms and the fast-changing nature of the underlying technologies. ParGuard, a smartphone-based app system, is designed to help parents to navigate through a maze of technologies and ultra-personalization algorithms, and guard children from the ill effects of new technologies.

2. PARENTAL PROBLEMS

Most concerned parents have signed up with social media platforms to provide any guidance needed for their children. While these practices have worked to a certain extent with current social media because of limited algorithmic personalization, the parents get to see the same posts and to a certain extent similar ads as their children [2]. However, personalization algorithms and immersive experiences of XR technologies are driven by the history of interactions, and implicit preferences of the users. Parents are served with totally different experiences because their interaction history and ongoing choices are different from that of their children. Also, parents need to understand the contextual background when their children are having a negative emotional response such as stress or loss of appetite.

3. PARGUARD

To help parents to track their children's XR platform activity and provide contextual guidance, I am developing ParGuard. There are three components to the solution in development: a harvester component for each of the XR platforms, a smartphone app for parents to enter the teen's emotional details and get insights, and a cloud-based server for classifying impactful events and actions.



Harvester component

The first component of the ParGuard system is the Harvester. The component is customized for each of the popular XR devices like Meta's Oculus Quest and is installed and activated with the child's account. Once activated, the component captures emotional information such as mental fatigue [3], cognitive load [4], toxic interactions, etc. using the cameras, sensors, and others. The Harvester sends the information securely to a cloud-based server for synthesis and generating insights [5].

Smart mobile app

The mobile app is primarily designed for parents to receive all contextual and event information along with Machine Learning (ML) based insights from cloud-based server. This information is mashed with the personal information of the concerned child, securely entered in the app, to get insights about the child's emotional wellness and identify any potential causes for concern. Parents can choose the type and details of personal information to enter about their children. For example, the parents can enter negative emotions the teenager is experiencing in the app and the information gets synthesized with XR experiences to generate personalized interventions and recommendations.

Cloud-based server

A cloud-based server facilitates and synthesizes the information received from the mobile app and harvester to identify potential problems and generate personalized recommendations. The data is encrypted and provided to authorized apps using ephemeral tokens.

4. DEMO

The demo, which lasts for about 3 min, will be principally focused on the mobile app and how it provides actionable insights for the parents. Also, the demo will highlight how the app could be used to customize for child's sensitivities such as flashlight sensitivities. Since the information captured and used in ParGuard is highly sensitive information, the demo will show how differential privacy is enabled to protect the privacy of all parties involved.

REFERENCES

[1] https://www.govtech.com/policy/conn-bill-would-restrict-teenagers-social-media-use

[2] https://www.wsj.com/articles/tiktok-to-adjust-its-algorithm-to-avoid-negative-reinforcement-11639661801

[3] https://eprint.iacr.org/2020/340.pdf

[4] https://www.sciencedirect.com/science/article/pii/S0933365717306140

[5] Ishimaru S, Kunze K, Kise K, Weppner J, Dengel A, Lukowicz P, et al. In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. In: ACM Augmented Human International Conference. New York, NY, USA: ACM; 2014. p. 15:1–15:4.

ABSTRACTS

Session 1: Some perspectives on future networks[1]	
S1.1	Integrated network control architecture for terrestrial and non-terrestrial network convergence in beyond 5G systems *Ved P. Kafle, Mariko Sekiguchi, Hitoshi Asaeda and Hiroaki Harai (National Institute of Information and Communications Technology, Japan)* To seamlessly cover urban and rural areas, mountains and deserts, as well as sea and air, with high quality ubiquitous communication services, the convergence of Terrestrial Networks (TNs) and Non-Terrestrial Networks (NTNs) such as satellites is essential in beyond 5G communication systems. This paper presents the scenarios and architectures of integrated and individual network control systems for TN and NTN convergence in the control plane. The integrated network control system enables end-to-end service design, orchestration, performance monitoring, closed-loop control, and automation. It promotes interoperability in the control plane of TN and NTN domains composed of virtualization-supporting infrastructures and possibly managed by different organizations. This paper is related to ITU-T Study Group 13's activities of standardizing fixed, mobile and satellite convergence technologies.
S1.2	Towards computing and network convergence: QoE-oriented service anycast based on SRv6 *Zicheng Wang, Yetong Wang, Xian Gao, Shuai Wu and Wei Lin (Inspur Communications Technology Co., Ltd., China)* The new mobile Internet services represented by extended Reality (XR) have brought new challenges to computing and networks. A new generation of Information and Communication Technology (ICT) infrastructure with both computing and communication capabilities is being promoted. Traditional load balancing technologies cannot sense the connection between users and services. It is also difficult to support large-scale distributed computing service instances, so the service experience provided can be poor. In this paper a service anycast system based on a Segment Routing over IPv6 (SRv6) data plane is proposed, which can provide a Quality of Experience (QoE)-oriented load balancing capability for the network according to the obtained status information of network and computing service instances. Besides, an IP address identification mechanism is also proposed to help the control plane handle traffic engineering policies, and to support efficient service resource discovery, seamless mobility, and service continuity.
S1.3	Towards a more flexible networking landscape *David Lou (Huawei Technologies Duesseldorf GmbH, Germany); Marinos Charalambides (Independent Researcher, United Kingdom)* Technological advancements leading to 5th generation networks mainly focused on improving coverage and performance in terms of bandwidth and latency. While these will likely remain aspects of continuous improvement, along with issues on reliability and security, this paper argues that flexibility is a key property that 6G networks should exhibit in order to overcome important limitations of the current networking landscape and fulfill emerging user needs and application requirements. We identify key areas that can contribute towards more flexibility, present existing efforts on relevant technologies and discuss the research challenges that need to be addressed in order to reach the desired level of flexibility.

[1] Papers marked with an "*" were nominated for the three best paper awards.

Session 2: Augmented reality systems: design and implementation	
S2.1	A framework for the design, implementation and evaluation of a multi-variant Augmented Reality application* *Sophie Westfahl and Dany Meyer-Renner (University of Applied Sciences Neu-Ulm, Germany); Antoine Bagula (University of the Western Cape, South Africa)* Augmented Reality (AR) is one of the key technologies of the fourth Industrial Revolution (4IR) and plays an increasingly important role in many companies. However, while the demand for new AR applications is rapidly increasing, fundamental best practices and frameworks for the industrial AR sector are still scarce or in their infancy stage. This paper addresses this gap by proposing a framework for the design and efficient implementation of AR applications with multiple models and variants. The proposed framework is built around: i) a development process that describes the different steps for the design of a model-based AR application and its implementation with Unity and Vuforia model targets; and ii) a multilayer orchestration model that describes the different interactions between a user and a server layer. The proposed framework is successfully implemented, and its performance analyzed using both quantitative and qualitative evaluation based on the Brooke's System Usability Scale.
S2.2	Enhancing user experience in pedestrian navigation based on Augmented Reality and landmark recognition* *Dhananjay Kumar, Shreayaas Iyer, Easwar Raja and Ragul Kumar (Anna University, MIT Campus, Chennai, India); Ved P. Kafle (National Institute of Information and Communications Technology, Japan)* Pedestrian navigation using traditional mapping systems is constrained by the inherent limitations of the existing digital online mapping services. The major challenges include complete reliance on GPS for user localization and inferior user experience caused by lack of information about the surroundings especially in unknown environments. In this paper, we design and develop a marker-less augmented reality based pedestrian navigation system which can handle navigation even in the absence of GPS as well as improve user experience by providing a novel landmark recognition feature, which allows users to identify nearby buildings or streets during navigation. To mitigate the absence of GPS signal, a user localization method utilizing a step-count based distance estimator is proposed. The performance comparison with existing state of the art techniques and devices shows locational accuracy of 2.5 meters on average and step-count detection accuracy increase of nearly 0.5% with latency of 70 milliseconds in an urban environment. The proposed solution is intended to be used as a mobile application on smartphones and has a potential to contribute to the smart city-related standardization activities of ITU-T Study Group 16.

S2.3	The knowledge graph as the interoperability foundation for an Augmented Reality application: The case at the Dutch Land Registry* *Alexandra Rowland and Erwin J.A. Folmer (University of Twente & Kadaster, The Netherlands); Tony Baving (Kadaster, The Netherlands)* The concept of the knowledge graph supports insight to a given context through the provision of standards-based mechanisms for accessing open and interoperable data. In doing so, the graph uses the power of the web to integrate data from distributed data sources and make this data available to end users in a transparent, flexible and application-independent manner, either by simply displaying data in the browser based on a dereference unique resource identifier or in an application built using the knowledge graph as the source. With the latter approach, the knowledge graph remains independent of the applications making use of it as a data source, where the connection between the graph and the application is achieved through interfaces which are completely based on open standards, most commonly through the use of a SPARQL endpoint. Indeed, chatbot applications often make use of the knowledge graph in this way but this paper aims to present the potential for Augmented Reality (AR) applications to be similarly built using knowledge graphs. By presenting this potential, this paper will argue that AR applications exemplify the potential opportunities that fully open, interoperable and standards-based approaches to data publication, such as the development of knowledge graphs, have and, therefore, will become key drivers within the organization in the investment of the further development of the concept of the knowledge graph in the future and improved accessibility of data for end users.

Session 3: Services in future networks	
S3.1	Research on asset administration shell standard system architecture *Quanbo Lu (China University of Geosciences, Beijing, China); Xinqi Shen (China Academy of Information and Communications Technology, China); Mei Li (China University of Geoscience, Beijing, China)* Asset Administration Shell (AAS) is an important enabling technology to implement digital twins for Industry 4.0(I4.0). It establishes cross-company interoperability. However, in the process of AAS application, there is a lack of AAS standard references. Aiming at the problem of lacking AAS standardization, this paper analyzes the requirement of AAS standards from three perspectives: AAS concept, the implementation of AAS key technologies, and AAS applications. As a basis for research to establish the three dimensions of the standardized architecture, AAS standard architecture is formed from the integration of architecture under different perspectives. This provides guidance for the study and formulation of standards related to AAS standardization.
S3.2	Research and standardization requirements for 5G network peak control technology in video transmission *Zhiji Deng (Zhejiang Provincial Key Laboratory of Harmonized Application of Vision & Transmission, China and Zhejiang Dahua Technology Co. Ltd, China); Zhewei Fu, Ming Liu and Xiangyu Qu (Zhejiang Dahua Technology Co. Ltd, China); Dong Ding (China Mobile Communications Corporation, China); Qi Ye, Weisheng Kong, Fei Wang, Jinyu Zhang, Hui Wan (Zhejiang Dahua Technology Co. Ltd, China); Jian Lou (China Mobile Communications Corporation, China)* The 5G network peak control technology in video transmission scenarios is used to solve the problem of violent fluctuations in network peaks caused by the collision of I-frames (key frames) during multichannel video transmission. Based on the research of key technologies such as I-frame collision detection and network peak shift scheduling, through the Multiaccess Edge Computing (MEC) external interface, the network data and video management system is opened to effectively smooth the network peak during multichannel video transmission, and realize the number and bandwidth of 5G terminal access under 5G base stations. The improved utilization rate is conducive to the large-scale promotion of 5G video surveillance scenarios.
S3.3	A comparative analysis of Augmented Reality frameworks aimed at diverse computing applications* *Mfundo A. Maneli and Omowunmi E. Isafiade (University of the Western Cape, South Africa)* Immersive systems such as Augmented Reality (AR) and Virtual Reality (VR) have proven useful in diverse computing domains. However, there is little effort on accuracy measurements within AR applications, which could greatly impact outcomes and decisions in certain domains, such as crime scene investigations, among others. This paper aims to analyze and evaluate two existing prominent AR frameworks, ARCore and ARKit, which support the development of diverse mobile computing applications for immersive systems. This research developed prototype applications and conducted comparison tests of measurement accuracy within the applications. The accuracy was tested using four distance criteria across six different devices, spanning ARCore and ARKit frameworks. A control experiment was used to benchmark the measurement accuracy. Relatively, an instance of the experiment presented ARCore as reliable. Overall, ARKit proved to be more accurate between the two frameworks, with an average accuracy of 99.36% as opposed to 89.42% scored by ARCore. The obtained results can give insight on the choice of framework to consider during AR application development for a specific domain, hence boosting quality of experience.

INDEX OF AUTHORS

Index of Authors

International
Telecommunication
Union
Place des Nations
CH-1211 Geneva 20
Switzerland

ISBN: 978-92-61-36151-8

9 789261 361518

Published in Switzerland
Geneva, 2022