Lecture Notes on Data Engineering
and Communications Technologies 205

Smain Femmam
Pascal Lorenz *Editors*

Recent Advances in Communication Networks and Embedded Systems

Proceedings of 6th International Conference on Communication and Network Technology

Springer

Lecture Notes on Data Engineering and Communications Technologies **205**

Series Editor

Fatos Xhafa, *Technical University of Catalonia, Barcelona, Spain*

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

Smain Femmam · Pascal Lorenz
Editors

Recent Advances in Communication Networks and Embedded Systems

Proceedings of 6th International Conference on Communication and Network Technology

Springer

Editors
Smain Femmam
UHA University France
Mulhouse, France

Pascal Lorenz
University of Haute Alsace
Mulhouse, France

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-3-031-59618-6 ISBN 978-3-031-59619-3 (eBook)
https://doi.org/10.1007/978-3-031-59619-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

The 2024 6th International Conference on Communication and Network Technology (ICCNT 2023) was held in Madrid, Spain, during September 18–20, 2023. As we embark on this intellectual journey, we are thrilled to bring together researchers, scholars, and industry professionals from around the globe to explore and exchange insights at the forefront of communication and network technology. In an era defined by rapid technological advancements, ICCNT has established itself as a pivotal platform for the exchange of cutting-edge research and innovative ideas. The conference serves as a nexus for interdisciplinary collaboration, fostering dialogue that transcends traditional boundaries and accelerates the pace of progress in the field. ICCNT 2023 aims to address the most pressing issues and anticipate the future landscape of our interconnected world. We extend our deepest appreciation to the organizing committee, keynote speakers, session chairs, and, most importantly, the contributors whose research forms the foundation of this conference. Your dedication to advancing the field is truly commendable. We also invite all participants to engage in the vibrant discussions, forge new collaborations, and contribute to the collective knowledge that will shape the trajectory of communication and network technology. Thank you for joining us at ICCNT 2023. Together, let us explore, innovate, and pave the way for a future connected by communication and network technology.

With my warmest regards,
ICCNT 2023 Organizing Committees

Committee

Conference Chair

Mahmoud Shafik	University of Derby, UK

Conference Co-chair

Maurici Ruiz	University of the Balearic Islands, Spain

Program Chairs

Alexey Vinel	Karlsruhe Institute of Technology (KIT), Germany
Rochdi Merzouki	University of Lille, France
Gang-Len Chang	University of Maryland, USA

Publicity Chairs

Philippe Martinet	École Centrale de Nantes, France
Mamoun Alazab	Charles Darwin University, Australia
Angel-Antonio San-Blas	Miguel Hernandez University of Elche, Spain

Technical Program Committees

Smain Femmam	Haute-Alsace University, France
Kwanho You	Sungkyunkwan University, Korea
Ljiljana Trajkovic	Simon Fraser University, Canada
Pascal Lorenz	University of Haute-Alsace, France
Dimitris Kanellopoulos	University of Patras, Greece
Apostolos Gkamas	University of Ioannina, Greece
Luca Davoli	University of Parma, Italy
Antonio De Nicola	ENEA CR Casaccia, Italy
Daniele Codetta Raiteri	Università del Piemonte Orientale, Italy
Xiaohui Zou	Sino-American Searle Research Center, China

Vivekananda Bhat K.	Manipal Institute of Technology, India
Xiaoxuan Wang	Beijing Jiaotong University, China
Jiajia Jiang	Tianjin University, China
Lounis Adouane	University of Technology of Compiègne, France
Gururaj H. L.	Vidyavardhaka College of Engineering, India
Gowtham M.	National Institute of Engineering, Mysuru, India
M. F. M. Firdhous	University of Moratuwa, Sri Lanka
Kasturi Vasudevan	IIT Kanpur, India
Shahzad Ashraf	Hohai University, China
Thien Wan Au	Universiti Teknologi Brunei, Brunei Darussalam
Napa Sae-Bae	SWU, Thailand
Qamar Nawaz	University of Agriculture, Pakistan
Stavros N. Shiaeles	University of Portsmouth, UK
Apostolos Xenakis	University of Thessaly, Greece
June Tay	Singapore University of Social Sciences, Singapore
Kunihiko Kaneko	Fukuyama University, Japan
Manijeh Keshtgari	University of Georgia, USA
Rinku Basak	American International University-Bangladesh, Bangladesh
Andreas Handojo	Petra Christian University, Indonesia
Huifang Chen	Zhejiang University, China
Hanafy Mahmoud Ali	Minia University, Egypt
Khondker Shajadul Hasan	University of Houston-Clear Lake, USA
Benjamin Aziz	University of Portsmouth, UK

Contents

Wireless Communication Network and Development Technology Based on Sensing

Data Distribution Method for the IoT System NAMI

Tadashi Ogino(✉)

Meisei University, Tokyo, Japan
tadashi.ogino@meisei-u.ac.jp

Abstract. With the spread of coronavirus disease (COVID-19), office workers and school students have transitioned to online platforms for work and education, respectively. During this period, the importance of face-to-face human interactions has been highlighted. In response, in a prior study, I proposed an advanced system based on the harmony between human capabilities and information technology (IT) (SHONAN), which restricts the functions implemented by IT to a minimum and limits the range of data distribution to nearby users, preventing the unnecessary spread of data over the Internet. This system was developed as a potentially new sustainable IT system. Additionally, I had implemented a short-range message-exchange system called narrow area communication system (NAMI) as a specific application of SHONAN to implement crowd detection alerts to lower the risk of COVID-19 infection. In the current study, I propose a method to provide services to users while reducing the amount of transferred data by placing the original data on an edge server and processing the data on the cloud in a COVID-19 infection warning system. The proposed method was implemented as a prototype, and its effectiveness was experimentally verified.

Keywords: SHONAN · NAMI · IoT · sustainable system · cloud · edge

1 Introduction

The coronavirus disease (COVID-19) pandemic has forced people to transition to a predominantly online working lifestyle. This was initially viewed as desirable for IT professionals, enabling them to work without commuting to their offices, and favorable for students, who could attend classes without going to school. However, this transition has not been as favorable as expected, which can be attributed to the lack of maturity of IT infrastructure and systems and the lack of awareness regarding the impact of this shift on the abilities of people. Despite online resources being valuable in gaining knowledge, there are many things that can only be learnt through in-person interactions. Additionally, it is not necessary to replace everything with IT, with many tasks being performed better by humans rather than IT-based systems.

Based on these experiences, I previously proposed a sustainable system based on harmony between human capabilities and IT (SHONAN), a novel concept that further extends the applicability of existing IT systems through the integration of IT and human

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 3–13, 2024.
https://doi.org/10.1007/978-3-031-59619-3_1

involvement [1]. I believe that SHONAN makes a great case for employing human-based solutions as opposed to utilizing complete IT-based solutions (Fig. 1). Additionally, using the proposed sustainable system, unnecessary developments and resource utilization in IT can be greatly reduced, lowering the need to inexhaustibly expand the scope of IT.

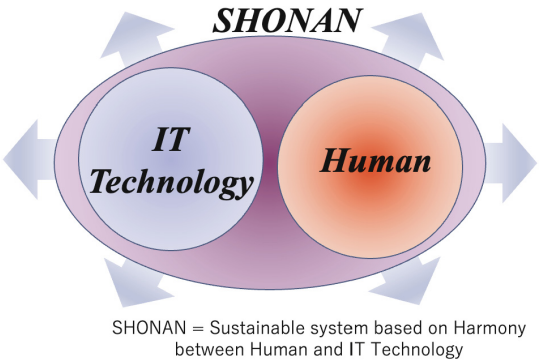


Fig. 1. Proposed sustainable system

In a previous study, I implemented a messaging system called neighborhood communication system (NAMI), which allows message exchanges in a narrow area, as a specific application of SHONAN. Additionally, a COVID-19 infection avoidance system was implemented with NAMI [2].

The COVID-19 infection avoidance system uses the short-range communication technology Bluetooth Low Energy (BLE) to estimate the distance and number of nearby users and warn nearby users when the risk of COVID-19 infection is high. Moreover, by recording warning information, it is possible to find a safer route with low infection risks (Fig. 2) [3].

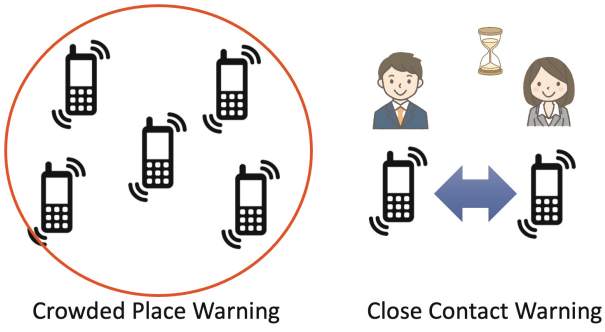


Fig. 2. Crowded place/close-contact warning system

Searching for routes that avoid infections is necessary not only for nearby users within the communication range of NAMI, but also for people who wish to commute

somewhere from a distance. Accordingly, all the data that can be handled by NAMI are provided not only to nearby users, raising the need for a mechanism that can provide useful data to a wider range of users when needed.

In this study, a mechanism to divide the data collected and processed by edges into edge and cloud servers on a network is proposed. The prototype of this mechanism was successfully implemented, verifying its effectiveness. Additionally, the proposed mechanism is more application-dependent compared to traditional methods in conventional research.

In the following section, the data distribution method for NAMI is examined. Then, Sect. 3 discusses data distribution in an IoT system, Sect. 4 discusses the implementation of the proposed method, and Sect. 5 presents the discussion of the results and the scope for future work. Finally, Sect. 6 concludes the paper.

2 Related Research

The increased proliferation and advanced functionalities of mobile terminals have enabled the collection of large amounts of data, which are collected and processed in a cloud via the high-speed Internet. The increased sophistication of applications has created a need for reducing the load on networks and shortening the delay time; this need can be met by placing an edge computer near a mobile terminal [4].

Placing the data sent from a mobile terminal on the edge and cloud and determining an appropriate location for data processing is a challenge. Regarding caching, it is a data storage process, and various cache arrangement locations and algorithms have been proposed. Regarding data processing, factors such as the location of processing and type of processing algorithm, along with the associated delay time and power consumption, have been studied [5–7].

There are many approaches being considered for supporting the provision of IT infrastructure and services to people worldwide. However, there are also human-based approaches that limit the scope of data distribution and processing to a small level and are potentially more favorable, especially considering the importance of human-based solutions during the COVID-19 pandemic.

3 Data Distribution in an Internet of Things (IoT) System

In an Internet of Things (IoT) system, the data distribution between clouds and edges depends on the system configuration and application characteristics. In this section, the proposed data distribution method that works based on the assumption of the short-distance message-exchange system NAMI is discussed.

3.1 Premise of NAMI

The following conditions are essential with respect to data distribution in NAMI.

- The system comprises devices that collect data, edge servers that store the collected data, and a cloud that stores the summary data processed from the data collected by the edges.

- Each device temporarily saves the collected log data itself.
- Each device exchanges messages with nearby devices through short-range wireless communication. The current implementation of the system uses BLE as the communication channel. The communication channel used for this system is a relatively low-speed and low-capacity communication channel that is unsuitable for sending large amounts of log data to edge servers.
- Edge servers store the data sent from the devices in their databases. Here, the information collected by the devices is stored without analysis or aggregation.
- In response to a request from the cloud, the edge server processes its own data into summary data and provides them to the cloud. The processing method is specified in the cloud and based on the application provided to users.
- Edge servers cannot always connect to the cloud. Additionally, because there are cases in which there is no global Internet Protocol (IP), accessing an edge server from the cloud may be infeasible.
- A wide range of summary data gathered from multiple edges are collected in the cloud. A general user can access the cloud via the Internet and request services from a corresponding application.

Based on these assumptions, the configuration for realizing the data distribution of the edges and cloud was examined. Figure 3 illustrates the steps involved in data distribution.

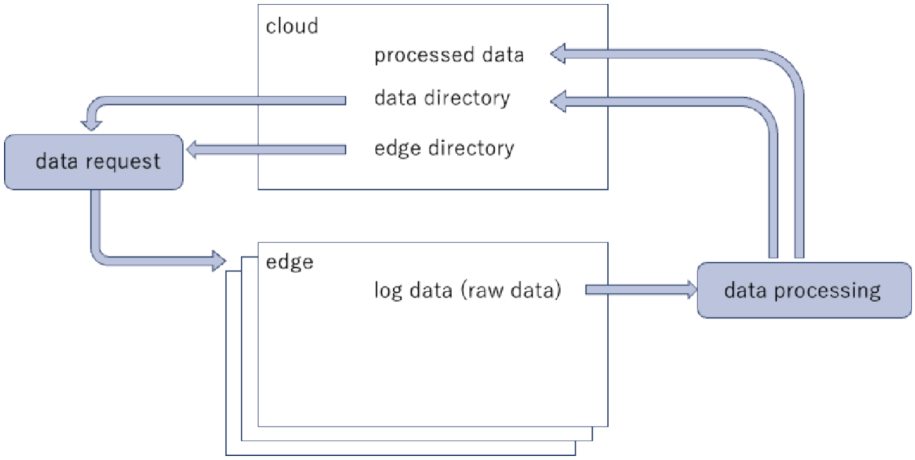


Fig. 3. Data distribution procedure

The edge servers store log data, while the cloud holds the processed data derived from the log data by the edge servers. Details of the processed data existing in the cloud (edge information and types of processing) are recorded in the data directory. Information related to the types of log data, their locations on the edges, and the means of accessing them, is stored in the edge directory. Upon a request from the user, the cloud directly checks the data to determine whether it contains the necessary data. In case it does not,

it checks the edge directory and requests the processed data from the edges that have the original raw data.

3.2 Processing Flow

The details of the processing flow are shown below in Fig. 4.

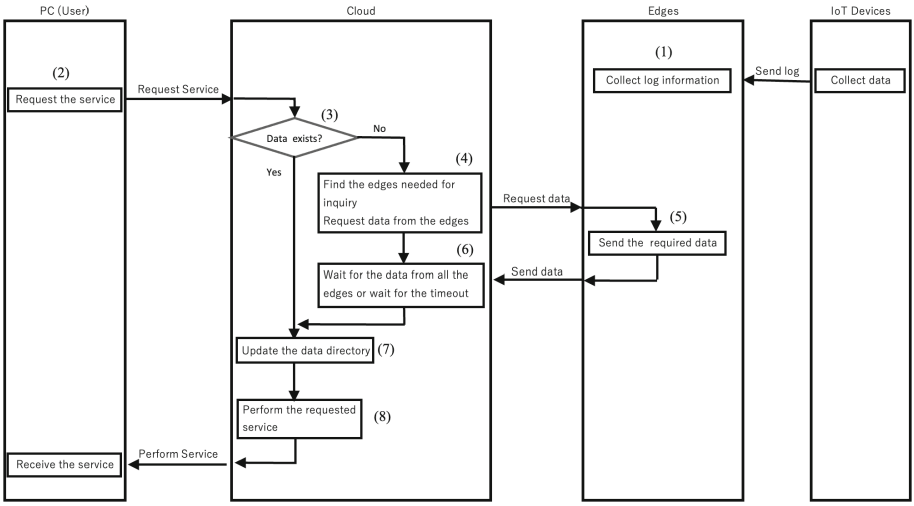


Fig. 4. Data processing flow

1. Raw data are collected by the devices and uploaded to a nearby edge (1).
2. Service requests from general users are sent to the cloud (2).
3. The cloud checks whether all the data required to provide the requested service exist in the cloud (3). If not, the cloud checks for the edge with missing data and requests that the data be sent to that edge (4).
4. When data are requested, the edge sends the processed data to the cloud (5). The edge has all the logs for which it holds responsibility; however, because the request from the cloud may be the processed result, it will perform some processing as necessary and send the result back to the cloud.
5. When the requested data are sent from all the edges (6), the cloud updates the data directory (7) and executes the service (8). However, because it is not guaranteed that the edge is always connected to the network, processing is interrupted at an appropriate point, even if data from some edges do not arrive. Finally, the cloud provides the results to the user with a range of available data.

4 System Implementation

The proposed data distribution method was implemented on a COVID-19 infection avoidance system.

4.1 Summary Information

The COVID-19 infection avoidance system collects the location information of both moving and nearby devices, examines the context surrounding each device, and provides alerts to the user when a high-risk situation arises. Information collected by the device, including warnings, is sent from the device to a nearby edge.

System users can easily display the locations and ratio of warnings on a Google map and search for routes with a lower risk of COVID-19 infection. Viewing all the collected data individually would be impractical; thus, users can access the summary statistics. The system configuration is illustrated in Fig. 5.

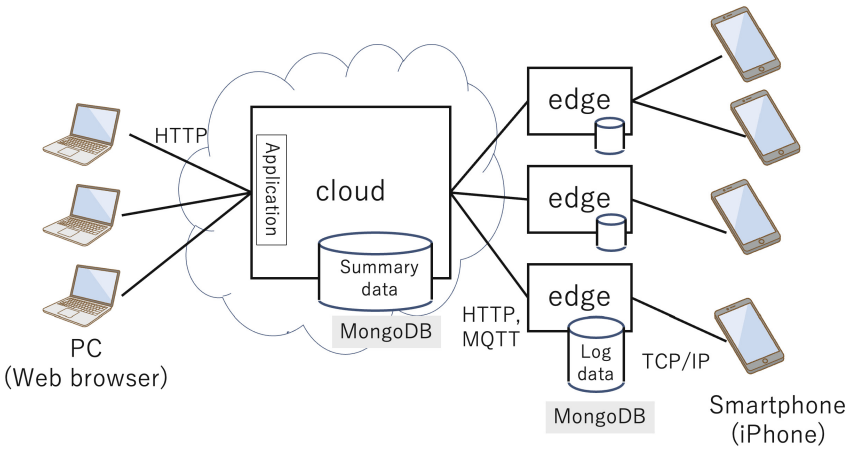


Fig. 5. System Configuration

As shown in the figure, the edges contained log data collected from the devices. When the cloud provided a certain function via an application, some of the log data were processed and used by the application. For example, the density value of each area was required when displaying the size of a crowd on a map; the cloud only required the data necessary to display this value [8].

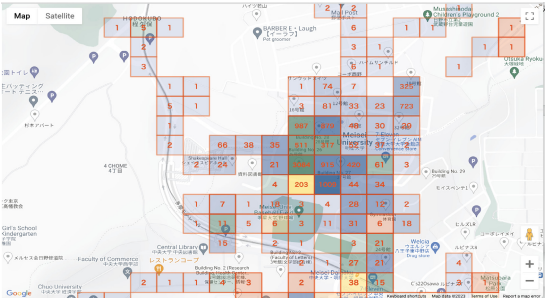


Fig. 6. Display of different degrees of warnings

Fig. 7. Display of safer routes

Two functions were implemented in this system; one displayed the ratio of the crowd on a map using color and darkness (Fig. 6) and the other displayed a route with low infection risk between two points on the map (Fig. 7). In the former case, the cloud required data regarding the number of collected data points and warning occurrences for each rectangular area (mesh). Since the route search function considered the risk in each area, it could be calculated using the same data.

Additionally, although the size of an area changed depending on the scale of the map, information on the smallest area (0.3 s for latitude, 0.45 s for longitude, about 10×10 m in the vicinity of Tokyo) enabled an even larger area to be displayed using the number of smallest areas.

4.2 Implementation Details

In the following section, details of the specific implementation method will be described.

Data Collection on Devices

The location information of the device (iPhone) was determined using the Global Positioning System (GPS). The statuses of the other devices were confirmed using BLE, which was also used for performing message exchanges.

The device was connected to a nearby edge server via WiFi using a private IP address. The collected log data were saved as textual data on the device and then sent to the edge server using a normal file-transfer function. The edge server stored the text–log data in the database using MongoDB.

Functions Provided to the Users

The functions provided by the cloud to the users included displaying the amount of collected data and the ratio of warnings on the map and a route with a low COVID-19 infection risk when specifying two points on the map. The map that was displayed to the users used Google Maps, and users could freely change the location and scale of this map.

The map was divided into rectangular areas that displayed the ratio of warnings, with the value of the warning information being displayed in color and darkness for each rectangular area. Data outside the map area were excluded. A low-infection risk route could be calculated from the same data used for display.

Communication
The cloud provided general web applications, and users could access these functions through a web browser. A request from the user would be sent to the cloud web server via HTTP.

The Message Queuing Telemetry Transport (MQTT) protocol was used when sending a request from the cloud to the edge servers. By becoming a MQTT publisher on the cloud side and a MQTT subscriber on the edge side, even an edge without a global IP could receive requests from the cloud through the Internet. The files were uploaded from the edges to the cloud using HTTP.

Data Processing
The data that the cloud required when providing this function included the number of collected data points and warnings for each rectangular area. The map was divided into rectangular areas according to the map range and scale specified by the user.

The scale could be modified by the user. Rather than separating the data for each scale, the data for the rectangle with the smallest scale were stored. The value for the larger scale was calculated by summing the numbers for the smallest rectangular areas.

The cloud requested these data from edge servers; the returned summary data were stored in a cloud. If this information was recorded in the data directory, the same information could not be requested.

Since log data reside on multiple edges, it is necessary to know the edges that contain the data. Each edge server registers the minimum and maximum values of the GPS data (i.e., latitude and longitude) of each edge in the edge directory. The area of the data for each edge could be recognized. The cloud only requested data from the edges that contained data for the extent of the display.

In this implementation, the items of each data dictionary are shown in Table 1.

Table 1. Contents of data and edge directories

Data Directory	latitude index, longitude index, data type, edge information
Edge Directory	edgeID, minimum/maximum latitude index, minimum/maximum longitude index
Processed Data	latitude index, longitude index, number of collected data, number of warnings, edge information, processing type, updated date, log existence flag

4.3 Evaluation

The use of the proposed method in the implemented system was examined. Four edges were provisionally set near the campus of the MEISEI University (Fig. 8). The collected data were arranged as data for each edge according to the corresponding area. Using this configuration, both the size of the database of each edge and the size of the database collected by the cloud (Fig. 9, Table 2) were measured.

Fig. 8. Area of the edges

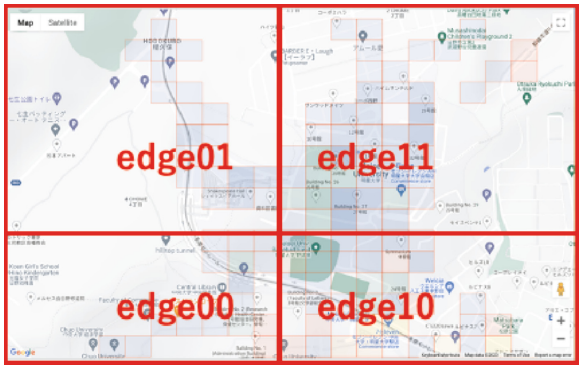


Fig. 9. Results shown by the displays

Table 2 shows that because the cloud only contained the collected data and issued warnings, the database size was smaller than the sum of the database sizes of the logs of each edge.

Table 2. Database sizes

edgeID	DB size (byte)
edge00	21,630
edge01	639,839
edge10	71,751
edge11	5,535,669
Total	**6,268,889**
Summary Data (cloud)	**806,963**

5 Discussion and Future Work

In this study, a method that could reduce the amount of transferred data and the data that a cloud should retain by storing the data collected by a device at the edge and providing only the results necessary for data processing to the cloud was developed and implemented on the COVID-19 infection avoidance system. Although the data processing was simple, it is necessary to consider extensions to the method that can enable it to handle more complicated processing. For example, methods to implement filtering by factors such as time and date on the cloud side are needed.

In addition to the COVID-19 infection detection system, a disaster information-sharing system that can support local areas during typhoons and floods is needed. Such a system works on the assumption that the details of such disasters will be shared through photos and videos, which are difficult to exchange through low-speed communication methods, such as BLE.

6 Conclusion

In this study, a data-division method between clouds and edges in the NAMI architecture was proposed. The method was implemented on the COVID-19 infection avoidance system, with a reduced amount of initial data transfer. In the future, I also plan to implement the method on other applications, such as an emergent information data sharing system that can provide support during natural disasters when high-speed Internet facilities are unavailable. I will also consider extending the range of information sharing by moving edge devices (or similar) with summary data even when the Internet is not available.

Acknowledgement. This study was partly supported by JSPS KAKENHI (grant number 21K12149).

References

1. Ogino, T.: Proposal of a sustainable system based on harmony between human and information technology. Int. J. Comput. Theory Eng. **14**(3), 135–140 (2021)
2. Ogino, T.: Crowded and close warning system for Covid-19 prevention on NAMI. In: Proceedings of 3rd International Conference on Applied Sciences, Engineering and Information Technology, Online (2022)
3. Ogino, T.: Safer route search to lower COVID-19 infection risk using NAMI. In: Proceedings of International Conference on Internet Technologies and Society, Online (2023)
4. Ren, J., Zhang, D., He, S., Zhang, Y., Li, T.: A survey on end-edge-cloud orchestrated network computing paradigms: transparent computing, mobile edge computing, fog computing, and cloudlet. ACM Comput. Surv. **52**(6), article no. 125 (2020)
5. Yang, X., Fei, Z., Zheng, J., Zhang, N., Anpalagan, A.: Joint multi-user computation offloading and data caching for hybrid mobile cloud/edge computing. IEEE Trans. Veh. Technol. **68**(11), 11018–11030 (2019)
6. Wu, Y.: Cloud-edge orchestration for the internet of things: architecture and AI-powered data processing. IEEE Internet Things J. **8**(16), 12792–12805 (2021)
7. Zhou, Z., Yu, S., Chen, W., Chen, X.: CE-IoT: cost-effective cloud-edge resource provisioning for heterogeneous IoT applications. IEEE Internet Things J. **7**(9), 8600–8614 (2020)
8. Ogino, T.: Data collection of crowded and close-contact warning system for Covid-19 prevention on NAMI. In: Proceedings of 3rd International Conference on Emerging Research in Engineering, Information Technology, Bioinformatics, Applied Sciences, Greece (2022)

Efficient Beam Selection for Increased Overall Wireless Network Capacity

Parmida Geranmayeh[1](✉), Ekaterina Sedunova[1], and Eckhard Grass[2]

[1] Humboldt University of Berlin, Berlin, Germany
{Parmida.Geranmayeh,ekaterina.sedunova}@hu-berlin.de
[2] Humboldt University of Berlin and IHP – Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder), Germany
grass@ihp-microelectronics.com

Abstract. Antenna beamforming is an increasingly utilized technology in various wireless systems, particularly in the realm of cellular telecommunications, such as 5G. By manipulating the spacing and phases of the antennas within an array, the direction and shape of the beam can be controlled. The principal goal of beamforming is to direct a wireless signal towards a specific receiving device instead of dispersing it in all directions from the transmitting antenna. This targeted approach enables the delivery of a higher quality signal to the intended receiver, resulting in faster and more reliable information transfer with reduced errors. For modelling a communication system, the beam pattern can be discretized into rays in angular and gain dimensions. Ray-tracing plays a crucial role in evaluating different beamforming techniques by accurately predicting signal strength, coverage, and quality at various locations within the environment. This analysis helps to select optimal beamforming parameters, including antenna placement, beam steering angles, and beamforming weights, to achieve the desired performance objectives. By effectively modeling the interaction between signals and objects/surfaces, ray-tracing aids in optimizing beamforming algorithms, improving signal quality, and maximizing overall system performance in wireless communication systems. The primary aim of this research is to determine, from a codebook, the ideal beam for each transmitter and receiver that maximizes the total channel capacity. However, due to the complexity and large number of angles, finding the best angles becomes challenging, time-consuming, and nearly impossible. Therefore, this study investigates various optimization methods to efficiently discover the optimal angles within a shorter timeframe. The accuracy and speed of different optimization techniques for identifying the best angles for achieving maximum total channel capacity are compared.

Keywords: Beamforming · Network capacity · Optimization · Ray-tracing

1 Introduction

By utilizing digital twins, wireless network design and deployment can be optimized through detailed analysis of the physical environment, propagation characteristics, and obstacles [1]. Beamforming, a technique that enhances wireless communication performance, can be incorporated into the digital twin framework to simulate and optimize

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 14–27, 2024.
https://doi.org/10.1007/978-3-031-59619-3_2

beamforming capabilities [2–5]. Ray-tracing algorithms play a crucial role in accurately modeling signal propagation characteristics, allowing for precise analysis of signal strength, coverage, and interference patterns [6–8]. Integrating beamforming and ray-tracing techniques into the digital twin framework provides a comprehensive approach to analyze and optimize network performance, leading to efficient wireless communication systems [3, 8]. Additionally, the selection of the appropriate antenna type is crucial, as it directly impacts system performance in terms of coverage, range, and capacity [9, 10]. Considering factors such as radiation pattern, gain, frequency compatibility, and polarization ensures optimal wireless communication system design [10].

We essentially decompose the beam pattern of beams from a codebook into rays and use the rays in a simulation model.

This article explores various optimization methods for transmitters and receivers in a specific room with involve the utilization of beamforming techniques, which employ a set of antenna angle patterns to direct the transmitted rays while minimizing interference. The aim of our examination is seeking to determine the optimal send-receive angle for improved performance and ultimately to enhance the total channel capacity and mitigate interference.

In Sect. 1.1, we provide a comprehensive review of the existing research and studies conducted in the field of beamforming and ray-tracing in wireless communication systems. Moving on to Sect. 2, we present detailed information about the experimental setup, including the room specifications and parameters, and elaborate on the methodology employed for generating the antenna patterns. The subsequent section, Sect. 2.1, focuses on the exploration of various optimization models, wherein we introduce and discuss each model in detail. To validate and compare these models, Sect. 3 presents the simulation results obtained from our analysis. In Sect. 4, we provide meaningful conclusions and insights based on our findings. We also outline potential avenues for future work in the field, emphasizing the areas that require further investigation and development.

1.1 State of the Art

In this section, we provide an overview of the state-of-the-art developments in digital twin technology and its role in conjunction with beamforming and ray-tracing in the context of 5G wireless communication.

The incorporation of digital twin technology in 5G networks, particularly in the industrial Internet of Things (IIoT) and mobile edge computing (MEC) environments, has been explored in multiple articles [11–14]. These articles emphasize the potential of digital twins in enabling data-driven modeling, real-time simulation, and a deeper understanding of physical objects. The proposed approaches leverage digital twin technology to optimize task offloading, reduce latency, improve edge server selection, and enhance system performance in industrial IoT networks. By considering constraints and computation resources, these approaches aim to minimize latency and enhance overall efficiency in digital twin-empowered networks, offering new possibilities for optimizing 5G internet networks in IoT and industrial settings. In [15], an adaptive beamforming algorithm is proposed for MIMO-OFDMA systems to address co-channel interference (CCI) and improve performance in wireless channels. The algorithm tracks the direction

of arrival (DOA) of each signal, nullifies interference directions, and preserves space-time code (STC) diversity. Simulation results demonstrate improved bit error rate (BER) performance in multipath fading channels with CCI. Another study by Maheshwari et al. focus on a flexible beamforming (FBF) antenna architecture for 5G networks, achieving advancements in data rates, coverage, scalability, and energy efficiency [16]. The FBF architecture shows superior scalability and 53% higher energy efficiency compared to maximum power beams. Wu et al. introduce coherent beamforming (CB) technology that is explored for Internet of Vehicles (IoV) and 5G networks, leading to the development of the Iterative Coherent Beamforming Node Design (ICBND) algorithm [17]. This algorithm optimizes task transfer between vehicles and roadside nodes, reducing communication network infrastructure costs. Sun et al. propose distributed collaborative beamforming (DCB) with a virtual node antenna array (VNAA) to enhance mobile wireless sensor networks (MWSNs) [18]. The objective is to optimize sidelobe level, transmission power, and motion energy consumption of DCB nodes, reducing communication infrastructure costs. Another work presents a novel SDN-based virtual Fog-RAN design for IIoT [19]. It jointly considers radio resource allocation and transmit beamforming to improve resource utilization and user satisfaction by minimizing power consumption and maximizing sum-rate. Lastly, beamforming optimization for intelligent reflecting surface-aided SWIPT IoT networks is explored [20]. The study focuses on optimizing passive IRS reflection coefficients and active base station beamforming to maximize signal-to-interference-plus-noise ratio and total harvested energy in MIMO SWIPT systems. Therefore, beamforming has a significant impact on network performance and efficiency. By intelligently directing beams towards desired users and nullifying interference directions, beamforming techniques enhance signal quality, increase coverage, and mitigate co-channel interference. This results in improved data rates, reduced latency, and enhanced overall network capacity. With the advent of technologies like 5G and beyond, beamforming plays a crucial role in enabling advanced applications and services, such as IoT, MIMO systems, edge computing, and industrial networks. The continued advancements in beamforming algorithms and hardware capabilities promise even greater gains in network performance, making it a key technology for the future of wireless communication networks.

Ji et al. [21] propose an efficient and accurate UHF band propagation prediction approach for indoor environments. Their model optimizes ray-tracing by reorganizing objects into irregular cells and incorporates reflection, refraction, and diffraction, leading to enhanced accuracy. Hsiao et al. in [22] focus on ray-tracing simulations for millimeter-wave propagation in 5G wireless communications, considering line-of-sight and non-line-of-sight rays. Ray-tracing for Radio Propagation Modeling [23] offers a comprehensive review of ray-tracing algorithms and radio propagation modeling, covering practical advancements and future perspectives. Tiberi et al. [24] focus on characterizing the indoor propagation channel in ultra-wideband communication system design through ray-tracing simulations at different frequencies. These articles provide valuable insights for optimizing wireless communication system design and planning across different scenarios, improving accuracy while conserving computational resources.

Overall, the combination of digital twin technology, beamforming, and ray-tracing in 5G wireless communication networks holds great promise for optimizing network performance, enhancing coverage, and improving the overall user experience. Continued research and development in these areas will further propel the advancements in wireless communication systems and enable the realization of the full potential of 5G and beyond to drive the next generation of intelligent and optimized wireless communication systems.

2 Design Parameters

In this section, we explain the basic implementation and an exhaustive search by iterating through 390,625 combinations. This number of iterations is obtained by considering a range of -60 to $60°$ with a step size of $5°$, resulting in 25 values for each of the four antennas. Therefore, the total number of combinations is $25 \times 25 \times 25 \times 25$. For each combination, the algorithm calculates the total channel capacity and stores the values. The objective is to find the maximum total channel capacity among all the combinations.

As an input for the map, we utilized the stereolithography (STL) file named "*office.stl*" for visualization. The STL file represents a 3D model of a room. By passing the file name to the "*siteviewer*" function with the "*SceneModel*" option, the viewer object is created, which allows to view and interact with the 3D scene. The dimensions of the room, as shown in Fig. 1-a, are given as 8 m $\times$ 5 m $\times$ 2.75 m. The carrier frequency is set to 60 GHz, and the wavelength is calculated based on the speed of light. The bandwidth has been fixed at 2 GHz. As shown in Fig. 1-b, the system uses 4 antennas and considers a range of angles from -60 to $60°$ for calculations.

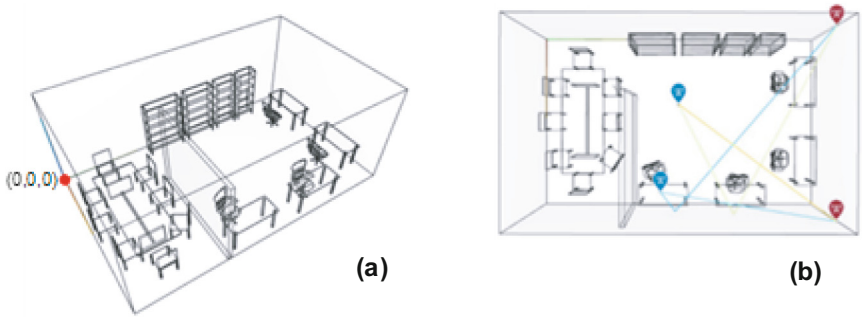


Fig. 1. **(a)** The room layout model. **(b)** The antennas placement in the model room.

This model involves creating a Uniform Rectangular Array (URA) using a rectangular patch antenna element. The azimuth and elevation angles are defined as -180 to $180°$ and -90 to $90°$, respectively. The URA is configured using the phased. URA function, with the specified antenna element. It has 4 rows and 8 columns, with half-wavelength spacing between elements. Coordinates for the receiver (Rx) and transmitter (Tx) positions are defined as (Rx_x, Rx_y, Rx_z) and (Tx_x, Tx_y, Tx_z), respectively. The

coordinates of the receiver (Rx) and transmitter (Tx) positions are respectively defined as follows:

$$\text{Tx1: } (0.02, 8, 2) \quad \text{Tx2 : } (5, 8, 2) \quad \text{Rx1 : } (4.5, 3.5, 0.85) \quad \text{Rx2 : } (2, 4, 0.85)$$

A propagation model is then defined using the propagation model function. The model is set to "ray-tracing" and utilizes a Cartesian coordinate system. Ray-tracing is performed using the shooting and bouncing ray method (SBR) with low angular separation. The maximum number of reflections is limited to 1, and the surface material is specified as "*metal*". The aim is the evaluation of the channel capacity of a wireless communication system for different combinations of transmitter (Tx) and receiver (Rx) angles. The main part of implementation begins by initializing counters on 390,625. It then enters a nested loop structure to iterate over various transmitter and receiver angles.

It's important to note that the setup we're discussing here doesn't involve any unique or specific room conditions. Instead, it's based on a standard layout that's readily available within the MATLAB environment. We've opted for a general room setup and layout with a generalized configuration. Additionally, the choice of antenna locations isn't tied to any particular scenario. As a result, we anticipate that the conclusions drawn won't be affected even if we were to alter the rooms and locations of the antennas. This underscores the robustness of our findings across different setups.

Inside the nested loop, the implementation calculates the antenna patterns for the current transmitter and receiver angles using the beam steer function from the Phased Array System Toolbox in MATLAB. This function calculates the radiation pattern of the antenna element or array based on the given frequency, propagation speed, and weights.

The process of generating the antenna patterns involves the following steps:

1. A steering vector is created using the phased. SteeringVector function. This vector represents the spatial response of the antenna array and takes the antenna array and propagation speed as input.
2. The desired scan angles for the main lobe beam are defined. In the code, the azimuth angle is set as the loop variable in the beam function. These scan angles determine the direction in which the antenna array focuses its radiation.
3. Using the steering vector and the specified scan angles, the weights for beamforming are computed. These weights determine the amplitude and phase of the signals fed to each antenna element in the array, enabling control over the direction of the main lobe beam.
4. The pattern function is called to compute the radiation pattern of the antenna array. This function takes several parameters as input, including the antenna array, frequency of operation, propagation speed, and weights. It calculates the total electric field pattern in decibels and provides the azimuth and elevation angles over which the pattern is computed.

By manipulating the scan angles and applying corresponding weights to the antenna array, various radiation patterns can be generated, representing the antenna's directionality and power distribution in 3D space. It's important to note that the specific patterns generated depend on the design and characteristics of the antenna elements used in the code. In this implementation, the "*patchMicrostrip*" and "*phased.CosineAntennaElement*"

classes are utilized for designing the antenna elements. The "*patchMicrostrip*" class represents a micro-strip patch antenna element, which is known for its compact size and ease of fabrication. The "*design*" function configures the antenna element with properties such as substrate material, patch dimensions, feed point, and ground plane, and it is tailored to the specific operating frequency.

On the other hand, the "*phased.CosineAntennaElement*" class represents a cosine-shaped antenna element. These elements possess a radiation pattern with a wider beam width compared to other designs. The constructor of "*phased.CosineAntennaElement*" takes parameters like the frequency range ("*FrequencyRange*") to configure the element. As shown in Fig. 2, by adjusting these parameters, the radiation pattern and characteristics of the antenna element can be customized. Both the "*patchMicrostrip*" and "*phased.CosineAntennaElement*" classes can be employed in array configurations to create desired radiation patterns and perform tasks such as beam steering and beamforming in phased array systems.

After generating the pattern, custom antenna elements ("*antennaTx1*", "*antennaTx2*", "*antennaRx1*"and "*antennaRx2*") are created using the calculated patterns and specified azimuth and elevation angles. Then we created transmitter ("*tx*") and receiver ("*rx*") objects using the custom antenna elements and specified positions.

The "*ray-trace*" function is used to calculate the ray-tracing-based channel characteristics between the transmitter and receiver. The resulting rays are stored in the "*raysPerfect*" variable. After that, it enters a nested loop to process each ray in the "*raysPerfect*" variable. For each transmitter and receiver combination, the channel parameters such as path loss, path phase, angle of departure (AoD), and angle of arrival (AoA) are extracted. The global spherical coordinates of AoD and AoA are converted to local spherical coordinates using transformation matrices. Then, the antenna patterns for each element in the transmitter and receiver arrays are calculated using the "*pattern*" function. Path gains are calculated using antenna patterns and path features, then saved with the corresponding path delays. The loop continues to process all the rays, and at the end, the total power received by each device and the channel capacity for each transmitter-receiver pair are calculated. The resulting channel capacity values are stored in the "*ChannelCapacityTx1Rx1*", "*ChannelCapacityTx2Rx2*", and "*Total_Channel_Capacity*" arrays.

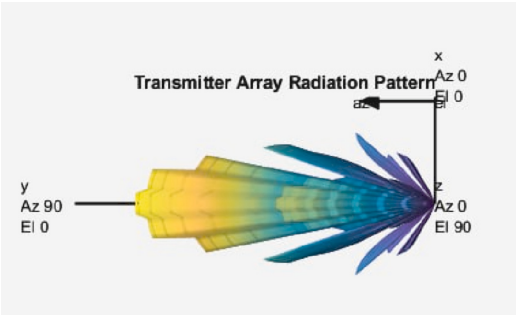


Fig. 2. The digitized antenna radiation pattern.

Finally, the loop variables and counters are updated, and the loop continues until all the desired combinations of transmitter and receiver angles have been evaluated. The best combination of angles refers to the combination that yields the maximal total channel capacity. This combination represents the set of angles that results in the maximum achievable capacity across all antennas.

2.1 Introduction of Methods

After executing the initial code, which required a computation time of 3 days, we obtained the optimal solution. Our model yielded two global maximum answers, with a total channel capacity of 2.9E+10 bps. The optimal angle for Tx1 was found to be −40°. For Tx2, the optimal angles were 60 and −60°, while Rx1 exhibited the best performance at an angle of 45°, and Rx2 at an angle of 55°. To further improve the runtime in this model, we proceeded to implement and evaluate seven optimization models. These models aimed to identify the best angles that maximize the total channel capacity while also assessing their accuracy and computational efficiency. By examining these models, we sought to refine our understanding of the system and identify the most effective optimization approach in terms of both performance and runtime.

Subsequently, we will present an overview of the following methods:

- Gradient Descent

  We utilized Gradient Descent, a widely recognized iterative optimization algorithm, to optimize our system parameters and maximize the channel capacity. By leveraging the power of Gradient Descent, we iteratively refined our parameters, achieving significant performance improvements without exhaustive search or complex computations. The algorithm's effectiveness is described in the book 'Algorithm for Optimization' [25].
- Simulated Annealing

  The Simulated Annealing algorithm was utilized as a practical optimization technique for finding near-optimal solutions. Simulated annealing is a well-known metaheuristic algorithm that combines random exploration and probabilistic acceptance, making it particularly effective for tackling complex optimization problems [26]. The algorithm starts with an initial solution and gradually reduces the acceptance of worse solutions by decreasing the temperature. We used a predefined schedule ($T = T_0 / (1 + \alpha * log(1+k))$) where T, represents the temperature, T_0 is the initial temperature, α denotes the decay rate, and k signifies the current iteration. For our study, we set the decay rate (alpha) to 0.995 and the initial temperature (T0) to 100. The choice of these parameters affects the cooling rate and the transition from exploring a wide range of solutions to refining the current solution.
- Tabu Search

  Tabu Search, an effective combinatorial optimization algorithm, was utilized in our research to optimize system parameters. It intelligently explores the solution space, avoids repetitive moves, and prevents revisiting previously explored solutions using a limited-size tabu list. By leveraging tabu search, we achieved notable performance enhancements and effectively navigated the search space, demonstrating its value in optimizing our system [26].

- Particle Swarm Optimization (PSO)

  Particle Swarm Optimization (PSO) is a heuristic search algorithm used to find optimal solutions in a search space. It maintains a population of particles that explore the space to identify the best solution. In our work, we apply PSO to optimize wireless communication system angles, aiming to maximize channel capacity. This approach efficiently optimizes system parameters, leading to significant performance improvements [25].
- Genetic Algorithm

  We utilized the Optimized Genetic Algorithm [25] as a practical solution. The genetic algorithm is a highly effective search and optimization technique that addresses complex problems where traditional algorithms fall short. By harnessing the power of the Optimized Genetic Algorithm, we successfully identified the best individual and its fitness, indicating the optimal or near-optimal solution achieved through the genetic algorithm. Notably, we employed a mutation rate of 0.05 and set the population size to 50, ensuring diversity and exploration within the algorithm.
- Hill Climbing

  In our study, we employed the Hill Climbing algorithm [27, 28] as a practical optimization technique. Known for its simplicity and effectiveness in finding local optima, the Hill Climbing algorithm allowed us to iteratively refine our system parameters and achieve improved performance. By leveraging the advantages of the Hill Climbing algorithm, we were able to fine-tune our system with minimal computational complexity and time requirements. By iteratively exploring neighboring solutions and selecting those that improve the objective, the algorithm aims to converge toward an optimal solution.
- Shotgun Hill Climbing

  Shotgun hill climbing is an optimization algorithm that mitigates the issue of getting trapped in local optima. Unlike traditional hill climbing, it simultaneously evaluates multiple neighbors and selects the best solution, utilizing randomness to explore a larger search space. In our project, we employed shotgun hill climbing to optimize a channel capacity problem [29]. With 5 random restarts, we initialized variables to store the best angles and capacity. The algorithm performed hill climbing for each restart, updating the best solution based on improved capacity. By combining shotgun hill climbing with random restarts, we successfully identified the optimal solution.

3 Simulation Results and Comparison of Efficiency

Figure 3 presents the average total channel capacity achieved in 100 iterations after 10 runs by each algorithm. The genetic algorithm stands out as the top-performing method, as it closely approaches the maximum total channel capacity with a value of 2.86E+10 bits per second (bps). This indicates that the genetic algorithm has demonstrated a high capability in optimizing the system and achieving a near-optimal channel capacity. The particle swarm optimization (PSO) algorithm ranks as the second-best method, with a total channel capacity of 2.63E+10 bps. While it falls slightly behind the genetic algorithm, it still exhibits strong performance in maximizing the channel capacity. Following

PSO, the order of preference for the remaining algorithms, in terms of channel capacity, is as follows: Hill climbing, Shotgun hill climbing, Tabu Search, Gradient descent, and Simulated annealing. These algorithms have shown relatively lower performance compared to the genetic algorithm and PSO, but they still provide viable optimization solutions for improving the channel capacity.

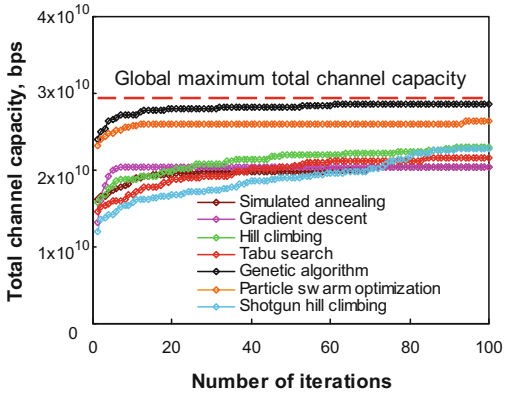


Fig. 3. Average channel capacity across 100 iterations (after 10 runs).

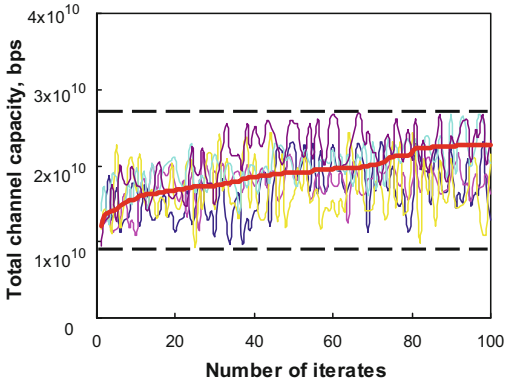


Fig. 4. Shotgun algorithm outputs and average curve across iterations.

Figure 4, represents the results of the shotgun hill climbing algorithm. In this figure, each of the 5 shots performed by the algorithm is depicted. Each shot consists of executing the code for 100 iterations and recording the output at each iteration. For each shot, the figure illustrates a sequence of 100 outputs obtained by the algorithm and we included a curve representing the average of these 5 shots as the main output of the shotgun hill climbing algorithm with a red curve. In Fig. 4, we present the main output of the shotgun hill climbing algorithm. This output is represented by a curve that shows the average performance across the 5 shots.

Figure 5 provides a comprehensive comparison of different optimization methods in terms of the average channel capacity achieved after 10 runs. Additionally, the figure

includes information on the corresponding error values and accuracy averages of each method. The left axis of the figure represents the channel capacity, measured in bits per second (bps). The right axis displays the accuracy of the optimization methods, represented as a percentage. Among the analyzed methods, the genetic algorithm demonstrates superior performance in both channel capacity and accuracy. It achieves the highest accuracy of 98% and delivers excellent channel capacity results. The genetic algorithm stands out as the best-performing method in terms of maximizing the channel capacity while maintaining a high level of accuracy. Following the genetic algorithm, the particle swarm optimization method obtains the second-highest accuracy of 90%. It demonstrates commendable performance in achieving channel capacity but falls slightly behind the genetic algorithm. The remaining methods are ranked in terms of accuracy, with shotgun hill climbing at 85%, hill climbing at 79%, tabu search at 74%, gradient descent at 70.4%, and simulated annealing at 70%. These methods provide varying levels of accuracy but still offer optimal solutions to improve the channel capacity.

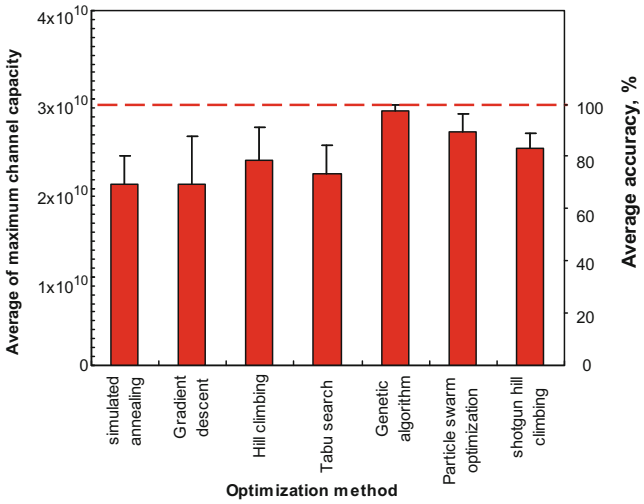


Fig. 5. Comparison of Optimization methods: Average of maximum channel capacity and average of accuracy.

Therefore, this figure provides valuable insights into the comparative performance of different methods, aiding in the selection of the most effective optimization approach for maximizing channel capacity.

Figure 6 provides insights into the ratio of the optimization model's computational time to the corresponding error values observed during the 10 runs of each algorithm. This figure highlights the trade-off of computational time and time error for each algorithm. Considering that the execution time of the algorithms varies significantly, it is beneficial to discuss their time requirements collectively. Among the algorithms, the genetic algorithm consumes the most significant amount of time, which is expected given its high accuracy. The execution time for the genetic algorithm is approximately 4302 s. Following the genetic algorithm, the Particle Swarm optimization method has

the second-longest execution time, approximately 1731 s. This method also exhibits favorable accuracy, making it a reasonable choice in terms of the trade-off between time and accuracy. The gradient descent algorithm ranks next in terms of execution time, followed by the shotgun hill climbing, tabu search, simulated annealing, and hill climbing algorithms. It is important to note that the duration of hill climbing is the shortest, taking approximately 86 s. The genetic algorithm takes a long time due to its iterative nature, which involves repeated evaluations, selection, and evolution of individuals within a population, leading to a more thorough search of the solution space at the expense of increased computational time.

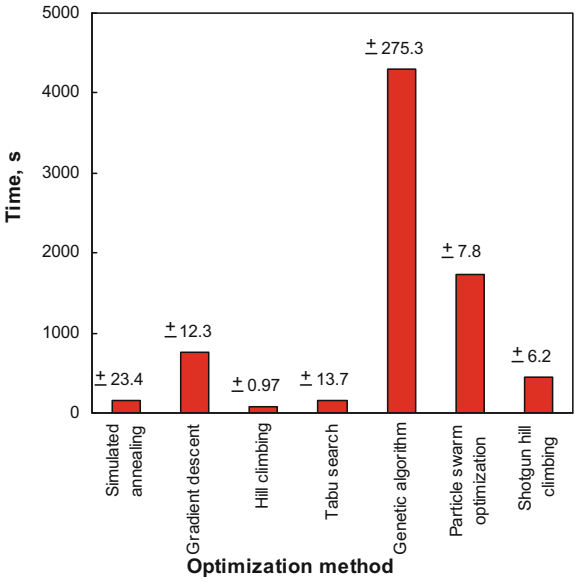


Fig. 6. Optimization model-time ratio.

According to Eq. 1, Fig. 7 incorporates an effect size of 0.7 for accuracy and 0.3 for execution time. This approach enabled us to assess the overall impact of accuracy and execution time on the observed outcomes. According to Fig. 7, it can be observed that the Shotgun Hill Climbing, Hill Climb, and Gradient Descent algorithms exhibit superior performance in terms of both time efficiency and accuracy. These algorithms require less computational time to achieve accurate results, making them favorable choices in terms of optimization. Particle Swarm optimization and Tabu Search, Simulated Annealing and finally the Genetic Algorithm, occupy the subsequent positions.

$$\begin{aligned} \textit{Weighted efficiency} = &(\textit{Normalized Accuracy of Algorithms} * \mathbf{0.7}) \\ &- (\textit{Normalized Execution Time of Algorithms} * \mathbf{0.3}) \end{aligned} \quad (1)$$

Selecting the effect size is contingent on the preference. In our case, prioritizing accuracy instead of time led us to assign a larger coefficient to it.

Based on the results we have obtained, we gain valuable insights into the effectiveness of different algorithms and optimal antenna angles when designing a real-world room with general conditions. The digital twin concept plays a pivotal role in translating this knowledge into practical applications, ensuring its reliability. Consequently, the findings presented in this article can be seamlessly integrated into a digital twin framework to generate alternative setups in the physical world.

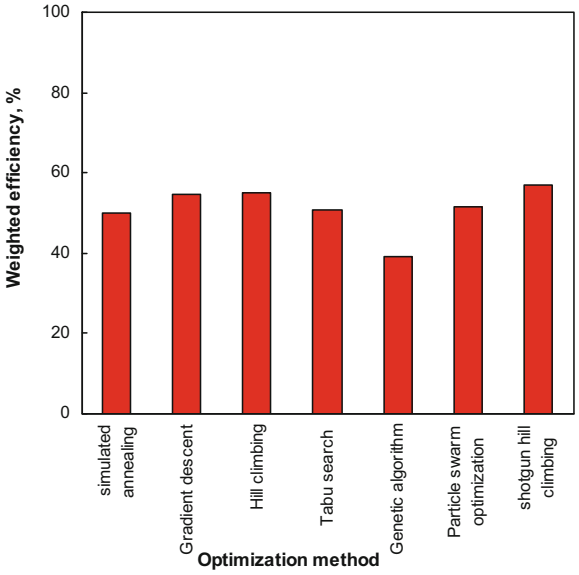


Fig. 7. Comparison of optimization methods: accuracy-to-time ratio.

4 Conclusion and Future Work

We investigated the importance of antenna beamforming in wireless systems, particularly in 5G cellular telecommunications, as it enables directed signal transmission for improved quality and reliability. Through the utilization of ray-tracing, which plays a crucial role in evaluating beamforming techniques and optimizing their parameters, we aimed to determine the optimal angles for transmitting and receiving signals to achieve maximum total channel capacity. This research focused on employing different optimization methods to optimize the angles. In order of preference, the genetic algorithm, particle swarm optimization (PSO), and shotgun hill climbing algorithms exhibited promising performance in terms of channel capacity, accuracy, and computational efficiency. These findings underscore the significance of optimization algorithms in achieving optimal outcomes in wireless communication systems. In the realm of future work, several promising directions can be pursued to further advance the field of beamforming and ray-tracing in wireless communication systems, including enhanced optimization techniques using machine learning and game theory principles, real-world implementation for validation

and necessary modifications, integration with emerging technologies such as IoT and 5G/6G networks, and investigation of additional performance metrics and trade-offs. By addressing these aspects, a deeper understanding of beamforming and ray-tracing can be achieved, leading to advanced and optimized systems in the future.

References

1. Rosen, R., Von Wichert, G., Lo, G., Bettenhausen, K.D.: About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine (2015). https://doi.org/10.1016/j.ifacol.2015.06.141
2. He, W., Zhang, C., Deng, J., Zheng, Q., Huang, Y., You, X.: Conditional generative adversarial network aided digital twin network modeling for massive MIMO optimization. IEEE Wirel. Commun. Netw. Conf. (WCNC) (2023). https://doi.org/10.1109/WCNC55385.2023.10118756
3. Jalali, J., Khalili, A., Rezaei, A., Famaey, J., Saad, W.: Power-efficient antenna switching and beamforming design for multi-user SWIPT with non-linear energy harvesting. In: IEEE 20th Consumer Communications & Networking Conference (CCNC) (2023). https://doi.org/10.1109/CCNC51644.2023.10059879
4. Charis, G., Showme, N.: Beamforming in wireless communication standard: a survey. Indian J. Sci. Technol. **10**(5), 1–5 (2017)
5. Salehi, B., et al.: Multiverse at the edge: interacting real world and digital twins for wireless beamforming.https://doi.org/10.48550/arXiv.2305.10350
6. Ding, C., Wang, I., Ho, H.: Digital-twin-enabled city-model-aware deep learning for dynamic channel estimation in urban vehicular environments. IEEE Trans. Green Commun. Netw. **6**(3), 1604–1612 (2022)
7. Geok, T.K., et al.: A comprehensive review of efficient ray-tracing techniques for wireless communication. Int. J. Commun. Antenna Propag. (IRECAP) (2018). https://doi.org/10.15866/irecap.v8i2.13797
8. Sheen, B., Yang, J., Feng, X., Chowdhury, M.: A digital twin for reconfigurable intelligent surface assisted wireless communication.https://doi.org/10.48550/arXiv.2009.00454
9. Ojaroudi-Parchin, N., et al.: Recent developments of reconfigurable antennas for current and future wireless communication systems (2019). https://doi.org/10.3390/electronics8020128
10. Mikki, S.: The antenna spacetime system theory of wireless communications. Proc. R. Soc. A (2018). https://doi.org/10.1098/rspa.2018.0699
11. Jagannath, J., Ramezanpour, K., Jagannath. A.: Digital twin virtualization with machine learning for IoT and beyond 5G networks: research directions for security and optimal control. In: ACM Workshop on Wireless Security and Machine Learning (WiseML) (2022). https://doi.org/10.48550/arXiv.2204.01950
12. Huynh, D.V., Nguyen, V.D., Sharma, V., Dobre, O.A., Duong, T.Q.: Digital twin empowered ultra-reliable and low-latency communications-based edge networks in industrial IoT environment. In: IEEE International Conference on Communications, ICC 2022 (2022). https://doi.org/10.1109/ICC45855.2022.9838860
13. Liu, T., Tang, L., Wang, W., Chen, Q., Zeng, X.: Digital-twin-assisted task offloading based on edge collaboration in the digital twin edge network. IEEE Internet Things J. **9**(2), 1427–1444 (2022)
14. Do-Duy, T., Huynh, D.V., Dobre, O.A., Canberk, B., Duong, T.Q.: Digital twin-aided intelligent offloading with edge selection in mobile edge computing. IEEE Wirel. Commun. Lett. **11**(4), 806–810 (2022)

15. Kim, C.K.: Performance improvement of MIMO-OFDMA system with beamformer. Int. J. Internet Broadcast. Commun. **11**(1), 60–68 (2019). https://doi.org/10.7236/IJIBC.2019.11.1.60
16. Maheshwari, M.K., Agiwal, M., Saxena, N.: Flexible beamforming in 5G wireless for internet of things. IETE Tech. Rev. **36**(1), 3–16 (2019)
17. Wu, L., Xu, J., Shi, L., Shi, Y., Zhou, W.: Optimize the communication cost of 5G internet of vehicles through coherent beamforming technology. Wirel. Commun. Mob. Comput. **2021**, 1–12 (2021)
18. Sun, G., Zhao, X., Shen, G., Liu, Y., Wang, A., Jayaprakasam, S.: Improving performance of distributed collaborative beamforming in mobile wireless sensor networks: a multiobjective optimization method. IEEE Internet Things J. **7**(8), 6787–6801 (2020)
19. Rahimi, P., Chrysostomou, C., Pervaiz, H., Vassiliou, V., Ni, Q.: Joint radio resource allocation and beamforming optimization for industrial internet of things in software-defined networking-based virtual fog-radio access network 5G-and-beyond wireless environments. IEEE Trans. Industr. Inf. **18**(6), 4198–4209 (2022)
20. Gong, S., Yang, Z., Xing, C., An, J., Hanzo, L.: Beamforming optimization for intelligent reflecting surface-aided SWIPT IoT networks relying on discrete phase shifts. IEEE Internet Things J. **8**(10), 8585–8602 (2021)
21. Ji, Z., Li, B.H., Wang, H.X., Chen, H.Y., Sarkar, T.K.: Efficient ray-tracing methods for propagation prediction for indoor wireless communications. IEEE Antennas Propag. Mag. **43**(2), 41–49 (2021)
22. Hsiao, A.Y., Yang, C.F., Wang, T.S., Lin, I., Liao, W.J.: Ray-tracing simulations for millimeter wave propagation in 5G wireless communications. In: IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting (2017).https://doi.org/10.1109/APUSNCURSINRSM.2017.8072993
23. Yun, Z., Iskander, M.F.: Ray-tracing for radio propagation modeling: principles and applications. IEEE Access **3**, 1089–1100 (2015). https://doi.org/10.1109/ACCESS.2015.2453991
24. Tiberi, G., Bertini, S., Malik, W.K., Monorchio, A., Edwards, D.J., Man, G.: Analysis of realistic ultrawideband indoor communication channels by using an efficient ray-tracing based method. IEEE Trans. Antennas Propag. **57**(3), 777–785 (2009)
25. Kochenderfer, M.J., Wheeler, T.A.: Algorithm for Optimization. Massachusetts Institute of Technology (2019)
26. Floudas, C.A., Pardalos, P.M.: Encyclopedia of Optimization. Springer (2009). ISBN 978-0-387-74760-6
27. Selman, B., Gomes, C.P.: Hill-climbing search. (2006). https://doi.org/10.1002/0470018860.S00015
28. Chinnasamy, S., Ramachandran, M., Amudha, M., Ramu, K.: A review on hill climbing optimization methodology. Recent Trends Manag. Commer. **3**(1), 1–7 (2022)
29. Bala, A., Padmaja, T., Gopisettry, G.K.D.: Auto-dialog systems: implementing automatic conversational man-machine agents by using artificial intelligence & neural networks. Int. J. Sci. Res. Rev. **7**(1), 1–5 (2018)

SVBOX: Switch Video BOX for Monitoring and Protection of the Elderly and Their Residences

José Paulo Lousado[1(✉)], Sandra Antunes[2], and Ivan Pires[3,4]

[1] Research Centre in Digital Services (CISeD), Polytechnic Institute of Viseu, Viseu, Portugal
jlousado@estgl.ipv.pt
[2] Centre for Studies in Education and Innovation (CI&DEI), Polytechnic Institute of Viseu, Viseu, Portugal
santunes@estgl.ipv.pt
[3] Instituto de Telecomunicações, Covilhã, Portugal
[4] Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Águeda, Portugal
impires@ua.pt

Abstract. The COVID-19 pandemic has shown poor communication between people, especially the elderly living in heavily deserted and depopulated agricultural regions. This is a reality present in Portugal and all over the world, where there are technological and physical barriers that prevent these older adults, the most vulnerable people, from having easy access to essential goods, food, and medicines, among others, and are also exposed to risky situations, such as theft and fraudulent activities. Bearing in mind that the best company for older adults is television, we propose the development of a system called SVBOX (Switch Video Box), which includes the design and proof of concept of an electronic system that allows the integration of various equipment, including video intercom, designed to be an active surveillance device, with the integration of the video intercom system with the internal network of sensors in the home, for monitoring living conditions and security, detecting fire, flooding, toxic gases, among others, using the television as the main means of communication and sensors to monitor the vital signs and health status of older adults. This system has growth potential, with several applications in real security and protection contexts, namely in security environments and monitoring of health conditions.

Keywords: Internet of Things · Domotics · Active Assisted Living · Sensor Networks

1 Introduction

Portugal is a country of older adults and is the second most-aged country in the European Union. According to the Census 2021 report, a study conducted by INE (National Statistics Institute), the aging index in Portugal translates into 182 elderly for every 100 young people, contrasting with 128 elderly for every 100 young people in 2011 [1].

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 28–41, 2024.
https://doi.org/10.1007/978-3-031-59619-3_3

The last Senior Census [2], conducted in 2021, under the responsibility of GNR (National Republican Guard), a security force that acts precisely in the prevention and protection of people and property, reported 44,484 elderly living alone and isolated, or in a situation of vulnerability, due to their physical, psychological, or other condition that could jeopardize their safety. These older adults are concentrated mainly in the country's interior, around 50%. Let's consider the highly depopulated Alentejo region. This figure rises to about 64%, i.e., the most depopulated areas of mainland Portugal concentrate the most significant number of older adults in situations of risk in terms of health, security, and social abandonment.

Compounding the already worrying situation, we had the pandemic of COVID-19, which had a significant impact on the lives of older adults, especially those living isolated in remote areas without home support or permanent medical care. These people faced additional challenges due to mobility restrictions, lack of access to medical and social services, as well as the increased risk of other diseases that high restrictions in terms of mobility and medical assistance may have contributed to the worsening health conditions of the population, who reside in such conditions.

The article is organized as follows: in Sect. 2 we analyze recent works related to the theme; in Sect. 3 we present the framework and design of the SVBox system; in Sect. 4 we show the proof of concept, using the development of the prototype, with the integration of several microcontrollers and sensors, in a controlled environment and in Sect. 5 we present the conclusions and point the way for future work.

2 Related Work

Several authors have contributed to monitoring people and goods, with relevant works in this field, with complex applications. In this field, one of the first works was proposed in [3], in which the authors developed a system for tracking and monitoring the health of older adults living at home. The system was composed of an infrared communication device connected to several sensors whose data were collected using software installed on a personal computer at home. The collected data was transferred to a server over the Internet using a cable television (TV) connection. In this study, one person was monitored for approximately 6 months. The objective was to detect, throughout this time, atypical days when the older woman went out of her daily routine, and the results were presented with the suggestion of the system as a complement for remote rehabilitation monitoring as assistive technology.

More recently, in [4], the authors present a system that proposes itself as a SecureHome TV ecosystem, a technical solution based on the interaction of older adults with their TV set, one of the most used appliances in their daily lives, acting as a non-invasive sensor that allows the detection of possible risk situations through an elaborated alert algorithm. The authors describe the SecureHome TV ecosystem, emphasizing the alert algorithm, and report on its validation process. The algorithm detects the most dangerous situations, contributing to the monitoring of the elderly's well-being at home, requiring, however, the contracting of a television and cable communications service, with the installation of the software in the television service box.

In previous works, namely [5] and [6], solutions were presented that allow monitoring of the health conditions of older adults using technologies supported by LoRa (Long Range) communication systems. LoRa (Long Range) communication networks can improve the living conditions of older adults living in remote and isolated areas. They are especially suitable for long-range and low-power consumption applications, making them a viable option for remote monitoring and communication in areas with little telecommunications infrastructure.

In [6], we proposed the development of an infrastructure based on LoRa networks supported by a LoRa gateway and an environmental conditions monitoring node only for signal strength testing. We showed the feasibility of this system for use in different contexts, particularly in monitoring older adults.

In [5], we extended the concept and its applicability to a practical case for outdoor monitoring of people, using a vest, with equipment with vital signs monitoring and GPS sensors, accelerometer, and temperature, among others, to prevent falls, disorientation, etc., sending SMS messages to family members with alert information, when problems occur, for example when it moves away from the residence beyond the configured radius or polygon. This system can be connected to TTN (The Things Network), allowing the data to be analyzed practically in real time. However, LoRa networks have several data transfer rates and sending frequency limitations. This leads to the fact that part of the system proposed above, namely sending SMS to family members or security and civil protection forces, is supported on 3G/4G networks.

3 System Design

Following the work that has been developed, in this section we present the framework and motivation that define the scope of this article, as well as the conceptual scheme and the monitoring areas on which this work focuses.

3.1 Framework and Motivation

The role of information systems and technologies, particularly monitoring systems, can play a key role in improving the living conditions of isolated older adults who, even after the end of the pandemic, will continue to suffer for a long time, not to say irreversibly, the consequences associated with the restrictions arising from COVID-19. These systems could benefit communities living in remote and isolated locations, especially the elderly population, notably through:

- Remote health monitoring: Monitoring systems can allow older adults to take measurements of vital signs, such as blood pressure, heart rate, and blood oxygen levels, among others, from home. This information can be transmitted to health professionals who can keep track of the user's health status and intervene in an emergency.
- Virtual communication: Communication technologies such as video calls and instant messaging can help combat social isolation. Seniors can connect with their family members, friends, and healthcare professionals, making them feel safer and more supported, even at a distance.
- Safety monitoring: Sensor networks linked with smart devices can be installed in homes to monitor the safety of older people. For example, motion sensors can detect falls or unusual activity and alert family members, emergency services, or law enforcement.
- Access to information and services: Information technology can provide up-to-date health information, relevant news, and local services available to seniors, such as pharmacy opening hours, health clinics, and more, helping them make decisions and find needed resources without leaving home.
- Telemedicine and remote consultations: One field that has had the most relevance during the COVID-19 pandemic has been virtual medical consultations, allowing seniors to get medical care and advice without needing physical travel. This primarily benefits those living in remote areas without easy access to medical services.

However, information and communication technologies naturally have excellent resistance in their adoption, both by the population in more depopulated and isolated areas, since these are older adults with high, not to say total digital illiteracy, having as their main point of contact with information and as a companion, open signal television or paid cable, fiber, ADSL or satellite television services, so this equipment naturally emerges as the device that motivated us in this work, and that led us to try to answer the main question: How can we support the monitoring of older adults and their homes, especially those living in remote and isolated areas, integrating security alerts and warnings, both locally and remotely?.

3.2 System Concept

The present work has developed a device called SVBOX for active surveillance, which integrates a video entry system with a network of internal sensors to the house and sensors for monitoring the vital signs of older adults. The system has an interconnection point with the home's television to maintain a permanent state of alert with resources of communication and telecommunication means.

The operational objectives of the device are:

- Integration of the housing sensor network for fire detection, flooding, and dangerous gases (CO2, CO, combustible gases), among others, with biometric sensors for individual use (Heartbeat, Temperature, fall detection), in a system of collection and processing of data in Real Time;
- Integration of the Video Intercom with the TV so that it works as an intercom (larger vision area, better audio capacity);

- Establish the concept of e-Neighbor, as an active neighborhood, permitting certain people to access and receive information from older adults whenever some anomalous situation occurs.

Figure 1 shows the operational scheme of the system, as well as the elements that constitute it. It should be noted that the proposed system is independent of any telecom operator, open and low-cost, and can be implemented with low-cost devices.

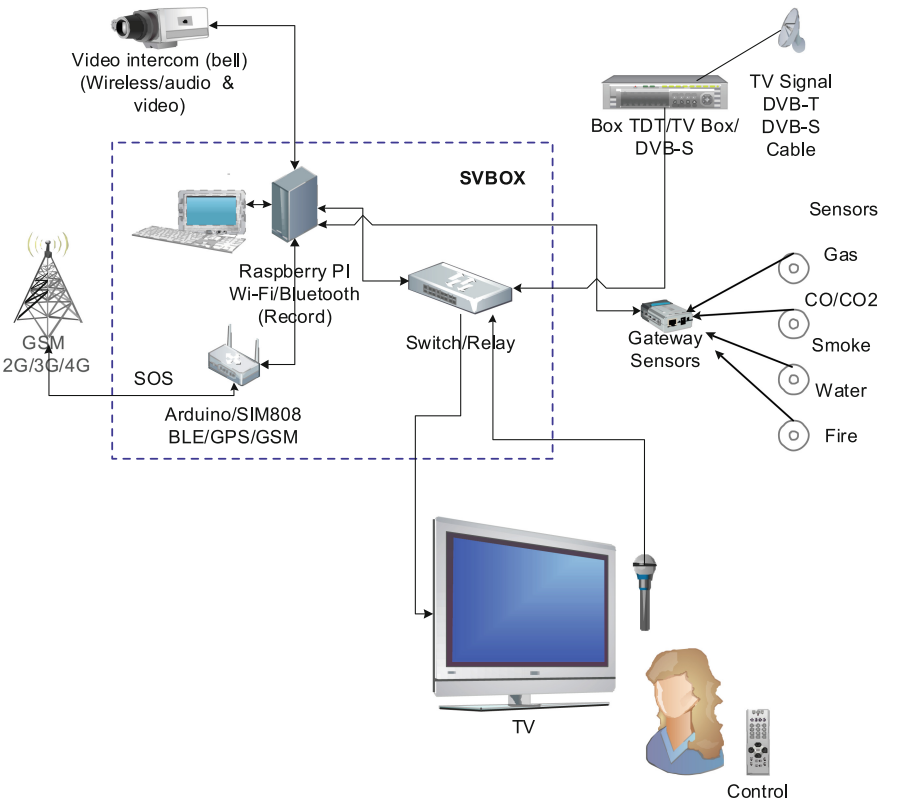


Fig. 1. Conceptual scheme

The operation of the proposed system is quite simple, focused on day-to-day functionalities, where older adults, in their daily routines, will have in the SVBOX a companion, expanding the concept of active surveillance to assistive technologies, supported in IoT (Internet of Things), integrating the concept of (e-caregiver) [7].

3.3 Operationalization of the External Surveillance System

Whenever someone touches the Video Doorbell, the call is established with the TV, and the answering and conversation can be made by the user and the outside, using the microphone built into the SVBOX or the remote control.

The SVBOX should have a specific command that allows sending an automatic call to the emergency (panic button type) if something is not going well, and in this case, may have a payment plan for the use of GSM low cost (similar to the devices provided in Portugal by law enforcement and civil protection). Alternatively, the possibility of sending a message via the LoRa network (free of charge) to active neighbors (e-Neighbor), in a policy of solidarity and proximity support, when the family members do not live close by. In this case, it should be considered that LoRa communication is low-speed and is recommended for sending small data packets of up to 5 km (in ideal situations). This critical scenario was analyzed in previous works, namely in [5, 6], as mentioned above.

The option of recording communication with the outside world can be activated, always subject to the rules and restrictions imposed by the GDPR (General Data Protection Regulation) in force in the European Union [8].

3.4 Indoor Monitoring

When any sensor (from the sensor network) detects an anomalous value, the SVBOX will switch to the TV and show the message on the screen with an image of the danger level (reminder, Warning, Danger, etc.), accompanied by a beep, which will only be turned off when the resident presses a suspend button, to correct the possible problem (configurable). If the problem and the level of risk remain, it will be activated again after a certain period. To control the alerts for possible false positives, the system turns off the alarm after some time, also configurable, even without human intervention, or if the situation has been restored to normal values. However, if the condition persists, the alerts will be activated again.

The requirement of a connected TV is complemented by a small LCD and sound equipment always on and integrated with the SVBOX, redundant, which will remain active even if the TV is turned off, also being a solution for cases where the TV has no HDMI interface (homes with old equipment, where it is not possible to integrate with the SVBOX).

A UPS should allow the SVBOX system and the sensor network in the house to remain operational for some time, even if the power fails (no connection to the TV).

The sensors for vital data monitoring and fall detection will be for individual use, being sent to the SVBOX via Wi-Fi/Bluetooth, which will analyze them in real-time, and, if it detects any anomalous occurrence, it should send an alert signal to connected civil protection entities, neighbors (e-Neighbor's) or family members, for help and emergency support. This communication may be via SMS or 112 calls, as already happens with the entities connected to civil protection in Portugal, which try to establish verbal communication with people needing assistance screening [9].

4 Proof of Concept

A prototype consisting of several components was developed for the proof of concept. The devices were interconnected in a local network and configured so that each one had a fixed IP address. Among these devices are a Raspberry PI3, an aggregator, and a main application server.

4.1 Video Intercommunication with ESP32-CAM

The ESP32-CAM [10] is an MCU (Micro Controller Unit) development board module with a small size with an ESP32-S chip and an OV2640 camera, a microSD card slot, and with several GPIOs to connect peripherals. The ESP32-CAM is widely used in various IoT applications, such as smart home equipment, wireless monitoring, and QR code identification (Fig. 2). Its specifications and the ESP32-CAM features can be found on the manufacturer's website [10].

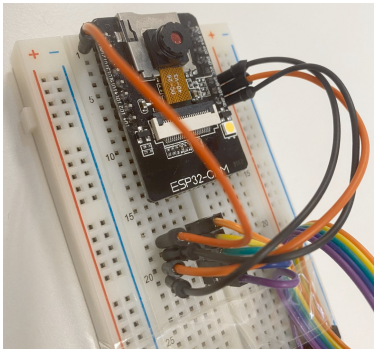


Fig. 2. ESP32-CAM in a protoboard

The main limitation is that it does not have built-in audio. However, given the features and capabilities built into the MCU, it can be used in remote monitoring and video surveillance systems.

For programming the ESP32-CAM, we used the PL2303 TTL USB-Serial converter. This adapter uses Prolific's PL2303HX chip to convert USB signals into TTL RS232 signals, which makes it worthwhile when we want the computer to communicate with microcontrollers in general.

This converter is a small board with a USB connector at one end to be connected directly to the computer and at the other end a pin bar for easy connection with the board whose communication will be established. Its voltages of 3.3 V and 5 V allow it to work in a broader number of microcontrollers regardless of their power supply.

4.2 AdaFruit M0

Another device we integrated into the prototype was the Feather M0 MCU from Adafruit [11], which has an ATSAMD21G18 ARM Cortex M0 processor with a 48 MHz clock and 3.3 V power supply, the same used in the new Arduino Zero. This chip has 256K FLASH and 32K RAM (Fig. 3). The chip has built-in USB, so it has USB-to-serial programming and built-in debugging without the need for an FTDI-type chip. Its specifications can be found on the manufacturer's website [11].

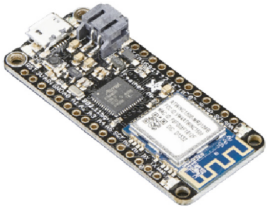


Fig. 3. Adafruit Feather M0 Wi-Fi - ATSAMD21 + ATWINC1500 [11]

Its application in the prototype focused on monitoring the sensors and sending the data in real-time to the server via Wi-Fi.

4.3 SVBOX System Prototype

In building the prototype, we started by working on the video intercom, the ESP32-CAM, mentioned earlier.

Since the ESP32-CAM does not contain a USB port, it was necessary to use this converter to upload the code via GPIOs. The connection between the microcontroller and the peripheral can be seen in the table below (Table 1).

Table 1. The connection between the microcontroller and the FTDI.

ESP32-CAM	FTDI Programmer
GND	GND
5 V	VCC (5 V)
U0R	TX
U0T	RX
GPIO 0	GND

After its installation and configuration, we resorted to the Arduino IDE to develop the web server source code for sending images in real time. Considering that at this stage of prototyping, it is unnecessary to configure the internal network, the IP addresses were set as static so that the various connected devices can be easily identified, from sensors and MCUs, among others. In the ESP32-CAM, through the IP address, the camera configuration page displays all the available settings of the OV2640 camera (Fig. 4).

It should be noted that the ESP32-CAM allows facial recognition, so this feature can be used to develop Machine Learning and Artificial Intelligence components that will be integrated into the system for the automatic identification of people.

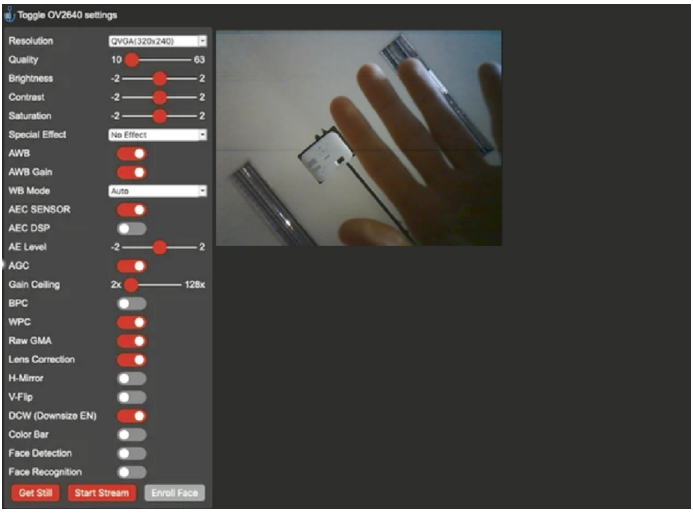


Fig. 4. ESP32-CAM Web Server

It was necessary to use the web application CyberChef [12] to convert code that the ESP32-CAM support software provides in hexadecimal to HTML. This conversion was required to make changes to the OV2640 camera so that it would transmit automatically without needing to be started by a command button.

We used another microcontroller, the Adafruit M0 MCU, to simulate the sensor network. We attached an MQ3 sensor to this microcontroller that allows the detection of certain types of flammable vapors (e.g., ethanol). Still, we could use another sensor, e.g., the MQ2, that enables the detection of toxic gases, such as carbon monoxide and dioxide, receiving different values. The MCU was configured and programmed to work as a web server permanently sending data, which will be parameterized by the Raspberry PI3 so that it can later detect the abnormal values and trigger the necessary alert and warning actions on the TV (Fig. 5).

It should be noted that the MQ3 sensor works efficiently at 5 V, becoming unstable with a voltage of 3.3 V, so it was necessary to guarantee this voltage, available on the MCU's USB pin.

During the implementation of the system with the ESP32-CAM, we found that the GPIOs that were supposed to be available for programming the call button (push button)

Fig. 5. MQ3 sensor Web Server

were all occupied for operating the camera and the one intended for the memory slot, so we had to study another solution. To circumvent the problem, we turned to another MCU Adafruit M0, which was programmed in the same way as we had designed for the ESP32 (Fig. 6). Its function is to provide a push button sensor, working as a buzzer, lighting an LED for 10 s. This sensor will also work as a web server, permanently sending data of "button state", state 0 or 1 (0 being for off and 1 for on).

The Raspberry reading function given by the button automatically makes the switch to its port (HDMI) and shows the camera transmitting what is on the street, with a warning sound signal that lasts for 20 s, that someone is ringing the bell.

In designing this work, we resorted to using the VS1838B sensor, an infrared receiver, on which we used a command to turn off the alerts. This way, we can give instructions to the Raspberry, turning the switch on and off, among others that can be added as needed. The TV Switch, given the technical limitations, which prevented us from using a video signal multiplexer, such as the TS3DV421 Multiplexer [13], available only for industry, for the proof of concept, we used the relay connections, having studied the different types of signal in the HDMI cable, opting for a parallel analog solution of 12 relays (only the ones needed for sending signal switching) controlled by the Raspberry which automatically switches to the equipment to be displayed on the TV. The TV Switch has two inputs and one output, the inputs are the TV box, and the other is the Raspberry itself.

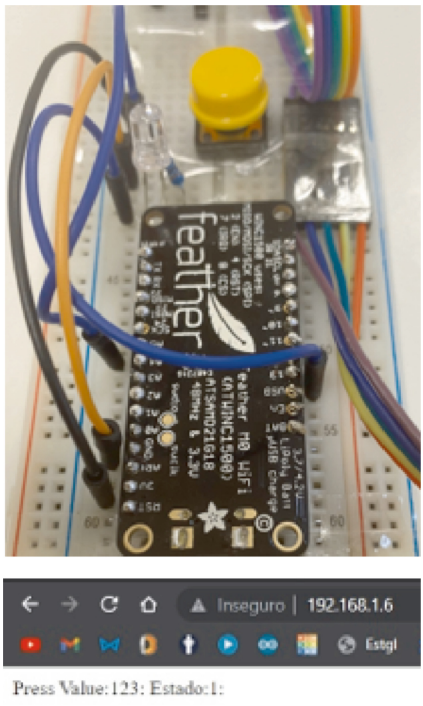


Fig. 6. Video Enable and Bell Activation Web Server

The output is connected to the TV set (Fig. 7).

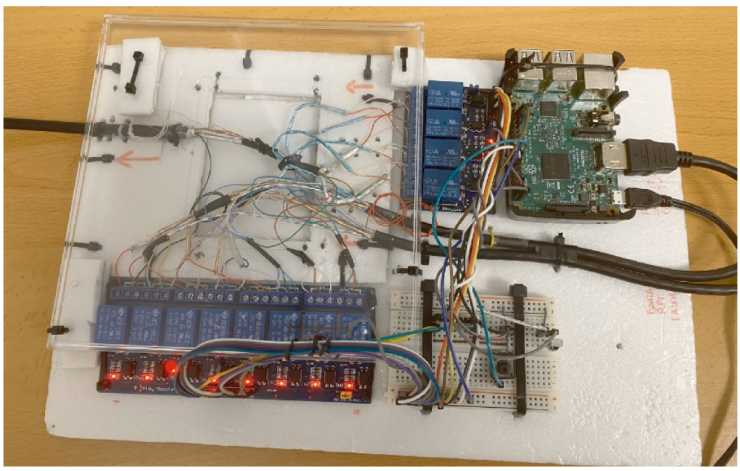


Fig. 7. Analog switch with relays

The core equipment of the prototype is the Raspberry PI. The model we used to develop the prototype was the Raspberry PI 3 Model B [14]. In programming the Raspberry PI, the Python language was used.

In the project prototype, the functional scheme has the ESP32-CAM always operational, constantly transmitting on the Raspberry through the graphical interface created using the QT5 libraries and the QT Designer.

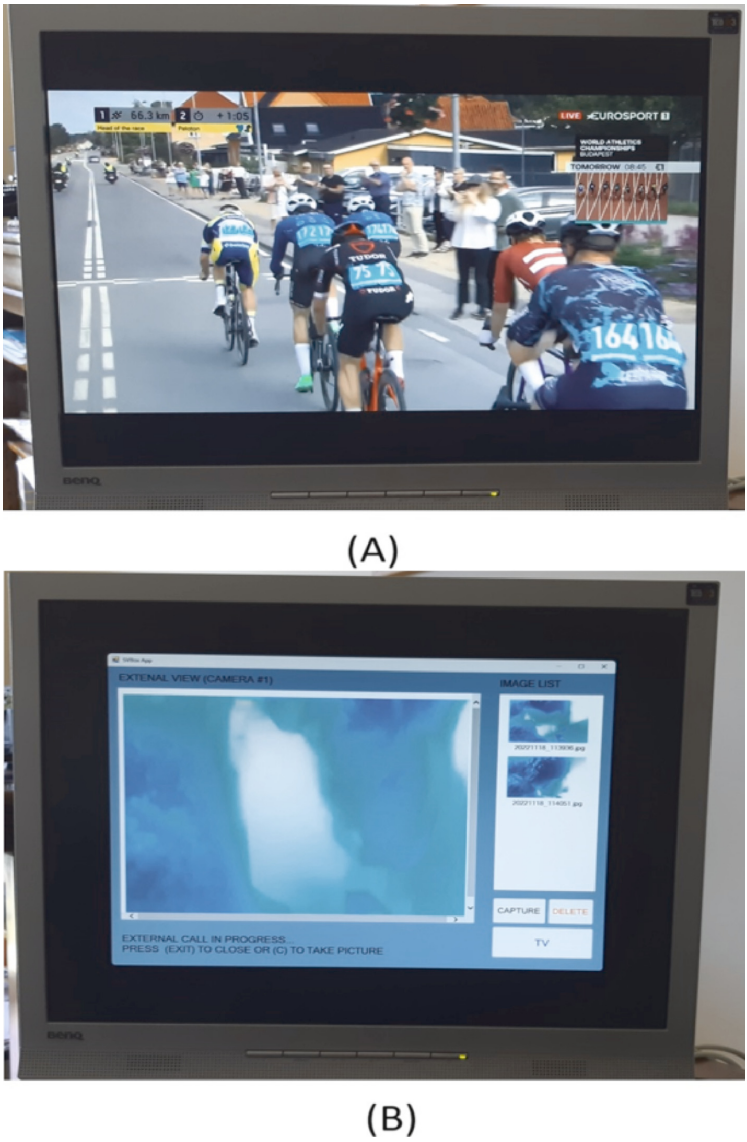


Fig. 8. TV screen image: (A) normal operation with DTT Box; (B) Interface with external call

We switch between the GUI (contained in a Raspberry) and the operator box through the remote command. The remote control was configured with four options so that it could also work manually. The system automatically switches to the Raspberry interface when we click the ON button. Consequently, when we click the OFF button, the system reverts the previously said process and goes back to the box.

A screen capture button has also been programmed in case the user wants to register some photograph from the outside (Fig. 8-B), for example, of someone who may be in the vicinity of the residence, within the capture range of the camera. When the sensors register an anomalous value, an "Alert" window will be displayed and overlay the graphical interface. Once the user knows the alert, they can close the window using the command button.

In Fig. 8-A and 8-B we can see the two connection interfaces in operation, the TV-DTT (Digital Terrestrial Television) Box and the SVBox in the reception of an external call respectively with 800 × 600 resolution. It should be noted that the image of the face in Fig. 8-B was intentionally distorted.

Once the system can be connected to the Internet, the Raspberry PI, the central heart of the system, can be used for remote access. A family member or the resident can access, via a remote computer or mobile device, the images and data of the sensors collected at any moment.

5 Conclusion and Future Work

The present work proved very important in developing an integrated system called SVBOX. It enables older adults inside their residences to control and monitor their safety and health conditions, using the television as a central monitor. Currently, with the expansion of IoT to particular areas such as home automation, and the monitoring of people and goods, the investigation of new equipment and modes of operation in different contexts becomes a constant challenge for the scientific community. However, the limitations imposed by external situations, as was the pandemic by COVID-19, showed us that we need to strengthen the means of communication and alternative methods, to meet the needs of keeping people safe.

In this work, we found some limitations and constraints for future work, namely the study of other MCUs existing in the market, which facilitate video communication and incorporate audio, which here was by using a conventional voice intercom activated by a push button. Using the TS3DV421 Multiplexer or similar could guarantee excellent quality in the video commutation, which we didn't get with the analog solution, having the maximum resolution achieved with 800 × 600 VGA. Another limitation is related to the equipment available on the market. Since the project started in the middle of the pandemic, much of the equipment we needed for the project's development was not available in time, preventing us from implementing the prototype with the features we wanted at the beginning.

In future work, in addition to audio, we intend to include a multiplexer with an I2C interface, such as TS3DV642EVM [15].

Finally, but equally important, the development of a mobile application that allows remote monitoring without using a browser, either by residents when they are away

from home or by relatives or neighbors (e-neighbors) who have a relationship of trust and responsibility in monitoring the most vulnerable older adults.

Acknowledgments. This work was funded by National Funds through the Foundation for Science and Technology (FCT), I.P., within the scope of the project Ref. **UIDB/05583/2020**. Furthermore, we would like to thank the Research Centre in Digital Services (CISeD) and the Polytechnic of Viseu for their support.

This work is funded by FCT/MEC through national funds and co-funded by FEDER – PT2020 partnership agreement under the project **UIDB/50008/2020**.

References

1. Fernanda Cerqueira: Censos 2021: seniores representam 23,4% da população portuguesa. https://impulsopositivo.com/censos-2021-seniores-representam-234-da-populacao-portuguesa/
2. GNR: Operação Censos Sénior 2021 – Balanço. https://www.gnr.pt/comunicado.aspx?linha=4625. Accessed 29 Apr 2023
3. Suzuki, R., et al.: Rhythm of daily living and detection of atypical days for elderly people living alone as determined with a monitoring system. J. Telemed. Telecare **12**, 208–214 (2006). https://doi.org/10.1258/135763306777488780
4. Abreu, J., Oliveira, R., Garcia-Crespo, A., Rodriguez-Goncalves, R.: TV interaction as a non-invasive sensor for monitoring elderly well-being at home. Sensors **21** (2021). https://doi.org/10.3390/s21206897
5. Lousado, J.P., Pires, I.M., Zdravevski, E., Antunes, S.: Monitoring the health and residence conditions of elderly people, using LoRa and the things network. Electronics **10**, 1729 (2021). https://doi.org/10.3390/electronics10141729
6. Lousado, J.P., Antunes, S.: Monitoring and support for elderly people using LoRa communication technologies: IoT concepts and applications (2020).https://doi.org/10.3390/fi12110206
7. Harish, B.R.: Medbox: a reliable e-caregiver smart system using IoT. Int. Innov. Res. J. Eng. Technol. **2**, 53–61 (2017). https://doi.org/10.32595/iirjet.org/v2i4.2017.40
8. Agarwal, S., Kirrane, S., Scharf, J.: Modelling the general data protection regulation. Jusletter IT **2014** (2017)
9. GNR: Programa Apoio 65 – Idosos em Segurança. https://www.gnr.pt/ProgEsp_idososSegurança.aspx. Accessed 21 Apr 2023
10. Thinker, A.: ESP32-CAM. https://docs.ai-thinker.com/en/esp32-cam. Accessed 30 Mar 2021
11. Adafruit: Adafruit Feather M0 WiFi - ATSAMD21 + ATWINC1500. https://www.adafruit.com/product/3010. Accessed 20 Mar 2021
12. Crown: CyberChef. https://gchq.github.io/CyberChef. Accessed 20 Mar 2023
13. Texas Instruments: 4-Channel Differential 8:16 Multiplexer Switch for DVI/HDMI Applications (2010)
14. Raspberry Pi Foundation: Raspberry Pi 3 Model B. https://www.raspberrypi.com/products/raspberry-pi-3-model-b/. Accessed 20 Mar 2022
15. Texas Instruments: TS3DV642EVM : 12-channel 1:2 mux & demux with 1.8-V compatible control & power-down mode evaluation module. https://www.ti.com/tool/TS3DV642EVM. Accessed 30 Apr 2023

Node Importance Evaluation Method for Heterogeneous Networks Based on Node Embedding

Hui Cui(✉), Linlan Liu, and Jian Shu

Nanchang Hangkong University, NanChang 33000, China
991136904@qq.com, {liulinlan,shujian}@nchu.edu.cn

Abstract. In reality, complex systems are often represented by networks, and heterogeneous networks are more effective in describing the interaction behaviors among various elements. Evaluating the importance of nodes in a heterogeneous network is beneficial for maintaining the stability of the network. This study proposes a node importance evaluation method, named TAGCN Auto-Encoder Comprehensive Influence (TAE-CI), for heterogeneous networks, which combines graph neural networks with centrality measures. The method uses Topology Adaptive Graph Convolutional Networks (TAGCN) to improve graph autoencoders, encode different semantic subgraphs, reconstruct adjacency matrices, and optimize reconstruction loss to obtain node embedding vectors. To obtain the comprehensive influence of the nodes, the node embedding vectors are incorporated into the topological potential function to calculate the global influence, which is then combined with the local influence. The proposed method is evaluated on three real network datasets using the Susceptible Infected Recovered (SIR) model, and the results demonstrate its efficacy in evaluating node importance.

Keywords: Heterogeneous networks · Node importance evaluation · Node embedding

1 Introduction

The integration of new technologies, such as artificial intelligence, cloud computing, and big data analysis, in the era of big data has led to increasingly closer connections between humans and various objects in society. Real-world networks are similar to interwoven webs, where data serves as a medium of communication and networks function as a means of transmission through various forms of mutual interaction and influence. Complex networks serve as an abstraction of real networks [1], where people and various objects are represented as nodes, and their connections form edges. Complex networks can be classified into homogeneous networks (HON) and heterogeneous networks (HEN) based on the characteristics of nodes and edges. Homogeneous networks are networks

Supported by The National Natural Science Foundation of China (62062050, 61962037), and the Innovation Foun-dation for Postgraduate Student of Jiangxi Province (YC2022-s727).

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 42–53, 2024.
https://doi.org/10.1007/978-3-031-59619-3_4

composed of nodes and connections of the same type, whereas heterogeneous networks contain a range of nodes and connections of different types [1]. Homogeneous networks are a simplified representation of real-world networks, while heterogeneous networks better reflect the dynamics and complex interactions of real-world systems.

An increasing number of scholars are studying heterogeneous networks, which encompasses research in link prediction, community detection, node classification and node importance evaluation. Node importance evaluation is the method of quantifying the influence of nodes within a network. Node influence can be assessed from two viewpoints. Firstly, the information dissemination ability of a node in the network, which measures whether information acquired by a node can be efficiently transmitted to a significant fraction of nodes in a timely manner. Secondly, the importance of a node in maintaining network stability, which assesses the impact of a node's disruption on the network connectivity in light of the network topology. The assessment of node importance in networks has a broad range of real-world applications. For instance, it is applicable in securing critical facilities in communication or power networks and responding promptly to emergencies. In academic communities, it can be used to gauge the impact of publications or researchers. The evaluation of node importance in online social networks allows the identification of influential users and promotes efficient management of public opinion while curbing the propagation of rumors.As a result, assessing node importance in a network is a valuable approach to ensure network stability and enhance operational efficiency.

This study focuses on the topological and semantic information in heterogeneous networks and utilizes graph autoencoders to encode various semantic subgraphs and reconstruct their corresponding adjacency matrices. By optimizing the reconstruction loss, this study obtains node embedded vectors and calculates the overall impact of nodes by combining topological potential [2] and centrality measures. The main contributions of this study are described as follows:

(1) The topology adaptive graph convolutional networks (TAGCN) [3] are utilized to enhance the graph autoencoder, which allows the encoder to have an expanded receptive field and aggregate more comprehensive node features in a single convolution. Additionally, TAGCN enables the efficient processing of large-scale networks with less computation cost.
(2) This paper proposes a node importance evaluation method TAGCN Auto-Encoder Comprehensive Influence (TAE-CI), which comprehensively assesses the importance of nodes from both local and global perspectives. Local influence is calculated based on node degree, while global influence is computed by combining network topological potential with the similarity between embedding vectors.

2 Related Work

The assessment of node importance is a fundamental challenge in complex network analysis. While research on homogeneous networks has reached a considerable level of maturity, research on the evaluation of heterogenous networks remains relatively limited. This paper presents a comprehensive overview of related work on node importance evaluation in complex networks, focusing on two primary approaches: indicator-based evaluation methods and node embedding-based evaluation methods.

2.1 Indicator-Based Evaluation Methods

The indicator-based evaluation method defines nodal importance indicators on either a local or global scale and employs one or more indicators to assess the significance of nodes. In homogeneous networks, traditional centrality indicators used in both local and global contexts include degree centrality (DC) [4], betweenness centrality [5], and PageRank [6]. Reference [7] proposes the k-shell based on gravity centrality (KSGC) method to rank the importance of nodes based on global indicators. This method considers the K-shell value of the node in the network and the topological structure and edge importance between its neighboring nodes. By using a gravity model to synthesize local and global information about the nodes, this method produces nodal importance rankings that are more precise than those obtained using the K-shell [8] method. On the other hand, reference [9] proposes the Global Structure Model (GSM), which integrates k-shell and path-based evaluation methods. In [10], the Local-and-Global Centrality (LGC) centrality index is proposed as a method that combines both local and global influence. To calculate local influence, LGC uses degree centrality, and to calculate global influence, it incorporates shortest paths between nodes and node degrees. The product of these two factors determines the importance score for each node. Although this key node evaluation method has demonstrated its effectiveness, applying it directly to heterogeneous networks would significantly reduce its performance. Furthermore, such methods tend to concentrate on the structural information of the network and overlook the semantic information.

To evaluate heterogeneous networks, in [11] the Entropy Rank Method (ERM) is proposed, an algorithm for sorting based on information entropy. ERM defines three types of probability entropies based on meta-paths and linearly combines them to determine the importance of a node within that meta-path. Reference [12] calculates the centrality of nodes on each layer using global indices K-shell and betweenness centrality on multilayer networks. By quantifying the neighbor's weights in multilayer heterogeneous networks, the method utilizes the multi-relationship coupling information and transmission mechanism to obtain the importance of nodes. Reference [13] collects information from multiple levels of neighborhoods, evaluating the super spreaders in the epidemic propagation network by considering different types of networks, including social networks, highly heterogeneous interpersonal contact networks, and epidemiological networks.

2.2 Node Embedding-Based Evaluation Methods

In recent years, researchers have leveraged machine learning and deep learning techniques, such as graph neural networks and reinforcement learning, to extract the essential features of nodes in a network. These methods are used to form evaluation methods based on node embeddings. In [14], node importance labels are generated using the Susceptible Infected Recovered (SIR) model, where the top 5% of nodes are ranked as the most influential, and the other nodes are classified as less influential. The node sampling is performed using a breadth-first search algorithm, and the graph convolutional network is utilized to aggregate features and obtain node embedding vectors. Finally, node importance evaluation is transformed into a node classification task. Reference [5] employed

betweenness centrality to label nodes on a small-scale network. Node embedding vectors were then generated using an encoder, transformed into scalars using a decoder, and used to fit the betweenness centrality. This approach was subsequently applied to evaluate crucial nodes in a large-scale network. Reference [15] starts with the robustness of the network, using graph resistance as the node importance label, and applies GraphSAGE to obtain node embedding vectors. The embedding vectors are input into a multi-layer perceptron to obtain node importance scores for evaluation. The aforementioned methods are all evaluation methods in homogeneous networks, and most of them require labels. Reference [16] applies metapath2vec to construct a node importance evaluation method for heterogeneous networks. Based on a series of metapaths with specific semantics, this method extracts heterogeneous features from different metapaths and integrates different structural features using a weighted mechanism. By adopting a relevance threshold, it identifies the most influential set of nodes. This kind of method relies heavily on prior knowledge because it highly depends on the discovery and selection of metapaths.

3 Preliminaries

3.1 Heterogeneous Network

Heterogeneous network is defined as a directed graph $G = (V, E, \phi, \psi)$ with two mapping functions [17], one for object types ϕ:V $\rightarrow$ A and one for relationship types $\psi : E \rightarrow R$. Every node v $\in$ V is categorized by a specific node type A:ϕ(v) $\in$ A and every edge e $\in$ E is categorized by a specific relationship type R : $\psi(e) \in R$, in which the sum of | in which the sum of || in which the sum of |A|and |R| greater than 2.

3.2 Metapath

Meta-path is defined as a sequence of alternating node types and relationships [17], denoted as P:$A_1 \rightarrow A_2 \rightarrow \ldots \rightarrow A_l$, it defines the composite relationship between node type A_1 and node type A_l, each meta-path can be regarded as Semantic information channel [1].

3.3 Semantic Subgraph

To extract semantic information in heterogeneous networks, meta-paths are utilized to divide the original network into several semantic subgraphs.The set of meta-paths is denoted as $MP = \{MP_1, MP_2, \ldots MP_S\}$, a heterogeneous network with s meta-paths is partitioned into s semantic subgraphs which can be denoted as $G = \{G_1, G_2, \ldots, G_s\}$, the i-th subgraph is constructed by traversing nodes and edges based on metapath MP_i. The adjacency matrix A_{G_i} of subgraph G_i is defined such that its elements represent the existence of edges between nodes in G_i. The initial feature matrix X_{G_i} of subgraph G_i is defined such that node features are randomly initialized.

4 The Proposed Method

This study proposes TAE-CI, a node importance evaluation method that combines graph neural networks, centrality measures, and topological potential. To improve the graph autoencoder, the study employs TAGCN, which encodes variant semantic subgraphs and gets embedded matrices of nodes. Graph convolutional network (GCN) is then used to decode and reconstruct them. Node embedded vectors are obtained by minimizing the reconstruction loss, The framework of node embedding model is depicted in Fig. 1. The global importance of nodes is calculated by providing the embedded vectors of nodes to the topological potential function [2], and their importance is estimated by combining them with local importance. TAE-CI includes two parts: the node embedding model and the node importance evaluation model.

4.1 Node Embedding Model

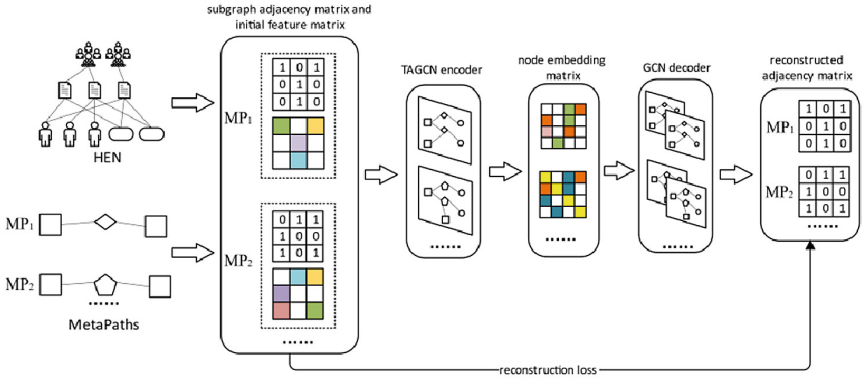


Fig. 1. Framework of node embedding model

Node Feature Encoding

This study utilizes TAGCN to encode the node features of diverse semantic subgraphs. As a variant of GCN, TAGCN employs k graph convolution kernels at each layer to extract local features of various sizes. This method can fully capture the graph's information and enhance the model's performance [3]. For a semantic subgraph G_i, its adjacency matrix is normalized using the degree matrix D_{G_i}, as shown in Eq. (1):

$$\tilde{A}_{G_i} = D_{G_i}^{-0.5}(A_{G_i} + I)D_{G_i}^{-0.5} \tag{1}$$

Then, the polynomial convolution kernel of the l-th layer for a given metapath is defined as $H_{P_i}^{l}$, which can expand the receptive field and adaptively adjust the size of node neighborhoods for different metapaths, as shown in Eq. (2):

$$H_{P_i} = \sum_{k=0}^{K} h_{k,p_i} A_{G_i}^{k} \tag{2}$$

where h^{l}_{k,p_i} is the polynomial coefficient and K is the kernel size. Finally, K convolutional kernels extract features on the graph structure, which are linearly combined, as shown in Eq. (3):

$$Z = \sigma\left(H_{P_i} X_{G_i} + b_{P_i}\right) \tag{3}$$

where Z is the embedded matrix of nodes, b_{P_i} is the bias and σ is the activation function.

Node Feature Decoding

Once the graph structure and initial feature matrix are fed into the encoder, the node embedded matrix can be generated. Given that node importance evaluation is an unsupervised task, this paper utilizes GCN [18] to decode the embedded vectors for reconstructing the semantic subgraph adjacency matrix and optimizing the reconstruction loss, thereby enhancing the representation power of the node embeddings. The decoding process of the embedded vectors of nodes is shown in Eqs. (4) and (5):

$$\hat{Z} = \tilde{A}_{G_i} Z \tag{4}$$

$$\hat{A}_{G_i} = \sigma(\hat{Z} W + b_{de}) \tag{5}$$

where $\hat{Z}$ is the node feature matrix after aggregation by GCN, $\hat{A}_{G_i}$ is the adjacency matrix of the reconstructed semantic subgraph G_i, W is the weight matrix, and b_{de} is the bias in the decoder. The loss function during network reconstruction is defined as Eq. (6):

$$L = \sum_{i=1}^{s} \left|\left|\hat{A}_{G_i} - A_{G_i}\right|\right|_2 + L_2 \tag{6}$$

where s is the number of semantic subgraphs, and L_2 is the regularization term to prevent overfitting.

4.2 Node Importance Evaluation Model

The network itself is an abstract system composed of nodes and edges, where the influence of a node itself is closely related to its neighboring nodes. In this paper, the local influence of a node is expressed by its degree. The local influence is defined as LI, and its calculation method is shown in Eq. (7):

$$LI(v_i) = \frac{d(v_i)}{N} \tag{7}$$

where $d(v_i)$ is the degree of node v_i and N is the total number of nodes in the graph.

Moreover, a node's global influence is usually associated with the shortest path between nodes. Normally, to determine the shortest path among all possible pairs of nodes, Dijkstra algorithm or Floyd algorithm is usually required. Nonetheless, for large-scale networks, these methods are impractical because of their time-consuming nature. In this paper, the global influence of a node is calculated by combining its embedded vector with topological potential. Similar to an electromagnetic field, a node itself forms a potential field that affects other nodes, and the effect of the influence is related to the

similarity between nodes. The calculation of similarity between two nodes is shown in Eq. (8):

$$c(v_i, v_j) = \frac{Z_{v_i} \cdot Z_{v_j}}{||Z_{v_i}||||Z_{v_j}||} \tag{8}$$

where Z_{v_i}, Z_{v_j} is the embedded vector of node v_i, v_j. Then, the similarity between nodes is incorporated into the topological potential function to calculate the global influence of a node using Eqs. (9) and (10):

$$m_{v_i} = \sum_{j=1}^{\phi(v_i)} d(v_i) \tag{9}$$

$$GI(v_i) = \frac{1}{N} \sum_{i \neq j}^{N} \left(m_{v_j} \cdot e^{-\frac{c(v_i, v_j)}{\lambda \bar{d}}} \right) \tag{10}$$

In this process, $\phi(v_i)$ is the set of neighboring nodes of node v_i, $\bar{d}$ is the mean degree of the network, and λ is a tunable parameter. The value of λ is determined by minimizing its entropy value as defined in Eq. (11):

$$\Pi(\lambda) = -\sum_{i=1}^{N} \frac{GI(v_i)}{U} \ln\left(\frac{GI(v_i)}{U}\right) \tag{11}$$

where $U = \sum_{i=1}^{N} GI(v_i)$ is normalization factor.

After the aforementioned process, the local and global influence of a node have been obtained. Therefore, the comprehensive influence of a node is defined as shown in Eq. (12):

$$CI(v_i) = LI(v_i) \cdot GI(v_i) \tag{12}$$

5 Experiment Result and Analysis

5.1 Datasets

In this paper, three real-world network datasets were selected for experimentation: ACM [19], a citation network containing three types of nodes (paper, author, and Subject); IMDB [20], a movie network subset from the Internet Movie Database containing three types of nodes (movie, actor, and director); and DBLP [19], an academic citation network containing four types of nodes (author, paper, term, and conference). The detail information about the datasets is shown in Table 1.

5.2 Evaluation Criterion

The assessment of node importance in complex networks often involves the use of propagation dynamics models to evaluate the assessment results. In this paper, the SIR model was selected for simulation experiments [21]. The SIR model contains three

Table 1. Detail Information about the Datasets

datasets	Number of nodes	node types	relationship types	metapaths
ACM	Paper:3025 Author:5995 Subject:56	3	2	APA APSPA
IMDB	Movie:4278 Director:2081 Actor:5257	3	2	MAM MDM
DBLP	Author:4057 Paper:14328 Term:7723 Conference:20	4	3	APA APTPA APCPA

types of node states: susceptible (S), infected (I), and recovered (R). There are two important parameters in the SIR model: the infection rate and the recovery rate.Nodes in the infected state have a probability of infecting first-order neighboring nodes in the susceptible state with infection rate, while nodes in the infected state have a probability of becoming nodes in the recovered state with recovery rate and nodes in the recovered state cannot be infected again.

To verify the effectiveness of node importance evaluation with the SIR model, all nodes are first initialized as susceptible. Then, a subset of nodes with the highest node importance values are selected to be infected at the initial stage according to a preset infection rate and recovery rate, and the entire network is infected accordingly. The proportion of infected and recovered nodes compared to the total number of nodes in the network when the spread reaches a steady state is then calculated as the spread range. If two methods have the same spread range, their propagation speeds will be compared by examining the time it takes for them to reach a steady state. For convenience in explaining the experimental results, the propagation coverage of nodes at the *t* time step is denoted as F(t), as shown in Eq. (13).

$$F(t)=\frac{\text{Propagation coverage at time step t}}{N} \tag{13}$$

As the infection and recovery rates affect the spread process, there might be some variations in experimental results even with uniform conditions. To mitigate this issue, this paper conducted 100 independent experiments and computed the average of the 100 experiments as the final evaluation metric.

5.3 Results Analysis

In order to compare the TAE-CI algorithm proposed in this paper with LGC, KSGC, GSM and traditional centrality indices PageRank and DC, the node importance assessed by these methods is ranked from highest to lowest, and the top 20 nodes are selected as

the key node set. The nodes in this set are set as the infected state in the SIR model. The propagation coverage rate F(t) at a given time step is counted, and the SIR transmission experimental curve is drawn. Figure 2 shows the SIR propagation curve of each method on three datasets.

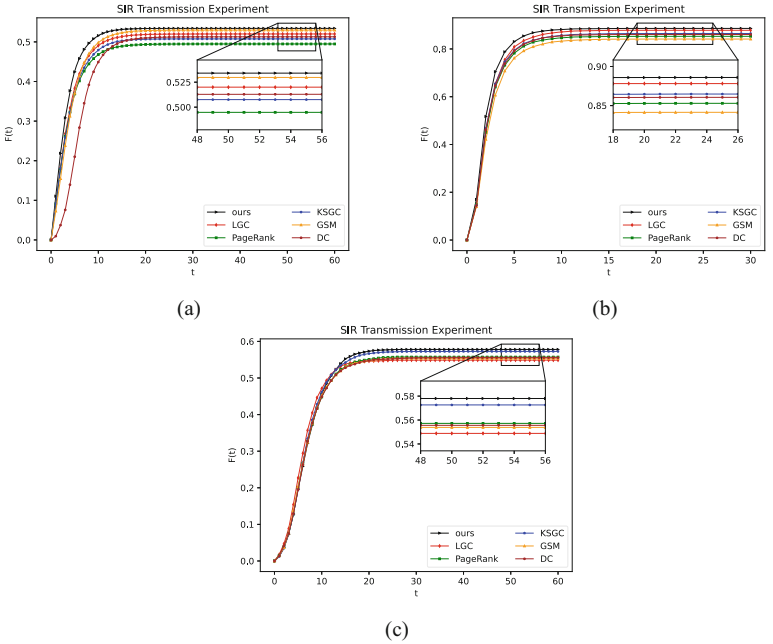


Fig. 2. The SIR transmission experiment. (a): ACM dataset;(b): DBLP dataset;(c): IMDB dataset.

Figure 2a is an experiment in the ACM dataset, in which the network is relatively sparse, so methods that focus more on the local importance, such as DC, are slower in the first 10 iterations. On the other hand, the proposed method, LGC, and GSM all pay attention to both local and global influence and perform well in the experiment. The networks of the DBLP and IMDB datasets are denser, making it difficult to distinguish the performance of the methods in the first few iterations. The DBLP is the densest network, and under the same infection and recovery rates, the infection range of all methods exceeds 0.8, leading the other two datasets by about 0.3.

In order to further validate the effectiveness of our method, we conducted experiments on the information propagation ability of nodes in different ranking positions, as shown in Fig. 3. The top 20, middle 20, and bottom 20 nodes ranked by node importance were selected, as the node importance decreases, the information propagation ability of the nodes weakens, and both the propagation range and the propagation speed become smaller. Generally speaking, the propagation range and propagation speed of the top nodes are higher than those of the middle and bottom nodes. In addition, due to the different density of the networks, the difference in the propagation ability between the top 20 and bottom 20 nodes vary. For example, in the case of the ACM network, which

is a sparse network, the difference in the propagation ability between the top 20 nodes and the bottom 20 nodes is the largest, with a difference of 44.7%, while for the IMDB network, which is a dense network, the difference between the top 20 nodes and the bottom 20 nodes is the smallest, with a difference of 19.2%.

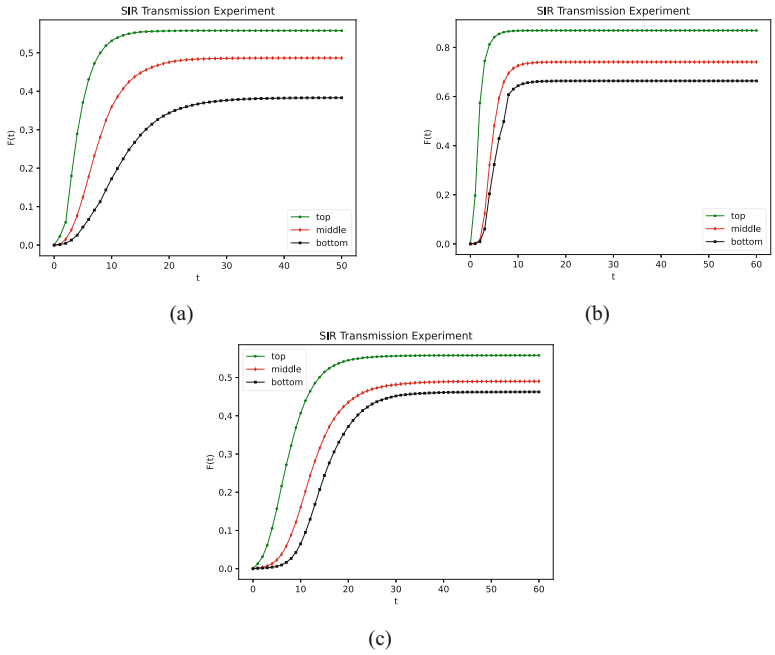


Fig. 3. The SIR transmission experiment of nodes with different ranking positions. (a): ACM dataset; (b): DBLP dataset; (c): IMDB dataset.

6 Conclusion

This paper applies node embedding with topological potential to the evaluation of node importance in heterogeneous networks, and proposes a node importance evaluation method called TAE-CI. The method evaluates the comprehensive influence of nodes both locally and globally, avoiding the partiality of single-index evaluation, and avoiding the time cost brought by computing the shortest path between nodes. As verified by SIR propagation experiments, the proposed method in this paper can better evaluate node importance.

Although the method proposed in this paper achieved better results in the evaluation of node importance in heterogeneous networks, the semantic information obtained by using metapaths in the node embedding model is not comprehensive enough. In the future, it is considered to introduce metagraphs to extract semantic information in order to further improve the accuracy of the evaluation method.

Acknowledgments. This paper is supported by the National Natural Science Foundation of China (62062050, 61962037), and the Innovation Foundation for Postgraduate Student of Jiangxi Province (YC2022-s727).

References

1. Maji, G., Mandal, S., Sen, S.: A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques. Expert Syst. Appl. **161**, 113681 (2020)
2. Yu, P., Fu, C., Yu, Y., et al.: Multiplex heterogeneous graph convolutional network. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2377–2387 (2022)
3. Li, M., Lu, Y., Wang, J., et al.: A topology potential-based method for identifying essential proteins from PPI networks. IEEE/ACM Trans. Comput. Biol. Bioinf. **12**(2), 372–383 (2014)
4. Du, J., Zhang, S., Wu, G., et al.: Topology adaptive graph convolutional networks. arXiv 2017[J]. arXiv preprint arXiv:1710.10370 (2017)
5. Ugurlu, O.: Comparative analysis of centrality measures for identifying critical nodes in complex networks. J. Comput. Sci. **62**, 101738 (2022)
6. Fan, C., Zeng, L., Ding, Y., et al.: Learning to identify high betweenness centrality nodes from scratch: a novel graph neural network approach. In: 2019 28th ACM International Conference on Information and Knowledge Management (CIKM), New York, USA, 559–568, 2019–11–03–2019–11–07 (2019)
7. Gleich, D.F.: PageRank beyond the web. siam Rev. **57**(3), 321–363 (2015)
8. Yang, X., Xiao, F.: An improved gravity model to identify influential nodes in complex networks based on k-shell method. Knowl.-Based Syst. **227**, 107198 (2021)
9. Kitsak, M., Gallos, L.K., Havlin, S., et al.: Identification of influential spreaders in complex networks L. Nat. Phys. **6**(11), 888–893 (2010)
10. Ullah, A., Wang, B., Sheng, J., et al.: Identification of nodes influence based on global structure model in complex networks. Sci. Rep. **11**(1), 6173 (2021)
11. Ullah, A., Wang, B., Sheng, J.F., et al.: Identifying vital nodes from local and global perspectives in complex networks. Expert Syst. Appl. **186**, 115778 (2021)
12. Molaei, S., Farahbakhsh, R., Salehi, M., et al.: Identifying influential nodes in heterogeneous networks. Expert Syst. Appl. **160**, 113580 (2020)
13. Wan, L., Zhang, M., Li, X., et al.: Identification of important nodes in multilayer heterogeneous networks incorporating multirelational information. IEEE Trans. Comput. Soc. Syst. **9**(6), 1715–1724 (2022)
14. Shetty, R.D., Bhattacharjee, S., Dutta, A., et al.: GSI: an influential node detection approach in heterogeneous network using covid-19 as use case. IEEE Trans. Comput. Soc. Syst. (2022)
15. Zhao, G., Jia, P., Zhou, A., et al.: InfGCN: identifying influential nodes in complex networks with graph convolutional networks. Neurocomputing **414**, 18–26 (2020)
16. Munikoti, S., Das, L., Natarajan, B.: Scalable graph neural network-based framework for identifying critical nodes and links in complex networks. Neurocomputing **468**, 211–221 (2022)
17. Li, Y., Li, L.L., Liu, Y.J., et al.: MAHE-IM: multiple aggregation of heterogeneous relation embedding for influence maximization on heterogeneous information network. Expert Syst. Appl. **202**, 117289 (2022)
18. Zhong, H., Wang, M., Zhang, X.: Unsupervised embedding learning for large-scale heterogeneous networks based on metapath graph sampling. Entropy **25**(2), 297 (2023)

19. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
20. Yang, Y., Guan, Z., Li, J., et al.: Interpretable and efficient heterogeneous graph convolutional network. IEEE Trans. Knowl. Data Eng. **35**(2), 1637–1650 (2021)
21. Fu, X., Zhang, J., Meng, Z., et al.: Magnn: metapath aggregated graph neural network for heterogeneous graph embedding. In: Proceedings of the Web Conference, pp. 2331–2341 (2020)
22. Zheng, C., Xia, C., Guo, Q., et al.: Interplay between SIR-based disease spreading and awareness diffusion on multiplex networks. J. Parallel Distrib. Comput. **115**, 20–28 (2018)

Research on Vibration Sensor Based on Cylindrical Resonator Structure

Haozhe Chen[1,2,3], Xiaojuan Zhang[1,2(✉)], and Xiaoxiao Xiang[1,2,3]

[1] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
xjzhang@aircas.ac.cn

[2] Key Laboratory of Electromagnetic Radiation and Sensing Technology, Chinese Academy of Sciences, Beijing, China

[3] School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

Abstract. In order to reduce the influence of co-frequency interference on the receiver in continuous wave radar remote micro motion detection, we design a new type of vibration sensor. The sensor is designed based on cylindrical resonator and TM_{010} and TE_{111} resonant modes. The relationship between the cavity geometry, tuning column geometry and the working frequency is studied by simulation. The results show that the sensor has good quality factor and antenna gain at two different frequencies, and can be used in the long-distance micro motion detection of continuous wave radar. Its work does not need power supply, which effectively improves the scope of its application scenarios.

Keywords: Micro motion detection · Co-frequency interference · Cylindrical resonator · Resonant mode

1 Introduction

With the development of modern microwave technology, in the life detection of earthquake collapse and other disaster rescue, or in the remote measurement of micro motion of engine, ship, spacecraft and other running surface, because the radar signal is not easily affected by environmental factors, and has strong penetration ability, the technology of using continuous wave radar to detect micro motion target has been widely used [1]. By using the continuous wave radar to transmit electromagnetic wave to the target, the frequency and phase of the echo will change due to the Doppler effect, which can reflect the movement rate of the distant object or analyze the vibration frequency of the micro moving object.

When using continuous wave radar to irradiate distant micro objects, if there is a close distance obstacle between antenna and object under test, it will lead to the signal with the same frequency as carrier in the received signal. These clutter interference with the same frequency will seriously affect the normal operation of the system, and cause the receiver sensitivity reduction, receiver saturation, flooded out of use signal and other problems. At present, the common technologies used to solve the same frequency

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 54–63, 2024.
https://doi.org/10.1007/978-3-031-59619-3_5

interference are clutter suppression technology based on digital signal processing [2], isolation technology based on receiving and receiving waveform control, etc. [3]. In this paper, a vibration sensor with different frequency is designed from the front-end sensor, which avoids the interference of the same frequency interference on the micro motion detection system of CW radar.

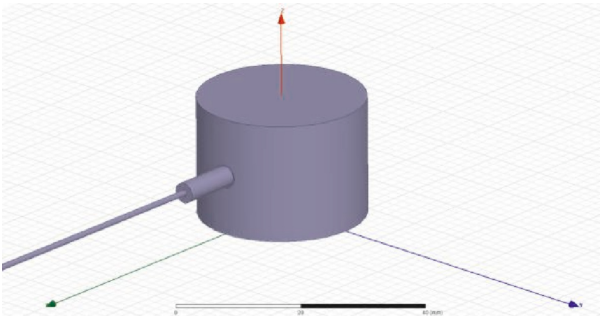


Fig. 1. Schematic diagram of resonant cavity vibration sensor.

2 Introduction

2.1 Principle of Resonant Cavity Vibration Sensor

The specific structure of the vibration sensor based on cylindrical resonator structure is shown in Fig. 1.

The vibration sensor based on cylindrical resonator is composed of cylindrical resonator and microwave transceiver antenna. When used, the resonant cavity is close to the object to be measured. When the object to be measured vibrates, the cavity wall of the resonant cavity will be slightly deformed and the resonant frequency will change slightly:

$$\frac{\omega - \omega_0}{\omega_0} = \frac{(\overline{w}_e - \overline{w}_m)\Delta V}{W} \tag{1}$$

Here, ω is the resonant frequency after the deformation of the cavity wall, and ω_0 is the resonant frequency before the deformation. $\overline{w}_e$ is the average energy density of electric field and the $\overline{w}_m$ is the average energy density of magnetic field, ΔV is the volume change before and after deformation. When the cavity wall is forced to deform into the cavity, $\Delta V < 0$. We select the lowest mode TM_{010} of the cylindrical resonator as the echo emission frequency of the sensor. When the main mode of the resonator is TM_{010} mode, there are strong electric field and weak magnetic field on the top and bottom of the resonator. At this time, $\overline{w}_e > \overline{w}_m$, the resonant frequency of the cavity decreases. This method is also called perturbation method [4].

2.2 Principle of Continuous Wave Remote Micro Motion Detection System

When a continuous wave is used to irradiate the micro antenna of the resonant cavity sensor, the received carrier wave has the same resonant frequency as the resonant cavity, which will excite different frequencies and modes of electromagnetic waves in the resonant cavity. When the measured object vibrates, the resonant frequency of the resonant cavity will decrease, and the carrier signal is modulated. The modulation mode includes both FM and AM, so the echo signal $S_m(t)$ is:

$$\begin{aligned} S_m(t) &= Acos(\omega_c t + \varphi(t)) \\ &= A_0\left[1 + k_a cos(2\pi f_m t)\right] cos\left(\omega_c t + k_f \int m(\tau) d\tau\right) \\ &= A_0\left[1 + k_a cos(2\pi f_m t)\right] cos\left(\omega_c t + k_f \int cos 2\pi f_m \tau d\tau\right) \end{aligned} \tag{2}$$

Here, k_a is the modulation degree of AM signal and k_f is the modulation degree of FM signal. f_m is the frequency information of the vibration of the measured object. ω_c is the carrier frequency of the echo signal. The carrier frequency here is not the same as that received by the resonator. The echo is transmitted through the micro antenna on the resonant cavity, amplified and received by the receiver, and then demodulated to recover the frequency information of the vibration of the measured object. The system block diagram of continuous wave remote micro motion detection is shown in Fig. 2.

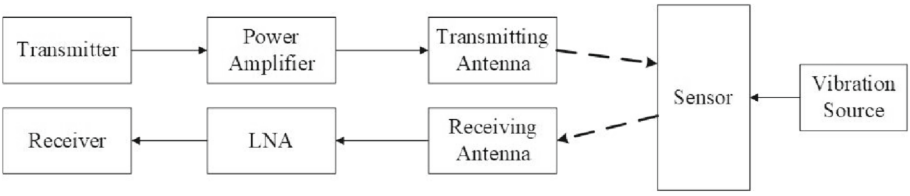


Fig. 2. Block diagram of continuous wave remote micro motion detection system.

2.3 Frequency Analysis

When a circular waveguide transmitting TM_{mn} mode or TE_{mn} mode is short circuited at $l = p\lambda_g/2$, where l is the length of the cylindrical waveguide and λ_g is the wavelength of the waveguide, we can get a cylindrical cavity with closed ends which can be used to transmit TM_{mnp} mode.

Here, m is the standing wave number of the field along the circumference, n is the semi standing wave number of the field along the radius, and p is the semi standing wave number of the field along the Z direction. Therefore, the resonant wavelength of the cylindrical resonator can be calculated by the cutoff wavelength λ_c corresponding to TM_{mn} mode and TE_{mn} mode:

$$\lambda_0 = \frac{1}{\sqrt{\left(\frac{1}{\lambda_c}\right)^2 + \left(\frac{p}{2l}\right)^2}} \tag{3}$$

The cutoff wavelengths of TE_{11}, TE_{01} and TM_{01} in circular waveguide are 3.41R, 2.62R and 1.64R respectively, where R is the radius of cylindrical cavity. Therefore, the resonant frequencies of corresponding oscillation modes TE_{111}, TE_{011} and TM_{010} are as follows:

$$\lambda_{TE_{111}} = \frac{1}{\sqrt{\left(\frac{1}{3.41R}\right)^2 + \left(\frac{1}{2l}\right)^2}} \tag{4}$$

$$\lambda_{TE_{011}} = \frac{1}{\sqrt{\left(\frac{1}{1.64R}\right)^2 + \left(\frac{1}{2l}\right)^2}} \tag{5}$$

$$\lambda_{TM_{010}} = 2.62R \tag{6}$$

According to the application requirements of the actual scene, we take $l = 4/3R$, and the lowest two modes of the resonator are TM_{010} mode and TE_{111} mode. Since the resonant wavelength of TM_{010} mode in the cylindrical cavity is $2.62R$, it is independent of the height l of the cavity. When tuning the cavity sensor, the resonant frequency of TM_{010} mode can be adjusted by adding a tuning column in the center of the cavity to change the electric field distribution of TM_{010} mode in the cavity.

3 Design and Simulation of the Resonant Cavity

When designing the resonator, the radius of the resonator should be calculated first. We set the receiving frequency of the resonator as 5 GHz and the transmitting frequency as 1 GHz. According to formula (6), the radius of the resonator is 114.5 mm. Since the sensor must be small in size, we set the radius of the cavity to 15 mm and the height of the cavity to 20 mm, and then adjust the resonant frequency of TM_{010} mode by adding a tuning column.

The resonant cavity used in this paper is made of gold with a wall thickness of 0.3 mm. The tuning structure consists of two parts, one is tuning column. The other part is a tuning disk with a thickness of 0.3 mm at the top of the tuning column, which is used to tune the lowest order mode f_1 and the second lowest order mode f_2 respectively. The specific structure is shown in Fig. 3.

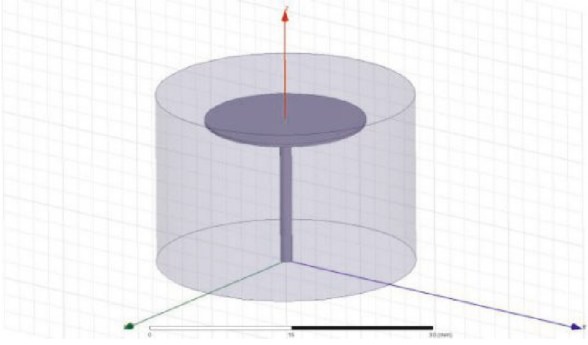


Fig. 3. Schematic diagram of the tuning structure.

When we design the size of the tuning structure, firstly we reduce the radius of the tuning disk to the same size as that of the tuning column, and then change the height of the tuning column with a certain step size to observe the variation of f_1 and f_2 with the height of the tuning column.

As shown in Fig. 4, the frequency of TM_{010} mode changes rapidly with height. Therefore, we first adjust the height of the tuning column to make f_2 close to the ideal value of 5 GHz, and then observe the variation of the resonant frequency of the two modes by adjusting the radius of the tuning dome.

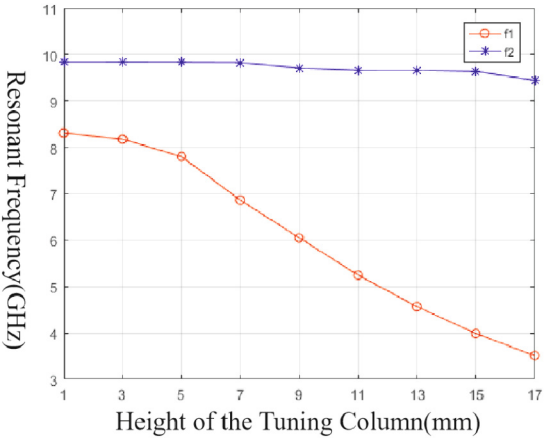


Fig. 4. The variation of two resonant frequencies with the height of the tuning column.

Due to the limitation of the height of the resonator, the tuning column can not be extended indefinitely. We take the height as 17 mm, and the frequencies of f_1 and f_2 are 3.5 GHz and 9.4 GHz respectively. After that, we gradually increase the radius of the tuning disk and get the variation law of f_1 and f_2 with the radius, as shown in Fig. 5.

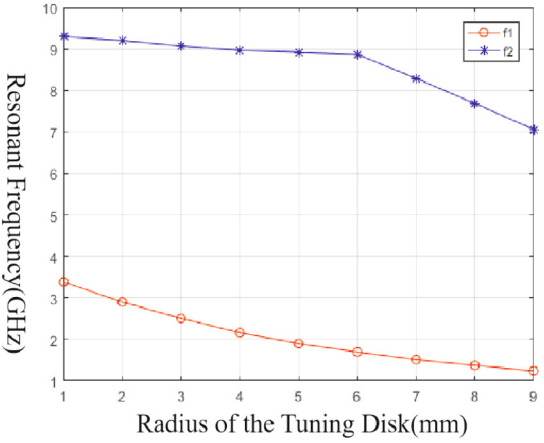


Fig. 5. The variation of two resonant frequencies with the radius of the tuning disk.

As shown in Fig. 5, by adjusting the radius of the resonant disk, we can further reduce the frequency of the lowest mode of the cylindrical resonator. f_1 varies greatly with radius, while f_2 is less affected by radius. Due to the limitation of the radius and height of the resonator, the size of the tuned structure can not be expanded infinitely. Therefore, by adjusting the tuning disk radius, we make f_1 approach the design frequency near 1 GHz, and finally determine that when the tuning column height and tuning disk radius are 18.3 mm and 9 mm respectively, the resonant frequencies of the two modes are 1.01 GHz and 7.69 GHz, as shown in Table 1.

Table 1. Resonant frequencies and quality factors of the lowest two modes of the resonator.

Eigenmode	Frequency(GHz)	Q
Mode 1	1.01	1325.06
Mode 2	7.69	3973.22

Due to the electric field component of TM_{010} mode is the largest at the central axis of the cavity, the resonant frequency of the lowest mode of the cylindrical cavity can be reduced from 7 GHz to 1 GHz by adding a tuning column. However, the resonant frequency of TE_{111} mode remains about 7.6 GHz. We can change the electric field distribution of TE_{111} mode by adding a tuning body on the side of the cylindrical resonator. The coupling effect between the probe and the cavity can be enhanced by adding a conductor disk at the top of the probe. By adjusting the size of the conductor disk, we can further adjust f_2. The specific process will be described in the next section.

4 Design and Simulation of the Coaxial Probe

4.1 Design of the Coaxial Antenna

The microwave transceiver antenna consists of coaxial antenna and coaxial probe. In order to meet the requirements of different transmitting and receiving frequencies, we need to design a dual band antenna. The simplest design of dual frequency antenna is to use the principle of frequency doubling. For the monopole antenna, if the fundamental frequency of the antenna is f_0, then the antenna will produce resonance at odd times of its fundamental frequency, such as $3f_0$, $5f_0$. At even times of the fundamental frequency, there is no resonance due to the harmonic cancellation effect. Since the receiving frequency is 5 times of the transmitting frequency, we can design a coaxial antenna with 1 GHz fundamental frequency to meet the requirements.

According to the design requirements of coaxial antenna, a coaxial cable with a length of $\lambda/2$ is peeled off the outer conductor and dielectric layer with a length of $\lambda/4$ to expose the inner core to form a coaxial antenna. Therefore, the conductor length is 75 mm. The characteristic impedance of the designed antenna is 50 Ω, Polyethylene

is used as the dielectric layer of the antenna. According to the calculation formula of antenna characteristic impedance:

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}} * ln\frac{A}{B} \tag{7}$$

Here, Z_0 is the characteristic impedance of the antenna, ε_r is the relative permittivity of the dielectric layer, B is the radius of the outer conductor, and A is the radius of the inner conductor. The ratio of the outer radius to the inner radius of the coaxial probe is 3.49. The return loss curve of the designed antenna is shown in Fig. 6. The return loss of the designed antenna is −14.7 dB at 1 GHz and −7 dB at 5 GHz, which has good return loss characteristics.

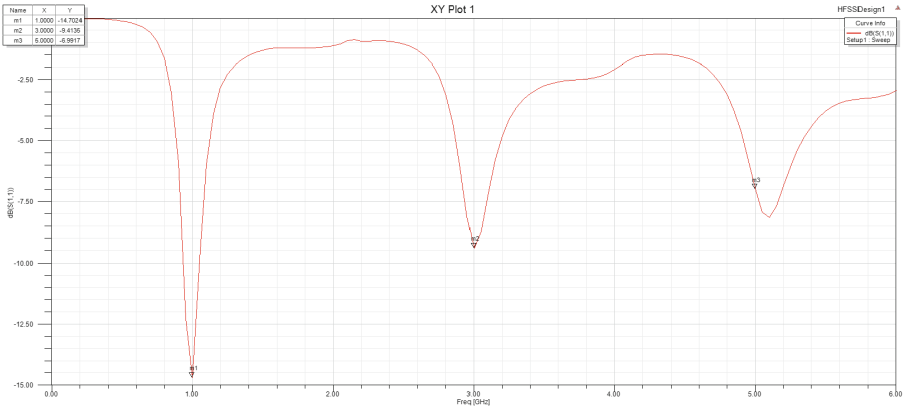


Fig. 6. Return loss curve of the coaxial antenna.

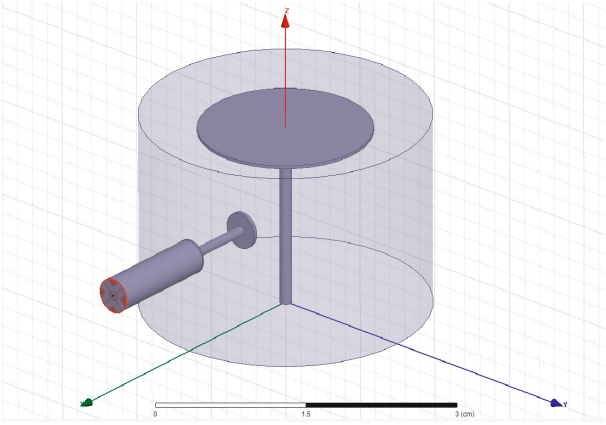


Fig. 7. Return loss curve of the coaxial antenna.

4.2 Design of the Coaxial Antenna

The model of coaxial probe feeding is shown in Fig. 7. The outer part of the coaxial probe is wrapped by polyethylene layer, and its radius is the same as that of the coaxial antenna. The inner core goes deep into the resonant cavity and can excite various modes in the resonant cavity [5, 6].

In the model of coaxial probe exciting the resonator, the two main factors that affect the exciting effect are the length d_1 of the probe deep into the resonator and the distance d_2 from the bottom of the resonator [7–10].

First, we placed the coaxial probe in the center of the chamber, then adjusted the d_1 to observe the return loss performance of the probe.

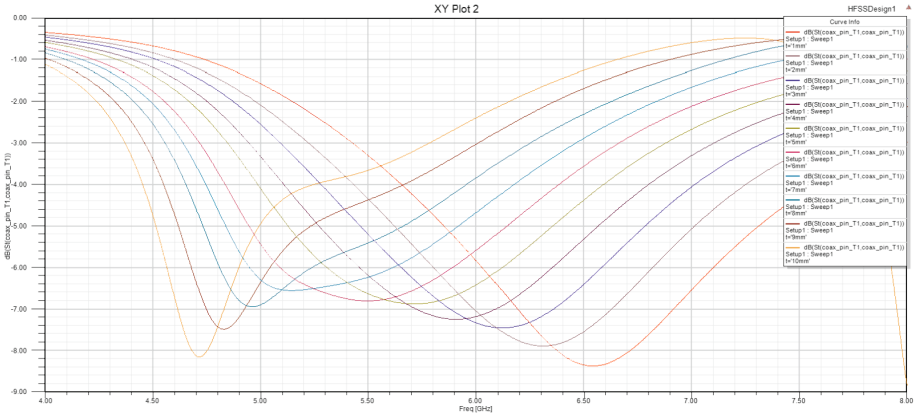


Fig. 8. Variation of return loss of probe with d_1.

As shown in Fig. 8, with the increase of the length of the probe into the resonator, the peak return loss of the probe has good performance when $d_1 = 8$ mm, and the peak return loss is −6.9 dB. Therefore, we probe the length of the probe into the cavity 8 mm, and then adjust the d_2 to observe the return loss performance of the probe.

As shown in Fig. 9, the peak return loss of the probe is −6.9 dB at the depth of 8 mm and the distance of 10 mm from the bottom of the resonator.

In addition, the coupling between the probe and the resonator can be enhanced by adding a conductor disk on the top of the probe, with a thickness of 0.4 mm. The length of the probe into the cavity and the radius of the conductor disk will affect the resonant frequency f_2 of the second lowest mode. We need to consider the coupling effect of the probe and the resonant effect of the resonator. Figure 10 shows the curves of the resonant frequencies of the two lowest modes varying with the radius of the conductor disk. By adjusting the radius of the conductor disk, we can further reduce the frequency of the second lowest order mode of the cylindrical resonator. The frequency f_2 of TE_{111} mode changes rapidly with the radius of the conductor disk, while the frequency f_1 of TM_{010} mode is less affected by the radius of the conductor disk. By adjusting the radius of the conductor disk, f_1 and f_2 can be adjusted to 1 GHz and 5 GHz. Finally, the length of the probe deep into the cavity is 8 mm, and the radius of the conductor disk is 1.83 mm.

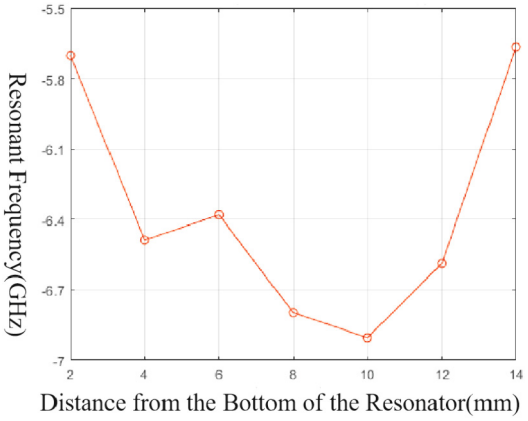


Fig. 9. Variation of return loss of probe with d_2.

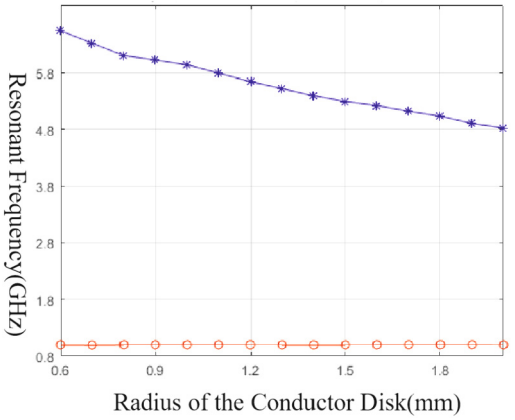


Fig. 10. The variation of two resonant frequencies with the radius of the conductor disk.

5 Summary

In this paper, a cylindrical resonator vibration sensor is designed, which can effectively avoid the influence of the same frequency interference on the performance of the receiver when applied to the continuous wave remote micro motion detection system. The resonant frequency of the sensor is related to the size of the tuning structure, the length of the probe into the resonant cavity, and the size of the conductor disk. Through software simulation, we finally determine that the resonant frequencies of the lowest two modes of the sensor designed in this paper correspond to 1 GHz and 5 GHz respectively. And the microwave transceiver antenna has a good response at the receiving frequency.

References

1. Zhang, Z., Yang, L., Pang, S.: Micro vibration dynamic environment analysis of high precision spacecraft. Spacecraft Environ. Eng. **6**, 528–534 (2009)
2. He, Y., Guan, J., Peng, Y.: Radar Automatic Detection and CFAR Processing. edtied by Tsinghua University Press, Beijing (1999). in press
3. Khan, R.H., Mitchell, D.K.: Waveform analysis for high-frequency FMICW radar. Radar Sig. Process. IEE Proceedings F **138**, 411–419 (1991)
4. Wang, D., Zhao, G.: Approximate calculation of TM_(010) mode in cylindrical resonator. J. Shaanxi Normal Univ. (Nat. Sci. Edition), vol.3, pp. 38–42 (1987)
5. Zhao, F., Pei, J., Pei, X., et al.: Resonant frequency calibration of complex permittivity measurement system by resonant cavity method. Aerosp. Meas. Technol. **39**(S1), 19–22 (2019)
6. Xie, Y., Liang, C.: Resonant frequency analysis of probe coupled resonator. J. Xi'an Univ. Electron. Sci. Technol. **2**, 263–265 (1996)
7. Gao, Y., Mao, C., La, Y., et al.: Study on mode suppression of microwave chemical resonator. J. Microwave **35**(05), 90–95 (2019)
8. You, Y., Lan, Z., Gao, Y., et al.: Glucose solution concentration measurement system based on microwave resonator perturbation method. Magn. Mater. Devices **4**, 38–41 (2020)
9. Yan, X., Zhao, L., Liu, F., et al.: Simulation design of rectangular waveguide to microwave resonator coaxial antenna. Vacuum Electron. Technol. **3** (2019)
10. Qian, J., Yan, X., Han, Z., et al.: Theoretical analysis of measuring liquid film thickness with microwave coaxial resonator. Steam Turbine Technol. **000**(002), 89–92 (2015)

Iterative Decision-Feedback Hybrid Equalization for CP-OTFS on Time-Varying Multipath Channels

Shuen-Yu Tsai[1], Po-Jen Chen[2(✉)], Wei-Chang Chen[3], and Char-Dir Chung[2]

[1] Communications Network Group, Realtek Inc, Hsinchu, Taiwan
[2] Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
r11942048@ntu.edu.tw
[3] Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan

Abstract. Cyclic-prefix orthogonal time frequency space (CP-OTFS) systems have gained great attention recently due to achievable prominent error performance characteristics in high-mobility wireless communication scenarios and the system realizability by use of efficient Discrete Fourier transform. In this paper, a symbol decision scheme is proposed for the CP-OTFS system by virtue of iterative decision-feedback hybrid equalization (IDFHE) using time-domain linear minimum mean-square-error (LMMSE) and constrained least square (CLS) equalizers at different iterations. The proposed IDFHE scheme can exploit the advantage that intersymbol interference is suppressed progressively over iterations and noise power enhancement is avoided at the final iteration, and thereby enhance the error performance. From simulation results, the proposed IDFHE scheme is shown to outperform non-iterative LMMSE decision scheme and iterative message-passing symbol decision scheme in average error performance on time-varying Rayleigh multipath channels.

Keywords: Orthogonal time frequency space · cyclic prefix · decision-feedback equalization · time-varying multipath channel

1 Introduction

Cyclic-prefix orthogonal time frequency space (CP-OTFS) modulation is a promising candidate for supporting reliable data transmission in high-mobility wireless communication scenarios [1–5]. Since data symbols are multiplexed on the delay-Doppler (DD) grid and then transformed onto the time-frequency (TF) grid by the inverse symplectic finite Fourier transform (ISFFT), CP-OTFS can be regarded as cyclic-prefix single carrier (CP-SC) modulation with inter-block inverse discrete-Fourier-transform (DFT) precoding. Therefore, CP-OTFS modulator can be efficiently realized by adding intra-block DFT precoding and inter-block inverse DFT precoding prior to cyclic-prefix orthogonal frequency division multiplexing (CP-OFDM) modulator. Due to such precoding structure, CP-OTFS can exploit joint TF diversity and is more robust to CP-OFDM on

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 64–76, 2024.
https://doi.org/10.1007/978-3-031-59619-3_6

static and time-varying multipath channels [1–10]. Like CP-OFDM, efficient DFT-based implementation is applicable to realize CP-OTFS systems.

Several symbol decision schemes have been studied for CP-OTFS systems, including iterative message passing algorithm (MPA) [6, 7], non-iterative linear minimum mean-square-error (LMMSE) equalization [8, 9], non-iterative zero-forcing (ZF) equalization [9], and iterative LMMSE equalization [10]. Although the ZF scheme can remove the intersymbol interference (ISI), it suffers from severe equalization-induced noise power enhancement and thus is outperformed by the other schemes. By exploiting the channel sparsity, the iterative MPA scheme and its variants were designed to achieve approximate maximum a posteriori probability (MAP) decision with affordable implementation complexity, and can perform comparably with the non-iterative LMMSE scheme when the channel response is sparse on the DD grid [6, 7]. However, when the channel response consists of a large number of resolvable paths or exhibits fractional Doppler shifts, MPA suffers from strong ISI and thus entails high error floor [7, 10]. In [8, 9], the non-iterative LMMSE scheme was shown to suffer from equalization-induced residual ISI and noise power enhancement jointly. To mitigate residual ISI, the iterative LMMSE scheme employing decision-feedback equalization (DFE) is shown in [10] to outperform the non-iterative LMMSE scheme in symbol decision on static multipath channels. However, since the identical LMMSE equalizer is adopted at all iterations in the iterative LMMSE scheme, the same level of noise power enhancement is suffered in all iterations [10]. This leaves the room for performance enhancement by reducing noise power enhancement in later iterations when residual ISI is largely suppressed by DFE.

In [10], the iterative decision-feedback hybrid equalization (IDFHE) method was recently proposed for CP-OTFS to enhance symbol decision performance on static multipath channels. By adopting hybrid equalizers in different iterations, IDFHE reduces residual ISI progressively over iterations and avoids noise power enhancement at the final iteration. Although the IDFHE scheme is shown to outperform MPA scheme and non-iterative LMMSE schemes in error performance on static multipath channels, the IDFHE scheme in [10] is not directly applicable to CP-OTFS transmission over time-varying multipath channels. This is because all block signals in a CP-OTFS frame experience the same channel response on static multipath channels but different channel responses on time-varying multipath channels. This paper is thus motivated to devise IDFHE schemes for CP-OTFS transmission over time-varying multipath channels.

In this paper, a symbol decision scheme is proposed for CP-OTFS transmission on time-varying multipath channels by means of IDFHE using time-domain LMMSE equalization and constrained least square (CLS) equalization [11] at different iterations. The proposed IDFHE scheme is shown to outperform MPA scheme and non-iterative LMMSE scheme in average error performance on time-varying Rayleigh multipath channels with a wide mobility range. The rest of the paper is organized as follows. Section 2 models the CP-OTFS system on time-varying multipath channels. The IDFHE scheme

is described in Sect. 3 and demonstrated for performance characteristics in Sect. 4. Section 5 concludes this paper.[1]

2 CP-OTFS System Model

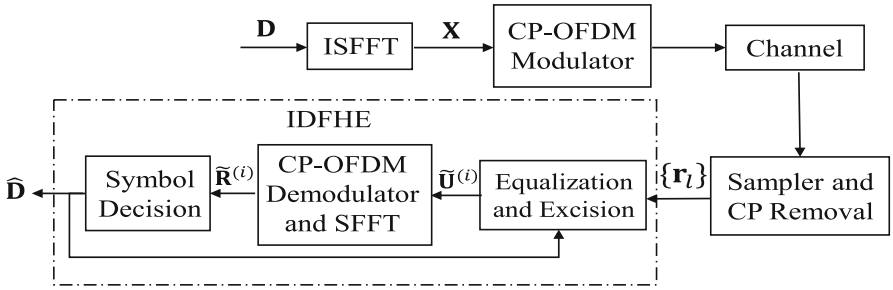


Fig. 1. The CP-OTFS system.

2.1 Transmitted Signal Model

Figure 1 illustrates the considered CP-OTFS system. In the nominal frame interval, the source produces a data matrix $\mathbf{D} \triangleq \left[d_{m,n}; m \in \mathcal{Z}_M, n \in \mathcal{Z}_N\right]$ containing NM complex-valued data symbols on the DD grid. All data symbols $d_{m,n}$'s are independent and identically distributed (i.i.d.) with $\mathcal{E}\left[d_{m,n}\right] = 0$, $\mathcal{E}\left[d_{m,n}^2\right] = 0$, and $\mathcal{E}\left[\left|d_{m,n}\right|^2\right] = 1$.

The CP-OTFS modulator realizes a cascade of the ISFFT and the CP-OFDM modulation. First, the ISFFT takes $\mathbf{D}$ as the input matrix discretized on the DD grid $\Lambda_{\mathrm{DD}} \triangleq \left\{\left(\frac{m}{M\Delta f}, \frac{n}{NT_{\mathrm{d}}}\right); m \in \mathcal{Z}_M, n \in \mathcal{Z}_N\right\}$, where $\frac{1}{M\Delta f}$ and $\frac{1}{NT_{\mathrm{d}}}$ are respectively the quantization steps in delay and Doppler, and produces the output matrix $\mathbf{X} \triangleq \left[x_{k,l}; k \in \mathcal{Z}_M, l \in \mathcal{Z}_N\right]$ discretized on the TF grid $\Lambda_{\mathrm{TF}} \triangleq \{(k\Delta f, lT_{\mathrm{d}}); k \in \mathcal{Z}_M, l \in \mathcal{Z}_N\}$, where Δf and T_{d} are respectively the quantization steps in frequency and time, as $\mathbf{X} = \mathbf{W}_M \mathbf{D} \mathbf{W}_N^h$. Next, all columns in $\mathbf{X}$ are successively modulated by the CP-OFDM modulator to produce a frame of N rectangularly-pulsed OFDM block waveforms on M Δf-spaced subcarriers. Before cyclic prefixes are inserted, the produced OTFS signaling matrix is represented by $\mathbf{S} \triangleq \mathbf{W}_M^h \mathbf{X} = \left[\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\right]$. For

[1] *Notations*: Boldface lower-case and upper-case letters denote column vectors and matrices, respectively. Superscripts t and h denote transpose and conjugate transpose, respectively. $\mathcal{Z}_K^+$ and $\mathcal{Z}_K$ denote the integer sets $\{1, 2, \ldots, K\}$ and $\{0, 1, \ldots, K-1\}$, respectively. $\mathbf{X}$ is the Frobenius norm of matrix $\mathbf{X}$. 0_K, $\mathbf{O}_{M,N}$ and $\mathbf{I}_K$ are a $K \times 1$ all-zero vector, an $M \times N$ all-zero matrix and a $K \times K$ identity matrix, respectively. $\left[x_{m,n}; m \in \mathcal{Z}_M, n \in \mathcal{Z}_N\right]$ is an $M \times N$ matrix having $x_{m,n}$ as the (m, n)-th entry and $\mathbf{x}_n \triangleq \left[x_{m,n}; m \in \mathcal{Z}_M\right]$ as the n-th column. $[\mathbf{x}_n; n \in \mathcal{Z}_N]$ is a matrix having $\mathbf{x}_n$ as the n-th column, and $\mathrm{diag}\left(\left[x_k; k \in \mathcal{Z}_K\right]\right)$ a $K \times K$ diagonal matrix having x_k as the k-th diagonal entry. $j \triangleq \sqrt{-1}$ is the imaginary unit. $\mathbf{W}_M$ is the unitary DFT matrix having $M^{-1/2} \exp\{-j2\pi nm/M\}$ as the (n, m)-th entry. $\mathcal{E}[\cdot]$ is the expectation operator.

each $\mathbf{s}_n$, N_{CP} CP symbols are inserted ahead of M samples in $\mathbf{s}_n$ to form $M + N_{CP}$ symbols transmitted in the n-th CP-OFDM block waveform over a block interval length $T = T_{\mathrm{g}} + T_d$, where T_{d} is the useful block subinterval length and $T_{\mathrm{g}} = \frac{N_{CP}T_d}{M}$ is the CP subinterval length. The transmitted CP-OTFS frame waveform extends over a total duration of length NT and is composed of N contiguous M-subcarrier CP-OFDM block waveforms. Throughout, $\Delta f T_d = 1$ is constrained to maintain the orthogonality among M subcarriers.

2.2 Received Signal Model

Consider the coherent demodulation of the CP-OTFS signal received over the time-varying multipath channel with L paths, where the ε-th path has complex response h_ε and induces a delay τ_ε and a Doppler frequency ν_ε with $0 \le \tau_\varepsilon \le \tau_{\max}$ and $0 \le |\nu_\varepsilon| \le \nu_{\max}$, $\varepsilon \in \mathcal{Z}_L$. The maximum delay $\tau_{\max}$ is assumed to be not longer than T_{g} so that ISI over useful blocks can be effectively rejected at the synchronous receiver.

In most high-mobility environments, the Doppler-variant channel response results from only few reflectors moving within one frame duration [1]. To avoid ambiguous identification of transmitted symbols on the TF grid, the useful block subinterval T_{d} and the frequency spacing Δf are chosen to be much larger than $\tau_{\max}$ and $2\nu_{\max}$, respectively. Within such parametric setup, τ_ε's and ν_ε's can be further specified by

$$\tau_\varepsilon = \frac{l_\varepsilon}{M\,\Delta f} \text{and } \nu_\varepsilon = \frac{\kappa_\varepsilon}{NT_{\mathrm{d}}} \tag{1}$$

where delay shifts l_ε's are integer-valued, Doppler shifts κ_ε's are real-valued, and they are limited respectively to $0 \le l_\varepsilon \le M\,\Delta f\,\tau_{\max}$ and $|\kappa_\varepsilon| \le NT_{\mathrm{d}}\nu_{\max}$. Notably, l_ε and $|\kappa_\varepsilon|$ are constrained to be much smaller than M and $N/2$, respectively.

When the received waveform is perfectly synchronized in timing and frequency, the receiver converts the received signal down to baseband, discards block CPs, and then samples the received baseband waveform at rate M/T_d. In this case, the l-th received baseband block signal can be modeled as [2]

$$\mathbf{r}_l = \rho \mathbf{H}_l \mathbf{s}_l + \mathbf{z}_l \tag{2}$$

with ρ an amplitude factor. Here, $\mathbf{z}_l$ contains i.i.d. circularly-symmetric complex Gaussian (CSCG) noise samples with $\mathcal{E}[\mathbf{z}_l] = 0_M$, $\mathcal{E}\left[\mathbf{z}_l \mathbf{z}_l^t\right] = \mathbf{O}_{M,M}$ and $\mathcal{E}\left[\mathbf{z}_l \mathbf{z}_l^h\right] = \sigma^2 \mathbf{I}_M$. $\mathbf{H}_l$ is the equivalent l-th channel matrix in time domain, given by [2]

$$\mathbf{H}_l = \sum\nolimits_{\varepsilon=0}^{L-1} h_\varepsilon \mathbf{\Delta}_{l,l_\varepsilon,\kappa_\varepsilon} \mathbf{\Pi}^{l_\varepsilon} \tag{3}$$

where $\mathbf{\Delta}_{l,l_\varepsilon,\kappa_\varepsilon} \triangleq \mathrm{diag}([\omega^{\kappa_\varepsilon[(M+N_{CP})l+N_{CP}-l_\varepsilon+m]}; m \in \mathcal{Z}_M])$ is a phase-shifting matrix with $\omega = e^{\frac{j2\pi}{(M+N_{CP})N}}$ and $\mathbf{\Pi}$ is a permutation matrix which shifts all rows in $\mathbf{I}_M$ downward by one position circularly.

2.3 IDFHE Receiver

The IDFHE receiver iteratively processes the received blocks $\{\mathbf{r}_l; l \in \mathcal{Z}_N\}$ under the *a priori* knowledge of channel matrices $\{\mathbf{H}_l; l \in \mathcal{Z}_N\}$, and produces a decision matrix $\widehat{\mathbf{D}}$ in the end. The process can be divided into three steps in each iteration: *equalization and ISI excision*, *DFT and SFFT*, and *symbol decision.*

At the first iteration, the IDFHE receiver only conducts time-domain equalization by the equalizers $\mathbf{E}_l^{(1)}$ to produce $\overset{\sim}{\mathbf{u}}_l^{(1)} \triangleq \rho^{-1}\mathbf{E}_l^{(1)}\mathbf{r}_l$ for all $l \in \mathcal{Z}_N$, and forms $\overset{\sim}{\mathbf{U}}^{(1)} \triangleq \left[\overset{\sim}{\mathbf{u}}_0^{(1)}, \overset{\sim}{\mathbf{u}}_1^{(1)}, \ldots, \overset{\sim}{\mathbf{u}}_{N-1}^{(1)}\right]$. At the remaining iterations, say the i-th iteration, equalization is first conducted by $\mathbf{E}_l^{(i)}$ for all $l \in \mathcal{Z}_N$ and the DFE-assisted ISI excision is then executed with the aid of the i-th bias matrices $\mathbf{Y}_l^{(i)} \triangleq \mathbf{E}_l^{(i)}\mathbf{H}_l - \mathbf{I}_M$ for all $l \in \mathcal{Z}_N$. Before ISI excision, the k-th symbol in the l-th equalized block $\rho^{-1}\mathbf{E}_l^{(i)}\mathbf{r}_l \triangleq [u_{k,l}^{(i)}; k \in \mathcal{Z}_M]$ is given by

$$u_{k,l}^{(i)} = \left(1 + y_{k,k,l}^{(i)}\right)s_{k,l} + \sum_{p\in\mathcal{Z}_M, p\neq k} y_{k,p,l}^{(i)} s_{p,l} + \tilde{z}_{k,l}^{(i)} \tag{4}$$

where $y_{k,p,l}^{(i)}$ represents the (k,p)-th entry of $\mathbf{Y}_l^{(i)}$. In (4), $\left(1 + y_{k,k,l}\right)s_{k,l}$, $\sum_{p\in\mathcal{Z}_M, p\neq k} y_{k,p,l}^{(i)} s_{p,l}$, and $\tilde{z}_{k,l}^{(i)}$ are the desired signal term, the ISI term and the equalized noise term, respectively. Each $\tilde{z}_{k,l}^{(i)}$ is a CSCG noise sample having mean zero and variance $\mathcal{E}\left[\left|\tilde{z}_{k,l}^{(i)}\right|^2\right] = \gamma_{\mathrm{d}}^{-1}\sum_{p=0}^{M-1}\left|e_{k,p,l}^{(i)}\right|^2$, where $e_{k,p,l}^{(i)}$ denotes the (k,p)-th entry of $\mathbf{E}_l^{(i)}$ and $\gamma_{\mathrm{d}} \triangleq \rho^2/\sigma^2$ is a signal-to-noise power ratio (SNR). To make reliable symbol decision, the ISI term in $u_{k,l}^{(i)}$ is excised with the aid of the decision $\widehat{\mathbf{D}}^{(i-1)}$ fed back from the previous iteration, and this yields $\tilde{u}_{k,l}^{(i)}$. After the first step in the i-th iteration, the joint DFT and SFFT process transforms the ISI-excised matrix $\overset{\sim}{\mathbf{U}}^{(i)} \triangleq [\tilde{u}_{k,l}^{(i)}; k \in \mathcal{Z}_M, l \in \mathcal{Z}_N]$ to

$$\overset{\sim}{\mathbf{R}}^{(i)} = \mathbf{W}_M^h\left(\mathbf{W}_M\overset{\sim}{\mathbf{U}}^{(i)}\right)\mathbf{W}_N \tag{5}$$

on receive Λ_{DD}. Lastly, the tentative symbol decision is made by applying decision function $F(\cdot)$ to each entry in $\overset{\sim}{\mathbf{R}}^{(i)}$, where $F(\cdot)$ realizes the minimum Euclidean distance rule [12]. Notably, with less ISI in time domain, tentative symbol decision can be reliably conducted based on $\overset{\sim}{\mathbf{R}}^{(i)}$. The tentative decision $\widehat{\mathbf{D}}^{(i)}$ is then fed back to next iteration for ISI excision. This iterative process terminates when I iterations are completed. The IDFHE algorithm is detailed in the following.

3 Iterative Decision-Feedback Hybrid Equalization Algorithm

The IDFHE algorithm that employs LMMSE and CLS equalizers in different iterations is considered here. Before introducing IDFHE, non-iterative LMMSE and CLS equalizers are first briefed.

3.1 LMMSE and CLS Equalizers

The LMMSE equalizer $\mathbf{E}_{l,\mathrm{LM}}$ minimizes the mean square error between $\mathbf{r}_l$ and $\rho\mathbf{s}_l$ for each l. Since $\mathcal{E}\left[\mathbf{s}_l\mathbf{s}_l^h\right] = \mathbf{I}_M$, $\mathbf{E}_{l,\mathrm{LM}}$ is given by

$$\mathbf{E}_{l,\mathrm{LM}} = \mathbf{H}_l^h\left(\mathbf{H}_l\mathbf{H}_l^h + \gamma_d^{-1}\mathbf{I}_M\right)^{-1} \tag{6}$$

Since $\mathbf{H}_l$ is a banded matrix, the low-complexity LMMSE method in [8] can be employed to compute $\rho^{-1}\mathbf{E}_{l,\mathrm{LM}}\mathbf{r}_l$.

The CLS equalizer $\mathbf{E}_{l,\mathrm{CLS}}$ aims to provide nearly unbiased equalized symbols without enhancing noise power in equalized noise samples, by following the criterion $\min_{\mathbf{E}} \|\mathbf{E}_l\mathbf{H}_l - \mathbf{I}_M\|^2 s.t. \mathbf{E}_l\mathbf{E}_l^h = \mathbf{I}_M$. The solution is obtained after applying the singular value decomposition (SVD) on $\mathbf{H}_l = \boldsymbol{\phi}_l\boldsymbol{\Sigma}_l\boldsymbol{\Omega}_l^h$ [11], as

$$\mathbf{E}_{l,\mathrm{CLS}} = \boldsymbol{\Omega}_l\boldsymbol{\phi}_l^h \tag{7}$$

where $\boldsymbol{\phi}_l$ and $\boldsymbol{\Omega}_l$ are unitary matrices and $\boldsymbol{\Sigma}_l$ is a diagonal matrix whose diagonal entries are the singular values in descending order. Notably, SVD requires high computational complexity and is a major tradeoff for using $\mathbf{E}_{l,\mathrm{CLS}}$.

3.2 IDFHE Scheme

The IDFHE concept appeared first in [11] for symbol decision in CP-SC transmission over static multipath channels. In [11], various non-iterative equalization criteria are used over iterations to cope with iteratively reduced residual ISI effectively while CLS equalization is adopted at the final iteration to avoid noise power enhancement and thus provide error performance prevalence. Motivated by the concept in [11], two IDFHE schemes were also proposed in [10] for CP-OTFS to enhance error performance on static multipath channels. In this paper, we consider the LM-IDFHE scheme proposed in [10], as follows.

In the first $I-1$ iterations, the LMMSE equalizer is adopted in the considered I-step IDFHE scheme because it can provide tentatively smaller error rate than ZF and CLS equalizers when ISI excision is not applied. The DFE-assisted ISI excision is executed in the last $I-1$ iterations to suppress ISI successively. Since the effect of ISI can be suppressed effectively after successive DFE-assisted ISI excisions, the CLS equalizer is used at the final iteration due to the avoidance of noise power enhancement. Such IDFHE scheme can cope with iteratively reduced residual ISI over iterations and avoid noise power enhancement when making final decision.

The I-step IDFHE algorithm in Fig. 2 operates on $\{\mathbf{r}_l; l \in \mathcal{Z}_N\}$ and is detailed below.

Initialization: The total iteration times I and the iteration count $i = 0$ are set *a priori*. For each l, $\mathbf{E}_{l,\mathrm{LM}}$ and $\mathbf{E}_{l,\mathrm{CLS}}$ are computed when the l-th channel matrix $\mathbf{H}_l$ is given. Accordingly, the equalization matrices at various iterations are set to $\mathbf{E}_l^{(1)} = \mathbf{E}_l^{(2)} = \cdots = \mathbf{E}_l^{(I-1)} = \mathbf{E}_{l,\mathrm{LM}}$ and $\mathbf{E}_l^{(I)} = \mathbf{E}_{l,\mathrm{CLS}}$ for $l \in \mathcal{Z}_N$. Then, the iterative process is launched.

Update: The iteration count i is set to $i+1$. If $i > 1$, compute $\widehat{\mathbf{S}}^{(i-1)} \triangleq \left[\hat{s}_{k,l}^{(i-1)}; k \in \mathcal{Z}_M, l \in \mathcal{Z}_N\right]$ as $\widehat{\mathbf{S}}^{(i-1)} = \widehat{\mathbf{D}}^{(i-1)}\mathbf{W}_N^h$.

Equalization and ISI Excision: For $i = 1$, only the equalization part is conducted, i.e., $\overset{\sim}{\mathbf{u}}_l^{(1)} \triangleq \rho^{-1}\mathbf{E}_l^{(1)}\mathbf{r}_l$ for all $l \in \mathcal{Z}_N$. Otherwise, this equalization is first performed and the DFE-assisted ISI excision is then executed. As discussed previously, the ISI term $\sum_{p\in\mathcal{Z}_M, p\neq k} y_{k,p,l}^{(i)} s_{p,l}$ in (4) needs to be removed from the equalized symbol $u_{k,l}^{(i)}$. Since the desired signaling matrix $\mathbf{S}$ is unknown, the removal is done with the aid of $\widehat{\mathbf{S}}^{(i-1)}$. Thus, the ISI-excised symbol $\tilde{u}_{k,l}^{(i)}$ can be obtained as $\tilde{u}_{k,l}^{(i)} = u_{k,l}^{(i)} - \sum_{p\in\mathcal{Z}_M, p\neq k} y_{k,p,l}^{(i)} \hat{s}_{p,l}^{(i-1)}$ symbol by symbol. With $\overset{\sim}{\mathbf{u}}_l^{(i)} \triangleq [\tilde{u}_{k,l}^{(i)}; k \in \mathcal{Z}_M]$, $\hat{\mathbf{s}}_l^{(i-1)} \triangleq [\hat{s}_{k,l}^{(i-1)}; k \in \mathcal{Z}_M]$, and (4), this ISI excision process can be neatly expressed in vector form as

$$\overset{\sim}{\mathbf{u}}_l^{(i)} = \rho^{-1}\mathbf{E}_l^{(i)}\mathbf{r}_l^{(i)} - \overline{\mathbf{Y}}_l^{(i)}\hat{\mathbf{s}}_l^{(i-1)}, \forall l \in \mathcal{Z}_N \tag{8}$$

where $\overline{\mathbf{Y}}_l^{(i)}$ is obtained from $\mathbf{Y}_l^{(i)}$ after setting all diagonal entries to zeros.

DFT and SFFT: The joint DFT and SFFT process in (5) transforms $\overset{\sim}{\mathbf{U}}^{(i)} \triangleq \left[\overset{\sim}{\mathbf{u}}_l^{(i)}; l \in \mathcal{Z}_N\right]$ in time domain to $\overset{\sim}{\mathbf{R}}^{(i)} \triangleq \left[\tilde{r}_{m,n}^{(i)}; m \in \mathcal{Z}_M, n \in \mathcal{Z}_N\right]$ on receive Λ_{DD} as $\overset{\sim}{\mathbf{R}}^{(i)} = \overset{\sim}{\mathbf{U}}^{(i)}\mathbf{W}_N$.

Decision: At the i-th iteration, the tentative decision $\widehat{\mathbf{D}}^{(i)} \triangleq \left[\hat{d}_{m,n}^{(i)}; m \in \mathcal{Z}_M, n \in \mathcal{Z}_N\right]$ is obtained from $\overset{\sim}{\mathbf{R}}^{(i)}$ by applying $\hat{d}_{m,n}^{(i)} = F(\tilde{r}_{m,n}^{(i)})$ for each $m \in \mathcal{Z}_M$ and $n \in \mathcal{Z}_N$.

Stopping Criterion: Go to **Finalization** if $i = I$ and return back to **Update** otherwise.

Finalization: The final decision is made as $\widehat{\mathbf{D}} = \widehat{\mathbf{D}}^{(I)}$.

4 Performance Results

In this section, performance results are obtained by simulation for various CP-OTFS symbol decision schemes using the system and channel parameters in Table 1 over time-varying random multipath channels with different mobility conditions. The carrier frequency $f_c = 3.5$ GHz and the subcarrier spacing $\Delta f = 60$ kHz are set as per n78 band's specification in the 5G NR standard [13]. The maximum delay shift is set to $l_{\max} = 7$. The maximum vehicular speeds are set to be 57.8, 231, and 578 in Kmph, leading to maximum Doppler shifts $\kappa_{\max} = 0.1, 0.4,$ and 1, respectively. The number of channel paths and the number of dominant path groups are L and L_d, respectively. With $L_d \leq L$, the channel paths with the identical delay shift $l_d^{(i)}$ belong to the i-th dominant group for $i \in \mathcal{Z}_{L_d}$, which may contain single or multiple subpaths with different Doppler shifts and arbitrary path responses, because of the isotropic reflection in mobile environments. The delay shifts $l_d^{(i)}$ for L_d dominant path groups are arranged

Algorithm IDFHE

Input: Received signal vector $\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{N-1}$, channel matrix $\mathbf{H}_0$, $\mathbf{H}_1, \ldots, \mathbf{H}_{N-1}$.

Initialization: Set iteration count $i = 0$, iteration times I. Compute $\mathbf{E}_{l,LM}$ and $\mathbf{E}_{l,CLS}$ by $\mathbf{H}_l$, and set $\mathbf{E}_l^{(1)} = \cdots = \mathbf{E}_l^{(I-1)} = \mathbf{E}_{l,LM}$, $\mathbf{E}_l^{(I)} = \mathbf{E}_{l,CLS}$, for $l = 0,1, \ldots, N-1$.

repeat

- **Update**: $i = i + 1$. And let $\hat{\mathbf{S}}^{(i-1)} = \widehat{\mathbf{D}}^{(i-1)}\mathbf{W}_N^h$ if $i > 1$.
- **Equalization and ISI Excision**:
  - For $i = 1$, $\tilde{\boldsymbol{u}}_l^{(1)} = \frac{1}{\rho}\mathbf{E}_l^{(1)}\mathbf{r}_l$, for $l = 0,1, \ldots, N-1$.
  - For $i > 1$, compute $\mathbf{Y}_l^{(i)} = \mathbf{E}_l^{(i)}\mathbf{H}_l - \mathbf{I}_M$ and $\tilde{\boldsymbol{u}}_l^{(i)} = \frac{1}{\rho}\mathbf{E}_l^{(i)}\mathbf{r}_l - \bar{\mathbf{Y}}_l^{(i)}\hat{\mathbf{s}}_l^{(i-1)}$, for $l = 0,1, \ldots, N-1$.
- **DFT and SFFT**: $\widetilde{\mathbf{R}}^{(i)} \triangleq \widetilde{\mathbf{U}}^{(i)}\mathbf{W}_N$.
- **Decision**: $\hat{d}_{m,n}^{(i)} = F(\tilde{r}_{m,n}^{(i)})$ for $m \in \mathcal{Z}_M$, $n \in \mathcal{Z}_N$ to form $\widehat{\mathbf{D}}^{(i)}$.

until **Stopping Criterion**

Finalization: $\widehat{\mathbf{D}} = \widehat{\mathbf{D}}^{(I)}$

Output: Data decision matrix $\widehat{\mathbf{D}}$

Fig. 2. The IDFHE algorithm.

in the ascending order $l_d^{(0)} < l_d^{(1)} < \ldots < l_d^{(L_d-1)}$ and drawn without repetition from $\mathcal{Z}_{l_{\max}+1}$. For the i-th dominant group, there are $L_d^{(i)}$ subpaths with identical delay shift $l_d^{(i)}$ and different Doppler shifts $\kappa_\varepsilon^{(i)}$ and arbitrary path responses $h_\varepsilon^{(i)}$ with $\varepsilon \in \mathcal{Z}_{L_d^{(i)}}$. The group sizes $L_d^{(i)}$'s satisfy $\sum_{i=0}^{L_d-1} L_d^{(i)} = L$ and $L_d^{(0)} = 1$ is fixed for single direct path. Doppler shifts $\kappa_\varepsilon^{(i)}$'s are drawn independently from $\kappa_\varepsilon^{(i)} = \kappa_{\max}\cos\left(\theta_\varepsilon^{(i)}\right)$ where the radian phases $\theta_\varepsilon^{(i)}$'s are i.i.d. and uniform in $[0, 2\pi)$. Path responses $h_\varepsilon^{(i)}$'s are independent and generated according to the following multipath modeling.

The multipath model consists of one direct path and $L-1$ diffuse paths. The dominant direct path group exhibits single direct path with random response $h_0^{(0)} = \sigma_{\text{direct}}\exp\{j\theta\} + \tilde{h}_0$, where the phase θ is arbitrary, σ^2_{direct} is the direct power, and $\tilde{h}_0$ is a CSCG with mean zero and diffuse power $\sigma^2_{\text{diff},0}$. Next, the remaining $L_d - 1$ dominant path groups consist of $L - 1$ diffuse paths with random responses $\{h_\varepsilon^{(i)}; \varepsilon \in \mathcal{Z}_{L_d^{(i)}}\}$ for $i \in \mathcal{Z}^+_{L_d-1}$, which are independent CSCGs with mean zero and diffuse path powers $\mathcal{E}\left[\left|h_\varepsilon^{(i)}\right|^2\right] = \sigma^2_{\text{diff},i}$. The channel power is normalized to one, i.e., $\sigma^2_{\text{direct}} + \sum_{i=0}^{L_d-1} L_d^{(i)}\sigma^2_{\text{diff},i} = 1$, and in this case γ_{d} represents the average received symbol SNR. The multipath model is also characterized by factors K_h and D_h for convenience, in which K_h is the ratio of direct power to diffuse power sum and D_h is an exponential decaying factor used to specify $\sigma^2_{\text{diff},i} = C_h\exp\{-l_d^{(i)}/D_h\}$ for $i \in \mathcal{Z}_{L_d}$ with C_h being a normalization factor. To show the performance prevalence of IDFHE over worse channel conditions, the

Table 1. Simulation Parameters.

Parameter	Value
Number of Subcarriers M	128
Number of Blocks N	32
Modulation Alphabet Size Q	QPSK ($Q = 4$)
Subcarrier Spacing Δf	60 KHz
Carrier Frequency f_c	3.5 GHz
Maximum Vehicular Speeds	57.8 Kmph ($\kappa_{\max} = 0.1$) 231 Kmph ($\kappa_{\max} = 0.4$) 578 Kmph ($\kappa_{\max} = 1$)
Channel Estimation	Ideal
Random Multipath Channel	L Random Paths ($l_{\max} = 7$)

multipath channel is specified by $K_h = 0$ (i.e., zero direct power) and $D_h = 2$ in the following demonstration. Also, $N_{CP} = l_{\max}+1$ is set to avoid ISI over frames and blocks. Gray coding of bits-to-symbol is used to represent quadrature-phase-shift-keyed (QPSK) symbols $d_{m,n}$. The path numbers are also set to $L = 5, L_d = 3, L_d^{(0)} = 1$, and$L_d^{(1)} = L_d^{(2)} = 2$ for demonstration.

Figures 3, 4 and 5 show the average bit-error-rate (BER) versus average received bit SNR $\gamma_d/\log_2 Q$ characteristics of various symbol decision schemes for CP-OTFS and CP-OFDM on time-varying Rayleigh multipath channels with $K_h = 0$ and three different $\kappa_{\max}$ values. For MPA, the number of iterations $n_{iter} = 20$ and the probability mass function threshold 1-$\zeta = 0.99$ are set for performance prevalence [6]. Due to uniform data spreading on the TF grid, all symbol decision schemes for CP-OTFS outperform the non-iterative LMMSE scheme for CP- OFDM significantly. As to the schemes for CP-OTFS, IDFHE outperforms non-iterative LMMSE and MPA remarkably when the BER is below 10^{-2}, and the performance prevalence is more significant for time-varying channels with narrower mobility ranges (or equivalently smaller $\kappa_{\max}$ values). Such performance prevalence for IDFHE results from the iterative symbol decision process, by which noise power enhancement is avoided at the final iteration and residual ISI is iteratively reduced by DFE.

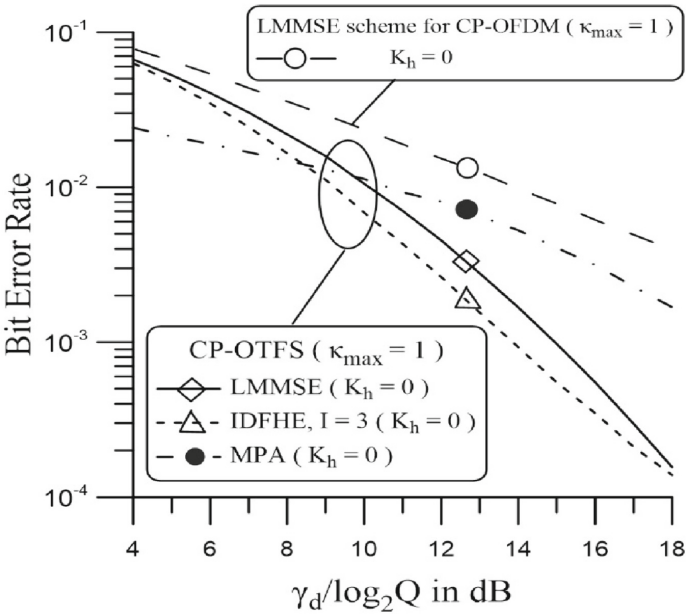


Fig. 3. Average BER characteristics of various symbol decision schemes for CP-OTFS and CP-OFDM over the time-varying Rayleigh multipath channel with $\kappa_{\max} = 1$, $L = 5$, $D_h = 2$, and $K_h = 0$.

The computational complexity is also compared in terms of the total number of complex multiplications required for implementing various CP-OTFS symbol decision algorithms that process $\{\mathbf{r}_l; l \in \mathcal{Z}_N\}$ as input and listed in Table 2. By employing the low-complexity LMMSE method in [8], $O(MNL^2)$ is required to process $\rho^{-1}\mathbf{E}_{l,\mathrm{LM}}\mathbf{r}_l$ for total N blocks. For MPA, $O(n_{\mathrm{iter}}MNLQ)$ is required to complete a maximum of n_{iter} iterations [6]. Both non-iterative LMMSE and MPA require $O(MNlog_2N)$ for the DFT and SFFT processes. For I-step IDFHE, $\rho^{-1}\mathbf{E}_{l,LM}\mathbf{r}_l$ is computed only once for leading I-1 iterations and requires $O(MNL^2)$. For the final iteration, SVD requires $O(M^3N)$ in generating $\mathbf{E}_{l,\mathrm{CLS}}$ and $\rho^{-1}\mathbf{E}_{l,\mathrm{CLS}}\mathbf{r}_l$ consumes $O(M^3N + M^2N)$. In addition, the computation of $\{\overline{\mathbf{Y}}_l^{(i)}; l \in \mathcal{Z}_N\}$ for $\mathbf{E}_{l,\mathrm{LM}}$ and $\mathbf{E}_{l,\mathrm{CLS}}$, along with $\{\overline{\mathbf{Y}}_l^{(i)}\hat{\mathbf{s}}_l^{(i-1)}; l \in \mathcal{Z}_N\}$ and $\widehat{\mathbf{D}}^{(i-1)}\mathbf{W}_N^h$ for the purpose of ISI excision in the last $I-1$ iterations, requires $O(2M^3N + [I-1]M^2N + [I-1]MNlog_2N)$. Last, joint DFT and SFFT process is conducted for all I steps and thereby requires $O(IMNlog_2N)$ in total. Combining all above complexities yields the total complexity for I-step IDFHE in Table 2. As shown in Table 2, non-iterative LMMSE requires less complexity than IDFHE obviously, and than MPA when $n_{\mathrm{iter}}Q > L$. For large MN values, IDFHE requires higher complexity than MPA and non-iterative LMMSE in exchange for better BER performance.

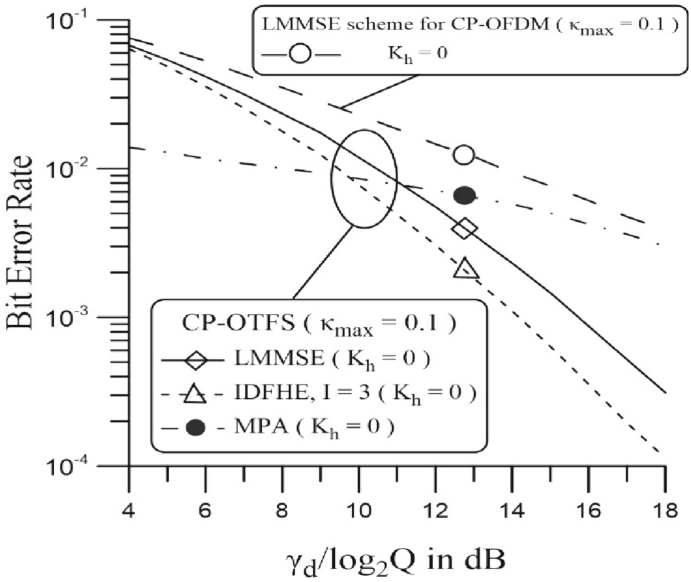


Fig. 4. Average BER characteristics of various symbol decision schemes for CP-OTFS and CP-OFDM over the time-varying Rayleigh multipath channel with $\kappa_{\max} = 0.4$, $L = 5$, $D_h = 2$ and $K_h = 0$.

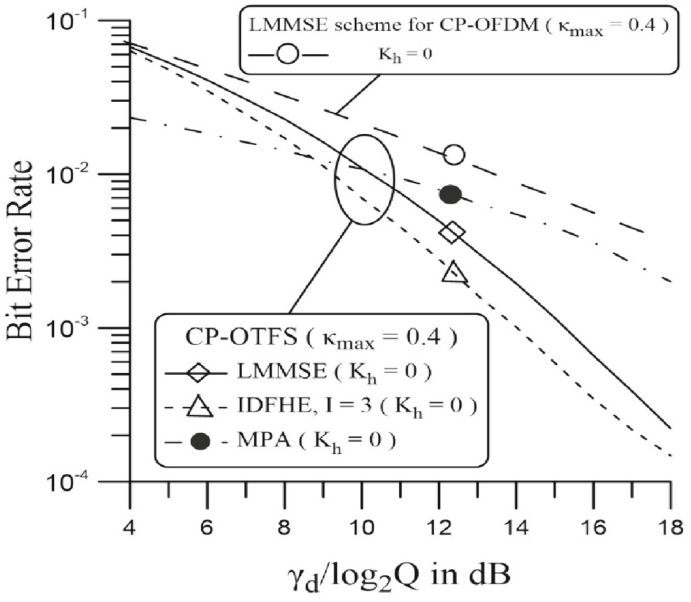


Fig. 5. Average BER characteristics of various symbol decision schemes for CP-OTFS and CP-OFDM over the time-varying Rayleigh multipath channel with $\kappa_{\max} = 0.1$, $L = 5$, $D_h = 2$ and $K_h = 0$.

Table 2. Number of Complex Multiplications Required to Realize Various CP-OTFS Symbol Decision Algorithms.

Algorithm	Computational Complexity
LMMSE	$O(MNL^2 + MNlog_2N)$
MPA	$O(n_{iter}MNLQ + MNlog_2N)$
I-step IDFHE	$O(4M^3N + IM^2N + MNL^2 + [2I - 1]MNlog_2N)$

5 Conclusion

In the paper, an IDFHE symbol decision scheme is proposed to enhance error performance for CP-OTFS on time-varying multipath channels. By adopting time-domain LMMSE and CLS equalizers in different iterations, the IDFHE scheme can suppress ISI progressively over iterations and avoid noise power enhancement at the final iteration. Such IDFHE scheme turns out to make symbol decision with better error performance on time-varying Rayleigh multipath channels than conventional non-iterative LMMSE scheme and iterative message-passing symbol decision scheme, at the cost of consuming higher computational complexity.

Acknowledgment. This work was supported by the Ministry of Science and Technology in Taiwan, under Grants MOST 109–2221-E-002–156-MY3 and MOST 111–2221-E-027–053.

References

1. Hadani, R., et al.: Orthogonal time frequency space modulation. In: Proceedings of IEEE Wireless Communication Networking Conference, pp. 1–6, San Francisco (2017)
2. Das, S.S., et al.: Time domain channel estimation and equalization of CP-OTFS under multiple fractional Dopplers and residual synchronization errors. IEEE Access **9**, 10561–10576 (2021)
3. Wei, Z., et al.: Orthogonal time-frequency space modulation: a promising next-generation waveform. IEEE Wireless Commun. **28**(4), 136–144 (2021)
4. Hashimoto, N., et al.: Channel estimation and equalization for CP-OFDM-based OTFS in fractional Doppler channels. In: Proceedings of IEEE International Conference on Communications Workshops, pp. 1–7. Virtual/Montreal (2021)
5. Qu, H., Liu, G., Zhang, L., Wen, S., Imran, M.A.: Low-complexity symbol detection and interference cancellation for OTFS system. IEEE Trans. Commun. **69**(3), 1524–1537 (2021)
6. Raviteja, P., et al.: Interference cancellation and iterative detection for orthogonal time frequency space modulation. IEEE Trans. Wireless Commun. **17**(10), 6501–6515 (2018)
7. Yuan, Z., et al.: Iterative detection for orthogonal time frequency space modulation with unitary approximate message passing. IEEE Trans. Wireless Commun. **21**(2), 714–725 (2022)
8. Tiwari, S., Das, S.S., Rangamgari, V.: Low complexity LMMSE receiver for OTFS. IEEE Commun. Lett. **23**(12), 2205–2209 (2019)
9. Zou, T.,et al.: Low-complexity linear equalization for OTFS systems with rectangular waveforms. In: Proceedings of IEEE International Conference on Communication Workshops, pp. 1–6. Virtual/Montreal (2021)

10. Tsai, S.-Y., Chen, W.-C., Chung, C.-D.: Iterative symbol decision schemes for CP-OTFS on static multipath channels. In: Proceedings of International Conference on Computing, Networking and Communications, pp. 1–6, Honolulu (2023)
11. Ma, L.-Y., Chen, W.-C., Chung, C.-D.: Iterative decision-feedback hybrid equalization in SC-FDM. In: Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications, pp. 617–622. Virtual Conference (2022)
12. Proakis, J.G.: Digital Communications, 5th edn. McGraw Hill, New York (2007)
13. 3GPP TS 38.211: NR; Physical channels and modulation (2020)

A 0.4 V 21.6 nW Duty Cycle Generator Based on Compact Pulsed Modulator for MEMs Sensing Interface

Xi Sung Loo[1(✉)], Wang Ling Goh[1], and Yuan Gao[2]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
xisung.loo.phd@ieee.org

[2] Institute of Microelectronics, Agency for Science, Technology and Research, Singapore, Singapore

Abstract. This paper presents a 0.4 V duty cycle generator based on compact pulsed modulator architecture for Micro-Electromechanical Systems (MEMS) sensing interface. It compares favorably against literature in power performance (21.6 nW) through adopting of current re-used amplifiers and Schmitt trigger based pulsed generator with self-biasing architecture in subthreshold operations. Further simulations confirm the circuit functionality which shows varying duty cycles with sensing frequencies below 1 kHz. The design only occupies active area of 0.031 mm^2 and is fabricated on 0.18 μm Complementary Metal Oxide Semiconductor (CMOS) technology.

Keywords: Analog integrated circuits · differential amplifiers · relaxation oscillators · pulsed width modulation

1 Introduction

The body sensor network (BSN) technology is one of the core technologies of IoT developments in healthcare system, where a patient can be monitored using a collection of tiny-powered and lightweight wireless sensor nodes. Such applications require durable sensing solution for the devices to remain attached to the body without charging. Micro-Electromechanical Systems (MEMS) technology plays an important role for the development of reliable, low cost, and low power sensor systems for body sensor network. It can be used to monitor the human body conditions such as motion acceleration, blood pressure, temperature, sound, humidity and so on. Nevertheless, additional power consumption is incurred as it must be integrated with signal conditioning circuits for conversion of change in sensing capacitance into measurable form. One way to optimize the power consumption is through the use of management circuits [1] but at the cost of additional control circuitry. Thus, It is more useful in large scale circuits where the power incurred by the control circuitry is relatively negligible.

Many interfacing circuits [2–9] have been reported in attempt to achieve ultra-low power consumption with small trade-off in the sensing performance. MEMS sensing

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 77–83, 2024.
https://doi.org/10.1007/978-3-031-59619-3_7

circuits based on capacitance to frequency (C2F) conversion [2–8] can achieve better performance than those based on voltage conversion [9] due to better noise immunity and lower DC offset. A frequency converter could be realized with as simple as ring oscillator [7, 10] with voltage control at one of the inverter stages. Although noise immunity could be improved with differential version, it is sensitive to PVT and supply variations. Meanwhile PWM configuration exploits semi-digital approach where the sensing signal is converted into duty cycle format that is proportional to source amplitude. The PWM signal is useful for driving LDOs in wake-up MEMS sensors. Various PWM architectures have been proposed but less attention has been given on operation at subthreshold region to achieve sub nano-watt power. Many require additional clocks [3–6] or external off-chip components [2, 7] for proper operations. Power consumption is further deteriorated in [5, 6, 8] due to higher number of active components needed.

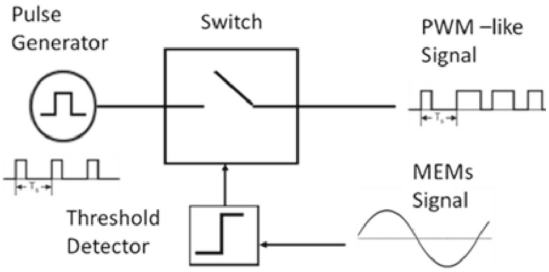


Fig. 1. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

In this paper, we demonstrate the use of switching modulation to generate PWM-like signal for the first time. Figure 1 illustrates the operation principle which rely on variation in sensing input to turn on/off the pulse generator output. It is similar to Amplitude Shift Keying (ASK) [11] used in communications with exception that only digital signals are mixed during the process. Therefore, pulse generator is used instead of sinusoidal carrier and threshold detector is used to digitize the sensing input signal before mixing, The advantage of such method is less components used as compared to conventional PWM scheme [8] and thus consuming less power for the same purpose. The circuit architecture of proposed duty cycle generator is further described in the following sections.

2 Circuit Architecture

The duty cycle generator proposed operates in clockless mode and is composed of only two active elements as compared to [8]. They are relaxation oscillator (pulse generator) and inverter amplifier which acts as a threshold detector. The switching component of the oscillator is made up of low resistive transmission gates and are controlled by amplified signal of sensing input. In order to achieve nano-watt level power consumption, the entire circuit is designed to work close to subthreshold region with power supply of 0.4 V.

2.1 Inverter Based Amplifier

The inverter-based (Current-reuse) amplifier is popularly used in ultra-low power operation due to extreme high gain in switching region even at low supply voltages or subthreshold operation. Comparatively, conventional amplifiers [12] consume higher power as they are designed to work at saturation region. Figure 2 shows the circuit schematic of the proposed self-biasing inverting amplifier. It consists of 2 stages of inverter-based amplifiers (M7, M8, M16, M17, M11, M12, M18 and M19). The bias current is auto adjusted through output DC feedback from M14 and M15 to the top and bottom transistors (M6, M9, M10 and M13). Although the self-biasing amplifier is less sensitive to process, voltage, or temperature (PVT) variations than simple inverter, still significant offset can occur due to large variation in biasing current [13]. Thus, additional Common Mode Feedback (CMFB) Amplifier (M1, M2, M3, M4 and M5) is tapped on the output node of 1st inverter stage to regulate the bias current through transistors M14 and M15 during operation. The CMFB Amplifier shown on the left of Fig. 2 is designed to be resistorless and clockless. Basically, it senses the voltages variations in differential paths through subtractions of output current from the reference before feedbacked to the amplifier.

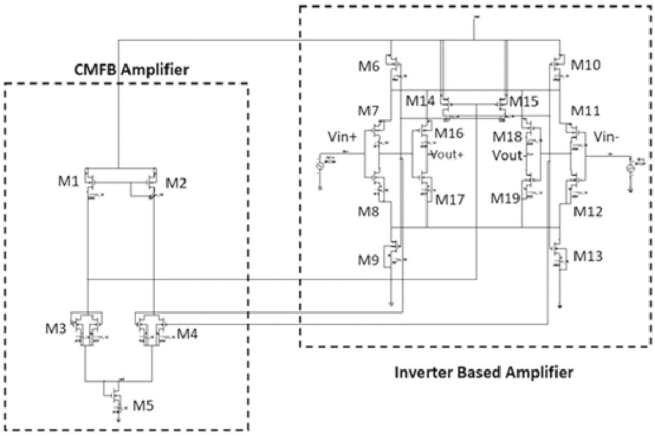


Fig. 2. Architecture of proposed duty cycle generator.

2.2 Schmitt Trigger Based Relaxation Oscillator

Meanwhile, the oscillator (Fig. 3) is formed by Schmitt trigger connected with active pseudo resistor (M22, M31) at feedback loop. The Schmitt trigger is designed for low power consumption with the use of self-biasing transistors (M23, M26, M27 and M30) next to the supply rails. Such technique is effective in reducing leakage power of Complementary Metal Oxide Semiconductor (CMOS) transistors [14]. Although self-biasing design is deployed in [15, 16], It is based on single ended configuration which has lower noise immunity. The proposed Schmitt trigger operates in differential manner with only

2 extra transistors needed as compared to [15]. Voltage mode operation is preferred for ultra-low power operation since the leakage current becomes comparable with biasing current for smaller size transistors. The feedback path of the oscillator is connected through switching network described earlier for turning on and off the oscillator.

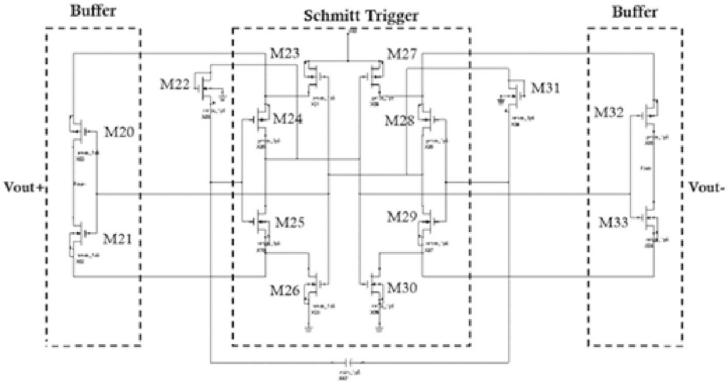


Fig. 3. Circuit schematic of proposed Schmitt trigger-based oscillator.

3 Results and Discussion

The proposed duty cycle generator has been designed and simulated in a standard commercial 0.18 μm CMOS technology. The chip layout is shown in Fig. 4 which occupies an active area of 0.031 mm^2. Under typical condition of 0.4 V supply and 27 °C, the entire circuit consumes an average power of 21.6 nW. Specifically, the inverting amplifier consumes only 18.4 nW and have DC gain of 65.3 dB with bandwidth of 70 kHz. Meanwhile, the carrier oscillator operates at 0.83 kHz and dissipate power of approximately 3.2 nW. The phase noise is − 40.6 dBc/decade at 2 kHz offset.

Figure 5 shows the transient simulation of the duty cycle generator output (Blue) when 200 Hz and 400 Hz sensing input signals (Red) are applied. At 200 Hz, the circuit output shows periodically shifts of duty cycle from 50% to 16.5%. Meanwhile, 27.9% duty cycle output is achieved when sensing input of 400 Hz is applied. With appropriate output filtering, these give different readings of average voltages which are needed to distinguish the sensing states. The duty cycle generator remains functional till 1.2 kHz. Nevertheless, it is recommended to operate below half of oscillator carrier frequency to avoid aliasing.

The proposed generator is benchmarked against other PWM circuits in Table 1. It compares favorably against others in power consumption with subthreshold circuit design that uses low supply voltage of 0.4 V. Other contributing factor is the modulation method used which requires only 2 active modules and avoid the need for clocks. Consequently, it consumes the least active area of 0.031 mm^2. Depending on the frequency of sensing input, the proposed PWM generator can produce up to 85.8% of duty cycle output.

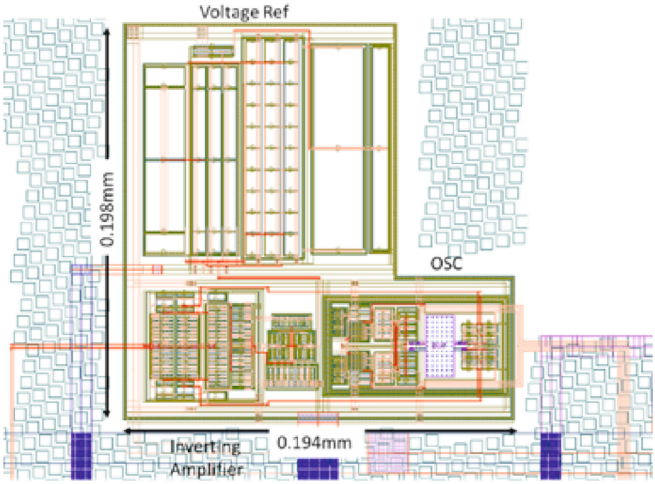
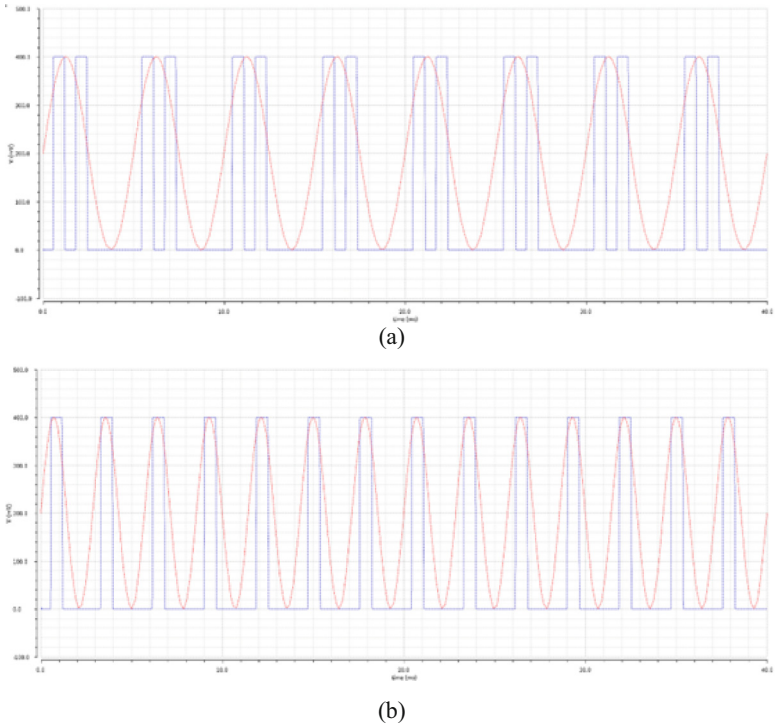


Fig. 4. Chip Layout of proposed duty cycle generator.



(a)

(b)

Fig. 5. Output voltage waveform of duty cycle generator when input frequency is at 200 Hz (a) and 400 Hz (b) for transient period of 40 ms.

Table 1. Performance Comparison of duty Cycle Generators.

	This Work	[4]	[5]	[6]	[8]
Type	Differential	Unipolar	Differential	Differential	Unipolar
Process	0.18 μm CMOS	0.35 μm CMOS	0.18 μm CMOS	0.32 μm CMOS	0.35 μm CMOS
Vdd	0.4 V	3 V	1.8V	3V	3V
Area	0.031 mm^2	0.17 mm^2	0.09 mm^2	0.52mm^2	0.57mm^2
Power	21.6 nW	54 μW	98 μW	84μW	0.4 mW
Operating Frequency	<1.2 kHz	25 kHz	12.5 kHz	<2.6 kHz	32 kHz
Duty Cycle	0%–85.8%	17.53%–20.65%	50%/96%	-	0%–72%
Active Components	2	4	5	5	4
Clock	No	Yes	Yes	Yes	Yes

4 Conclusion

A ultra-low power modulator has been presented to generate PWM-like signal through modification of conventional ASK architecture [9]. It exhibits distinct duty cycle patterns at different frequencies with variations up to 85.8%. Low voltage headroom coupled with compact architecture enable it to achieve less than 22 nW. This make it attractive for use in batteryless sensing interface where high durability is desirable.

Acknowledgment. The authors are grateful to grant and support from Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR) on this work.

References

1. Kumar, R.: Reduce power consumption for digital CMOS circuits using DVTS algoritham. Int. J. Electr. Electron. Eng. Telecommun. **1**(1), 376–383 (2015)
2. Kokolanski, Z., Gavrovski, C., Dimcev, V., Makraduli, M.: Simple interface for resistive sensors based on pulse width modulation. IEEE Trans. Instrum. Meas. **62**(11), 2983–2992 (2013)
3. Yazdi, N., Mason, A., Najafi, K., Wise, K.D.: A generic interface chip for capacitive sensors in low-power multiparameter microsystems ensors and actuators a. Physical **84**(3), 351–361 (2000)
4. Sheu, M.-L., Hsu, W.-H., Tsao, L.-J.: A capacitance-ratio-modulated current front-end circuit with pulsewidth modulation output for a capacitive sensor interface. IEEE Trans. Instrum. Meas. **61**(2), 447–455 (2012)
5. Arefin, M.S., Redouté, J.M., Yuce, M.R.: A low-power and wide-range MEMS capacitive sensors interface IC using pulse-width modulation for biomedical applications. IEEE Sens. J. **16**(17), 6745–6754 (2016)

6. Nizza, N., Dei, M., Butti, F., Bruschi, P.: A low-power interface for capacitive sensors with PWM output and intrinsic low pass characteristic. IEEE Trans. Circuits Syst. I Regul. Pap. **60**(6), 1419–1431 (2013)
7. Al Kadi Jazairli, M., Flandre, D.: Low power pulse generator as a capacitive interface for MEMS applications. 2009 Ph.D. Research in Microelectronics and Electronics, Cork, Ireland (2009)
8. Lee, C.-H., Chuang, W.-Y., Cowan, M.A., Wu, W.-J., Lin, C.-T.: A low-power integrated humidity CMOS sensor by printing-on-chip technology. Sensors **14**(5), 9247–9255 (2014)
9. Yazdi, N., Kulah, H., Najafi, K.: Precision readout circuits for capacitive microaccelerometers. In: Proceedings of IEEE Sensors, Vienna, Austria, pp. 28–31 (2004)
10. Sachdeva, P., Aggarwal, A.: Design of CMOS ring VCO for PLL based frequency synthesizer. Int. J. Electr. Electron. Eng. Telecommun. **5**(2), 73–79 (2016)
11. , Miller, F.P., Vandome, A.F., John, M.B.: Amplitude-Shift Keying VDM Publishing (2010)
12. Elsayed, F., Rashdan, M., Salman, M.: CMOS operational floating current conveyor circuit for instrumentation amplifier application. Int. J. Electr. Electron. Eng. Telecommun. **9**(5), 317–323 (2020). https://doi.org/10.18178/ijeetc.9.5.317-323
13. Baltolu, A., Albinet, X., Chalet, F., Dallet, D., Begueret, J.-B.: A robust inverter-based amplifier versus PVT for discrete-time integrators. Int. J. Circuit Theory Appl. **46**, 2160–2169 (2018)
14. Goel, A., Mazhari, B.: Gate leakage and its reduction in deep submicron SRAM. In: 18th International Conference on VLSI Design held jointly with 4th International Conference on Embedded Systems Design, pp. 606–611 (2005). https://doi.org/10.1109/ICVD.2005.103
15. Al-Sarawi, S.F.: Low-power Schmitt trigger circuit. Electron. Lett. **38**(18), 1009–1010 (2002)
16. Nowbahari, A., Marchetti, L., Azadmehr, M.: Analysis of a low power inverting CMOS Schmitt trigger operating in weak inversion. Int. J. Electr. Electron. Eng. Telecommun. **11**(6), 392–397 (2022)

Exploring Usability Challenges of E-Services in University Academic Portal: An Eye-Tracking Analysis of Participant's Navigation and Searching Behavior

Mohamed Basel Almourad[1(✉)], Emad Bataineh[1], and Zeal Wattar[2]

[1] College of Technological Innovation, Zayed University, Dubai, UAE
basel.almourad@zu.ac.ae

[2] College of Communication and Media Sciences, Zayed University, Dubai, UAE

Abstract. This paper discusses the shift from paper-based academic services to e-services, which has become prevalent in College of Technological Innovation (CTI), UAE. However, the usability of these e-services is a challenge due to design issues, and This paper introduces a usability assessment study focusing on the e-services available through the CTI's academic portal. The study employs eye tracking to examine the viewing, searching, and navigation behavior of college students, along with the factors that impact their searching behavior. The study focuses on the students' visual patterns in choosing an e-service. Eye-tracking experiments were conducted, and data were collected from a group of CTI students. According to the findings of the heatmap and eye gaze analysis, students had trouble locating the right link for the desired service and displayed intense, disoriented scattered, and erratic visual behaviors when performing assignments. We discovered several user interface design issues that hinder student productivity. Based on the results of the usability assessment study, the research recommends changes to the CTI portal's design to better suit students' needs, preferences, and expectations.

Keywords: Academic portal · Student E-services · Usability testing · Eye tracking technology · Heat map analysis

1 Introduction

A website is a collection of web pages and digital material, such as text, pictures, videos, and other media [1]. Organizations need websites to interact with their stakeholders and to communicate their values, vision, and mission. They facilitate different jobs and offer a quick and simple way to distribute content to a broad audience [2, 3]. Website usability refers to how effectively and efficiently a website can fulfill the information needs of its intended users while ensuring their satisfaction with its usage [4]. It is a crucial determinant of user engagement and the likelihood of revisiting the site. If universities or colleges do not have user-friendly websites for international students,

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 84–96, 2024.
https://doi.org/10.1007/978-3-031-59619-3_8

they risk losing their competitive edge to other institutions, which could negatively impact their enrollment rates [5, 6]. Therefore, it is challenging and unclear to measure user satisfaction and interface usability. Designers therefore need tools to help them get the best level of user happiness and make intelligent choices.

The eye-tracking tool is a technology that presents objective and qualitative data on to where and how long users look at somewhere and, in this way, provides an opportunity to explore users' information searching behaviors, the points that are focused on screen and surfing durations [7, 8]. In recent years, there has been a notable rise in the adoption of eye tracker methodology within usability studies. Numerous research endeavors have emphasized the significance of individuals in determining the effectiveness of academic web portals [9–11].

This study's main goal is to examine how students in CTI use the academic portal at the university to look for information. The study will track and evaluate the users' eye movements and gather data on their behavior and experiences using eye tracking equipment and a combination of quantitative and qualitative methodologies. The academic portal underwent an extensive usability study involving students as participants. The study encompassed planning, development, execution, and analysis phases. The aim of this investigation is to illuminate the scanning, navigation, and interaction patterns of CTI students with e-service links. In addition, it explores how their visual and clicking behavior correlates with decision-making judgments. Moreover, the study seeks to identify the factors influencing students' viewing habits, including the specific areas of focus on the e-service page and the duration of time spent before initiating a click. Data from respondents was gathered to ascertain which graphical interface elements attracted users' attention the most in terms of initial fixation of the eye-gaze and fixation count. Heat maps and eye gaze analysis were used to display the results. The results of this study will assist e-service designers in presenting the connections in a way that takes users' requirements, interests, and expectations into account and improves the academic portal's usability.

The paper's structure is detailed as follows. Section 2 encompasses diverse subjects, including recent advancements in academic portals and e-services, a comprehensive literature review, and the utilization of eye tracking methodology in studying information seeking behavior and web user interface. The subsequent section explores the methodology and design of the eye tracking experiment. Following this, another section presents the findings of the eye gaze plot and heat map study, accompanied by data analysis and discussions. The concluding section summarizes the study, outlining its implications, significance, limitations, and offering suggestions for future research.

2 Related Works

Many research studies have been conducted to explore users' assessments of academic portals and their usability features [6, 12–14]. In [6], the authors employed a distinctive combination of qualitative and observational research methods to assess the usability and interactivity of a university's website for international student services. The authors found that providing usable and interactive services in a manner that is pleasing to users can increase the likelihood of user acceptance of systems like university websites. Their

study emphasizes the importance of eliciting positive perceptions from users. In [13], the authors conducted research on the elements that impact how often students utilize Saudi university portals. They proposed a model that fulfills student expectations and boosts their contentment, with the aim of enabling universities to communicate more effectively with their students. This model is expected to enhance the efficiency of communication between universities and students. Study that aimed to evaluate the effect of university portals on student satisfaction. In [14], the authors created the University Portal Usability Assessment Index as a tool to gauge the usability of university portals. Their findings underscored the paramount importance of structural design and interactive capabilities in influencing student satisfaction, where structural design pertains to layout, navigation, and visual aspects, while interactive capabilities involve facilitating online tasks like course registration and accessing academic resources.

Eye-tracking has been used often for many years to monitor and analyze human visual behavior in industries like e-commerce, online portals, medicine, and psychology [8, 15–17]. The main benefits of eye-tracking include determining the user's regions of interest, spotting fixations, and measuring how long it takes a user to find what they are looking for. This can be important for the re-design recommendations as it can highlight the locations of crucial parts and the elements that needlessly divert users' attention [8]. Eye fixations represent a critical measure for evaluating information processing and acquisition in an online search and viewing environment. Eye tracking tests can provide quantitative data on a user's visual behavior, including which information items immediately capture their attention, how long they look at an item, how frequently they look at a particular item, the order in which they visually navigate the page, and the areas (high-attention areas) on the page they viewed the most for information. Eye-tracking technology is used to record eye movement while a person looks at a visual stimulus. The eyes move frequently, with micro-movements that sometimes only cover a few pixels, and saccades occur when the eye swiftly moves from one object to another during a fixation. Eye-tracking software uses the data collected to identify fixations and saccades. Heat maps show how long each area of a screen has been viewed, while gaze plots can display the sequence of fixations and saccades on a screen or webpage for a specific user [16, 17].

Eye tracking has been employed by a small number of studies to measure the usability of websites [17–19]. In [17], the authors evaluated the usability of two websites in the e-commerce and educational sectors: jamb.org.ng and jumia.ng using eye tracking. The research found that web apps for e-commerce and education had quite different user activity patterns. Although the e-commerce website performed better in terms of aesthetics and the use of multimedia, the use of distracting colors and an abundance of media on the site's design prevented users from making a purchase. In [18], the authors employed eye-tracking technology to assess user efficacy, efficiency, eye-movement patterns, and quantitative indicators related to the area of interest. Ten tasks were assigned to six students as part of an experiment to examine the usability of Moodle, an open-source learning management system. The results showed that familiarity with the course led to increased Area of Interest (AOI) measures and greater performance on half of the tasks. Although correlations were not always present, eye movement patterns were discovered to be related to AOI metrics, efficiency, and effectiveness in completing tasks.

When looking at user interface elements like text hyperlinks or images, participants' eye movements varied. In [19], the authors assessed the usability of the SimplyTick e-commerce system using eye tracking. Metrics including the first mouse clicks delay, the total number of clicks, and the interval between the first fixation and the subsequent click were examined. According to the study's findings, a given element's placement and presentation were accountable for any problems users had finding it. Users' attention was scattered around the SimplyTick webpage and was not drawn to the key elements.

The literature review identified research gaps such as statistical methods are infrequently applied for data analysis and only a small set of website features are used for eye-tracking usability testing. By using eye-tracking technology to undertake a thorough usability assessment of all the elements on the CTI web portal, the current study aims to close these gaps. The study explores the factors influencing students' searching behavior, focusing on the observation, search, and navigation patterns of students interacting with the CTI web site. A mixed-methods approach combining qualitative analysis of heat maps and eye gaze patterns with quantitative analysis of metrics such as search time and time to first observation was employed to investigate the gazing pattern.

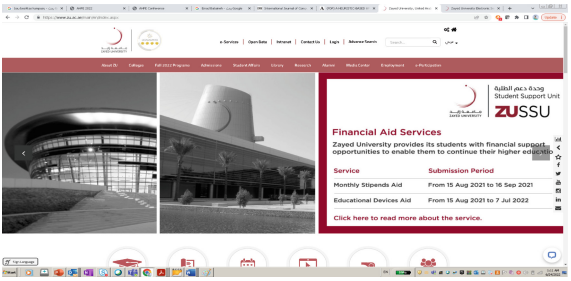
3 Materials and Methods

3.1 Subjects

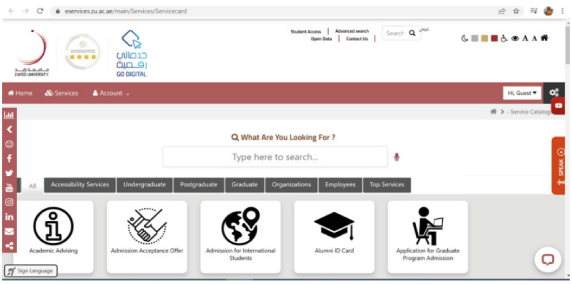
The research included the involvement of 36 female undergraduate students pursuing diverse degree programs at CTI. These students, aged between 18 and 24, with an average age of 21, were selected randomly and possessed a minimum of three years of experience utilizing the Internet for various information searches. Furthermore, the majority of respondents regarded themselves as proficient in accessing online services. Prior to commencing the study, all respondents were obliged to sign a consent form delineating the study's purpose and outlining their rights as contributors.

3.2 Study Components and Activities

The research employed a collection of twenty-five brief assignments centered on information retrieval. These tasks differed in complexity and subject, each accompanied by explicit instructions guiding respondents on what to search for. To maintain consistency among respondents, we supplied an initial search query for each task, ensuring comparability of the initial search engine results page. A visual representation of the stimulus sequence utilized in the eye-tracking experiment is depicted in Fig. 1.



(a)



(b)

Fig. 1. (a) Screenshot of Homepage E-Services GUI (b) Scrnshot of E-Services Link Interface

3.3 Approach

The study took place in a standardized and consistent setting, employing non-intrusive mobile eye-tracking technology to gather eye movement data. Before commencing the test, the moderator informed participants about the study's main objective and calibrated the eye tracker for each individual. Throughout the experiment, participants operated within a controlled environment, having the freedom to browse, click, and advance to the next page. The screen displayed a sequence of screenshots featuring various CTI web site-services, prompting participants to fulfill a series of navigational tasks. Following the test, participants provided feedback and evaluated their user experience through ten post-test satisfaction questions, using a Likert scale ranging from one to five.

4 Findings and Interpretation

The research study gathered quantitative and qualitative data from the experiment, utilizing a range of usability specifications and criteria in its analysis. These criteria encompassed effectiveness, proficiency, and user satisfaction, allowing for a comprehensive evaluation of the study's outcomes. Effectiveness measured the user's ability to successfully find information and complete tasks, while proficiency measured the user's ability to quickly and easily complete tasks. User satisfaction measured how much a user enjoyed using the e-services portal.

To analyze the collected eye-tracking data, the researchers used three different analysis tools. Descriptive data analysis tools were used to measure a user's performance in completing the given tasks, while visualization analysis tools like heat maps and gaze plots were used to measure individual user visual behaviors on the university e-services portal page. Heat maps provided insights into the areas of the page that garnered the highest user attention, whereas gaze plots tracked the trajectory of participants' eye movements, pinpointing the sequence of eye movement, including fixation points. Statistical analysis tools, such as areas of interest (AOIs) and various metrics data like time for first fixation, number of fixations, fixation duration on specific content areas, and number of visits, were used to measure user dwelling time and decision-making behavior for mouse-clicking activities. AOIs were delineated on every e-services, predicated on which interface elements students would need to utilize to proficiently fulfill the twenty-five navigational tasks. In this study, one AOI was created to represent the link for e-service on each screen, and we can see the visual attention of participants scattered across the stimuli from these heat maps.

4.1 Eye Gaze Plots

The researchers utilized eye gaze analysis [20, 21] as a visualization technique to gain comprehensive insights into individual users' search behaviors on the e-services homepage. By examining the gaze plots of individual students, common search patterns were identified, displaying a wide range from basic to intensive and random search behaviors. Eye gaze plots provide valuable insights by illustrating a step-by-step representation of the entire search process. The size of each circle indicates the duration spent at each location, providing further context into user behavior.

The study utilized gaze plots to investigate how students navigated the e-services webpage. Examples of individual and group eye gaze plots are depicted in Fig. 2. The results revealed that most students initially directed their attention towards the top middle position of the webpage. Within five seconds, approximately 65% of students displayed a disorganized and random pattern of eye movement across the online portal, while the remaining students maintained a consistent and focused attention on the top links (see Columns 1 and 2 in Fig. 2).

Participants who displayed irregular and dispersed visual fixations were characterized as nervous and unsettling because they were unable to locate the links in a logical way. The concentrated and regular eye gaze group, on the other hand, continued to retain comparable gaze patterns since they were accustomed to the CTI website. Around the 15-s point and beyond, this pattern becomes apparent. Additionally, the researchers created individual gaze plots to illustrate the various ways in which participants searched within the same task domain. These individual gaze plots, which range from simple to random, are shown in Fig. 2 Columns 3 and 4. The majority of students started their search at the top of the SERP, working their way up to the high middle spot. The gaze plots show how people skim over and assess search results before selecting a link to click on first.

According to the study, a sizable percentage of students had trouble finding and utilizing the e-services provided by the CTI main portal. The research result in [22] shown that during the first ten seconds of viewing the web site, consumers often choose

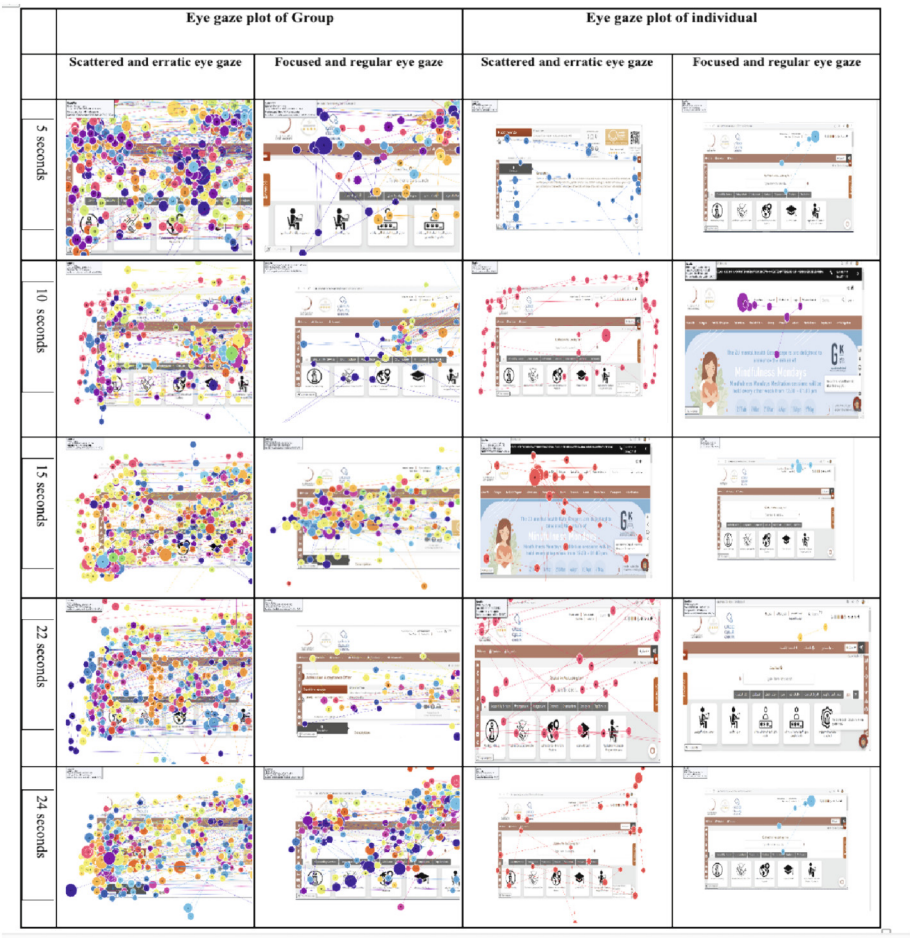


Fig. 2. Eye Gaze Plot Exploration of Individual and Group Search Behavior

a search result. In contrary, our research outcome demonstrated that before making their first click, students typically lingered longer and read more content on the results page. This suggests that students are spending more time looking for the appropriate link to the requested service on the e-services homepage. Overall, the results point to the necessity of enhancing the CTI web portal's e-services' accessibility and usability.

4.2 Heat Maps

The gathered data revealed a wide variety of search patterns, from simple to intense and erratic search behaviors. The eye tracker generated objective heat maps, with higher heat indicating more time spent. These heat maps could be helpful in locating possible problems with things like placement, tab size, color coordination, and interface complexity. The heat map depicted in Fig. 3 illustrates the activity of a group of people on

SERPSs for each domain. The area with the highest heat surrounding the left fold menu signifies interface complexity. Numerous elements, such as the menu's use of related terms, asymmetrical content, and font size, could contribute to its complexity.

After 5 s, an investigation revealed that approximately 65% of students were visually fixated in an unfocused and scattered manner over the online portal, while the remaining students were focused and directed in their attention towards the top links on the homepage (Refer to rows 2 in Fig. 3). Due to their inability to locate the links in a way that was acceptable, the majority of participants were characterized as erratic because their visual fixations were dispersed and unpredictable throughout the course of the experiment. However, because they were accustomed to the CTI website, the group that made regular and concentrated eye contact continued to gaze in a similar manner. This becomes visible at 15 s and beyond.

The research found that a large percentage of CTI students were unaware of the e-services provided by the main portal; this ignorance, along with the challenges associated with locating and utilizing these services, were the main challenges to their access to these services. Conversely, a previous study [22] revealed that users frequently click on a search engine result within the initial ten seconds of viewing the search engine result page. According to the results of the current study, CTI students, on average, took longer to click on a link after viewing the results page. Furthermore, as shown in Fig. 3, rows 1–3, the study's results show that students took longer to identify the right link on the e-services homepage that led to the requested service.

4.3 Statistical Overview and Evaluation

The goal of the research was to examine the students' visual attention by analyzing the gaze data. Each e-services link on the website was identified as an AOI and examined in order to accomplish this. Each e-service screenshot, treated as a single stimulus, contained at least one Area of Interest, and the analysis involved eight measures to assess and compare the group's gaze patterns. A description of each metric utilized in the analysis of the study is given in Table 1.

The statistical analysis facilitated the examination of participant behavior alterations based on the home page services and revealed significant disparities in the usability of the CTI web portal. In terms of total fixation time, participants spent 1.86 s on the "Search box," contrasting with only 0.62 s on "Login English." The text link "Search box" attracted the most fixations, accounting for 83% of the total, while "Login English" received the fewest (17%). Additionally, during the experiment, 94% of users clicked the text link "Change language (English)" at least once, whereas none clicked on "Login English". Among the observed elements, the "search box" garnered the least initial fixation time from participants (0.74 s), whereas the textual link "Login English" received the highest (2.97 s). Participants took the longest time (0.91 s) to click on the text link "Change language (English)" and the shortest time (0.31 s) to click on the text link "Login English". Furthermore, the most frequently clicked text link was "Login Arabic" (5.52 s), whereas "Login English" remained unused throughout the experiment. "Login Arabic" received the highest attention from participants, with 5.52 s, whereas "Login English" did not attract any attention.

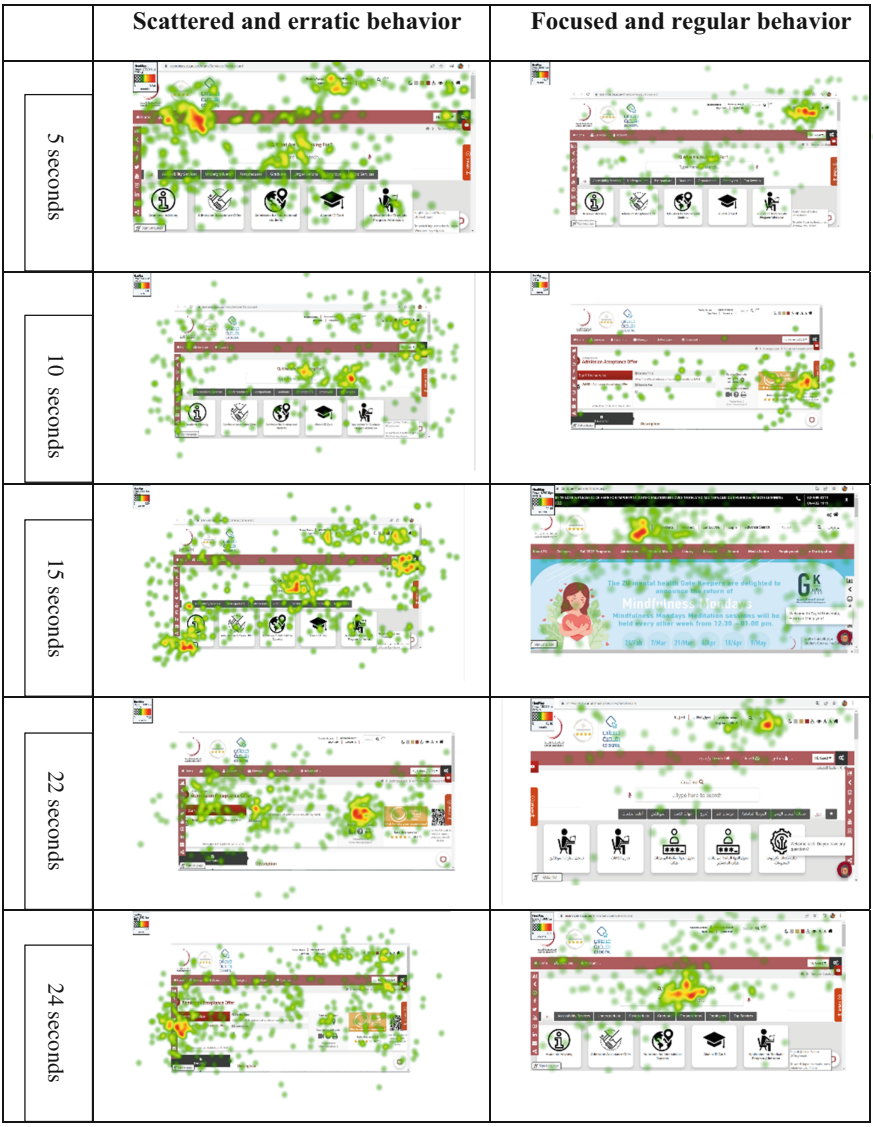


Fig. 3. Heat Map analysis of group of individuals

The statistical analysis indicates that "Login Arabic" is the most frequently utilized textual link on the CTI homepage, with an average usage time of 2.57 s. In contrast, "Login English" exhibits the lowest usage rate, with an average usage time of only 0.74 s. The usability ratings for the remaining text links on the CTI homepage are presented in Table 2.

Table 1. Statistics for gaze plots

Metric Variable	Metric Name	Brief Description
V1	Time to first fixation	The time from the start of the stimulus display until the participant fixates on the AOI
V2	First Fixation Duration	Duration of the first fixation on the AOI
V3	Total Fixation Duration	Duration of all fixations within the AOI
V4	Fixation count	Number of times the participant fixates on an AOI
V5	Percentage Fixated	Percentage of participants that fixated at least once with an AOI
V6	Percentage clicked	Percentage of participants that clicked at least once within an AOI
V7	Time to First mouse click	The time in seconds until the first click is made within an AOI
V8	Time from First fixation to next mouse clicks	The time in seconds from the first fixation in the AOI to the next mouse click

4.4 Future Directions and Implications

The study's outcomes propose enhancements to the usability of the CTI e-services portal. One suggestion involves conducting regular usability testing with diverse student representation across majors and colleges to ensure alignment with user satisfaction. Additionally, it is advised that CTI web designers prioritize the needs, interests, preferences, and expectations of intended users during webpage design, content organization, and navigation planning. The study highlights a notable correlation between the usability of the e-services portal and the uptake of online services. Thus, enhancing portal usability is posited to elevate service utilization.

Another significant finding is the considerable lack of awareness among a notable portion of students regarding the available e-services through the CTI web site. To address this awareness deficit, it is recommended that educational campaigns be targeted at informing current students about these e-services. Implementation of these measures can render the CTI e-services portal more user-friendly and accessible, thereby enriching the overall student experience.

Table 2. Textual Link Evaluation Through Eight Metric Variables

Textual Links	V1 (Seconds)	V2	V3	V4	V5	V6	V7	V8	Average
Account	2.25	0.6	0.99	1.33	0.42	0.83	5.53	2.14	2
Change language (Arabic)	2.51	0.75	0.94	1.45	0.31	0.83	3.44	1.59	1.48
Change language (English)	1.42	0.91	1.07	1.20	0.28	0.94	1.85	1.27	1.12
e-services English	2.25	0.71	1.73	2.31	0.36	0.86	5.24	3.32	1.82
e-services Arabic	2.21	0.33	1.21	2.38	0.44	0.86	3.84	2.59	1.45
Login English	2.97	0.31	0.62	1.83	0.17	0	0	0	0.74
Login Arabic	2.61	0.33	1.17	2.5	0.33	0.58	7.52	5.52	2.57
Search Box	0.74	0.38	1.86	4.2	0.83	0.83	3.29	2.32	1.81
Average	2.12	0.54	1.20	2.15	0.39	0.72	3.83	2.34	

5 Conclusion

The study's goal was to investigate user search activity on the CTI web portal using eye-tracking methodology. Analysis of the eye-tracking data revealed various usability concerns impacting student satisfaction and performance. Surprisingly, sixty five percent of students had never utilized the university's online services, resulting in inconsistent and erratic visual behavior during the experiment.

The study's heatmap analysis revealed that the students' visual behavior while performing tasks was intense and disorganized, with their visual fixations being erratic and scattered. This implies that their search for the right links was not consistent throughout the experiment. Interestingly, this finding contradicts the results of a previous study [22] that found participants spent less time on the homepage. Instead, our research outcome demonstrates that students spent more time searching for the appropriate link to the required service on the e-services homepage. The eye tracking data from the study indicated that the most utilized textual links on the CTI online portal were "Login Arabic," "Account," and "e-services English," with "Login English" being less frequently used.

Based on the study's findings, the researchers propose several recommendations to improve the usability of the CTI web portal. They propose conducting regular usability testing with students from various majors and colleges to guarantee user satisfaction with the platform's features, services, and overall student experience. Furthermore, they advise web designers to consider the needs, interests, preferences, and expectations

of the target users when designing webpage layouts, arranging content, and structuring navigation. They also suggest that creating more awareness campaigns to educate present students about the e-services offered through the CTI main portal could improve usage and adoption of online services.

References

1. Gee, L.L.S., Dasan, J., Hasan, C.H.C.: Investigating the usability of universities' websites: upgrading visualization preference and system performance. Int. J. Interact. Mob. Technol. **16**(2) (2022)
2. Ahmi, A., Mohamad, R.: Evaluating accessibility of Malaysian public universities websites using AChecker and WAVE. J. Inf. Commun. Technol. (2016)
3. Undu, A., Akuma, S.: Investigating the usability of a university website from the users' perspective: an empirical study of Benue State University Website. Int. J. Comput. Inf. Eng. **12**(10), 922–929 (2018)
4. Iso, W.: 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). Int. Organ. Standardization **45**(9) (1998)
5. Caglar, E., Mentes, S.A.: The usability of university websites–a study on European University of Lefke. Int. J. Bus. Inf. Syst. **11**(1), 22–40 (2012)
6. Diwanji, V.S.: Improving accessibility and inclusiveness of university websites for international students: a mixed-methods usability assessment. Technol. Pedagog. Educ. **32**(1), 65–90 (2023)
7. Yilmaz, F.G.K., Yilmaz, R., Durak, H.Y., Keser, H.: Examination of students processes of searching information in education informatics network via eye tracking. World J. Educ. Technol. Curr. Issues **11**(1), 65–73 (2019)
8. Vlachogianni, P., Tselios, N.: Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. J. Res. Technol. Educ. 1–18 (2021)
9. Bataineh, E., Al-Bataineh, B.: An analysis study on how female college students view the web search results using eye tracking methodology. In: Proceedings of the International Conference on Human-Computer Interaction, Crete, Greece, pp. 22–27 (2014)
10. Nielsen, J., Pernice, K.: Eye Tracking Web Usability. Pearson, Berkeley (2010)
11. Rayner, K.: Eye movements in reading and information processing. Psychol. Bull. **124**, 372 (1998)
12. TDRA. https://tdra.gov.ae/en/media/press-release/2022/2021-the-golden-year-of-comprehensive-digital-transformation-in-the-uae
13. Alatawi, S.S.T., et al.: A new model for enhancing student portal usage in Saudi Arabia universities. Eng. Technol. Appl. Sci. Res. **11**(3), 7158–7171 (2021)
14. Abdelhakim, M.N., Carmichael, J.N., Ahmad, S.: Quality evaluation of university web portals: a student perspective. Int. J. Inf. Oper. Manag. Educ. **4**(3), 229–243 (2012)
15. Pan, B., Hembrooke, H.A., Gay, G.K., Granka, L.A., Feusner, M.K., Newman, J.K.: The determinants of web page viewing behavior: an eye-tracking study. In: ETRA 2004, pp. 147–154. ACM (2004)
16. Ehmke, C., Wilson, S.: Identifying web usability problems from eye tracking data (2007)
17. Jankovski, C., Schofield, D.: The eyes have it: using eye tracking technology to assess the usability of learning management systems in elementary schools. Eur. J. Educ. Stud. (2017)
18. Oyekunle, R., Bello, O., Jubril, Q., Sikiru, I., Balogun, A.: Usability evaluation using eye-tracking on E-commerce and education domains. J. Inf. Technol. Comput. **1**(1), 1–13 (2020)
19. Maslov, I., Nikou, S.: Usability and UX of learning management systems: an eye-tracking approach. In: 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–9. IEEE (2020)

20. Chynał, P., Falkowska, J., Sobecki, J.: Web page graphic design usability testing enhanced with eye-tracking. In: Karwowski, W., Ahram, T. (eds.) IHSI 2018. AISC, vol. 722, pp. 515–520. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73888-8_80
21. Almourad, M.B., Bataineh, E., Hussain, M., Wattar, Z.: Usability assessment of a university academic portal using eye tracking technology. Procedia Comput. Sci. **220**, 323–330 (2023)
22. Shahab, M.A., Iqbal, M.U., Srinivasan, B., Srinivasan, R.: Metrics for objectively assessing operator training using eye gaze patterns. Process. Saf. Environ. Prot. **156**, 508–520 (2021)
23. Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in WWW search. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 478–479 (2004)

Next Generation Communication System and Network Security

A Strategy of Joint Service Placement and Request Dispatching for LEO Satellite MEC Networks

Qian Tan, Mengying Li, Hao Wang, Kanglian Zhao, Wenfeng Li(✉), and Yuan Fang

School of Electronic Science and Engineering, Nanjing University, Nanjing 210033, China
leewf_cn@hotmail.com

Abstract. Low Earth Orbit (LEO) satellite Multi-Access Edge computing (MEC) is expected to realize the ubiquitous coverage for global users and provide low-delay MEC service. In LEO satellite MEC networks, although recent studies have solved the problem of requests dispatched to the appropriate edge servers, they mainly focus on the effective use of computing resources and ignore the need to pre-store a large amount of data to provide services. Due to the limited storage and computing resources of satellites, it is impossible to place all services and meet the requests of all users. Considering the significant correlation between service placement and request dispatching, we investigate the joint optimization of service placement and request dispatching in LEO satellite MEC networks, which is formulated as a Mixed Integer Linear Programming (MILP) problem. In order to avoid overloading of some satellites while other satellites are not fully utilized, the load balancing factor is introduced in this paper. Our goal is to improve the user served ratio, minimize the cost of request dispatching and improve the load balancing among satellites while ensuring the constraints of computation and storage resources. The simulation results verify that, compared with the baseline mechanism, the scheme of joint service placement and request scheduling with loading balancing (SPRDLB) that we proposed has better performance.

Keywords: LEO satellite · multi-access edge computing · service placement · request dispatching · load balancing

1 Introduction

With the rapid development of 5G networks, the number of the promising applications have emerged, such as virtual reality, augmented reality, industry automation, automatic driving, and so forth, which require massive computing capabilities. These computation-intensive applications pose a huge challenge to the computing capacity of resources-constrained terminal devices, which promotes the development of cloud computing [1]. Although cloud computing can significantly reduce the computing latency, the long transmission distance between users and cloud servers leads to high latency, which may fail to meet the needs of delay-sensitive applications, such as augmented reality. To address this problem, Multi-Access Edge computing (MEC) has been widely studied,

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 99–108, 2024.
https://doi.org/10.1007/978-3-031-59619-3_9

where the computation resources in the network edge are utilized to provide efficient computing services [2–5]. Although 5G networks have made great progress in providing fast and convenient communication services for mobile terminals, rural areas and remote islands still lack high-speed Internet access services, in which multi-access edge computing cannot be applied. Since the satellite network can provide seamless global coverage, the multi-access edge computing servers are deployed in the LEO satellite network to ensure that all users on the ground can access MEC services. The users can obtain MEC services through communication with LEO satellites.

Recently, satellite edge computing has attracted numerous attentions and become an emerging research direction. The research work in satellite edge computing system mostly focuses on computing offloading and resources allocation. The request dispatching is similar to computing offloading. In [6], considering the intermittent communication caused by the LEO satellite orbit, Wang et al. propose a game-based computing offloading framework to optimize the response time and energy consumption of all users. The resources allocation problem in the multi-users MEC system is studied in [7] and [8], which adopt the regression and deep Q learning algorithm respectively to tackle the problem of reasonable allocation of communication and computing resources. To fully exploit both communication and computing resources in LEO satellite networks, in [9], Jin et al. propose a C-RAN dynamic resources allocation framework based on deep learning to jointly optimize computing offloading and resources allocation. In order to reduce computational complexity, Wang et al. [10] propose to tackle the two sub-problems as a whole.

Although the above research considers the multi-user MEC system, they ignore the multi-service scenario. Placing services on a satellite node needs to occupy some storage resources [11], where the problem of service placement should be tackled. In [12], service placement in the multi-user MEC scenario with multi-services is investigated, however, they ignore the load balancing between nodes. In addition, the satellite network has the characteristics of high dynamic topology and uneven distribution of ground users, and the network load is prone to imbalance [13, 14]. Motivated by these views, this paper comprehensively investigates the service placement, request dispatching, as well as load balancing in LEO satellite MEC networks with multi-user, multi-MEC nodes and multi-service. We aim to optimize the total cost of all users, the user served ratio and the load balancing among satellites, where the formulated problem is a mixed integer linear programming (MILP) problem. Moreover, it is possible to dispatch the service request to that satellite only if the service is placed on the satellite. Given the closed relation between service placement and request dispatching, we propose a strategy of joint service placement and request dispatching with load balancing (SPRDLB).

The rest of this paper is organized as follows. Section 2 describes system model. Section 3 depicts SPRDLB formulation. Simulation and result analysis are presented in Sect. 4. Finally, Sect. 5 summarizes this paper.

2 System Model

In this paper, we consider a LEO multi-access edge computing network, as shown in Fig. 1. The multi-access edge computing server is deployed on the LEO satellite, which is called LEO-MEC node. The LEO constellation consists of many LEO satellites and can achieve globally seamless coverage.

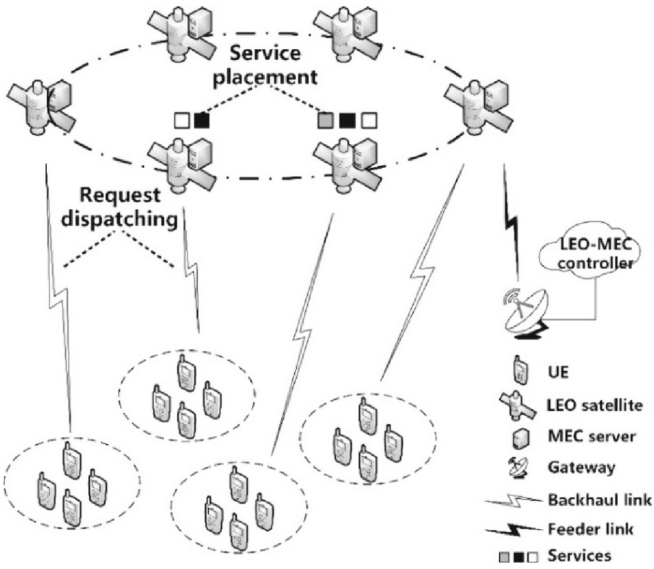


Fig. 1. System model of LEO satellite MEC networks.

Multiple services are placed on the LEO-MEC node, and users can obtain MEC services through communication with LEO satellites. If the requested service is placed on a nearby LEO-MEC node, it can be obtained by accessing the user or other satellites in the neighborhood. Moreover, the communication path doesn't involve many links, which reduces the traffic in the LEO network.

We assume that the user equipment (UE) on the ground accesses the LEO network only through one satellite during the snapshot [15], which is easy to realize in engineering. The access satellite is chosen according to some criterion [16, 17], which is out of the scope of this paper. Users send messages to the LEO-MEC global controller, including the required service and computing resources, and the global controller makes decisions by collecting information about users and LEO-MEC node. The satellite coverage area refers to the area where the user accesses the LEO constellation network through the satellite, and the entire area is divided into many coverage areas.

3 Problem Formulation

We consider that there are N areas and N LEO-MEC nodes, and K types of services that can be placed on the LEO-MEC node. We use n_i^k to represent the number of requests for service k in region i.

We define the service placement decision variable as the binary variable α_j^k, where $\alpha_j^k = 1$ represents service k is placed on LEO-MEC node j, otherwise service k is not placed on LEO-MEC node j. The request scheduling decision variable β_{ij}^k indicates the proportion of requests for service k in region i that are dispatched to LEO-MEC node j. Therefore, the amount of requests for service k in region i that are dispatched to LEO-MEC node j can be calculated as $n_i^k \beta_{ij}^k$, where $\beta_{ij}^k \in [0, 1]$.

We denote μ_j as the storage resources of LEO-MEC node j. Placing a service on a LEO-MEC node need to occupy some storage resource, and we assume that the storage resources required to place service k on LEO-MEC node is s_k. Therefore, the storage resource of LEO-MEC node j occupied by service k is $s_k \alpha_j^k$.

The computation resource of LEO-MEC node j is c_j. The number of requests that each LEO-MEC node can handle is limited due to constrained computation resource. The computation resource required to process requests for service k are denoted by η_k. Since the amount of requests for service k in region i which are dispatched to LEO-MEC node j is $n_i^k \beta_{ij}^k$, the total amount of computation resource required to process these requests is $n_i^k \beta_{ij}^k \eta_k$.

3.1 Constraints in LEO Satellite MEC Networks

There are three types of constraints in a LEO satellite MEC network, i.e., feasibility constraints, resources constraints, service request constraints.

Feasibility Constraints. The first basic constraint is caused by the scheme of service placement.

$$\beta_{ij}^k \leq \alpha_j^k, \forall i \in [1, N], \forall j \in [1, N], \forall k \in [1, K] \tag{1}$$

For a specific for the service k, it can be dispatched to a LEO-MEC node only if the service k is placed on that node. And the decision variables have to meet the following two constraints:

$$\alpha_j^k = \{0, 1\}, \forall j \in [1, N], \forall k \in [1, K] \tag{2}$$

$$0 \leq \beta_{ij}^k \leq 1, \forall i \in [1, N], \forall j \in [1, N], \forall k \in [1, K] \tag{3}$$

in which (2) ensures that the service placement decision is a binary variable and the value of that is 1 if the service is placed on the LEO-MEC node, otherwise is 0. In addition, (3) states that the request scheduling decision is a continuous variable, which indicates the ratio of requests for some service is between 0 and 1.

Resources Constraints. Because the storage and computation resources of a node are limited, so we have the following two constraints:

$$\sum_{k=1}^{K} s_k \alpha_j^k \leq \mu_j, \forall j \in [1, N] \tag{4}$$

$$\sum_{k=1}^{K} \sum_{i=1}^{N} n_i^k \beta_{ij}^k \eta_k \leq c_j, \forall j \in [1, N] \tag{5}$$

in which (4) is the storage resource constraint, which ensures that the storage resource required by the service placed on the node can't exceed the total storage resource of the LEO-MEC node. Moreover, (5) is the computation resource constraint ensures the computation resource required for requests to be dispatched to a node for processing can't exceed the total computation resource of that node.

Service Request Constraints. In a specific area, the services the user requests for are so different that we need to schedule the request to the LEO-MEC node with the corresponding service. We have

$$\sum_{j=1}^{N} \beta_{ij}^k \leq 1, \forall i \in [1, N], \forall k \in [1, K] \tag{6}$$

It ensures the sum of ratio of requests for some service from some region dispatched to all the LEO-MEC nodes can't be larger than 1.

3.2 SPRDLB Formulation

In order to improve the efficiency of utilizing the resources of the LEO-MEC network and to optimize the performance of the whole system, we introduce a load balancing factor for joint optimization of service placement and request scheduling. Three performance metrics are taken into account to optimize, which are request dispatching cost, the proportion of unserved users, and load balancing factor. The cost of request dispatching R_{ij} represents the cost of dispatching each request from region i to LEO-MEC node j. The cost of all requests to be served is $\sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} n_i^k \beta_{ij}^k R_{ij}$. The proportion of unserved users is $\sum_{k=1}^{K} \sum_{i=1}^{N} (1 - \sum_{j=1}^{N} \beta_{ij}^k)$. The load balancing factor is $\sigma = \left[\sum_{j=1}^{N} (p_j - p_{avg})^2 / N\right]^{1/2}$, where the computation resource required by LEO-MEC node j to process requests is $p_j = \sum_{k=1}^{K} \sum_{i=1}^{N} n_i^k \beta_{ij}^k \eta_k$, p_{avg} is the average computation resource required by each LEO-MEC node to process requests. We use σ to denote the variance of the computation resource required by the LEO-MEC nodes to process the request, therefore, a smaller value of σ indicates a more balanced load on each node. We formulate the problem of joint optimization service placement and request dispatching with load balancing factor as follows.

$$\begin{gathered} min_{\alpha_j^k, \beta_{ij}^k} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} n_i^k \beta_{ij}^k R_{ij} + r \sum_{k=1}^{K} \sum_{i=1}^{N} (1 - \sum_{j=1}^{N} \beta_{ij}^k) + \gamma \sigma \\ s.t.(1) - (6) \end{gathered} \tag{7}$$

The objective function (7) is composed of three parts. The first part is the total cost for all users to get service from LEO-MEC node. The second part is the proportion of users that can't be satisfied. The third part is the load balancing factor of LEO-MEC system.

For reader's convenience, useful notations used throughout the paper are listed in Table 1.

Table 1. Notations used in the paper.

Notation	Description
N	Number of LEO-MEC nodes
K	Number of service types
n_i^k	Number of requests for service k in the region i
α_j^k	Decision variables on service placement
β_{ij}^k	The proportion of requests for service k in region i that are dispatched to LEO-MEC node j
μ_j	The storage resources of LEO-MEC node j
s_k	The storage resources required for service k to be placed on LEO-MEC node
c_j	The computation resources of LEO-MEC node j
η_k	The computation resources required to process a request for service k
R_{ij}	The cost of a request in area i dispatched to LEO-MEC node j
r	The weight of unserved user ratio
γ	The weight of load balancing factor
σ	Load balancing factor

4 Simulation and Result Analysis

In this section, we evaluate the performance of the strategy we proposed. Firstly, we describe the simulation parameters setup. Secondly, we analyze the performance evaluation results.

4.1 Simulation Setup

The scenario we simulate is a LEO constellation network with 9 LEO satellites, which consists of a grid topology of 3 adjacent satellites in 3 adjacent orbits. The parameters in the simulation scenario are set as shown in Table 2. The types of service is 10, the number of users requesting service in each coverage area is a random value between 0 and 50. The storage resource for each LEO-MEC node is 10 units and the storage resource required for service placed in the LEO-MEC node is a random value between

1 and 4. The computation resource for a LEO-MEC node is set to 50 units, and the computation resource required by every service's single user is also a random value between 1 and 4.

We refer to the reference [14] to set the value of r. The weight of the load balancing factor γ is a constant. The larger the γ value is, the more balanced the load of the system, while it leads to a higher cost of the request dispatching. By changing the value of γ, it is found that too large a γ will lead to some undispatched requests. We also need to pay attention to the cost of request dispatching and the proportion of unserved users, so γ is set as 5.

From the above, it is easy to see that the problem we investigate is a mixed integer linear programming problem. Considering that there are already many MILP solvers, we utilize the mosek solver to solve it and obtain the optimal decision variables for service placement and request dispatching.

Table 2. Simulation parameters.

Parameter	Value
N	9
K	10
n_i^k	0–50
μ_j	10
s_k	1–4
c_j	50
η_k	1–4

The baseline mechanism is divided into two main parts. Firstly, for each LEO-MEC node, the services are sorted by the number of users requesting the services, and then it places sorted services on the LEO-MEC node one by one until the node don't have enough storage resources. Secondly, the request dispatching decision is the only variable to solve the problem. It is clear that the baseline mechanism solves the problem of service placement and request scheduling separately in sequence, while our proposed mechanism SPRDLB optimizes jointly.

4.2 Simulation Results

As shown in Fig. 2, with different values of r, the ratio of served users is higher than baseline mechanism when using SPRDLB, except $r = 10, 20, 30$. The simulation results show that SPRDLB doesn't outperform the baseline mechanism very much in terms of the proportion of users served.

We use the number of hops from users in coverage area i to LEO-MEC node j as the request dispatching cost R_{ij}. Figure 3 shows the request dispatching cost with different values of r, the average hop count when using SPRDLB is less than baseline mechanism, except $r = 40$.

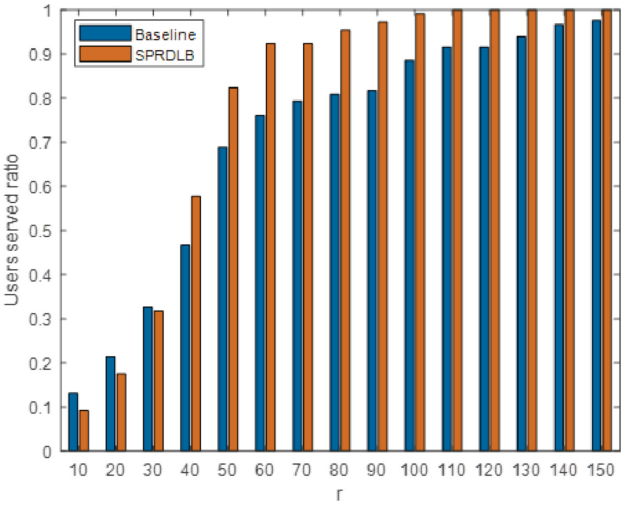


Fig. 2. Users served ration with different value of r.

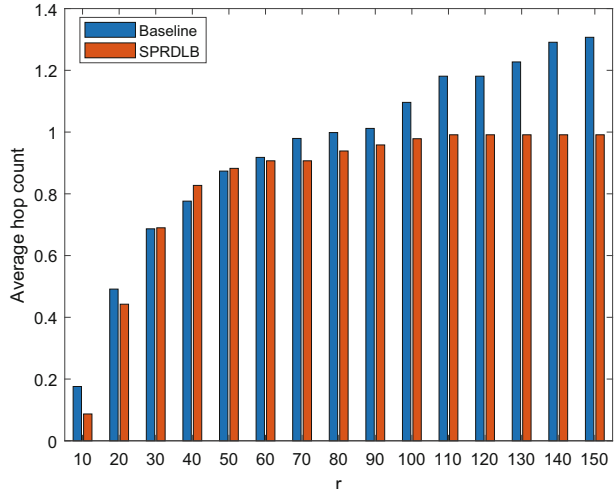


Fig. 3. The cost of request dispatching with different value of r.

Figure 4 depicts the objective function value solved by baseline mechanism and SPRDLB respectively with different values of r. It is clear that the objective value solved by SPRDLB is always less than baseline mechanism which verifies that SPRDLB can get the better objective value.

As can be seen in Fig. 5, with different values of r, the node load of SPRDLB is lower than or equal to that of the baseline mechanism, except $r = 30, 150$. On the whole, the simulation results show that SPRDLB outperforms the baseline mechanism in terms of the node load.

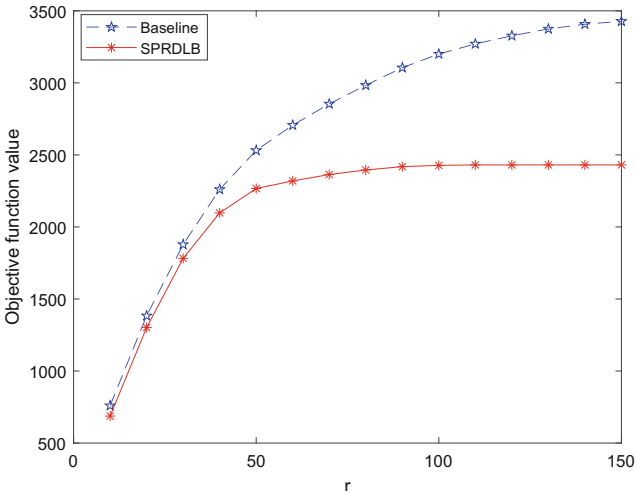


Fig. 4. Objective function value with different value of r

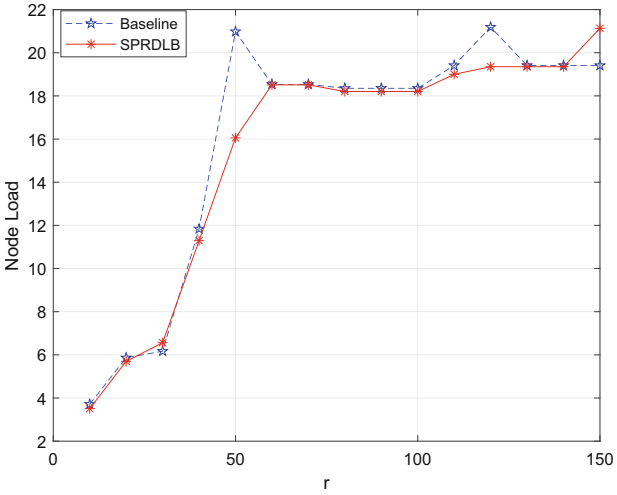


Fig. 5. LEO-MEC node loads with different value of r.

5 Conclusion

In this paper, we study the service placement and request dispatching in LEO-MEC networks with the constraint of computation and storage resources. Service placement and request scheduling are significantly related problems. Moreover, the highly dynamic topology of the satellite network is prone to the unbalanced load of system. To avoid wasting a large amount of resources, we propose a joint optimization mechanism for service placement and request scheduling by introducing a load balancing factor. Simulation results verify that the SPRDLB mechanism has better performance in terms of

served user ratio and dispatching cost compared with the baseline mechanism. With respect to node load, the SPRDLB also outperforms baseline mechanism in most cases.

References

1. Cheng, N., et al.: Space/aerial-assisted computing offloading for IoT applications: a learning-based approach. IEEE J. Sel. Areas Commun. **37**(5), 1117–1129 (2019)
2. Wang, Y., Zhou, J., Feng, G., Niu, X., Qin, S.: Blockchain assisted federated learning for enabling network edge intelligence. IEEE Network (2022)
3. Wang, L., Zhou, J., Wang, Y., Lei, B.: Energy conserved computation offloading for O-RAN based IoT systems. In: IEEE International Conference on Communications, ICC 2022, pp. 4043–4048 (2022)
4. Zhou, J., Feng, G., Yum, T.-S.P., Yan, M., Qin, S.: Online learning-based discontinuous reception (DRX) for machine-type communications. IEEE Internet Things J. **6**(3), 5550–5561 (2019)
5. Zhou, J., Feng, G., Yum, T.-S.P., Qin, S.: Actor-critic algorithm based discontinuous reception (DRX) for machine-type communications. In: 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–7 (2018)
6. Wang, Y., Yang, J., Guo, X., Qu, Z.: A game-theoretic approach to computation offloading in satellite edge computing. IEEE Access **8**, 12510–12520 (2020)
7. Nishiyama, H., Kudoh, D., Kato, N., Kadowaki, N.: Load balancing and QoS provisioning based on congestion prediction for GEO/LEO hybrid satellite networks. Proc. IEEE **99**(11), 1998–2007 (2011)
8. Zhang, Y., Zhou, X., Teng, Y., Fang, J., Zheng, W.: Resources allocation for multi-user MEC system: machine learning approaches. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp. 794–799 (2018)
9. Hussein, M., Abu-Issa, A., Elayyan, I.: Location-aware load balancing routing protocol for LEO satellite networks. In: 2018 International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, pp. 1–7 (2018)
10. Jin, X., Zhang, J., Sun, X., Zhang, P., Cai, S.: Computation offloading and resources allocation for MEC in C-RAN: a deep reinforcement learning approach. In: 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, pp. 902–907 (2019)
11. Wang, B., Feng, T., Huang, D.: A joint computation offloading and resources allocation strategy for LEO satellite edge computing system. In: 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, pp. 649–655 (2020)
12. Nishiyama, H., Tada, Y., Kato, N., Yoshimura, N., Toyoshima, M., Kadowaki, N.: Toward optimized traffic distribution for efficient network capacity utilization in two-layered satellite networks. IEEE Trans. Veh. Technol. **62**(3), 1303–1313 (2013)
13. Qiu, C., Yao, H., Yu, F.R., Xu, F., Zhao, C.: Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks. IEEE Trans. Veh. Technol. **68**(6), 5871–5883 (2019)
14. Li, C., Zhang, Y., Hao, X., Huang, T.: Jointly optimized request dispatching and service placement for MEC in LEO network. China Commun. **17**(8), 199–208 (2020)
15. Yang, L., Cao, J., Liang, G., Han, X.: Cost aware service placement and load dispatching in mobile cloud systems. IEEE Trans. Comput. **65**(5), 1440–1452 (2016)
16. Sağ, E., Kavas, A.: Modelling and performance analysis of 2.5 Gbps inter-satellite optical wireless communication (IsOWC) system in LEO constellation. J. Commun. **13**(10), 553–558 (2018)
17. Zaghloul, A., Shaalan, A.A., Kasban, H., Ashraf, A.: Evaluation of free space optics uplink availability to LEO satellite using climatic data in Cairo. J. Commun. **16**(8), 301–310 (2021)

A Dual Phase Genetic Algorithm with Aggregated Search for Fast Initial Access in 5G Millimeter Wave Communication

Krishnan B. Iyengar, Raghavendra Pal, and Upena Dalal(✉)

Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India
{ds21ec001,raghavendrapal,udd}@eced.svnit.ac.in

Abstract. 5G Millimeter Wave (mmWave) Communication between the base station (BS) and the user equipment (UE) involves a Multiple-Input Multiple-Output (MIMO) system where both BS and UE have many antennas. Initial Access (IA) in this context is the problem of establishing a directional link between the BS and UE, but finding the optimal beams can be prohibitively expensive in terms of delay and computation. Genetic Algorithms (GAs) can solve complex problems effectively, and in this case, they can be used to iteratively search for the optimal beams. We propose a dual phase GA that splits the GA process into two successive phases that uses different operations in each phase. It also navigates the search space in a smart manner, increasing the convergence rate to the optimal beamformer per iteration. We have analyzed the effect of this approach in terms of Capacity achieved vs number of transmit and receive antennas at BS and UE, total transmitted power, and number of iterations. It shows improved performance in terms of maximum Capacity achieved, reduced power consumption, and especially reduced IA delay.

Keywords: 5G · mmWave · Beamforming · Initial Access · GA

1 Introduction

5G millimeter wave (mmWave) communication is a promising technology that has attracted a lot of interest due to high bandwidth available in that spectrum compared to conventional sub-6 Gigahertz (GHz) frequencies [1]. However, it faces significant hurdles in the form of propagation losses and especially blockage from obstacles in the environment. Due to the small wavelength of the electromagnetic waves used in mmWave communication, a considerably larger number of antennas can be used compared to sub-6 GHz communication in a given physical space. These antennas can then be used to form highly directional beams to compensate for the propagation losses and blockage encountered by mmWaves. There may be a large number of antennas at the base station (BS), or user equipment (UE) or both. An additional way to compensate for these issues is to use intelligent reflective surfaces (IRS) [2]. Filter bank multi-carrier (FBMC) modulation may also be used in mmWave communication [3].

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 109–120, 2024.
https://doi.org/10.1007/978-3-031-59619-3_10

In the context of mmWave communication, Initial Access (IA) is the problem of establishing a high capacity communication link between the BS and the UE [4]. However, due to the large number of antennas, the corresponding beamspace is large as well, and there may significant delays or interruptions due to the BS and UE attempting to find the optimal set of beams by searching the beamspace, to achieve the required capacity. This work focuses on one particular approach of solving this problem.

Due to the large beamspace, we may use Genetic Algorithms (GAs) as a possible way to solve the IA problem. GAs are a set of search and optimization techniques inspired by natural selection and evolution in biology [5]. They use directed randomized searches with modifications and many iterations to find close-to-optimal solutions in complex problems with large search spaces, especially ones where no efficient universal algorithm exists, or is computationally complex to compute.

1.1 Related Work

The techniques for solving the IA problem can be categorized into six types:

1. Probabilistic/Statistical approach
2. Hierarchical/Iterative approach
3. Exhaustive Search approach
4. Machine Learning approach
5. Meta-Heuristics approach
6. Context-Information Based approach

The Exhaustive Search techniques perform a brute-force search of the whole beamspace. This results in considerable delay, and is not feasible for large search spaces, but guarantees an optimal solution and can be used if delay is not a concern [6].

Hierarchical techniques vary the size of the search space itself, and use the variation to iteratively search for the optimal solution at each step, however this method does not achieve a high beamforming gain [6].

Probabilistic/Statistical techniques use statistical information of mobility and connection probability of the user to improve beamforming gain, but require prior information regarding user and system behavior [7].

Context-Information based techniques use information regarding UE and BS locations, quality-of-service (QoS) demanded, behavior of user, and other details which are specific to a given context to reduce IA delay and improve performance, though getting these specific details may not be possible [8]. A recent work that falls in this category is [9]. It involves creation of a knowledge database that stores information regarding beam-width, beam-steering and received power to reduce detection time.

Machine Learning (ML) techniques can be applied by training a given ML model on the dataset containing user behavior and system parameter information. This can be particularly effective in very complex system models containing many users [10]. However, the performance of the ML model is highly dependent on the accuracy of the dataset, and the information contained within may change in a dynamic environment. A recent ML based approach involves training a deep neural network (DNN) to convert the received signal strength for a few test beams to the beam with the maximum received signal strength [11].

Meta-Heuristic techniques can solve optimization problems using the system model only, without needing specific context information or other knowledge. They include genetic algorithms, which have been used by [12–15] to solve the IA problem.

In [12], the authors propose a beamforming scheme in multi-user setup using a genetic algorithm which uses the concept of a fittest individual and generation of random individuals which are the beamformer matrices.

In [13], the authors use a similar GA used in [12], and analyzed the effect of user mobility, and used the high spatial correlation between successive user positions to improve beamformer gain in future GA iterations. They also analyze the effect of user collaboration in the multi-user case.

In [14], the authors propose a modified genetic algorithm that uses two additional operators, the discrete crossover operator and the real mutation operator, in each iteration of the GA process. They show this to increase capacity achieved, lower power consumption, and lower outage probability compared with the GA used in [13].

In [15], another modified genetic algorithm has been proposed. It involves using advanced elite tournament selection, a hybrid of two mutation operators, reverse sequence mutation and partial shuffle mutation, and a directional crossover operator. This approach shows even further increase in capacity achieved, and lower power consumption and outage probability compared to [14].

1.2 Contribution

This work proposes a meta-heuristic technique based on genetic algorithms. It uses a Dual-Phase GA which splits the total number of iterations into two phases. Phase 1 consists of the same GA algorithm used in [14], while Phase 2 consists of minor directional modifications to the beamformer matrices, and aggregating/combining the effect of multiple modifications so long as each of them individually give an improvement in capacity achieved. The result of using this dual phase approach has been analyzed in terms of Capacity Achieved vs Number of BS/UE Antennas, capacity achieved vs transmitted power at BS, and capacity achieved vs number of iterations. This approach shows improved capacity achieved, especially at large number of BS/UE antennas, and reduced transmit power required at BS for a given capacity.

The effect of this approach can most clearly be seen in the results obtained on capacity vs number of iterations, as the second phase rapidly converges on the local optimum in less than a 100 GA iterations.

The rest of the paper consists of 4 Sections, Sect. 2 explains the system model, Sect. 3 explains GAs and the modified GA technique used in [14], Sect. 4 explains the proposed dual phase approach and the 2^{nd} phase in particular. Section 5 analyzes the simulation results, and Sect. 6 states the conclusion.

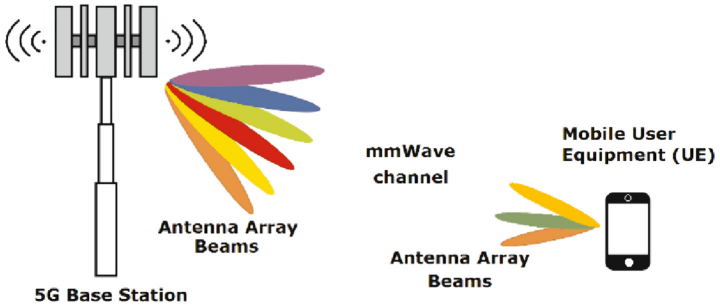


Fig. 1. System Model.

2 System Model

We consider a system model shown in Fig. 1. It contains a BS with M antennas and a UE with N antennas. A beamforming matrix is used by both the BS and the UE, the BS transmits the signal and the UE receives it. The received signal at the UE is given by [14]:

$$\boldsymbol{y}(t) = \sqrt{\frac{P}{M}}\boldsymbol{U}(t)^H \boldsymbol{H}(t)\boldsymbol{V}(t)\boldsymbol{x}(t) + \boldsymbol{z}(t) \tag{1}$$

P is the total transmitted power, $\boldsymbol{V}(t) \in C^{M\times M}$, $\boldsymbol{U}(t) \in C^{N\times N}$ are the beamforming matrices used by the BS and UE respectively. $\boldsymbol{x}(t) \in C^{M\times 1}$ is the message signal, $\boldsymbol{z}(t) \in C^{N\times 1}$ is the IID gaussian noise vector, and $\boldsymbol{H}(t) \in C^{N\times M}$ is the channel matrix. The time index is unnecessary as we are considering the channel to remain constant for the duration and is omitted in the rest of the paper.

The channel model used is the Saleh-Valenzuela extended geometric channel model [16], which is also the model used in [14]. It consists of N_c clusters of propagation paths and N_l propagation paths per cluster. The total number of paths between the BS and UE is hence $N_c \times N_l$. The channel matrix $H(t)$ is given by:

$$H = \sqrt{\left(\frac{MN}{N_c N_l}\right)} \sum_{i=1}^{N_c} \sum_{l=1}^{N_l} \beta_{il} A_r(\phi_{il}^r, \theta_{il}^r) A_t^H(\phi_{il}^t, \theta_{il}^t) \tag{2}$$

where β_{il} is the small scale fading complex coefficient associated with the i^{th} cluster and the l^{th} sub path. A_r and A_t are the received and transmit antenna array responses respectively. They are a function of the angles of arrival (AoA) and angles of departure (AoD) given by $\phi_{il}^r, \theta_{il}^r$ and $\phi_{il}^t, \theta_{il}^t$ respectively. The expressions for them are given by [17]:

$$A_r\left(\phi_{il}^r, \theta_{il}^r\right) = \frac{1}{N}[1, e^{\frac{j2\pi}{\lambda} d\left(\sin\left(\phi_{il}^r\right)\sin\left(\theta_{il}^r\right)+\cos\left(\theta_{il}^r\right)\right)}, \ldots, e^{\frac{j2\pi}{\lambda} d\left((N-1)\sin\left(\phi_{il}^r\right)\sin\left(\theta_{il}^r\right)+(N-1)\cos\left(\theta_{il}^r\right)\right)}] \tag{3}$$

$$A_t\left(\phi_{il}^t, \theta_{il}^t\right) = \frac{1}{M}[1, e^{\frac{j2\pi}{\lambda} d\left(\sin\left(\phi_{il}^t\right)\sin\left(\theta_{il}^t\right)+\cos\left(\theta_{il}^t\right)\right)}, \ldots, e^{\frac{j2\pi}{\lambda} d\left((M-1)\sin\left(\phi_{il}^t\right)\sin\left(\theta_{il}^t\right)+(M-1)\cos\left(\theta_{il}^t\right)\right)}] \tag{4}$$

where d is the distance between antenna array elements and λ is the wavelength.

The beamformer matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are constructed by a DFT-based codebook [18]. The BS codebook $\boldsymbol{W_t} \in C^{M \times N}{}_{vec}$ and the UE codebook $\boldsymbol{W_r} \in C^{N \times N}{}_{vec}$ are given by:

$$W_t(m, u) = e^{\left(-j\frac{2\pi}{N_{vec}}(u-1)(m-1)\right)} \tag{5}$$

$$W_r(n, u) = e^{\left(-j\frac{2\pi}{N_{vec}}(u-1)(n-1)\right)} \tag{6}$$

Table 1. Channel Model Parameters.

Parameters	Value Assigned
N_c	5
N_l	10
β_{il}	~CN(0,1)
ϕ_{il}^r	~U [0,2π]
θ_{il}^r	~Laplace(μ,1),
ϕ_{il}^t	~U [0,2π]
θ_{il}^t	~Laplace(μ,1),
d	λ/2

Here, $m = 1, \cdots, M$, $n = 1, \cdots, N$, and $u = 1, \cdots, N_{vec}$. N_{vec} denotes the codebook size and must be selected appropriately to satisfy $N_{vec} \geq \max(M, N)$. Larger codebook sizes result in higher capacity achieved by the optimal beamformers, but it also increases the search space (beamspace), and so cannot be increased arbitrarily as higher values lead to correspondingly high IA delay. The parameters used for this system model are given in Table 1.

2.1 Performance Metrics

The metric we use to analyze and evaluate the proposal is Capacity. If the maximum number of GA iterations is given by N_{it}, the Shannon capacity (Bits/second) for the kth iteration is given by [12]:

$$C(k) = (1 - \alpha k) \cdot BW \cdot \log_2(1 + SNR_k) \tag{7}$$

Here, $k = 1, 2, \cdots N_{it}$, and α is the cost factor for running each iteration of the GA, and it is assumed that $\alpha \cdot N_{it} < 1$. BW is the bandwidth of the system. SNR_k denotes the signal-to-noise ratio for the kth iteration. It is given by:

$$SNR_k = \frac{\frac{P}{M}\|U_k^H H V_k\|^2}{BWN_0} \tag{8}$$

where $||\cdot||$ is the Frobenius norm of a matrix, and N_0 is the Noise Power Spectral Density. The value of N_0 is set such that $BWN_0 = 1$. An important note to make is that the capacity equation with a non-zero α results in decreasing capacity for each additional iteration used by the GA algorithm. Thus, it is important to stop the algorithm once no more improvements in capacity are being found.

3 Genetic Algorithms

Genetic Algorithms can solve complex problems using iterated directed searches of a given search space [5]. In general, they involve initializing a random 'population' of 'individuals', then measuring the 'fitness' of each individual and applying various operators to some of them based on some criteria to generate a new population of different individuals, and then repeating the process. The process stops when the maximum number of iterations is reached or a stopping criterion is met. The fittest individual of the last generation is the output of the GA.

The terms used in GAs are mostly derived from evolutionary biology. They are: (1) Gene, (2) Chromosome, (3) Individual, (4) Parent, (5) Child, (6) Fitness, (7) Fitness Function, (8) Generation and (9) Search Space.

In the context of our system model, gene corresponds to a given column of the $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices, chromosome corresponds to either $\boldsymbol{U}$ or the $\boldsymbol{V}$ matrix, individual corresponds to a pair of $\boldsymbol{U}$,$\boldsymbol{V}$ matrices that create a beamforming pair used for communication between BS and UE. Fitness corresponds to the Capacity achieved by a given $\boldsymbol{U}$,$\boldsymbol{V}$ pair, Fitness function is a Shannon capacity equation given by (7). Generation corresponds to a given iteration of the algorithm, and Search space corresponds to the codebooks for BS and UE, as the columns for $\boldsymbol{U}$,$\boldsymbol{V}$ are selected from columns in their respective codebooks.

The steps of the Genetic Algorithm for the first phase of this dual phase approach are the same as the one in [14]:

1. Randomly Generate N_{pop} pairs beamforming matrices $\boldsymbol{U,V}$, these form the individuals of the first generation.
2. Compute fitness of all individuals in the generation.
3. Select N_{elite} fittest individuals and copy them into the next generation of individuals, this is termed Elitism.
4. Conduct a tournament to select $(N_{pop} - N_{elite})/2$ parents for the next step.
5. Apply crossover operator with probability P_{cross} to each pair of parents. These 'children' are copied into the next generation.
6. Apply mutation operator with P_{mut} probability to a selected child from previous step, or, with $1 - P_{mut}$ probability, modify 10% of columns of the fittest individual. Repeat this to generate $(N_{pop} - N_{elite})/2$ individuals. These individuals are copied into the next generation.
7. Form the next generation via individuals generated from the previous steps and repeat the process from step 2 until the stopping criterion is reached or N_{it} iterations have been performed.
8. The fittest individual in the last generation is the output of the algorithm.

The $\boldsymbol{U}$, $\boldsymbol{V}$ pair of the fittest individual is then used for communication between the BS and the UE.

4 Proposed Dual Phase GA with Aggregated Search

The GA algorithm given in the previous section is the one proposed by [14]. That algorithm, as well as the ones given in [12] and [13], use the same operations for all iterations and performs a wide search of the beamspace which requires testing with highly random column modifications for each individual. This is useful for when we have a lot of iterations over which to search for the optimal beamformer, but when the current iteration k is approaching the maximum number of iterations N_{it} or when we are no longer gaining improved capacity with each iteration due to a non-zero α, it is better to switch to a different set of operations in order to converge to the local optimum quickly, rather than continuing to perform a wide search of the beamspace.

This leads to the proposed Dual Phase GA, which splits the total iteration 'budget' into two phases with different number of iterations allotted to each phase, Phase 1 has N_{it} iterations, and Phase 2 has N_{itsec} iterations.

Phase 1 uses the same elitism, crossover and mutation operators that were described in the previous section and the one used in [14]. When Phase 1 iterations are complete, Phase 2 begins, which uses a different operator and an aggregation approach to quickly search for the individual with the highest local fitness.

The steps of this second phase are:

1. Initialize a set of modification matrices $\boldsymbol{U_{mod}} \in \mathbb{Z}^{N_{pop} \times N}$ and $\boldsymbol{V_{mod}} \in \mathbb{Z}^{N_{pop} \times M}$ which stores the modification made to each column of every individual in the previous iteration.
2. Compute fitness of all individuals in the generation.
3. Compare the fitness of each individual with the first individual in the population.
4. Add the rows of $\boldsymbol{U_{mod}}$,$\boldsymbol{V_{mod}}$ corresponding to each individual that was fitter than the first individual to give $\boldsymbol{U_{sum}} \in \mathbb{Z}^{1 \times N}$ and $\boldsymbol{V_{sum}} \in \mathbb{Z}^{1 \times M}$. This is the step where we perform the aggregation.
5. Modify the columns of the $\boldsymbol{V}$ matrix of the first individual using $\boldsymbol{V_{sum}}$ as follows: for each element in $\boldsymbol{V_{sum}}$, if it is negative, shift the corresponding column with the same index in $\boldsymbol{V}$ left w.r.t. the BS codebook. Apply a similar procedure to the columns of $\boldsymbol{U}$ and $\boldsymbol{U_{sum}}$. This is the step where we apply the effect of aggregation to improve the fitness of the first individual.
6. Generate $N_{pop} - 1$ individuals by copying the first individual, then shifting each column, with P_{mutsec} probability, of the $\boldsymbol{U}$,$\boldsymbol{V}$ matrices of the copied individual, either left or right (with 50% probability) w.r.t. the UE and BS codebooks respectively.
7. The first individual and the other $N_{pop} - 1$ individuals together form the next generation. Repeat the process from step 2 for N_{itsec} iterations.
8. The fittest individual after the last generation is the output of the second phase of the algorithm.

After both Phase 1 and Phase 2 are completed, the fittest individual's pair of beamforming matrices $\boldsymbol{U_{best}}$,$\boldsymbol{V_{best}}$ are used for communication between the BS and UE.

Table 2. Simulation Parameters.

Parameters	Value Assigned
Channel realizations	500
Number of Antennas at BS M	4–64
Total Transmit Power	10 dB
Tournament Competitors T_{tourn}	2
Population *Npop*	30 individuals
Number of Elites *Nelite*	2
Number of Secondary Iterations *Nitsec*	100, 0
Size of Codebook *Nvec*	120
Mutation Probability *Pmut*	5%
Number of GA Iterations N_{it}	900, 1000
Number of Antennas at UE N	4–64
Crossover Probability *Pcross*	95%
Phase 2 Mutation Probability *Pmutsec*	1–5%
Bandwidth (BW)	1 GHz

5 Performance Analysis and Simulation Results

We have analyzed the performance of this approach via simulations. We consider 500 channel realizations, and 10^3 GA iterations for each one with Phase 1 having 900 iterations and Phase 2 having 100 iterations. N_{pop} is set to 30, and $N_{elite} = 2$. Other simulation parameters are given in Table 2.

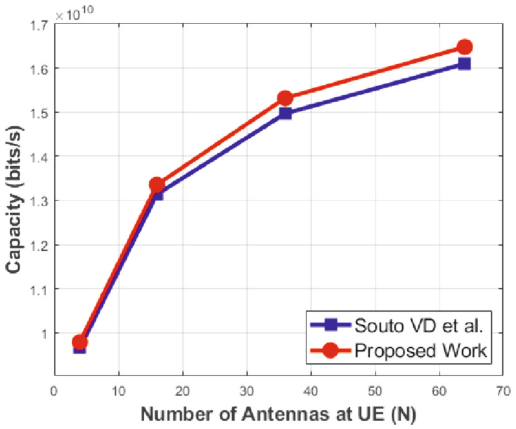


Fig. 2. Capacity vs Number of antennas at UE, with $M = 64$, $N_{vec} = 120$, $P = 10$ dB, $P_{cross} = 95\%$, $P_{mut} = 5\%$, $P_{mutsec} = 2\%$ and $\alpha = 0$.

From Fig. 2 we can see that increasing the number of antennas improves the effectiveness of the dual phase approach compared to the algorithm in [14], this is due to larger beamspace causing the final output of Phase 1 to be further away from the local optima, which Phase 2 can rapidly converge to in 100 iterations. This results in a *~2.3%* improved capacity achieved at $M = 64$, $N = 64$.

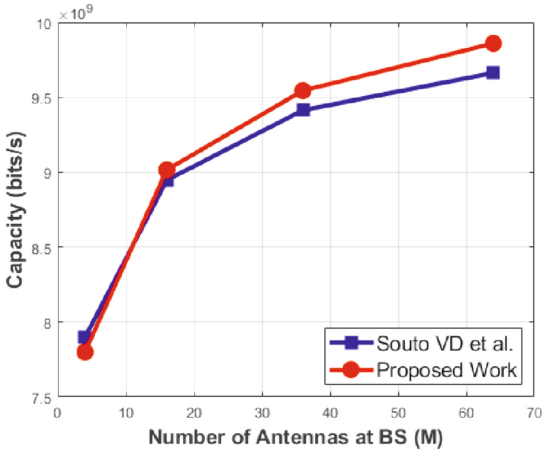


Fig. 3. Capacity vs Number of antennas at BS, with $N = 4$, $N_{vec} = 120$, $P = 10$ dB, $P_{cross} = 95\%$, $P_{mut} = 5\%$, and $\alpha = 0$.

From Fig. 3, we can see that low number of antennas, for example, $M = 4$, $N = 4$ hardly leads to an improvement. This approach requires a large beamspace, for example, $M = 64$, $N = 4$ for a significant capacity increase. Additionally, the value of P_{mutsec} for this figure in particular is not constant but is a function of M and N, i.e., $P_{mutsec} =$

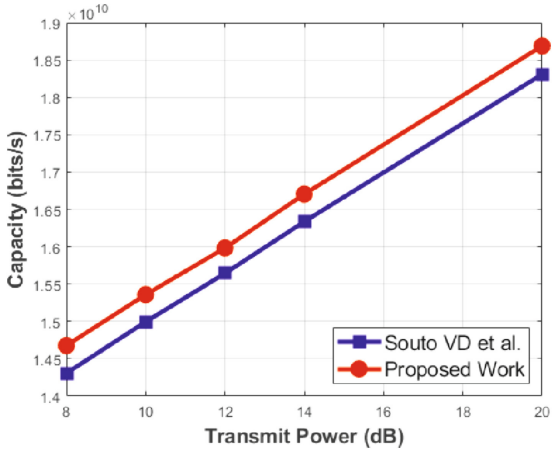


Fig. 4. Capacity vs Transmitted Power at BS, with $M = 64$, $N = 36$ $N_{vec} = 120$, $P_{cross} = 95\%$, $P_{mut} = 5\%$, $P_{mutsec} = 2\%$, and $\alpha = 0$.

$1/(M + N)$. This is required because a high value of P_{mutsec} relative to the beamspace causes the Phase 2 algorithm to be unable to detect which of the small random changes produced the improvement in capacity achieved. While a low value of P_{mutsec} relative to the beamspace may result in no changes at all for a given set of $\boldsymbol{U}$,$\boldsymbol{V}$ matrices, which results in wasted computation.

From Fig. 4 we can see that for the same transmit power, the proposed dual phase approach has a higher capacity achieved, and conversely, for the same capacity achieved, the proposed work requires lower transmitted power.

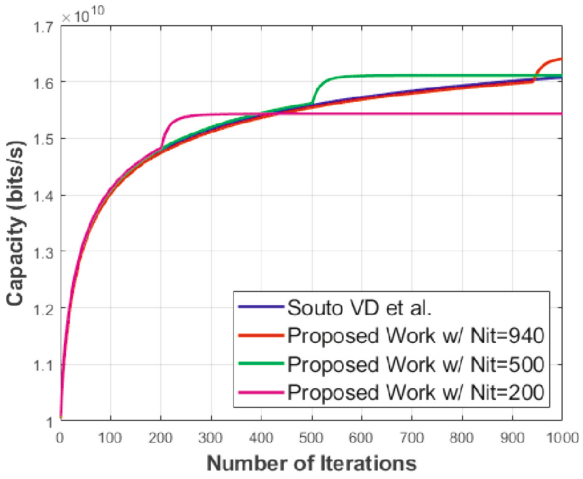


Fig. 5. Capacity vs Number of Iterations at various Phase 2 start points with $\alpha = 0$, $M = 64$, $N = 64$, $N_{vec} = 120$, $P = 10$ dB, $P_{cross} = 95\%$, $P_{mutsec} = 1\%$, and $P_{mut} = 5\%$.

From Fig. 5, we can see that the effect of Phase 2 is to quickly converge to the local optimum, but after 100 iterations or so, we observe that it has reached the optimum and can improve no further, whereas Phase 1 continues to improve via its wider, more random searches. Therefore, it is very important to not set the 'switch-over' point from Phase 1 to Phase 2 very low, as that can lead to highly suboptimal results.

In Fig. 6, we see the effect of having a non-zero α, where we observe that there is a point after which no significant improvements are being found by the GA process, and the penalty factor α causes the capacity to decrease at a linear rate. This implies an optimum point at which to switch the algorithm from Phase 1 to Phase 2, which is at the capacity peak of Phase 1. We can also observe that the optimum switch-over point is now lower compared to the case of $\alpha = 0$.

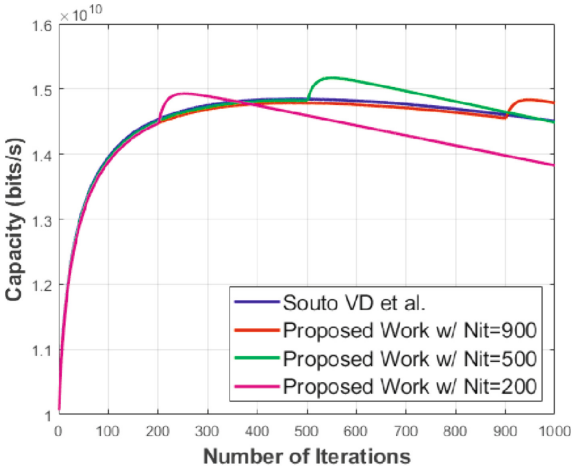


Fig. 6. Capacity vs Number of Iterations at various Phase 2 start points with $\alpha = 10^{-4}$, $M = 64$, $N = 64$, $N_{vec} = 120$, $P = 10$ dB, $P_{cross} = 95\%$, $P_{mutsec} = 1\%$, and $P_{mut} = 5\%$.

6 Conclusion

In this paper, we have proposed a Dual Phase Genetic Algorithm that splits the GA into two phases where the second phase uses small changes to the beamformer columns and an aggregation mechanism that adds only the beamformer changes that are beneficial in terms of fitness to quickly converge to the local optimum. It has shown an improvement in capacity achieved (~2.3% at M = 64, N = 64), lower transmit power required, and in particular, a very high convergence rate to the optimum value as long as we choose the switch-over point from Phase 1 to Phase 2 appropriately. We have also seen that for a non-zero α (10^{-4}), the switch-over point which has the maximum capacity reduces. Future work may include analyzing the effect of the aggregation mechanism on a mobile UE with a corresponding time varying channel.

References

1. Uwaechia, A.N., Mahyuddin, N.M.: A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: feasibility and challenges. IEEE Access **8**, 62367–62414 (2020)
2. Nandan, S., Abdul Rahiman, M.: Intelligent reflecting surface (IRS) assisted mmWave wireless communication systems: a survey. J. Commun. **17**(9), 745–760 (2022). https://doi.org/10.12720/jcm.17.9.745-760
3. Khudhair, S.A., Singh, M.J.: Performance evaluation of the use of filter bank multi-carrier waveform in different mmWave frequency bands. J. Commun. **16**(1), 36–41 (2021). https://doi.org/10.12720/jcm.16.1.36-41
4. Giordani, M., Mezzavilla, M., Zorzi, M.: Initial access in 5G mmWave cellular networks. IEEE Commun. Mag. **54**(11), 40–47 (2016)
5. Sivanandam, S.N., Deepa, S.N.: Genetic Algorithms, pp. 15–37. Springer, Heidelberg (2008)

6. Wei, L., Li, Q., Wu, G.: Exhaustive, iterative and hybrid initial access techniques in mmWave communications. In: 2017 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE (2017)
7. Soleimani, H., Parada, R., Tomasin, S., Zorzi, M.: Statistical approaches for initial access in mmWave 5G systems. Trans. Emerg. Telecommun. Technol. **32** (2021)
8. Abbas, W.B., Zorzi, M.: Context information based initial cell search for millimeter wave 5G cellular networks. In: 2016 European Conference on Networks and Communications (EuCNC), pp. 111–116 (2016)
9. Leoni, E., Guidi, F., Dardari, D.: A low-latency initial access technique for next 5G systems. In: 2020 IEEE International Conference on Communications, ICC 2020 Dublin, Ireland, pp. 1–6 (2020). https://doi.org/10.1109/ICC40277.2020.9149329
10. Alkhateeb, A., Alex, S., Varkey, P., Li, Y., Qi, Q., Tujkovic, D.: Deep learning coordinated beamforming for highly-mobile millimeter wave systems. IEEE Access **6**, 37328–37348 (2018)
11. Cousik, T.S., Shah, V.K., Reed, J.H., Erpek, T., Sagduyu, Y.E.: Fast initial access with deep learning for beam prediction in 5G mmWave networks. In: 2021 IEEE Military Communications Conference, MILCOM 2021, San Diego, CA, USA, pp. 664–669 (2021). https://doi.org/10.1109/MILCOM52596.2021.9653011
12. Guo, H., Makki, B., Svensson, T.: A genetic algorithm-based beamforming approach for delay-constrained networks. In: 2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pp. 1–7. IEEE (2017)
13. Guo, H., Makki, B., Svensson, T.: Genetic algorithm based beam refinement for initial access in millimeter wave mobile networks. Wirel. Commun. Mob. Comput. (2018)
14. Souto, V.D.P., Souza, R.D., Uchôa-Filho, B.F., Li, Y.: A novel efficient initial access method for 5G millimeter wave communications using genetic algorithm. IEEE Trans. Veh. Technol. **68**(10), 9908–9919 (2019)
15. Rasheed, I.: An effective approach for initial access in 5G-millimeter wave-based vehicle to everything (V2X) communication using improved genetic algorithm. Phys. Commun. **52**(C) (2022)
16. Samimi, M.K., Rappaport, T.S.: 3-D statistical channel model for millimeter-wave outdoor mobile broadband communications. In: 2015 IEEE International Conference on Communications (ICC), pp. 2430–2436 (2015)
17. Stutzman, W.L., Thiele, G.A.: Antenna Theory and Design. Antenna Theory and Design. Wiley (2012)
18. Wan, L., Zhong, X., Zheng, Y., Mei, S.: Adaptive codebook for limited feedback MIMO system. In: 2009 IFIP International Conference on Wireless and Optical Communications Networks, pp. 1–5 (2009)

Channel Estimation for RIS-Assisted Massive MIMO with Diffusion Model

Xiaofeng Liu[1,2(✉)], Xiao Fu[1,2], Xinrui Gong[1,2], Jiyuan Yang[1,2], and Xiqi Gao[1,2]

[1] National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
{xf_liu,fu_xiao,xinruigong,jyyang,xqgao}@seu.edu.cn
[2] Purple Mountain Laboratories, Nanjing 211100, China

Abstract. Reconfigurable intelligent surface (RIS) has received widespread attention as a critical enabling technology for next generation wireless communications. Accurate channel state information (CSI) is fundamental for RIS to reach its full potential. Thanks to the powerful latent representation from data, generative artificial intelligence (AI) has the potential as a strong driver for the highly intelligent and digital twin of 6G. In this paper, we propose a generative diffusion model (DM)-based cascaded channel estimation (CE) algorithm through unsupervised learning for RIS-aided massive multiple-input multiple-output (MIMO) systems. The DM can effectively capture the implicit prior of the cascaded channel, and the received signal is used as conditional information to precisely guide the channel recovery. Simulation results validate that the proposed algorithm achieves superior estimation accuracy with reduced pilot overhead.

Keywords: RIS · channel estimation · deep learning · diffusion model

1 Introduction

Reconfigurable intelligent surface (RIS) has been envisioned as a promising technology for future wireless communications, which can reshape the electromagnetic propagation environment with a large number of passive reflecting elements to enhance the range and capacity of communication systems [1]. Massive multiple-input multiple-output (MIMO), as one of the key technologies for 5G, brings a considerable improvement in spectral and energy efficiency and will also perform an essential role in the 6G-and-beyond wireless communications [2–4]. Due to the hardware-friendly, power-efficient and deployment-easily features, RIS can be well incorporated into massive MIMO systems. The performance gains provided by RIS-aided massive MIMO are highly dependent on

This work was supported by the National Key R&D Program of China under Grant 2018YFB1801103, the Jiangsu Province Basic Research Project under Grant BK20192002, the Key R&D Plan of Jiangsu Province under Grant BE2022067, and the Huawei Cooperation Project.

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 121–135, 2024.
https://doi.org/10.1007/978-3-031-59619-3_11

accurate channel state information (CSI). Since passive elements do not have signal processing capabilities, it is always necessary to estimate the cascaded transmitter-RIS-receiver channel. The cascaded channel does not follow the classical Rayleigh fading model, which limits the existing algorithms' performance [5]. Therefore, accurate cascaded channel estimation (CE) with the reduced pilot resource remains the focus of research to date.

To address the challenges of cascaded CE in RIS-aided communication systems, many pioneering efforts have recently emerged. By exploiting the sparse nature of cascaded channels in the transform domain, compressive sensing (CS) techniques have been widely used. In [6], the authors discover a sparse representation of the cascade channel utilizing the properties of Katri-Rao and Kronecker products, and orthogonal matching pursuit (OMP) and approximate message passing (AMP) are used. In [7], the authors model a Kronecker-structured channel prior, and solve the cascaded CE problem based on sparse Bayesian learning (SBL). Thanks to the powerful non-linear mapping ability of deep neural networks, deep learning has become a key enabler for future wireless communications. In [8], the authors construct a super-resolution convolutional neural network (CNN) to recover the RIS cascaded channel from the interpolated least square (LS) estimation results. The work in [9] adopts a deep residual learning method to implicitly learn the residual noise for cascaded CE. In [10], the time-varying cascaded channel is efficiently estimated by merging the recurrent neural network (RNN) with the neural ordinary differential equation (ODE).

Deep generative models have shown excellent latent representational power by learning data-driven implicit priors and have attracted great interest of researchers, such as variational autoencoders (VAEs) [11], generative adversarial networks (GANs) [12], and normalizing flows (NFs) [13]. Deep generative models can capture the complex and rich structure from natural signals, avoiding the performance penalty possible caused by manual prior design. Diffusion models (DMs) [14,15] have become a powerful rising class of deep generative models in the last two years, which demonstrate the impressive results in multi-modal learning as the fundamental driver for applications such as DALL-E2 [16] and Stable Diffusion [17]. DMs do not need to align posterior distributions as VAEs, train extra discriminators as GANs, or enforce network constraints as NFs [18]. Due to these virtues and record-breaking performance in many areas, DMs have kicked off the explosion of artificially intelligent-generated content (AIGC).

RIS cascaded CE generally has to cope with a large number of degradation matrices caused by phase variations of passive elements. Unsupervised learning may be a more appropriate option for this problem, as it only uses the degradation model during inference and can be flexibly adapted to different degradation matrices without re-training as in supervised methods [19]. Deep generative models are the natural candidate for unsupervised learning since they intrinsically involve the latent features of the generated data.

Accordingly, in this paper, we propose a DM-based unsupervised method for cascaded CE in RIS-aided massive MIMO systems. Specifically, we construct a conditional DM, which runs the diffusion process in the spectral domain of the degradation matrix. Since the loss function is designed to require no information

other than the training data, the training process is not affected by the specific degradation model in the same way as unconditional DMs. In the sampling process, the received signal is used as conditional information to guide the cascaded channel recovery. We investigate the performance effect of DM parameters in simulations, and the numerical results validate the superiority of the proposed algorithm.

2 System Model

In this section, we give the channel model and the received signal model, which formulates the cascaded CE problem.

2.1 Channel Model

We consider a single-cell uplink RIS-aided massive MIMO system, where the base station (BS) and the RIS are equipped with a half-wavelength spacing uniform planer array (UPA) comprised of $M = M_1 \times M_2$ antennas and $N = N_1 \times N_2$ elements to serve K single-antenna user terminals (UTs). By using $\mathbf{H}_{\mathrm{BR}} \in \mathbb{C}^{M \times N}$ to denote the channel between the BS and the RIS, we can represent it as

$$\mathbf{H}_{\mathrm{BR}} = \frac{1}{\sqrt{L_{\mathrm{BR}}}} \sum_{\ell=1}^{L_{\mathrm{BR}}} \xi_\ell \mathbf{a}_{\mathrm{B}} \left(\vartheta_{\mathrm{B},\ell}, \psi_{\mathrm{B},\ell}\right) \mathbf{a}_{\mathrm{R}}^H (\vartheta_{\mathrm{R},\ell}, \psi_{\mathrm{R},\ell}), \tag{1}$$

where L_{BR} is the number of paths between the BS and the RIS, ξ_ℓ is the complex channel gain of the ℓ-th path, $\vartheta_{\mathrm{B},\ell}$ $(\psi_{\mathrm{B},\ell}) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is the elevation (azimuth) angle of the ℓ-th path at the BS, and $\vartheta_{\mathrm{R},\ell}$ $(\psi_{\mathrm{R},\ell}) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is the elevation (azimuth) angle of the ℓ-th path at the RIS. The steering vectors $\mathbf{a}_{\mathrm{B}} \left(\vartheta_{\mathrm{B},\ell}, \psi_{\mathrm{B},\ell}\right)$ and $\mathbf{a}_{\mathrm{R}} \left(\vartheta_{\mathrm{R},\ell}, \psi_{\mathrm{R},\ell}\right)$ can be respectively expressed as

$$\mathbf{a}_{\mathrm{B}} \left(\vartheta_{\mathrm{B},\ell}, \psi_{\mathrm{B},\ell}\right) = \left[e^{-j\pi \cos(\vartheta_{\mathrm{B},\ell}) \sin(\psi_{\mathrm{B},\ell}) \mathbf{m}_1}\right] \otimes \left[e^{-j\pi \sin(\vartheta_{\mathrm{B},\ell}) \mathbf{m}_2}\right], \tag{2}$$

$$\mathbf{a}_{\mathrm{R}} \left(\vartheta_{\mathrm{R},\ell}, \psi_{\mathrm{R},\ell}\right) = \left[e^{-j\pi \cos(\vartheta_{\mathrm{R},\ell}) \sin(\psi_{\mathrm{R},\ell}) \mathbf{n}_1}\right] \otimes \left[e^{-j\pi \sin(\vartheta_{\mathrm{R},\ell}) \mathbf{n}_2}\right], \tag{3}$$

where $\mathbf{m}_1 = [0, 1, \cdots, M_1 - 1]^T$, $\mathbf{m}_2 = [0, 1, \cdots, M_2 - 1]^T$, $\mathbf{n}_1 = [0, 1, \cdots, N_1 - 1]^T$, and $\mathbf{n}_2 = [0, 1, \cdots, N_2 - 1]^T$. Let $\mathbf{h}_{\mathrm{RU},k} \in \mathbb{C}^{N \times 1}$ denote the channel between the RIS and the k-th UT, and we can also model it as

$$\mathbf{h}_{\mathrm{RU},k} = \frac{1}{\sqrt{L_{\mathrm{RU},k}}} \sum_{i=1}^{L_{\mathrm{RU},k}} \xi_{k,i} \mathbf{a}_{\mathrm{R}} \left(\vartheta_{\mathrm{R},k,i}, \psi_{\mathrm{R},k,i}\right), \tag{4}$$

where $L_{\mathrm{RU},k}$, $\xi_{k,i}$, and $\theta_{\mathrm{R},k,i}$ $(\psi_{\mathrm{R},k,i}) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right)$ are the number of paths, the complex channel gain of the i-th path, and the elevation (azimuth) angle of the i-th path at the RIS between the RIS and the k-th UT. Similarly, the steering

vector $\mathbf{a}_{\mathrm{R}}(\vartheta_{\mathrm{R},k,i}, \psi_{\mathrm{R},k,i})$ can be obtained by substituting $\vartheta_{\mathrm{R},k,i}$ and $\psi_{\mathrm{R},k,i}$ for $\vartheta_{\mathrm{R},\ell}$ and $\psi_{\mathrm{R},\ell}$ in (3).

Then, we define $\widetilde{\mathbf{H}}_k \triangleq \mathbf{H}_{\mathrm{BR}}\mathrm{diag}(\mathbf{h}_{\mathrm{RU},k}) \in \mathbb{C}^{M\times N}$ as the cascaded channel for the k-th UT, which can be re-expressed in the virtual angle domain via the discrete Fourier transform (DFT) as [6]

$$\widetilde{\mathbf{H}}_k = \mathbf{A}_{\mathrm{B}}\widetilde{\mathbf{W}}_k\mathbf{A}_{\mathrm{R}}^T, \tag{5}$$

where $\widetilde{\mathbf{W}}_k$ denotes the angle domain cascaded channel matrix, $\mathbf{A}_{\mathrm{B}} = \mathbf{F}_{M_1} \otimes \mathbf{F}_{M_2}$, $\mathbf{A}_{\mathrm{R}} = \mathbf{F}_{N_1} \otimes \mathbf{F}_{N_2}$, and $\mathbf{F}_M$ denotes the M-dimensional DFT matrix. Since $\widetilde{\mathbf{H}}_k$ and $\widetilde{\mathbf{W}}_k$ are one-to-one mappings, we can estimate $\widetilde{\mathbf{W}}_k$ instead of $\widetilde{\mathbf{H}}_k$.

2.2 Signal Model

We concentrate on the RIS cascaded CE problem, since the direct channel between the BS and UTs can be obtained from classic CE methods by turning off the RIS. We consider the time resource orthogonality implemented among different UTs. In each time slot, all K UTs transmit the orthogonal pilot sequence simultaneously. Let $\mathbf{p}_{k,q}^T \in \mathbb{C}^K$ denote the pilot sequence of the k-th UT in the q-th time slot with $|\mathbf{p}_{k,q}|_2^2 = 1$. Assuming that the RIS phase shift varies with the time slot, the RIS reflecting vector in the q-th time slot can be denoted as $\boldsymbol{\varphi}_q = \left[e^{j\varphi_{1,q}}, \cdots, e^{j\varphi_{N,q}}\right]^T \in \mathbb{C}^N$, where $\varphi_{n,q}$ is the phase shift of the n-th RIS element. By using $\mathbf{p}_{k,q}^H$ to decouple the received signal at the BS in the q-th time slot, we can have the measured signal $\widetilde{\mathbf{y}}_{k,q} \in \mathbb{C}^M$ for the k-th UT as

$$\begin{aligned}\widetilde{\mathbf{y}}_{k,q} &= \sqrt{\rho}\mathbf{H}_{\mathrm{BR}}\mathrm{diag}\left(\boldsymbol{\varphi}_q\right)\mathbf{h}_{\mathrm{RU},k} + \widetilde{\mathbf{n}}_{k,q} \\ &= \sqrt{\rho}\mathbf{A}_{\mathrm{B}}\widetilde{\mathbf{W}}_k\mathbf{A}_{\mathrm{R}}^T\boldsymbol{\varphi}_q + \widetilde{\mathbf{n}}_{k,q},\end{aligned} \tag{6}$$

where ρ is the pilot transmit power assumed to be identical for each UT, $\widetilde{\mathbf{n}}_{k,q} \in \mathbb{C}^M$ is the additive Gaussian noise following $\mathcal{CN}(\widetilde{\mathbf{n}}_{k,q}; \mathbf{0}, 2\sigma\mathbf{I}_M)$. Then, the measured signal over Q time slots for the k-th UT can be represented as

$$\widetilde{\mathbf{Y}}_k = \sqrt{\rho}\widetilde{\mathbf{W}}_k\widetilde{\boldsymbol{\Phi}} + \widetilde{\mathbf{N}}_k, \tag{7}$$

where $\widetilde{\mathbf{Y}}_k = \mathbf{A}_{\mathrm{B}}^H[\widetilde{\mathbf{y}}_{k,1}, \cdots, \widetilde{\mathbf{y}}_{k,Q}] \in \mathbb{C}^{M\times Q}$, $\widetilde{\boldsymbol{\Phi}} = \mathbf{A}_{\mathrm{R}}^T\left[\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_Q\right] \in \mathbb{C}^{N\times Q}$, and $\widetilde{\mathbf{N}}_k = \mathbf{A}_{\mathrm{B}}^H[\widetilde{\mathbf{n}}_{k,1}, \cdots, \widetilde{\mathbf{n}}_{k,Q}] \in \mathbb{C}^{M\times Q}$. Then, we can rewrite (7) in the real form as

$$\mathbf{Y}_k = \mathbf{W}_k\boldsymbol{\Phi} + \mathbf{N}_k, \tag{8}$$

where $\mathbf{Y}_k = \left[\Re(\widetilde{\mathbf{Y}}_k), \Im(\widetilde{\mathbf{Y}}_k)\right] \in \mathbb{R}^{M\times 2Q}$, $\mathbf{W}_k = \left[\Re(\widetilde{\mathbf{W}}_k), \Im(\widetilde{\mathbf{W}}_k)\right] \in \mathbb{R}^{M\times 2N}$, $\boldsymbol{\Phi} = \sqrt{\rho}\left[\Re(\widetilde{\boldsymbol{\Phi}}), \Im(\widetilde{\boldsymbol{\Phi}}); -\Im(\widetilde{\boldsymbol{\Phi}}), \Re(\widetilde{\boldsymbol{\Phi}})\right] \in \mathbb{R}^{2N\times 2Q}$, and $\mathbf{N}_k = \left[\Re(\widetilde{\mathbf{N}}_k), \Im(\widetilde{\mathbf{N}}_k)\right] \in \mathbb{R}^{M\times 2Q}$. For ease of elaboration, we use the equivalent form of (8) as

$$\mathbf{y}_k = \boldsymbol{\Theta}\mathbf{w}_k + \mathbf{n}_k, \tag{9}$$

where $\mathbf{y}_k = \mathrm{vec}(\mathbf{Y}_k) \in \mathbb{R}^{2MQ}$, $\mathbf{\Theta} = \mathbf{\Phi}^T \otimes \mathbf{I}_M \in \mathbb{R}^{2MQ \times 2MN}$, $\mathbf{w}_k = \mathrm{vec}(\mathbf{W}_k) \in \mathbb{R}^{2MN}$, and $\mathbf{n}_k = \mathrm{vec}(\mathbf{N}_k) \in \mathbb{R}^{2MQ}$ following $\mathcal{N}(\mathbf{n}_k; \mathbf{0}, \sigma \mathbf{I}_{2MQ})$. For practical system demands, the number of time slots is always smaller than that of RIS elements, i.e., $Q < N$, which makes (9) ill-defined. The UT index k is omitted later for notation clarity.

3 Methodology Design

In this section, we first introduce DMs. Then we give a RIS cascaded CE method based on the conditional DM.

3.1 Generative Diffusion Models

DMs aim to learn an easily sampled distribution $p_\theta(\mathbf{x}_0)$ to approximate a given target distribution $q(\mathbf{x}_0)$ for $\mathbf{x}_0 \in \mathbb{R}^G$, which contain both forward and backward processes [14]. In the forward process, the target is converted to the Gaussian noise by gradually adding a small noise. In the backward process, the Gaussian noise is recovered to the target by stepwise denoising. Define a series of latent variables $\mathbf{x}_{1:T}$ in the same dimension as $\mathbf{x}_0$. The forward process, also termed the diffusion process, is defined as

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{10}$$

with

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}_G\right), \tag{11}$$

where $\beta_{1:T}$ is a known variance schedule to control the additive noise variance at each step. A remarkable property of the forward process is that we can sample $\mathbf{x}_t$ from $\mathbf{x}_0$ in the closed form as

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\big(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1-\alpha_t)\mathbf{I}_G\big), \tag{12}$$

where $\alpha_t = \prod_{s=1}^{t}(1-\beta_t)$, enabling the parametrization as

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, \tag{13}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I}_G)$. When T is large enough, it is clear that $\alpha_T \doteq 0$ and $q(\mathbf{x}_T|\mathbf{x}_0) \doteq \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I}_G)$, thus $\mathbf{x}_T$ can be treated as a standard Gaussian noise. According to the Bayes' rule $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)\, q(\mathbf{x}_t|\mathbf{x}_0) / q(\mathbf{x}_{t-1}|\mathbf{x}_0)$, the forward process (10) can be rewritten as

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = q(\mathbf{x}_T|\mathbf{x}_0) \prod_{t=2}^{T} q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0), \tag{14}$$

in which for $t \geq 2$, we have

$$q\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0\right) = \mathcal{N}\left(\mathbf{x}_{t-1}; \widetilde{\boldsymbol{\mu}}_t\left(\mathbf{x}_t, \mathbf{x}_0\right), \widetilde{\beta}_t \mathbf{I}_G\right), \tag{15}$$

where

$$\widetilde{\boldsymbol{\mu}}_t\left(\mathbf{x}_t, \mathbf{x}_0\right) = \frac{\sqrt{\alpha_{t-1}}\beta_t}{1-\alpha_t}\mathbf{x}_0 + \frac{\sqrt{1-\beta_t}\left(1-\alpha_{t-1}\right)}{1-\alpha_t}\mathbf{x}_t, \tag{16}$$

$$\widetilde{\beta}_t = \frac{1-\alpha_{t-1}}{1-\alpha_t}\beta_t. \tag{17}$$

Then, the backward process, also termed the generative process, is defined as a Markov chain starting from $p\left(\mathbf{x}_T\right) \sim \mathcal{N}\left(\mathbf{x}_T; \mathbf{0}, \mathbf{I}_G\right)$ as

$$p_\theta\left(\mathbf{x}_{0:T}\right) = p\left(\mathbf{x}_T\right) \prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right) \tag{18}$$

with learnable Gaussian transitions as

$$p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right) = \mathcal{N}\big(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t\right), \boldsymbol{\Sigma}_\theta\left(\mathbf{x}_t, t\right)\big). \tag{19}$$

Parameters θ in (19) are trained by minimizing the variational upper bound on negative log likelihood $\mathbb{E}_{q(\mathbf{x}_0)}\left[-\log p_\theta\left(\mathbf{x}_0\right)\right]$ as [14]

$$\begin{aligned}
\mathcal{L}_D &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\log q\left(\mathbf{x}_{1:T}|\mathbf{x}_0\right) - \log p_\theta\left(\mathbf{x}_{0:T}\right)\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\mathrm{D}\big(q\left(\mathbf{x}_T|\mathbf{x}_0\right) || \, p\left(\mathbf{x}_T\right)\big)\right] \\
&+ \sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\mathrm{D}\big(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0\right) || \, p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)\big)\right] \\
&- \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\log p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)\right],
\end{aligned} \tag{20}$$

where $\mathrm{D}\left(\cdot||\cdot\right)$ denotes the Kullback-Leibler (KL) divergence.

For ease of training, the objective (20) guides us to design $p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)$ in (19) according to $q\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0\right)$ in (15). Since each generation step (19) has only the knowledge of $\mathbf{x}_t$, it is appropriate to model the mean of (19) following (16) as

$$\boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t\right) = \frac{\sqrt{\alpha_{t-1}}\beta_t}{1-\alpha_t}\mathbf{x}_\theta\left(\mathbf{x}_t, t\right) + \frac{\sqrt{1-\beta_t}\left(1-\alpha_{t-1}\right)}{1-\alpha_t}\mathbf{x}_t, \tag{21}$$

where $\mathbf{x}_\theta\left(\mathbf{x}_t, t\right)$ is the prediction function for $\mathbf{x}_0$ given the observation $\mathbf{x}_t$, which is formulated according to (13) as

$$\mathbf{x}_\theta\left(\mathbf{x}_t, t\right) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \sqrt{\frac{1-\alpha_t}{\alpha_t}}\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t\right), \tag{22}$$

where $\boldsymbol{\epsilon}_\theta\left(\cdot\right)$ is a function of $\mathbb{R}^G \times \mathbb{R} \rightarrow \mathbb{R}^G$ implemented with a θ-parameterized neural network. When $t \geq 2$, substituting (22) into (21), we can have the final mean function as

$$\boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t\right) = \frac{1}{\sqrt{1-\beta_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}}\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t\right)\right). \tag{23}$$

Further, the variance of (19) can be fixed following (17) as

$$\boldsymbol{\Sigma}_\theta\left(\mathbf{x}_t, t\right) = \widetilde{\beta}_t \mathbf{I}_G. \tag{24}$$

When $t = 1$, for consistency with $t \geq 2$, the mean function $\boldsymbol{\mu}_\theta\left(\mathbf{x}_1, 1\right)$ and the covariance function $\boldsymbol{\Sigma}_\theta\left(\mathbf{x}_1, 1\right)$ are respectively modeled as

$$\begin{aligned}\boldsymbol{\mu}_\theta\left(\mathbf{x}_1, 1\right) &= \frac{1}{\sqrt{1-\beta_1}}\Big(\mathbf{x}_1 - \frac{\beta_1}{\sqrt{1-\alpha_1}}\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_1, 1\right)\Big) \\ &= \frac{1}{\sqrt{\alpha_1}}\left(\mathbf{x}_1 - \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_1, 1\right)\right), \end{aligned} \tag{25}$$

$$\boldsymbol{\Sigma}_\theta\left(\mathbf{x}_1, 1\right) = \frac{\beta_1\left(1-\alpha_0\right)}{1-\alpha_1}\mathbf{I}_G = \left(1-\alpha_0\right)\mathbf{I}_G = \beta_0 \mathbf{I}_G, \tag{26}$$

where the second equality sign in the two equations above comes from the definition of $\alpha_1 = 1 - \beta_1$ in Eq. (12), and we additionally define

$$\beta_0 = 1 - \alpha_0 < \beta_1, \tag{27}$$

and we can set $\beta_0 = \beta_1 / 10^3$.

Since the KL divergence between univariate Gaussians has

$$\mathrm{D}\big(\mathcal{N}\left(\mu_1, v_1\right) || \mathcal{N}\left(\mu_2, v_2\right)\big) = \frac{1}{2}\log\frac{v_2}{v_1} + \frac{v_1 + \left(\mu_1 - \mu_2\right)^2}{2v_2} - \frac{1}{2}, \tag{28}$$

substituting (13), (23), (24) into (20) and ignoring constant terms can reduce the training loss (20) into

$$\begin{aligned}\mathcal{L}_D - C &= \sum_{t=1}^{T} \mathbb{E}_{\substack{\mathbf{x}_0 \sim q(\mathbf{x}_0), \mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0), \\ \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I}_G)}} \left[\gamma_t \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t\right)\right\|_2^2\right] \\ &= \sum_{t=1}^{T} \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I}_G)} \left[\gamma_t \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, t\right)\right\|_2^2\right], \end{aligned} \tag{29}$$

where $\gamma_t = \frac{\beta_t}{2(1-\alpha_{t-1})(1-\beta_t)}$. Moreover, experiments in [14] demonstrate that the following simplified variant of (29) is more helpful to generation quality and simpler to implement by discarding the weighting factor γ_t and sampling t uniformly from 1 to T, which is also considered later as

$$\begin{aligned}&\mathcal{L}_{\text{simple}} = \\ &\mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), t \sim \mathrm{Uni}(\{1, \cdots, T\}), \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I}_G)} \left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, t\right)\right\|_2^2\right]. \end{aligned} \tag{30}$$

3.2 Conditional DM Based RIS Cascaded CE

Given the signal model (9), we consider extracting samples from the conditional probability $p\left(\mathbf{x}_0 \equiv \mathbf{w} | \mathbf{y}\right)$ with the conditional DM as the posterior sampling

estimator for cascaded channels. Similar to DMs, the forward process of the conditional DM is defined as

$$q\left(\mathbf{x}_{1:T}|\mathbf{x}_0,\mathbf{y}\right)=q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)\prod_{t=2}^{T}q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right), \tag{31}$$

the backward process is defined as

$$p_\theta\left(\mathbf{x}_{0:T}|\mathbf{y}\right)=p\left(\mathbf{x}_T|\mathbf{y}\right)\prod_{t=1}^{T}p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right), \tag{32}$$

and the optimization objective is given by

$$\begin{aligned}\mathcal{L}_C=&\mathbb{E}_{q(\mathbf{x}_{0:T},\mathbf{y})}\left[\mathrm{D}\big(q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)||\,p\left(\mathbf{x}_T|\mathbf{y}\right)\big)\right]-\mathbb{E}_{q(\mathbf{x}_{0:T},\mathbf{y})}\left[\log p_\theta\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right]\\&+\sum_{t=2}^{T}\mathbb{E}_{q(\mathbf{x}_{0:T},\mathbf{y})}\left[\mathrm{D}\big(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)||\,p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\big)\right].\end{aligned} \tag{33}$$

To guarantee that the posterior sampling result $\mathbf{x}_0$ is guided faithfully by the condition $\mathbf{y}$, the diffusion process takes place in the spectral domain of the degradation matrix $\boldsymbol{\Theta}$ [20,21]. Since $\boldsymbol{\Theta}=\boldsymbol{\Phi}^T\otimes\mathbf{I}_M$, the singular value decomposition (SVD) $\mathrm{svd}\left(\boldsymbol{\Theta}\right)=\mathbf{U}\mathbf{S}\mathbf{V}^T$ can be obtained by $\mathrm{svd}\big(\boldsymbol{\Phi}^T\big)=\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^T$, where $\mathbf{U}=\widetilde{\mathbf{U}}\otimes\mathbf{I}_M$ and $\mathbf{V}=\widetilde{\mathbf{V}}\otimes\mathbf{I}_M$ are unitary matrice, $\mathbf{S}=\widetilde{\mathbf{S}}\otimes\mathbf{I}_M$ is a rectangular diagonal matrix whose main diagonal elements are non-increasingly ordered singular values $\mathbf{s}\in\mathbb{R}^{2MQ}$. Define spectral domain representations with $\bar{\mathbf{y}}=\mathbf{S}^\dagger\mathbf{U}^T\mathbf{y}$, $\bar{\mathbf{x}}_t=\mathbf{V}^T\mathbf{x}_t$, $\bar{\mathbf{n}}=\mathbf{S}^\dagger\mathbf{U}^T\mathbf{n}$, where $(\cdot)^\dagger$ stands for the Moore-Penrose generalized inverse, and the signal model (9) can be rewritten as $\bar{\mathbf{y}}=\bar{\mathbf{x}}_0+\bar{\mathbf{n}}$.

Then, define the auxiliary singular vector $\bar{\mathbf{s}}=[\mathbf{s};\mathbf{0}]\in\mathbb{R}^G$ with $G=2MN$. In order to apply the parameterization in (13), the forward process (31) also needs to satisfy the marginal probability (12). For simplicity, we first assume that the forward process (31) has a special conditional probability $q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)=q\left(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{y}\right)$, which leads to $q\left(\mathbf{x}_{1:T}|\mathbf{x}_0,\mathbf{y}\right)=\prod_{t=1}^{T}q\left(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y}\right)$, and can be designed as

$$q\left(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y}\right)=\prod_{i=1}^{G}q\left(\left[\bar{\mathbf{x}}_t\right]_i|\mathbf{x}_0,\mathbf{y}\right), \tag{34}$$

where $[\cdot]_i$ denotes the i-th entry in the vector, and $q\left(\left[\bar{\mathbf{x}}_t\right]_i|\mathbf{x}_0,\mathbf{y}\right)$ in (36) is given by

$$\begin{aligned}&q\left(\left[\bar{\mathbf{x}}_t\right]_i|\mathbf{x}_0,\mathbf{y}\right)\\&=\begin{cases}\mathcal{N}\left(\left[\bar{\mathbf{x}}_t\right]_i;\sqrt{\alpha_t}\left[\bar{\mathbf{x}}_0\right]_i,1-\alpha_t\right) & \text{if } \left[\bar{\mathbf{s}}\right]_i=0 \text{ or } \alpha_t\frac{\sigma}{\left[\bar{\mathbf{s}}\right]_i^2}>1-\alpha_t,\\ \mathcal{N}\left(\left[\bar{\mathbf{x}}_t\right]_i;\left(1-\left[\boldsymbol{\eta}_t\right]_i\right)\sqrt{\alpha_t}\left[\bar{\mathbf{x}}_0\right]_i+\left[\boldsymbol{\eta}_t\right]_i\sqrt{\alpha_t}\left[\bar{\mathbf{y}}\right]_i,1-\alpha_t-\left[\boldsymbol{\eta}_t\right]_i^2\alpha_t\frac{\sigma}{\left[\bar{\mathbf{s}}\right]_i^2}\right) & \text{if } \alpha_t\frac{\sigma}{\left[\bar{\mathbf{s}}\right]_i^2}\le 1-\alpha_t,\end{cases}\end{aligned} \tag{35}$$

which makes the marginal probability (12) hold, where $\boldsymbol{\eta}_{1:T}$ are a series of hyper-parameters, $\alpha_t \frac{\sigma}{[\bar{\mathbf{s}}]_i^2}$ and $1-\alpha_t$ are the spectral domain noise variance and the diffusion noise variance in (12), respectively. If $[\bar{\mathbf{s}}]_i = 0$ or $\alpha_t \frac{\sigma}{[\bar{\mathbf{s}}]_i^2} > 1-\alpha_t$, the diffusion step (35) keeps a standard form obviously satisfying (12). Otherwise, the diffusion step (35) uses a trimmed form, which can be proven to satisfy (12) from the fact $q\left([\bar{\mathbf{y}}]_i|\mathbf{x}_0\right) = \mathcal{N}\left([\bar{\mathbf{y}}]_i; [\bar{\mathbf{x}}_0]_i, \sigma/[\bar{\mathbf{s}}]_i^2\right)$ and the property of Gaussian distributions:

$$\mathcal{N}\left(z_2; m_2 + \alpha m_1, v_2 + \alpha^2 v_1\right) \propto \int \mathcal{N}\left(z_2; m_2 + \alpha z_1, v_2\right) \mathcal{N}\left(z_1; m_1, v_1\right) dz_1. \quad (36)$$

Then, the backward process of this conditional DM can be designed based on its forward process as

$$p\left(\mathbf{x}_T|\mathbf{y}\right) = \prod_{i=1}^{G} p\left([\bar{\mathbf{x}}_T]_i|\mathbf{y}\right), \quad (33)$$

$$p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}\right) = \prod_{i=1}^{G} p_\theta\left([\bar{\mathbf{x}}_{t-1}]_i|\mathbf{x}_t, \mathbf{y}\right). \quad (34)$$

Similar to DMs, the initial generation (33) is modeled as a standard Gaussian distribution with $\alpha_T \doteq 0$ as

$$p\left([\bar{\mathbf{x}}_T]_i|\mathbf{y}\right) = \mathcal{N}\left([\bar{\mathbf{x}}_T]_i; 0, 1\right). \quad (35)$$

Since the forward process (35) obeys the marginal probability (12), the backward process can utilize the function $\mathbf{x}_\theta\left(\mathbf{x}_t, t\right)$ in (22) to predict $\mathbf{x}_0$ at each generation step. Considering the parameterization in (13), naturally,

$$\bar{\mathbf{x}}_0 = \mathbf{V}^T\mathbf{x}_0 = \frac{1}{\sqrt{\alpha_t}}\left(\bar{\mathbf{x}}_t - \sqrt{1-\alpha_t}\mathbf{V}^T\boldsymbol{\epsilon}\right) \quad (36)$$

can be predicted by the following function as

$$\begin{aligned}\bar{\mathbf{x}}_{\theta,t} &\triangleq \bar{\mathbf{x}}_\theta\left(\mathbf{x}_t, t\right) = \mathbf{V}^T\mathbf{x}_\theta\left(\mathbf{x}_t, t\right) \\ &= \frac{1}{\sqrt{\alpha_t}}\left(\bar{\mathbf{x}}_t - \sqrt{1-\alpha_t}\mathbf{V}^T\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t\right)\right).\end{aligned} \quad (37)$$

Therefore, when $t \geq 2$, the generation step (34) can be modeled by replacing $\bar{\mathbf{x}}_0$ with $\bar{\mathbf{x}}_{\theta,t}$ as

$$\begin{aligned}&p_\theta\left([\bar{\mathbf{x}}_{t-1}]_i|\mathbf{x}_t, \mathbf{y}\right) \quad (38)\\ &= \begin{cases} \mathcal{N}\left([\bar{\mathbf{x}}_{t-1}]_i; \sqrt{\alpha_{t-1}}[\bar{\mathbf{x}}_{\theta,t}]_i, 1-\alpha_{t-1}\right) & \text{if } [\bar{\mathbf{s}}]_i = 0 \text{ or } \alpha_{t-1}\frac{\sigma}{[\bar{\mathbf{s}}]_i^2} > 1-\alpha_{t-1}, \\ \mathcal{N}\Big([\bar{\mathbf{x}}_{t-1}]_i; \left(1-\left[\boldsymbol{\eta}_{t-1}\right]_i\right)\sqrt{\alpha_{t-1}}[\bar{\mathbf{x}}_{\theta,t}]_i + \left[\boldsymbol{\eta}_{t-1}\right]_i\sqrt{\alpha_{t-1}}[\bar{\mathbf{y}}]_i, \\ \qquad 1-\alpha_{t-1}-\left[\boldsymbol{\eta}_{t-1}\right]_i^2\alpha_{t-1}\frac{\sigma}{[\bar{\mathbf{s}}]_i^2}\Big) & \text{if } \alpha_{t-1}\frac{\sigma}{[\bar{\mathbf{s}}]_i^2} \leq 1-\alpha_{t-1}. \end{cases}\end{aligned}$$

Algorithm 1. Training Process

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uni}(\{1, \cdots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I}_G)$
5: Update θ with $\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, t \right) \right\|_2^2$
6: **until** the termination condition is fulfilled

When $t = 1$, $p_\theta\left([\bar{\mathbf{x}}_0]_i \,|\, \mathbf{x}_1, \mathbf{y}\right)$ is designed as follows

$$p_\theta\left([\bar{\mathbf{x}}_0]_i \,|\, \mathbf{x}_1, \mathbf{y}\right) = p_\theta\left([\bar{\mathbf{x}}_0]_i \,|\, \mathbf{x}_1\right) = \mathcal{N}\left([\bar{\mathbf{x}}_0]_i; [\bar{\mathbf{x}}_{\theta,1}]_i, 1-\alpha_0\right), \tag{39}$$

where the definition of α_0 has been given in (27).

To improve the generalisation of the method, the training loss (33) is expected to be further simplified as (29), which makes the training process free from the degradation matrix $\boldsymbol{\Theta}$ and the condition $\mathbf{y}$. Since (35) and (38) are Gaussians, in order to make both KL divergences have a consistent form, according to (28) and after simplification, $[\boldsymbol{\eta}_t]_i$ should satisfy

$$[\boldsymbol{\eta}_t]_i = \frac{2(1-\alpha_t)[\bar{\mathbf{s}}]_i^2}{(1-\alpha_t)[\bar{\mathbf{s}}]_i^2 + \alpha_t \sigma}. \tag{39}$$

Finally, the training loss (33) can be reduced to

$$\begin{aligned} &\mathcal{L}_C - C \\ &= \sum_{t=1}^{T} \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I}_G)} \left[\widetilde{\gamma}_t \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, t \right) \right\|_2^2 \right], \end{aligned} \tag{40}$$

where $\widetilde{\gamma}_t = \frac{(1-\alpha_t)\alpha_{t-1}}{2(1-\alpha_{t-1})\alpha_t}$ when $t \geq 2$, and $\widetilde{\gamma}_t = \frac{1-\alpha_t}{2(1-\alpha_{t-1})\alpha_t}$ when $t = 1$. Thus, $\mathcal{L}_C$ in (40) can be naturally transferred to $\mathcal{L}_{\text{simple}}$ in (30), and the training process for DMs can also be used for the conditional DM to update its neural network parameters in its backward process, which is given in Algorithm 1. The sampling process of the conditional DM is used for RIS cascaded CE, which is determined by the backward process and given in Algorithm 2. $\mathbf{G}_{\text{est},k}$,$\widetilde{\mathbf{G}}_{\text{est},k}$ and $\widetilde{\mathbf{H}}_{\text{est},k}$ represent the estimation of the angle-domain channel $\mathbf{G}_k$ in the real form, the angle-domain channel $\widetilde{\mathbf{G}}_k$ in the complex form and the space-domain channel $\widetilde{\mathbf{H}}_k$ in the complex form, respectively. $\mathrm{Reshape}(\cdot, M_1, M_2)$ denotes the reorganization of the vectors into matrices of the dimension $M_1 \times M_2$ by columns, and $[\cdot]_{(:,N_1:N_2)}$ denotes the new matrix formed by the N_1-th to N_2-th columns of the matrix.

For sampling, the diffusion step number T is large and brings costly computational overhead, the successful engineering practice in [15] is to select a subsequence $\{t' = T', T' - \Delta t, \cdots, 1 + \Delta t, 1\}$ to accelerate sampling, where

Algorithm 2. Sampling Process

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I}_G)$, $\bar{\mathbf{x}}_T = \mathbf{V}^T \mathbf{x}_T$
for $k = 1 : K$ **do**
  for $t = T : 1$ **do**
    $\mathbf{z} \sim \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I}_G)$
    $\bar{\mathbf{x}}_{\theta,t} = \frac{1}{\sqrt{\alpha_t}} \left(\bar{\mathbf{x}}_t - \sqrt{1-\alpha_t} \mathbf{V}^T \boldsymbol{\epsilon}_\theta (\mathbf{x}_t, t) \right)$
    if $t > 1$ **then**
      if $[\bar{\mathbf{s}}]_i = 0$ **or** $[\bar{\mathbf{s}}]_i < \sqrt{\frac{\alpha_{t-1}\sigma}{1-\alpha_{t-1}}}, \forall i$ **then**
        $[\bar{\mathbf{x}}_{t-1}]_i = \sqrt{\alpha_{t-1}} [\bar{\mathbf{x}}_{\theta,t}]_i + (1-\alpha_{t-1}) [\mathbf{z}]_i$
      else if $[\bar{\mathbf{s}}]_i \geq \sqrt{\frac{\alpha_{t-1}\sigma}{1-\alpha_{t-1}}}, \forall i$ **then**
        $[\bar{\mathbf{x}}_{t-1}]_i = \left(1 - [\boldsymbol{\eta}_{t-1}]_i\right) \sqrt{\alpha_{t-1}} [\bar{\mathbf{x}}_{\theta,t}]_i + [\boldsymbol{\eta}_{t-1}]_i \sqrt{\alpha_{t-1}} [\bar{\mathbf{y}}]_i$
        $+ \sqrt{1 - \alpha_{t-1} - [\boldsymbol{\eta}_{t-1}]_i^2 \alpha_{t-1}\sigma / [\bar{\mathbf{s}}]_i^2} [\mathbf{z}]_i$
      end if
    else if $t = 1$ **then**
      $[\bar{\mathbf{x}}_{t-1}]_i = [\bar{\mathbf{x}}_{\theta,t}]_i + \sqrt{1-\alpha_{t-1}} [\mathbf{z}]_i, \forall i$
    end if
    $\mathbf{x}_{t-1} = \mathbf{V} \bar{\mathbf{x}}_{t-1}$
  end for
  $\mathbf{G}_{\text{est},k} = \text{Reshape}(\mathbf{x}_0, M, 2N)$
  $\widetilde{\mathbf{G}}_{\text{est},k} = [\mathbf{G}_{\text{est},k}]_{(:,1:N)} + j [\mathbf{G}_{\text{est},k}]_{(:,N+1:2N)}$
  $\widetilde{\mathbf{H}}_{\text{est},k} = \mathbf{A}_{\text{B}} \widetilde{\mathbf{G}}_{\text{est},k} \mathbf{A}_{\text{R}}^T$
end for
return $\widetilde{\mathbf{H}}_{\text{est},k}, \forall k$

$T - \Delta t < T' \leq T$. Accordingly, we can obtain the accelerated version of Algorithm 2 by changing its t, $t-1$, T, and $T-1$ to t', $t' - \Delta t$, T', and $T' - \Delta t$, respectively, and the step number for the accelerated sampling is $\lfloor \frac{T}{\Delta t} \rfloor + 1$.

4 Simulation Results

In this section, we present simulation results to evaluate the algorithm performance. The numbers of BS antennas and RIS elements are set to $M = 64$ ($M_1 = 8, M_2 = 8$) and $N = 64$ ($N_1 = 8, N_2 = 8$). The number of UTs is set to $K = 20$. The pilot transmit power is set to $\rho = 1$. The numbers of paths are set to $L_{\text{BR}} = 6$ and $L_{\text{RU},k} = 8$ for $\forall k$. The channel gains ξ_ℓ and $\xi_{k,i}$ follow the standard complex Gaussian distribution, and the angles $\vartheta_{\text{B},\ell}, \psi_{\text{B},\ell}, \vartheta_{\text{R},\ell}, \psi_{\text{R},\ell}, \vartheta_{\text{R},k,i}, \psi_{\text{R},k,i}$ follow the uniform distribution on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right)$ for $\forall \ell, k, i$. The RIS phase shift $\varphi_{n,q}$ is uniformly distributed from $[0, 2\pi)$ for $\forall n, q$ [22].

Our neural network architecture follows that in [14], which is a U-Net based on a Wide ResNet [23]. We have considered two neural networks with 32 and

64 base channels, which are named DM32 and DM64, respectively. The channel multipliers for five feature map resolutions are set to $[1, 2, 2, 2, 4]$. Each resolution level has two convolutional residual blocks, and self-attention blocks at the 16×16 resolution are between the convolutional residual blocks. The dropout and EMA rate are set to 0.1 and 0.9999, respectively. The diffusion step number T is set to 1000, and we have considered three sampling stepsizes Δt as 25, 20, and 16, which have 41, 51, and 63 sampling steps, respectively. The diffusion noise variance follows an equispaced linear schedule from $\beta_1 = 0.0001$ to $\beta_T = 0.02$. The training and testing datasets consist of 50,000 and 10000 randomly generated cascaded channels based on the presented channel model. We use the Adam optimizer and set the learning rate to 0.0002. The batch size is set to 512, and 5000 epochs are used for training. Every 20 testing data share a degradation matrix $\boldsymbol{\Phi}$ since the RIS reflecting vector $\boldsymbol{\varphi}_q$ is the same for all 20 UTs at the same time slot, which means a total of $M_c = 500$ degradation matrices for testing. We use 4 RTX 4090 GPUs in training and 1 RTX 4090 GPU in sampling. The normalized mean-squared error (NMSE) is used as the performance metric, which is formulated as

$$\mathrm{NMSE} = \frac{1}{M_c K} \sum_{m_c=1}^{M_c} \sum_{k=1}^{K} \frac{\left\| \widetilde{\mathbf{H}}_{\mathrm{est},k}^{(m_c)} - \widetilde{\mathbf{H}}_{k}^{(m_c)} \right\|_{\mathrm{F}}^2}{\left\| \widetilde{\mathbf{H}}_{k}^{(m_c)} \right\|_{\mathrm{F}}^2}. \tag{41}$$

Figure 1 presents the NMSE performance of our proposed DM method and three Bayesian baselines including expectation maximization-SBL (EM-SBL) [24], EM-Bernoulli Gaussian-AMP (EM-BG-AMP) [25], and EM-BG-vector AMP (EM-BG-VAMP) [26] versus SNR under different time slots, where DM32 (41) denotes DM32 with 41 sampling steps, and the others are similar. We can see that for the DM method, DM64 performs better than DM32, since the increase in base channels improves the model capacity. The effect of sampling steps on the performance of the DM method presents different results under various SNR conditions. At low SNRs, an appropriate reduction in sampling steps improves the performance, which is mainly for two reasons. First, a large number of sampling steps, while making the sampling trajectory more complete, amplifies the effect of the spectral domain noise. Second, although step 8 in the sampling process avoids the effect of the spectral domain noise, the excessive absence of the conditional signal can also be detrimental to the signal recovery. However, at high SNRs, we can see that too few sampling steps makes it difficult to obtain the performance gain from the increased SNR, while an appropriate increase in sampling steps can significantly improve CE accuracy. This is mainly because the spectral domain noise has less effect on the signal recovery and the increase in sampling steps helps the conditioned signal to better guide the recovered signal.

Moreover, we can see that only the performance of DM32 with 51 and 63 steps is slightly weaker than that of EM-BG-AMP and EM-BG-VAMP when the SNR is 0 dB in Fig. 1a, and the performance of DM32 with 63 steps is slightly weaker than that of EM-BG-VAMP when the SNR is 0 dB in Fig. 1b. However, the above observations indicate that at low SNRs, it is beneficial to reduce sampling steps appropriately, and thus we can use 41 steps with better

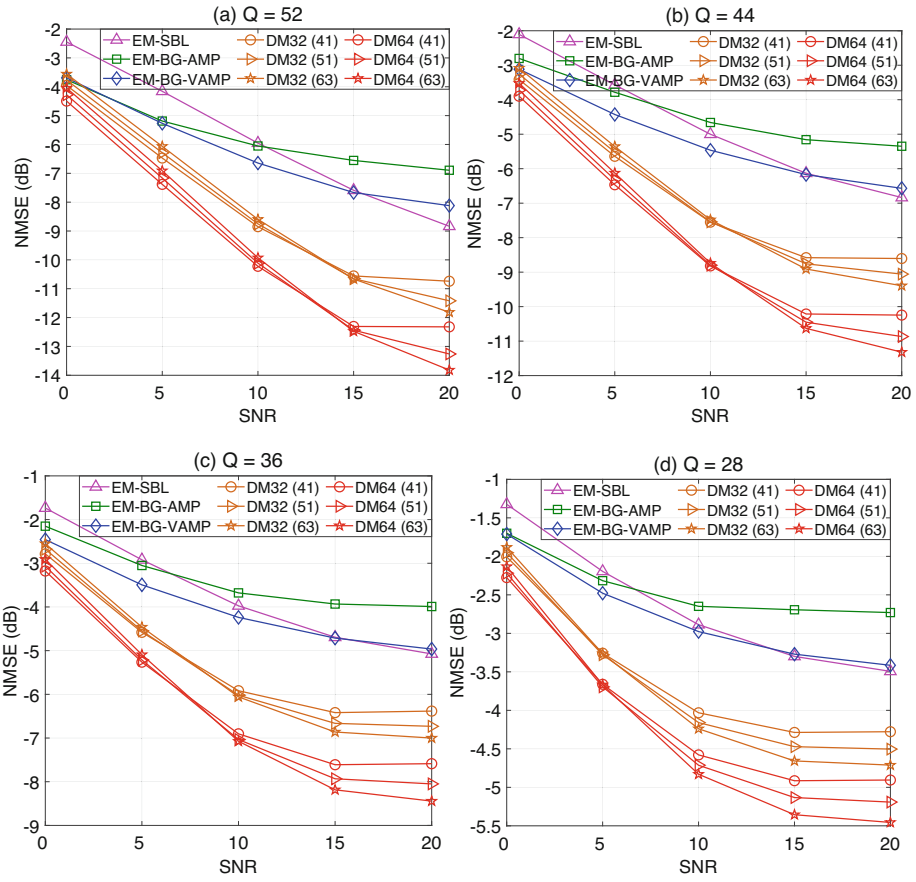


Fig. 1. NMSE performance of DM method and three Bayesian baselines versus SNR when the time slot number Q is (a) 52; (b) 44; (c) 36; (d) 28.

performance than EM-BG-AMP and EM-BG-VAMP. While in other cases, all DM schemes outperform three Bayesian baselines across the board, and this advantage becomes significant as the SNR increases. The results verify that our algorithm can achieve excellent performance and effectively reduce the pilot overhead.

5 Conclusion

This paper studies the cascaded CE problem for RIS-aided massive MIMO systems. An emerging deep generative model called diffusion model is used to effectively capture the latent representation of the cascaded channel through unsupervised learning. The loss function is carefully designed to make the training

process independent of different degradation matrices caused by RIS phase variations. During sampling, the received signal is used as conditional information to precisely guide the channel recovery. Simulation results validate the superiority of the proposed algorithm with reduced pilot overhead.

References

1. Pan, C., et al.: Reconfigurable intelligent surfaces for 6G systems: principles, applications, and research directions. IEEE Commun. Mag. **59**(6), 14–20 (2021)
2. Liu, X., Wang, W., Song, X., Gao, X., Fettweis, G.: Sparse channel estimation via hierarchical hybrid message passing for massive MIMO-OFDM systems. IEEE Trans. on Wirel. Commun. **20**(11), 7118–7134 (2021)
3. You, X., et al.: Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci. China Inf. Sci. **64**(1), 1–74 (2021)
4. Liu, X., Wang, W., Gong, X., Fu, X., Gao, X., Xia, X.G.: Structured hybrid message passing based channel estimation for massive MIMO-OFDM systems. IEEE Trans. Veh. Technol. (2023, early access)
5. Liu, C., Liu, X., Ng, D.W.K., Yuan, J.: Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications. IEEE Trans. Wirel. Commun. **21**(2), 898–912 (2021)
6. Wang, P., Fang, J., Duan, H., Li, H.: Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems. IEEE Signal Process. Lett. **27**, 905–909 (2020)
7. Xu, X., Zhang, S., Gao, F., Wang, J.: Sparse Bayesian learning based channel extrapolation for RIS assisted MIMO-OFDM. IEEE Trans. Commun. **70**(8), 5498–5513 (2022)
8. Wang, Y., Lu, H., Sun, H.: Channel estimation in IRS-enhanced mmWave system with super-resolution network. IEEE Commun. Lett. **25**(8), 2599–2603 (2021)
9. Liu, C., Liu, X., Ng, D.W.K., Yuan, J.: Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications. IEEE Trans. Wirel. Commun. **21**(2), 898–912 (2022)
10. Xu, M., Zhang, S., Ma, J., Dobre, O.A.: Deep learning-based time-varying channel estimation for RIS assisted communication. IEEE Commun. Lett. **26**(1), 94–98 (2021)
11. Pu, Y., et al.: Variational autoencoder for deep learning of images, labels and captions. In: Proceedings of the Advances Neural Information Processing Systems, vol. 29 (2016)
12. Goodfellow, I.J., et al.: Generative adversarial nets. In: Proceedings of the Advances Neural Information Processing Systems (2014)
13. Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: Proceedings of the International Conference on Machine Learning, pp. 1530–1538 (2015)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Proceedings of the Advances Neural Information Processing System, vol. 33, pp. 6840–6851 (2020)
15. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: Proceedings of the International Conference on Learning Representations (2020)
16. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125 (2022)

17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
18. Cao, H., Tan, C., Gao, Z., Chen, G., Heng, P.A., Li, S.Z.: A survey on generative diffusion model. arXiv preprint arXiv:2209.02646 (2022)
19. Venkatakrishnan, S.V., Bouman, C.A., Wohlberg, B.: Plug-and-play priors for model based reconstruction. In: Proceedings of the IEEE Global Conference on Signal and Information Processing, pp. 945–948 (2013)
20. Kawar, B., Vaksman, G., Elad, M.: SNIPS: solving noisy inverse problems stochastically. In: Proceedings of the Advances Neural Information Processing Systems, vol. 34, pp. 21757–21769 (2021)
21. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. In: Proceedings of the ICLR Workshop on Deep Generative Models for Highly Structured Data (2022)
22. An, J., Wu, Q., Yuen, C.: Scalable channel estimation and reflection optimization for reconfigurable intelligent surface-enhanced OFDM systems. IEEE Wirel. Commun. Lett. **11**(4), 796–800 (2022)
23. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: Proceedings of the British Machine Vision Conference (2016)
24. Wipf, D.P., Rao, B.D.: Sparse Bayesian learning for basis selection. IEEE Trans. Signal Process. **52**(8), 2153–2164 (2004)
25. Vila, J.P., Schniter, P.: Expectation-maximization Gaussian-mixture approximate message passing. IEEE Trans. Signal Process. **61**(19), 4658–4672 (2013)
26. Rangan, S., Schniter, P., Fletcher, A.K.: Vector approximate message passing. IEEE Trans. Inf. Theory **65**(10), 6664–6684 (2019)

Quantum Permutation Pad with Qiskit Runtime

Alain Chancé(✉)

Quantalain SASU, 38 rue des Mathurins, 75008 Paris, France
alain.chance@quantalain.com

Abstract. We demonstrate an efficient implementation of the Kuang and Barbeau's Quantum Permutation pad (QPP) symmetric cryptographic algorithm with Qiskit Runtime, a new architecture offered by IBM Quantum that streamlines quantum computations. We have implemented a Python class QPP and template Jupyter notebooks with Qiskit code for encrypting and decrypting with n-qubit QPP any text file in UTF-16 format or any image file in .png format. We offer the option of running either a quantum circuit with n qubits, or an alternate one with 2^n qubits which only uses swap gates and has a circuit depth of $O(n)$. It is inherently extremely fast and could be run efficiently on currently available noisy quantum computers. Our implementation leverages the new Qiskit Sampler primitive in localized mode which dramatically improves performance. We offer a highly efficient classical implementation which performs permutation gate matrix multiplication with information state vectors. We illustrate the use with two agents Alice and Bob who exchange a text file and an image file using 2-qubit QPP and 4-qubit QPP.

Keywords: Quantum · Communication · Encryption · Qiskit

1 Introduction

We present an efficient implementation of the Quantum Permutation Pad proposed by Kuang et al. in 2020 [1] that we have derived from the 2-qubit Qiskit code by Kuang and Perepechaenko [2, 3]. We have extended their work to deal with the general n-qubit QPP. We have developed an alternate way of building a quantum circuit for n-qubit QPP with 2^n qubits and a depth of $O(n)$ which is extremely fast and could be run on noisy quantum computers. Our QPP class method permutation_pad() prepares every quantum circuit corresponding to a permutation in the QPP pad, initializes it in turn for each of the possible 2^n input state vectors and submits all circuits for execution by the new Qiskit Runtime Sampler [4, 5] in a single job. Then it stores in a Python dictionary the input state vector as a key and the corresponding most frequent output state vector as value. These Python dictionaries are stored in a list. It also offers a classical implementation which performs matrix multiplication with each of the possible 2^n input state vectors and stores the results in a list of Python dictionaries, one for each permutation. The encryption of a message or the decryption of a ciphertext is achieved very efficiently by selecting the Python dictionary pertaining to the current permutation, then using

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 136–147, 2024.
https://doi.org/10.1007/978-3-031-59619-3_12

the input message or cipher chunk as a key and retrieving the value which is the most frequent outcome already computed. Our approach is very efficient, the duration of the encryption or decryption depend mostly on the number of permutations in the QPP pad and not much on the length of the message.

2 Permutation Operators

2.1 Classical Permutation Operators

A permutation operator changes the linear order of an ordered set S. There is factorial n, usually written n! permutations of a set of n distinct elements. The entire set of permutations of a set S with n elements form a group called the symmetric group S_n of the set. A permutation matrix can be associated with a permutation by permuting the columns of the identity matrix I_n. Multiplying the permutation matrix by a column vector will permute the rows of a vector. In below Eq. 1, multiplying the matrix associated with a permutation of columns 1 and 2 by state vector '01' gives the state vector '10':

$$\begin{bmatrix} 1\,0\,0\,0 \\ 0\,0\,1\,0 \\ 0\,1\,0\,0 \\ 0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{1}$$

2.2 Quantum Permutation Operators

A quantum permutation operator permutes a computational basis set of a n-qubit register which comprises 2^n state vectors. The entire set of permutations forms the symmetric group S_{2^n} or so-called Quantum Permutation Space or QPS.

3 Implementation of n-qubit QPP

Quantum Permutation pad or QPP is a symmetric cryptographic algorithm proposed by Kuang and Bettenburg in 2020 that uses a set of permutation matrices implemented by quantum gates to encrypt any plaintext into a ciphertext and the corresponding transposed matrices to decrypt the ciphertext.

The parameter version "V0" or "V1" is used to select the way our QPP class implements a n-qubit QPP.

3.1 Version "V0" Selects n-qubit QPP with n Qubits

For each permutation to be created in the QPP pad, the randomize() method of the QPP class performs Fisher Yates shuffling [6] with n permutations of columns in an array of size 2^n by 2^n. The permutation_pad() method uses the Qiskit Operator class to create a unitary matrix operator from this permutation matrix. Then it uses the QuantumCircuit.append method to convert the operator into a UnitaryGate object and to add it to the quantum circuit. Last it transpiles the quantum circuit with optimization level 2 and adds it to a list of quantum circuits, one for each permutation. Each quantum circuit is only transpiled once.

3.2 Version "V1" Selects n-qubit QPP with 2^n Qubits

The randomize() method of the QPP class, adds a swap gate between qubits in a quantum circuit with 2^n qubits whenever it performs a permutation between two different columns in an array of size n by n. By setting the parameter trace to 2 in the parameter file of the QPP instance, we see in the trace file what permutations of columns randomize has performed. The following Table 1 shows an example for a 2-qubit QPP of a permutation which is set up as three successive permutations of two columns in an array of size 2^2 which is 4.

Table 1. A permutation in a 2-qubit QPP set up as 3 permutations of two columns.

Keyword in trace file	Permutation
randomize	Permutation number: 3
randomize	Permuting columns 3 and 0
randomize	Permuting columns 2 and 1
randomize	Permuting columns 1 and 3

The permutation matrix that matches the three permutations of columns follows:

$$\begin{bmatrix} 0\,0\,0\,1 \\ 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \end{bmatrix} \tag{2}$$

The depth of the quantum circuit is 2 and the permutation dictionary is as follows:

$$\{0:'01',\ 1:'10',\ 2:'11',\ 3:'00'\} \tag{3}$$

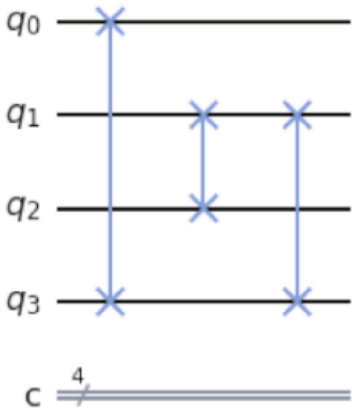


Fig. 1. A quantum circuit with a swap gate between qubits matching the corresponding permutations between two columns (2).

The corresponding quantum circuit with swap gates is shown in Fig. 1.

The operator matrix corresponding to the quantum circuit with 2^n qubits has a size of 2^N by 2^N where $N = 2^n$.

3.3 Qiskit Runtime Primitives

Our QPP class currently only uses a uses the localized version of the Sampler which resides in Qiskit Terra [7, 8] and uses the Statevector simulator to compute probabilities. The Qiskit Statevector simulator can simulate a quantum circuit of up to 16 qubits with 8 Gb available memory.

3.4 Classical Implementation

The QPP class offers a very efficient classical n-qubit QPP implementation. For each permutation it performs a matrix multiplication with each of the 2^n states in the computational basis and it stores results in a list of Python dictionaries, one dictionary per permutation. No computation is needed afterwards to find the outcome of applying a permutation.

4 Use Cases

4.1 Alice Encrypts a File Using 2-qubit QPP and then Sends It to Bob

Alice wants to encrypt a file 'Hello.txt' which contains the following text in UTF-16 format: ' Hello ☺'. Her Jupyter notebook contains the statements shown in Table 2.

Table 2. Qiskit statements in Alice's Jupyter notebook.

Action	Qiskit statement
Import QPP Class from QPP_Alain	from QPP_Alain import QPP
Create an instance of the QPP class	Alice_QPP = QPP("QPP_param_2-qubits_V0_Hello")
Convert plaintext into bitstring message	message = Alice_QPP.file_to_bitstring()
Encrypt bitstring message	ciphertext = Alice_QPP.encrypt(message = message)
Alice sends parameter and ciphertext files to Bob	

The Json file name is passed as a parameter. Its content is set using Table 1 in Ref. [2] as reference and is shown in Fig. 2. It contains the following list of parameters:

- **num_of_bits**: Classical key length (bit).
- **num_of_qubits**: Number of qubits.

- **num_of_perm_in_pad**: Number of permutation matrices in pad.
- **pad_selection_key_size**: Pad selection key size.
- **opt_level**: Optimization level used by Transpile, 2 is a good option.
- **resilience level**: resilience option used for error mitigation [9].
- **plaintext_file**: name of the plaintext file to be encrypted, only Microsoft Office text files in Unicode UTF-16 format with extension.txt and image files with extension .png are currently supported.
- **token_file**: name of the file containing the token for execution in the IBM Quantum Lab environment.
- **trace**: set to 0 for no trace, 1 minimum trace, 2 more detailed trace.
- **job_trigger**: number of quantum circuits to be grouped by Sampler in a single job execution.
- **print_trigger**: number of iterations after which a print is performed according to trace level.
- **draw_circuit**: if 'True' and 'trace' is greater than or equal to 1, then quantum circuits drawings are shown on the console using LaTeX formatting.
- **do-sampler**: 'True' - use Sampler primitive to submit several quantum circuits in a single job and then sample the results. Else, perform matrix multiplication with information state vectors for all 2^n possible states and store the results in a list of dictionaries, one for each pad in the QPP pad.
- **version**: "V0" - unitary matrix operators are created from permutation matrices in the pad and then appended in a quantum circuit. "V1" - for each permutation of two columns in a permutation matrix, a corresponding swap gate is inserted in a quantum circuit with 2^n qubits.
- **len_message**: automatically set to the length of a message in bytes read from the plaintext file.
- **len_ciphertext**: automatically set to the length of the ciphertext in bits by the encrypt method.

```
{
    "num_of_bits": 448,
    "num_of_qubits": 2,
    "num_of_perm_in_pad": 56,
    "pad_selection_key_size": 6,
    "opt_level": 2,
    "resilience_level": 1,
    "plaintext_file": "Hello.txt",
    "token_file": "Token_Alain.txt",
    "trace": 1,
    "job_trigger": 10,
    "print_trigger": 10,
    "draw_circuit": "True",
    "do_sampler": "True",
    "version": "V0",
    "len_message": 0,
    "len_ciphertext": 0
}
```

Fig. 2. On Alice's side, Json file passed as a parameter.

Convert Plaintext File into a Bitstring Message. A binary string is created and automatically padded if necessary, according to the number of qubits.

Encrypt Bitstring Message. The encrypt method creates a secret key if none is given and stores it in a secret key file. Alice shares it with Bob using a secure channel. With trace set to 1, Alice follows the execution of the encryption in her Jupyter notebook. An example of a permutation matrix is shown below:

$$\begin{bmatrix} 0\,0\,1\,0 \\ 0\,0\,0\,1 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \end{bmatrix} \tag{4}$$

Permutation pad - Permutation number: 17, Depth of quantum circuit: 3

The corresponding quantum circuit and its depth are shown in Fig. 3.

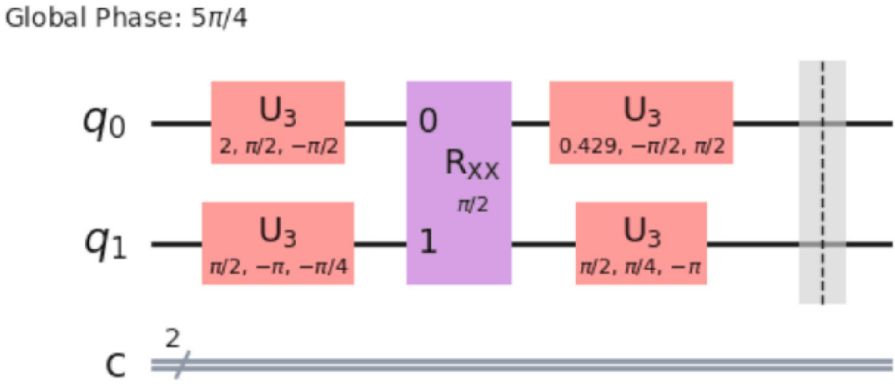


Fig. 3. On Alice's side, quantum circuit corresponding to permutation in Eq. 4.

Alice follows the last steps of the execution of the encryption process in her Jupyter notebook as shown in Fig. 4.

The permutation_pad() method runs all the quantum circuits in around 10 s and the encryption of the message is achieved in a very small amount of time.

```
encrypt - x : 9,  Permutation_Pad[24], State vector: 11, Most frequent: 11
encrypt - Elapsed time: 0:00:00.000017
encrypt - x : 19,  Permutation_Pad[20], State vector: 01, Most frequent: 11
encrypt - Elapsed time: 0:00:00.000029
encrypt - x : 29,  Permutation_Pad[50], State vector: 10, Most frequent: 11
encrypt - Elapsed time: 0:00:00.000038
encrypt - x : 39,  Permutation_Pad[6], State vector: 01, Most frequent: 11
encrypt - Elapsed time: 0:00:00.000046
encrypt - x : 49,  Permutation_Pad[13], State vector: 00, Most frequent: 00
encrypt - Elapsed time: 0:00:00.000053
encrypt - x : 59,  Permutation_Pad[15], State vector: 00, Most frequent: 11
encrypt - Elapsed time: 0:00:00.000060

encrypt - Elapsed time for encryption of message: 0:00:00.000066
```

Fig. 4. On Alice's side, trace of the last steps of the encryption process.

4.2 Bob Decrypts the File Using 2-qubit QPP

Bob's Jupyter notebook contains the following statements in Table 3:

Table 3. Qiskit statements in Bob's Jupyter notebook.

Action	Qiskit statement
Import QPP Class from QPP_Alain	from QPP_Alain import QPP
Create an instance of the QPP class	Bob_QPP = QPP("QPP_param_2-qubits_V1_Hello")
Read ciphertext binary file and extract the content to be transformed into a binary string	ciphertext = Bob_QPP.binary_to_ciphertext()
Decrypt the ciphertext	decrypted_message = Bob_QPP.decrypt(ciphertext = ciphertext)
Convert the decrypted message and save it into the decrypted file	Bob_QPP.bitstring_to_file(decrypted_message = decrypted_message)

Bob has updated the Json parameter file that he received from Alice, and he has selected option "V1" which sets up a quantum circuit with $2^2 = 4$ qubits for each permutation matrix in the QPP pad. The parameter file name is passed as a parameter and its content is illustrated in Fig. 5.

With trace set to 1, Bob follows the execution of the encryption in his Jupyter notebook. An example of a permutation matrix is shown below.

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{5}$$

Permutation pad - Permutation number: 17, Depth of quantum circuit: 2

```
{
    "num_of_bits": 448,
    "num_of_qubits": 2,
    "num_of_perm_in_pad": 56,
    "pad_selection_key_size": 6,
    "opt_level": 2,
    "resilience_level": 1,
    "plaintext_file": "Hello.txt",
    "token_file": "Token_Alain.txt",
    "trace": 1,
    "job_trigger": 10,
    "print_trigger": 10,
    "draw_circuit": "True",
    "do_sampler": "True",
    "version": "V1",
    "len_message": 8,
    "len_ciphertext": 128
}
```


Fig. 5. On Bob's side, Json file passed as parameter.

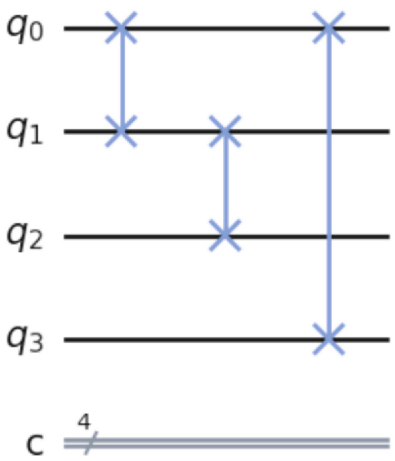


Fig. 6. On Bob's side, quantum circuit corresponding to permutation in Eq. 5.

The corresponding quantum circuit and its depth are shown in Fig. 6.

On Bob's side, the permutation matrix in Eq. 5 in the QPP pad is the transpose of the one used on Alice's side for encryption, thus achieving the inverse permutation.

Bob has selected option "V1" which creates a quantum circuit with $2^2 = 4$ qubits with only swap quantum gates and a low depth of 3.

Bob has selected trace level 1 and follows the execution of the decryption process in his Jupyter notebook as illustrated in Fig. 7.

Bob opens the file named 'Decrypted_Hello.txt' and reads the following text in UTF-16 format: ' Hello ☺'

Bob has successfully decrypted the ciphertext sent by Alice.

```
decrypt - x : 9,  Permutation_Pad[24], State vector: 11, Most frequent: 11
decrypt - Elapsed time for decryption: 0:00:00.000035
decrypt - x : 19,  Permutation_Pad[20], State vector: 11, Most frequent: 01
decrypt - Elapsed time for decryption: 0:00:00.000050
decrypt - x : 29,  Permutation_Pad[50], State vector: 11, Most frequent: 10
decrypt - Elapsed time for decryption: 0:00:00.000058
decrypt - x : 39,  Permutation_Pad[6], State vector: 11, Most frequent: 01
decrypt - Elapsed time for decryption: 0:00:00.000068
decrypt - x : 49,  Permutation_Pad[13], State vector: 00, Most frequent: 00
decrypt - Elapsed time for decryption: 0:00:00.000075
decrypt - x : 59,  Permutation_Pad[15], State vector: 11, Most frequent: 00
decrypt - Elapsed time for decryption: 0:00:00.000083

decrypt - Length of decrypted message in bits: 128

decrypt - Elapsed time for decryption of ciphertext: 0:00:00.000120
```

Fig. 7. On Bob's side, trace of the last steps of the decryption process.

4.3 Alice Encrypts an Image File Using 4-qubit QPP and then Sends It to Bob

Alice now wants to encrypt a small image (12758 bytes) of a Christmas tree using 4-qubit QPP. Her Jupyter notebook contains the following statements shown in Table 4:

Table 4. Qiskit statements in Alice's Jupyter notebook.

Action	Qiskit statement
Import QPP Class from QPP_Alain	from QPP_Alain import QPP
Create an instance of the QPP class	AliceQPP = QPP("QPP_param_4-qubits_V0_Christmas_tree")
Convert plaintext into bitstring message	message = Alice_QPP.file_to_bitstring()
Encrypt bitstring message	ciphertext = Alice_QPP.encrypt(message = message)
Alice sends parameter and ciphertext files to Bob	

Alice has selected 4 qubits and option "V0" as shown in Fig. 8.

With trace set to 1, Alice follows the execution of the encryption in her Jupyter notebook. An example of a permutation matrix and the depth of a quantum circuit is

```
{
    "num_of_bits": 384,
    "num_of_qubits": 4,
    "num_of_perm_in_pad": 6,
    "pad_selection_key_size": 6,
    "opt_level": 2,
    "resilience_level": 1,
    "plaintext_file": "Christmas_tree.png",
    "token_file": "Token_Alain.txt",
    "trace": 1,
    "job_trigger": 10000,
    "print_trigger": 10000,
    "draw_circuit": "True",
    "do_sampler": "True",
    "version": "V0",
    "len_message": 0,
    "len_ciphertext": 0
}
```


Fig. 8. On Alice's side, Json file passed as a parameter.

shown below. The drawing of the quantum circuit is too large to be shown.

$$\begin{bmatrix} 0\,0\,0\,1 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,1\,0 \vdots \ddots \vdots 0\,0 \\ \vdots\;\vdots\;\vdots\;\vdots \cdots \vdots\;\vdots\;\vdots \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \end{bmatrix} \tag{6}$$

Permutation pad - Permutation number: 5, Depth of quantum circuit: 187.

Alice follows the last steps of the encryption in her Jupyter notebook as shown in Fig. 9.

```
permutation_pad - permutation number: 5, dictionary:
{0: '0000', 1: '0111', 2: '0011', 3: '1010', 4: '0001', 5: '0110', 6: '1011', 7: '1100', 8: '0100',
9: '1111', 10: '0010', 11: '1101', 12: '1001', 13: '1110', 14: '0101', 15: '1000'}
permutation pad - Elapsed time: 0:00:21.012235
permutation pad - Length of Permutation_Pad: 6

encrypt - x : 9999,  Permutation_Pad[2], State vector: 0100, Most frequent: 0000
encrypt - Elapsed time: 0:00:00.002568
encrypt - x : 19999,  Permutation_Pad[1], State vector: 1111, Most frequent: 0101
encrypt - Elapsed time: 0:00:00.005223

encrypt - Elapsed time for encryption of message: 0:00:00.006664

encrypt - First 192 bits in ciphertext string
101111001101101010000101000100010000110010001101100000010101011011000010011010101100101101010100011
111101110110100101011000111011111110100100001100110000001011011001101001010011010100101110010
```

Fig. 9. On Alice's side, trace of the last steps of the encryption process.

4.4 Bob Decrypts the Image with 4-qubit QPP

Bob's Jupyter notebook contains the following statements shown in Table 5:

Table 5. Qiskit statements in Bob's Jupyter notebook.

Action	Qiskit statement
Import QPP Class from QPP_Alain	from QPP_Alain import QPP
Create an instance of the QPP class	Bob_QPP = QPP("QPP_param_4-qubits_V0_Christmas_tree")
Read ciphertext binary file and extract the content to be transformed into a binary string	ciphertext = Bob_QPP.binary_to_ciphertext()
Decrypt the ciphertext	decrypted_message = Bob_QPP.decrypt(ciphertext = ciphertext)
Convert the decrypted message and save it into the decrypted file	Bob_QPP.bitstring_to_file(decrypted_message = decrypted_message)

With trace set to 1, Bob follows the execution of the encryption in his Jupyter notebook. An example of a permutation matrix follows:

$$\begin{bmatrix} 0\,0\,0\,1 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,1\,0 \vdots \ddots \vdots 0\,0 \\ \vdots\;\vdots\;\vdots\;\vdots \cdots \vdots\;\vdots\;\vdots \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \\ 0\,0\,0\,0 \cdots 0\,0\,0 \end{bmatrix} \tag{7}$$

With trace set to 1, Bob follows the execution of the decryption in his Jupyter notebook as illustrated in Fig. 10.

```
permutation_pad - permutation number: 5, dictionary:
{0: '0000', 1: '0100', 2: '1010', 3: '0010', 4: '1000', 5: '1110', 6: '0101', 7: '0001', 8: '1111',
9: '1100', 10: '0011', 11: '0110', 12: '0111', 13: '1011', 14: '1101', 15: '1001'}
permutation pad - Elapsed time: 0:00:22.684784
permutation pad - Length of Permutation_Pad: 6

decrypt - Length of ciphertext: 102064
decrypt - Remainder of dividing (Length of cipher chunks) by (Job trigger): 5516

decrypt - x : 9999,  Permutation_Pad[2], State vector: 0000, Most frequent: 0100
decrypt - Elapsed time for decryption: 0:00:00.003083
decrypt - x : 19999,  Permutation_Pad[1], State vector: 0101, Most frequent: 1111
decrypt - Elapsed time for decryption: 0:00:00.006014

decrypt - Length of decrypted message in bits: 102064

decrypt - Elapsed time for decryption of ciphertext: 0:00:00.043991
```

Fig. 10. On Bob's side, trace of the last steps of the decryption process.

5 Conclusion

We have presented a very efficient and easy to use a toy implementation of the n-qubit QPP that we have tested for n from 2 to 5 with small text files in UTF-16 format and image files in .png format. The source code for the QPP class and the template Jupyter notebooks are publicly available under a MIT License in the public GitHub repository: https://github.com/AlainChance/QPP-Alain.

References

1. Kuang, R., Bettenburg, N.: Shannon perfect secrecy in a discrete Hilbert space. In: Proceedings of the IEEE International Conference on Quantum Computing and Engineering (QCE), pp. 249–255 (2020). https://doi.org/10.1109/QCE49297.2020.00039
2. Kuang, R., Perepechaenko, M.: Quantum encryption with quantum permutation pad in IBMQ systems. EPJ Quantum Technol. **9**, 26 (2022). https://doi.org/10.1140/epjqt/s40507-022-00145-y
3. Perepechaenko, M., Kuang, R.: Quantum encryption and decryption in IBMQ systems using quantum permutation pad. J. Commun. **17**(12), 972–978 (2022). https://doi.org/10.12720/jcm.17.12.972978
4. Introducing new Qiskit Runtime capabilities—and how our clients are integrating them into their use. https://research.ibm.com/blog/qiskit-runtime-capabilities-integration
5. Qiskit Runtime IBM Client. https://github.com/Qiskit/qiskit-ibm-runtime
6. Fisher–Yates shuffle, Wikipedia. https://en.wikipedia.org/wiki/Fisher%E2%80%93Yates_shuffle
7. Qiskit Terra API Reference, Primitives, Sampler. https://qiskit.org/documentation/stubs/qiskit.primitives.Sampler.html
8. Qiskit Terra API Reference, Primitives, Source code for qiskit.primitives.sampler. https://qiskit.org/documentation/_modules/qiskit/primitives/sampler.html#Sampler
9. Qiskit IBM runtime, Configure error mitigation, Advanced resilience options. https://github.com/Qiskit/qiskit-ibm-runtime/blob/ab7486d6837652d54cb60b83cfaa9165f5d0484c/docs/how_to/error-mitigation.rst#advanced-resilience-options

Post-Quantum Cryptography Key Exchange to Extend a High-Security QKD Platform into the Mobile 5G/6G Networks

Ronny Döring[1(✉)], Marc Geitz[1], and Ralf-Peter Braun[2]

[1] Deutsche Telekom AG, T-Labs, Winterfeldtstr. 21, 10781 Berlin, Germany
{ronny.doering,marc.geitz}@telekom.de
[2] ORBIT Gesellschaft für Applikations- und Informationssysteme mbH, Mildred-Scheel-Str. 1, 53175 Bonn, Germany
ralf-peter.braun@orbit.de

Abstract. This paper presents a way to integrate a Post-Quantum Cryptography key exchange mechanism based on the encryption and signature algorithms currently being standardized by the National Institute of Standards and Technology into a Quantum Key Distribution platform. In contrast to Quantum Key Distribution, the security of Post-Quantum Cryptography is based on the mathematical complexity of the cryptographic algorithms. The unique feature of the presented solution is that the Post-Quantum Cryptography key exchange mimics a Quantum Key Distribution system. As implemented, the encryption keys are continuously exchanged between the nodes that are part of the OpenQKD testbed Berlin and then stored in a secure key store. The testbed's key management can distinguish between Quantum Key Distribution and Post-Quantum Cryptography keys based on metadata information, allowing applications to select the appropriate key type. This architecture enables interoperability between the two technologies and may also provide a means to deliver quantum-secure keys to the end user by leveraging Post-Quantum Cryptography to secure the last mile.

Keywords: Post-Quantum Cryptography (PQC) · Quantum Key Distribution (QKD) · Asymmetric Key Exchange · 5G · 6G · Integration of PQC and QKD

1 Introduction

Many of today's asynchronous cryptographic systems will not be secure if a quantum computer powerful enough to run Shor's algorithm [13] becomes a reality. Network providers are closely monitoring the development of such machines,

This work is partly funded by the European Research and Innovation Program Horizon 2020 under the contract number 857156 (OpenQKD). Further details can be found at https://openqkd.eu.

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 148–158, 2024.
https://doi.org/10.1007/978-3-031-59619-3_13

because quantum computers could pose a threat to their business model of providing trustworthy and secure communication networks and services. One of the quantum-secure methods used for key exchange, the process of securely sharing a symmetric encryption key between two parties, is Quantum Key Distribution (QKD). QKD has been proven [2] and accepted as a secure means of key exchange between two parties. However, security is highly dependent on the technical implementation of QKD. The existence of technical side channels [9], the range limitation to a few hundred kilometers, the requirement of trusted network nodes for longer distances [8], and the need for special hardware are well known drawbacks of QKD networks. Because of these reasons and the practical challenges in QKD [5], the technology would only be feasible for core networks or special infrastructure to secure the physical layer. Another method of quantum-secure key exchange is Post-Quantum Cryptography (PQC). PQC algorithms provide similar functionalities to the established key exchange and authentication mechanisms Rivest-Shamir-Adleman (RSA [12]) and elliptic-curve cryptography (ECC). Thus, PQC can be used as a replacement for today's asymmetric cryptography and can be integrated into existing systems. The advantages of PQC over QKD are remarkable: PQC algorithms run as software, are network independent, and have no distance limitations. The downside is that PQC security has not been proven.

The work presented in this paper has been carried out within the OpenQKD project, one of the EU Horizon 2020 quantum flagship projects. The goal of OpenQKD is to demonstrate the feasibility of QKD in everyday use cases to enable the industry to adopt QKD technology. One of its defined goals was to investigate the interoperability of QKD with PQC. Deutsche Telekom has set up one of four major European testbeds for the OpenQKD project in the Berlin metropolitan area in order to demonstrate that QKD and PQC can be integrated into a single, quantum-secure platform. We have implemented a PQC key exchange mechanism, where security is based on mathematical complexity, that mimics a traditional QKD key exchange process using existing PQC algorithms. By relying on PQC, we can extend a QKD security platform, such as that deployed by a network operator, to cellular networks and even integrate mobile clients like smartphones and cars.

The paper is organized as follows. First, we briefly outline the architecture of the OpenQKD testbed Berlin in Sect. 2. The next section, Sect. 3, describes key exchange and authentication using PQC to mimic QKD and highlights some advantages of this method. In Sect. 4, we perform a security analysis of the proposed method and discuss some things that need to be considered when implementing such a solution. The extensibility of the platform and methods are presented in Sect. 5. Finally, we conclude this paper in Sect. 6 and mention some plans for future work.

2 Architecture

Our Berlin OpenQKD testbed architecture, presented in [3], is easily extensible due to its layered structure, which allows for a separation of functionalities between the different layers. Three layers play a key role in the architectural design of the testbed:

- Quantum layer
- Key management layer
- Application layer

The quantum layer ensures that symmetric keys are continuously exchanged between connected nodes, while the key management layer is responsible for storing the keys and providing them to the software and hardware in the application layer.

In the Berlin testbed, we deployed three nodes and different types of applications, but the number of nodes and applications is not limited. As deployed, the architecture and the interaction between the layers can be seen in Fig. 1.

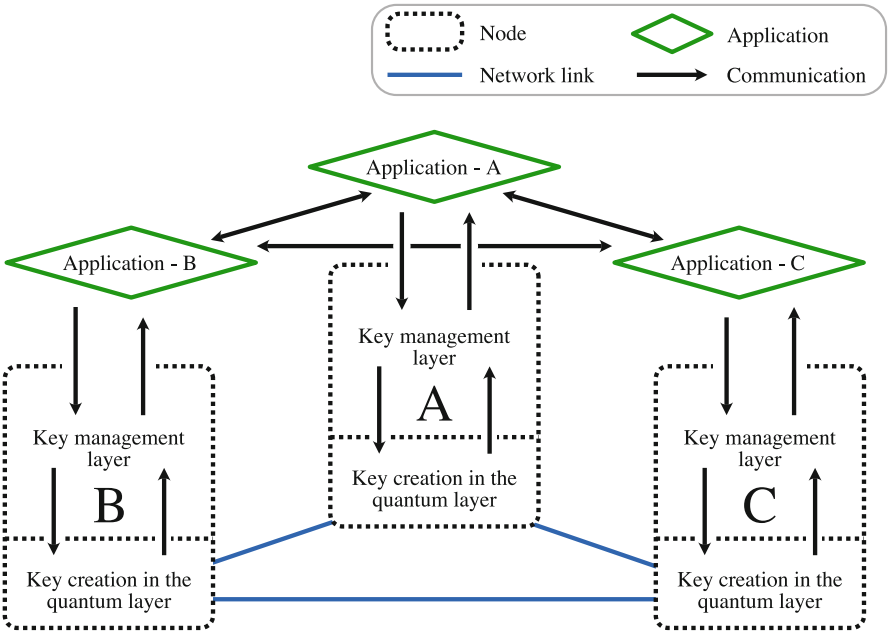
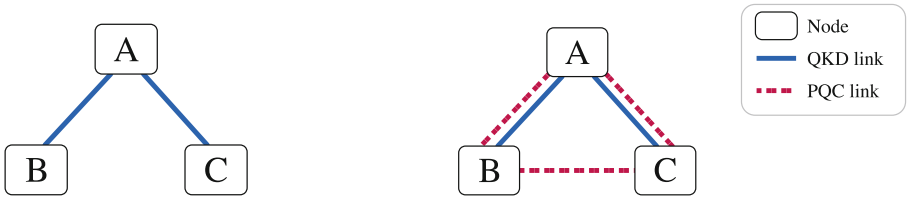


Fig. 1. OpenQKD testbed layered architecture and the interaction between the quantum, key management, and application layer.

2.1 PQC Key Creation in the Quantum Layer

In the OpenQKD testbed, the quantum layer consists of nodes A, B, and C with QKD links between A and B and A and C, as illustrated in Fig. 2a. Note that there is no direct QKD link from node B to C.



(a) QKD links (b) QKD and PQC links

Fig. 2. Types of links in the quantum layer going from pure QKD links (a) to a PQC-connected full mesh network (b).

The QKD systems continuously exchange keys between the connected nodes with a certain key rate, which is mainly influenced by the link attenuation [14]. In this example, we will assume that it is not possible to add a QKD link between B and C because there can't be a direct fiber connection due to the distance between the nodes being too large or for other reasons that prevent a QKD installation. Instead, a PQC link is established. Using PQC links (Fig. 2b), all nodes in the testbed are fully meshed, with the advantage that an additional node does not require a dedicated physical link to all other nodes, as would be the case with QKD. This allows for better scalability.

2.2 Key Management Layer

The whole idea of the key management layer is to abstract the quantum layer and make it a generic interface for external applications. This is useful because it allows an integration of QKD systems from different vendors regardless of their interface implementations. The key management layer consists of three core components: a secure key store (SKS) that acts as a buffer for the keys, a key export that retrieves the keys from the QKD systems and imports them into the SKS, and a user key management interface that ensures that the requesting party accesses the keys from the SKS in an appropriate manner. The key management ensures that the same secret encryption key cannot be retrieved multiple times.

2.3 Application Layer

When two application instances want to communicate securely, they contact their user key management based on the ETSI GS QKD 014 [18] standard. An instance starts by requesting a random key from its user key management. It

must then communicate the id of the received key to the other instance in some form of preliminary communication. This preliminary communication between the two instances does not need to be encrypted, only authenticated. With the key id, the other instance is then able to retrieve the same key from its own user key management. From this point on, both instances are able to encrypt the communication between them symmetrically using the keys they received from the user key management.

3 PQC Key Exchange Mechanism

The PQC key exchange system is part of the quantum layer of the infrastructure because the system exchanges encryption keys just like QKD systems. For encryption key persistence, the PQC key exchange system integrates with the key management by storing keys in the SKS. To facilitate integration, the PQC key exchange interface is identical to the interfaces of QKD systems using the ETSI GS QKD 014 standard. The PQC key exchange uses asymmetric cryptography to exchange keys and authenticate message endpoints. Asymmetric cryptography schemes require the generation of two key pairs for each node, with one public and one private key per pair. The public keys are distributed to remote network nodes, while the private keys are kept local and secret.

In the PQC domain, key exchange is performed using Key Encapsulation Mechanisms (KEM), which generate a shared encryption key and a ciphertext that can be transmitted to distribute the encryption key. Message verification can be achieved using signature algorithms (SIG). The diagram shown in Fig. 3 describes the processing and communication that two nodes (Alice and Bob) must perform to successfully derive the shared key on each node.

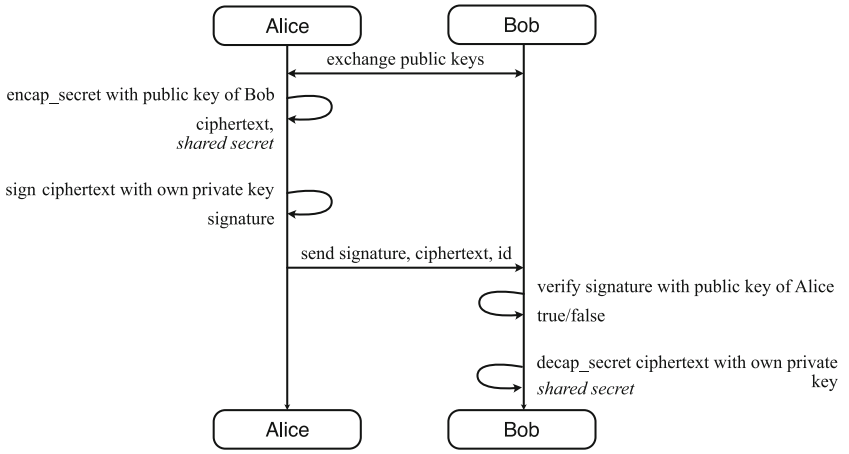


Fig. 3. PQC key exchange and authentication between two parties (Alice and Bob) and their communication.

On Alice's side, we encapsulate a secret key (hence the name KEM) with Bob's public key, producing a ciphertext. The secret key itself is generated as part of the encapsulation process. We then sign the ciphertext with Alice's own private key, creating a signature. The signature, ciphertext, and a uniquely generated identifier (id), such as a UUID [10], are then sent to Bob. On Bob, we verify the signature with Alice's public key. Once we know that our communication partner is Alice, we derive the secret key by performing a decryption operation on the ciphertext with Bob's own private key. This allows Bob to obtain the same secret key that Alice previously encapsulated. They store these shared keys with the corresponding ids in their local SKS, allowing applications to retrieve the same key on different nodes by id. This protocol has no replay attack protection [1], meaning that an attacker could send the same message Alice sent to Bob over and over again without Bob noticing. Bob could mitigate this, e.g., by hashing the shared key and keeping track of these hashes to detect duplicate keys, which would certainly mean that the ciphertext has been received before.

Performing key exchange in the manner described above requires many communication steps to exchange keys continuously because the messages sent are very small. To increase transmission speed, we can repeat the encapsulation step on Alice several times and then send all those ciphertexts and ids to Bob with only one signature inside a larger message. This method could also be used bidirectionally, allowing both Alice and Bob to initiate the communication, potentially doubling the key exchange rate.

3.1 Berlin Testbed Implementation

For our Berlin testbed implementation, we used an HP Enterprise DL380 Gen9 server with two Intel Xeon E5-2640 v4 CPUs at 2.40 GHz, running Ubuntu 20.04 on each node. In addition, KVM allowed us to create virtual machines separating various software services in the three architectural layers to integrate hardware appliances such as the SKS and to run our user key management. We connected the three nodes using a Virtual Private Network (VPN). The PQC key exchange and message verification implemented in Python was achieved using a client-server architecture. The key exchange process was initiated from the client side. Each node runs a server and a client, allowing bi-directional key exchange. For our implementations, we used the OpenQuantumSafe (OQS) project [16] and its liboqs library [15], which integrates well with industry standards such as openssl for cryptography and Python for general-purpose programming with quantum-secure communication capabilities. The liboqs library provides an implementation of a variety of different PQC algorithms and supports all of the finalists of the NIST standardization initiative [11]. This means that the specific PQC key exchange and authentication algorithms used in our implementation can be easily varied according to security and performance requirements.

In our initial implementation, we used Kyber1024 for key exchange and Falcon1024 for digital signatures. We chose the most performant algorithms from the standardization initiative based on our previous findings [6]. When running our PQC solution, we decided to use a block size of 10 PQC keys per key exchange

session every 5 min. In continuous operation, the bitrate increases significantly. Compared to existing QKD systems, the bitrates achieved with PQC are similar and can even be higher, especially when we started to optimize our solution, as can be seen in Table 1.

Table 1. Average key exchange rates of different QKD systems and our initial PQC implementations.

System	Bitrate [kbps]
IDQuantique Cerberis3 QKD @ 1,550 nm	2.1
Toshiba QKD @ 1,550 nm	1,970.0
PQC implementation (naive)	3.7
PQC implementation (optimized)	4,506.7
PQC implementation (optimized, multithreaded)	15,123.5

The naive solution, implemented in Python 3.8 using liboqs, sends messages over HTTP and exchanges only one key per message, as described in Sect. 3. Since HTTP is a stateless protocol, the signature is required in each message for authentication. This signature and the otherwise small message sizes limit the key rate to about 4 kbps. Our optimized solution, implemented in Rust 1.64 with liboqs, uses bare TCP. Since a TCP connection is stateful, both communicating parties only need to authenticate once, reducing the number of signatures that need to be sent. Each key exchange message also bundles about 1000 ciphertexts and corresponding ids, allowing a rate of 4.5 Mbps between each pair and up to 15 Mbps with multithreading. This is not the end of the optimizations, as the client and server software is designed to scale horizontally, allowing it to be run on multiple processes. Not only does this allow the key rate to increase even further, but we have done all of this while maintaining a full mesh between nodes. Since PQC key exchange is not limited to wired network connections, we also included two 5G access points in our network. In particular, we connected two nodes over the 5G cellular network. In our experiments, we did not observe a reduced key rate in this scenario, indicating that the computation required by PQC was the bottleneck, not the network itself. This may change if multiple key exchange services are run in parallel or if a less powerful network connection is used.

3.2 Advantages

The architecture of the testbed presented in Sect. 2 is agnostic to how keys are generated and thus can modularly incorporate different key exchange methods. Depending on the required security between two nodes, a QKD link, a PQC link, or any other innovative key exchange technology can be used in the quantum layer. The key management layer can then distinguish between different types

of keys and provide the type of key that best suits the application scenario through a standardized ETSI GS QKD 014 interface. PQC as a key exchange method also means that the node network can be easily extended, even using wireless connections. All we need to do is generate PQC certificates, deploy them on all nodes, and ensure there is a connection between the nodes. By leveraging PQC algorithms monitored by standardization bodies such as NIST, the likelihood of fundamental security flaws is also reduced. Unlike QKD, PQC key exchange is network-agnostic and not limited by distance. This means it is also network topology agnostic, allowing for ring, star, line, bus, and mesh networks. It is possible to transmit network traffic over cellular, line-of-sight, or even non-terrestrial networks, enabling intercontinental key exchange.

The architecture enables a whole new range of quantum-secure key exchange networks, including Metropolitan Area Networks where nodes do not have fiber connectivity and Wide Area Networks with nodes in rural areas. In practice, we see PQC key exchange as excellent for isolated subnets, networks where QKD range limitations become important, or networks requiring integration into a security platform, i.e., where PQC is complemented by other key exchange methods. In Fig. 4, we show an example of such interconnected subnets. Nodes within subnets are connected via QKD, wired PQC, 5G, or even satellite links, depending on the network's security and range requirements. Redundant links can be added to meet the specific requirements of quantum-secure networks.

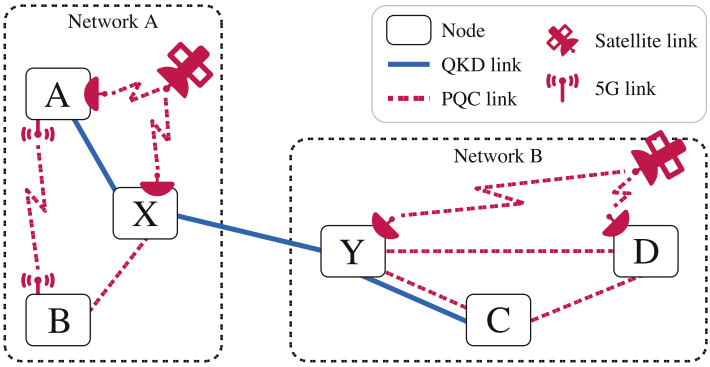


Fig. 4. Complex network example combining PQC and QKD key exchange using wireless and wired connections.

4 Security Analysis

Due to the fact that key exchange is based on PQC, there are limitations that need to be considered when it comes to the security of the method. First, there is no security proof for any PQC algorithm, and security depends strongly on

mathematical complexity. This has huge implications and manifests itself in recent candidates of the PQC standardization process being broken by security researchers [4]. This also means that the PQC algorithms in use must be continuously monitored to respond to possible breaches. To reduce the impact of such breaches, cryptographic agility must be at the heart of the implementation, meaning that cryptographic algorithms can be easily replaced with similar ones. Since the way we use PQC key exchange is implemented purely as software, it is also susceptible to implementation flaws. In the future, monitoring and certification of the software by security experts could significantly increase trust in this method. This is especially important for network operators, who must follow the recommendations of their respective security agencies (e.g., Federal Office for Information Security, Germany and National Institute of Standards and Technology, United States). Since cryptographic agility is the only response to security flaws in the design of a PQC algorithm, we could also try to implement hybrid protocols that use both QKD and PQC or a combination of different PQC algorithms to strengthen the overall security. Some of these methods are described in more detail in [7]. If the security of PQC-generated keys is deemed insufficient for the application scenario, QKD-generated keys can be used. For example, a secret of the European Commission may be subject to QKD transmission only. For other types of applications, such as communication between banking institutions and their customers, PQC may be sufficient. Therefore, the choice of the appropriate key type depends entirely on the application and should be made according to the security requirements as well as the distance limitations, where a risk assessment should always be performed.

5 Extensibility of the Security Platform

Both the proposed method for key exchange via PQC and the platform is highly extensible, allowing several use cases to be integrated. To demonstrate the extensibility, we integrated a mobile device via a 5G dongle as a node to our Berlin testbed to exchange keys with the other nodes. The keys were then stored in the local memory of the device. This is especially interesting considering that the Milenage system, which currently contributes to the security of 5G+ networks, was recently analyzed as potentially vulnerable to a quantum attack [17]. That would mean that all applications requiring cellular networks, such as connected cars, robotics, Industry 4.0, mobile clients, as well as IoT devices, would be vulnerable. As a solution to the potential threat, these applications could be integrated into the platform and participate in the PQC key exchange to enable quantum-secure communication capabilities.

When considering PQC for wireless networks, key forwarding within the platform could also be beneficial. This means that QKD keys secured by encryption with PQC keys can be forwarded to mobile devices. Especially in secure and private network environments, this could enable end-to-end quantum-secure network connections, including the last-mile and end-user applications over wired and wireless networks. An example is shown in Fig. 5.

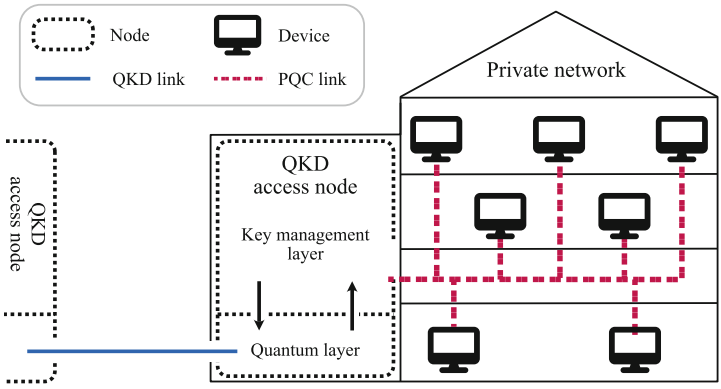


Fig. 5. Example of a private network where the only entry point from outside the network is through a QKD link. Internal communication is secured by PQC to bring the QKD-generated keys to other devices within the network for end-to-end quantum-secure encryption.

In this example network, an office building, there is a QKD node that connects another network via QKD to a separate, private, wired network inside the building. The PQC-secured network is used to deliver the QKD keys to other devices in the office. Security is then based on two factors: PQC and a disjoint, private network that can only be accessed via fiber inside the building. In the best case, an initial authentication, such as a smart card, is required to use the keys to establish a quantum-secure connection to peers in the other network. This would optimize the balance between security and flexibility by using PQC key exchange only within the private network.

6 Conclusion

In this paper, we proposed a possible use of a PQC key exchange mechanism using existing PQC algorithms, which is believed to be resistant to quantum computers and overcomes network distance and media limitations, towards a global quantum key exchange system for wired and wireless networks. The method is extensible and is expected to improve network efficiency, security, and resilience. It is perfectly integrated into key exchange architectures based on QKD technologies. We have demonstrated the feasibility of the proposed scheme in the Berlin OpenQKD testbed, enabling quantum-secure communication between network nodes alongside existing QKD solutions and extending the solution to wireless, last-mile, and end-user applications. In the future, we plan to integrate even more quantum-secure key exchange solutions into the platform to demonstrate its potential. This will also allow us to thoroughly investigate other quantum-secure technologies, especially in terms of their security and the challenges of integrating them.

Acknowledgements. The authors would like to thank their OpenQKD partners for providing the necessary QKD, network, and user equipment and for assisting with the integration of the devices into the testbed. This includes 1,310 nm and 1,550 nm QKD systems from IDQuantique and Toshiba, 10 Gbps network encryptors from ADVA, and IPSec devices from Thales.

References

1. Aura, T.: Strategies against replay attacks. In: Proceedings 10th Computer Security Foundations Workshop, pp. 59–68 (1997). https://doi.org/10.1109/CSFW.1997.596787
2. Biham, E., Boyer, M., Boykin, P.O., Mor, T., Roychowdhury, V.: A proof of the security of quantum key distribution. In: Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing, pp. 715–724 (2000)
3. Braun, R.P., Geitz, M., Döring, R., Ritter, M.: Berlin openqkd testbed evaluating quantum key distribution in provider networks (2022). in press
4. Castryck, W., Decru, T.: An efficient key recovery attack on SIDH (preliminary version). Cryptology ePrint Archive (2022)
5. Diamanti, E., Lo, H.K., Qi, B., Yuan, Z.: Practical challenges in quantum key distribution. npj Quantum Inf. **2**(1), 1–12 (2016)
6. Döring, R., Geitz, M.: Post-quantum cryptography in use: empirical analysis of the TLS handshake performance. In: NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, pp. 1–5 (2022). https://doi.org/10.1109/NOMS54207.2022.9789913
7. Geitz, M., Döring, R., Braun, R.P.: Hybrid QKD and PQC protocols implemented in the berlin OpenQKD testbed (2022). in press
8. Huttner, B., et al.: Long-range QKD without trusted nodes is not possible with current technology. npj Quantum Inf. **8**(1), 1–5 (2022)
9. Lamas-Linares, A., Kurtsiefer, C.: Breaking a quantum key distribution system through a timing side channel. Opt. Express **15**(15), 9388–9393 (2007)
10. Leach, P., Mealling, M., Salz, R.: A universally unique identifier (uuid) urn namespace. Technical report (2005)
11. Moody, D.: Let's get ready to rumble. the nist pqc "competition". In: Proceedings of First PQC Standardization Conference, pp. 11–13 (2018)
12. Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**(2), 120–126 (1978)
13. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev. **41**(2), 303–332 (1999)
14. Takeoka, M., Guha, S., Wilde, M.M.: Fundamental rate-loss tradeoff for optical quantum key distribution. Nat. Commun. **5**(1), 1–7 (2014)
15. The Open Quantum Safe project: liboqs - an open source c library for quantum-safe cryptographic algorithms (2022). https://github.com/open-quantum-safe/liboqs. Accessed 1 Dec 2022
16. The Open Quantum Safe Project: Software for prototyping quantum-resistant cryptography (2022). https://openquantumsafe.org/. Accessed 1 Dec 2022
17. Ulitzsch, V., Seifert, J.P.: Breaking the quadratic barrier: Quantum cryptanalysis of milenage, telecommunications' cryptographic backbone. Cryptology ePrint Archive (2022)
18. Yoshimizi, T., et al.: Quantum key distribution (QKD); protocol and data format of rest-based key delivery API (2019)

Quantum-Secure Autonomous Factories: Hybrid TLS 1.3 for Inter- and Intra-plant Communication

Wolfgang Rohde[1], Maria Perepechaenko[2], and Randy Kuang[2(✉)]

[1] Digital Factory Division, Siemens Digital Industries Software, Charlotte, NC, USA
wolfgang.rohde@siemens.com
[2] Quantropi Inc., Ottawa, Canada
randy.kuang@quantropi.com

Abstract. Given recent advances in the field of quantum cryptography, cryptographic agility is essential to protect data at rest and in transit. We consider hybrid key exchange and authentication in the framework of the TLS1.3 protocol. Traditional hybrid mode, which uses classical and quantum-safe algorithms together to secure data in the event of one of the algorithms being broken, has been a subject of research for a few years now. However, there are certain engineering challenges that come with using more than one algorithm in the framework of the TLS protocol. In this work, we propose a different hybrid mode that allows seamless negotiation between classical and quantum-safe algorithms. The proposed hybrid mode has different goals from the conventional hybrid mode and could be an excellent solution for entities not required to use standardized cryptographic algorithms and can be used in the timeframe when quantum-safe algorithms have been standardized but certain endpoints have not yet transitioned to post-quantum. We discuss conventional and proposed hybrid modes, described the proposed solution in detail, and briefly review engineering challenges associated with each hybrid mode.

Keywords: TLS protocol · Transport Layer Security · hybrid TLS1.3 · hybrid cryptographic protocol · quantum-safe TLS1.3 · quantum-safe algorithms · PQC · post-quantum cryptography · autonomous factories · secure communication · quantum-safe communication

1 Introduction

In the past few years, the field of quantum computing has been advancing at a rapid speed. Indeed, in its quantum roadmap, IBM has announced that the company is preparing to offer a system with over a thousand qubits in 2023 [1]. Gilbert *et al.* proposed silicon-based quantum computers, allowing for the possibility of a quantum system with billions of qubits in the near future [2]. These advancements highlight the urgency of deploying quantum-safe algorithms for key exchange and digital signatures (PQC). NIST has recently announced candidate PQC algorithms for standardization. However, the transition guidelines to the PQC standards have not been defined yet [3].

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, pp. 159–169, 2024.
https://doi.org/10.1007/978-3-031-59619-3_14

Historically, there are examples of cryptographic algorithms transitions that took a long time [4–6]. NSA projects that the complete migration of their cryptographic suite will take up until the year 2033 [7]. Accounting for the advancement in quantum computing, NIST is planning to provide transition guidelines and propose different stages of migration [3]. However, a few questions remain. For instance, we must understand that various entities might not follow the same timeline. Moreover, how do we account for the 'steal now and crack later' tactics? A feasible solution in this situation is hybrid modes.

NIST SP800-56C REV. 2 allows for a particular hybrid mode in the framework of the key establishment schemes [8]. They propose to use a shared secret Z', which is a concatenation of a classical secret Z, shared using a conventional classical algorithm, with a secret T, that was shared using some other method. There are, however, other ways to create hybrid modes. For instance, the two keys can be XOR'ed instead of concatenated. In fact, there are many possible solutions examined in the literature [5, 9–14]. However, not all hybrid modes are equivalent in terms of engineering challenges that come with implementation. Network protocols or certain areas such as IoT might not benefit from a conventional hybrid mode, where more than one algorithm is used together. Indeed, due to significant resource constraints, IoT devices might not support certain hybrid modes. Meanwhile, they might not require standard algorithms. They might benefit from a hybrid mode that allows them to seamlessly negotiate between classical and other algorithms to use for key exchange and authentication.

2 Motivation

Our global economy and social life are heavily depending on the secure and reliable exchange of information through digital media. With rise of augmented reality, artificial intelligence, wireless communication (5G) and the Metaverse the dependency on secure encryption increases constantly. In combination with the decrease costs to store large amount of data, the vulnerability regarding breaking the classical encryption in near future is immanent.

Given the omnipresence of encryption, a change in the encryption infrastructure (e.g., public key infrastructure) is basically a make-over of the entire IT infrastructure. Every encryption endpoint must be touched to a more secure way to encrypt data. Besides the costs involved in this endeavor, time is of the essence.

Looking at the complex IT infrastructure of a global supply chain ecosystems, there is no normative authority across the entire supply chain that could mandate a minimum technological level. In general, in manufacturing, there are different domains involved.

2.1 Intra-plant Communication

In terms of security, one of the biggest challenges for intra-plant communication is that once an attacker breaches the firewall into the plant, many systems within the plant are not held to the same security level as ERP systems or cannot be sufficiently protected because of the lack of computing power, for example, CNC machines.

2.2 Inter-plant Communication

For inter-plant communication, the security standard is often higher because of the obvious exposure to the outside. However, the severity of a potential breach is significantly higher, because once the communication is breached, potentially, all plants are open for attack.

Even if the industry would start to replace devices today, it would take years to replace/update every communication point. A solution with few or no physical replacements is demanded from a business perspective to minimize the migration time to a quantum-secure environment. A pure software solution as a hybrid approach is the preferred way. From a technical perspective, this solution could be rolled out into the field by a routine software update. All critical aspects of switching from classical to quantum would be fully transparent to the users.

Both aspects lowering the costs for migration and ease of transition, are critical for decision-makers in business to avoid the Waiting Problem. It enables decision-makers to transition from classical to quantum encryption as fast as possible with minimal financial and technical risks.

3 Contributions

In this work, we consider hybrid key exchange and authentication in the framework of the TLS1.3 protocol. Hybrid TLS protocol has been the subject of scrutinous study [5, 9–11, 13, 14]. A conventional hybrid TLS solution combines classical and quantum-safe algorithms to protect data in the event of one of the algorithms being cracked. In this work, we propose a new TLS1.3 hybrid mode – *HybridB*. The *HybridB* mode differs from the conventional TLS hybrid mode, as defined in [9–11, 13]. *HybridB* does not have the same objectives as the conventional hybrid mode. That is, it is not designed to use a combination of different algorithms to protect data, but rather provides a choice between classical and quantum-safe algorithms to use for key exchange and authentication. We do not claim that one of the hybrid modes is superior to the other but rather give the reader a new perspective and a choice between the two hybrid modes, considering use-cases, engineering limitations, security, government regulations, and other aspects.

4 Hybrid TLS 1.3

With advances in quantum computing, cryptographic agility became an essential practice of information security. It is particularly relevant in the context of encrypted communication over the internet, namely in the framework of the TLS protocol.

The TLS protocol is designed to be agile, supporting multiple cryptographic primitives and algorithms at once within each category of functionality [15]. The collection of these algorithms is referred to as a Cipher Suite. The Client and the Server can negotiate which Cipher Suite they choose to use and establish parameters of communication accordingly. These flexibilities allow us to consider more than one hybrid key-exchange and authentication in the framework of TLS. Furthermore, we can define more than one interpretation of what a hybrid mode in TLS can encompass.

4.1 Hybrid Modes

We consider two distinct interpretations of a hybrid mode. The first hybrid mode entails using two (or more) algorithms simultaneously, one classical and one quantum-safe, providing security even if one of the algorithms is broken. This conventional definition of a hybrid TLS mode has been subjected to research as in [9–11, 13]. In the remainder of this work, we refer to this hybrid mode as *HybridA*.

We propose another interpretation of a hybrid mode, emphasizing a choice between quantum-safe or classical algorithms, instead of using them together. The second interpretation involves using either classical or quantum-safe algorithms, allowing for the negotiation of either a classical or quantum-safe Cipher Suite. We denote this mode *HybridB*.

In network protocols such as TLS, hybrid modes bring about particular challenges that need to be addressed through engineering modifications. For instance, the TLS protocol might not support novel complex algorithms due to the inherent limitations of the protocol itself. In addition, modification of the TLS algorithm to allow for the negotiation of combinations of algorithms, and multiple shared secrets, public keys, etc., might lead to latency, reduction in performance, and other vulnerabilities. Note that some hybrid modes bring more engineering challenges and modifications to the protocol itself than others. On the other hand, some hybrid modes increase the security of the protocol while others do not.

4.2 Goals of a Hybrid Mode

The two hybrid modes that we have discussed above address different goals. The primary goal of the *HybridA* mode is to ensure the security of the protocol in the event of one of the component protocols being broken. Indeed, in the framework of the *HybridA*, the data is protected with two different cryptosystems, one of which is quantum-safe. *HybridA* is an excellent solution for entities that must comply with the current industry or government regulations, and thus must use classical certified and standardized algorithms, while looking to add an extra layer of protection against quantum adversaries. Having classical and quantum-safe algorithms used in conjunction ensures that if the classical component gets broken, the data remains secure with a quantum-safe component, and vice versa.

The main objective of a *HybridB* approach is to provide an efficient protocol that seamlessly allows two parties to interact regardless of whether each party supports quantum-safe algorithms. *HybridB* does not share the same engineering challenges with *HybridA* and requires fewer modifications to the existing framework of the TLS 1.3 protocol. For instance, if only a single algorithm is used for key exchange as opposed to multiple algorithms, the parties do not need to share multiple public keys, ciphertexts, and signatures. Moreover, the protocol does not need to be modified to support multiple cryptographic data or a combination of different algorithms with specific features. *HybridB* is a desirable solution for entities not required to use standardized cryptographic algorithms, as well as for the timeframe when quantum-safe algorithms have been standardized, but some endpoints have not yet transitioned to post-quantum algorithms.

Both hybrid modes share some common engineering objectives that include reducing compatibility risks, ensuring that the performance improves or remains unchanged, no extra round trips should be introduced, and the encryption and authentication latency should remain low. For instance, TLS 1.3 requires only a single round-trip compared to the older versions, which cuts the encryption latency in half [15]. Hybridizing the protocol should not affect the improvements of the TLS protocol achieved compared to the older versions.

5 Proposed Solution

We propose TLS 1.3 with hybrid key exchange and authentication introduced in this work as the *HybridB* mode. In this section, we describe the proposed solution in detail, starting with the key exchange phase. We consider the scenario where the client supports quantum-safe algorithms, and the server may or may not support quantum-safe algorithms. Note that if the client does not support quantum-safe algorithms, then the TLS handshake is carried out as usual.

5.1 Key Exchange

The key exchange phase establishes shared keying material and selects cryptographic parameters. The key exchange procedure starts with the client sending the ClientHello message to share its cryptographic capabilities. The ClientHello message in the framework of the *HybridB* mode contains the following:

- cipher_suites: this field remains the same as the conventional TLS 1.3, containing a list of the symmetric schemes supported by the client and a hash used in the key derivation function.
- extensions: the ClientHello extensions contain but are not limited to the following:
- supported_groups: this extension is used to share a list of conventional classical key exchange algorithms supported by the client

  supported_pqc: this is an additional extension used to share a list of quantum-safe algorithms supported by the client. Note also that the NamedGroup enum must be extended to include algorithm identifiers for the quantum-safe algorithms key_share: key exchange information, and in particular public keys corresponding to the conventional classical algorithms are included in this extension

- key_share_pqc: this is an additional extension that contains key exchange information corresponding to the quantum-safe algorithms supported by the client, in particular public keys
- client_random: client random will only be used if the server does not support quantum-safe algorithms or if the quantum-safe algorithms are key exchange algorithms as opposed to key encapsulation algorithms. Otherwise, the server will ignore the client_random.

The server receives the ClientHello message and creates the ServerHello message containing the following fields:

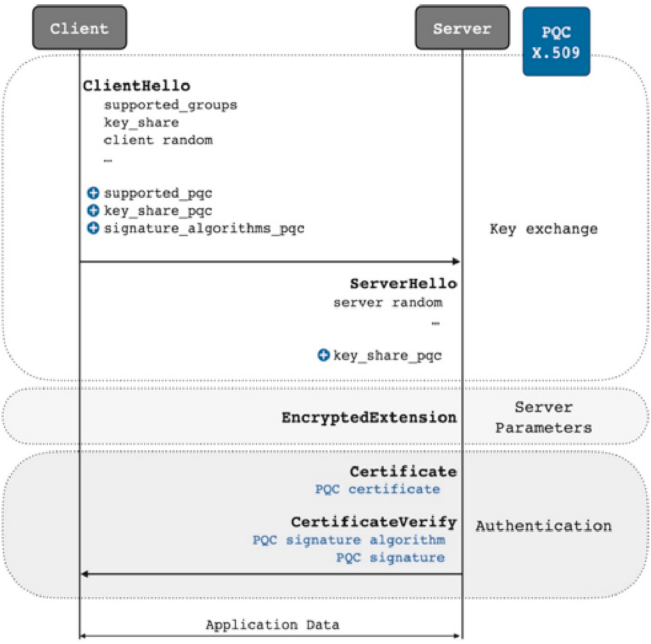


Fig. 1. A scheme of a TLS Handshake in the framework of the *HybridB* mode, given that both the client and the server support post-quantum algorithms.

- cipher_suites: this field remains the same with the conventional TLS 1.3, containing a single Cipher Suite selected by the server.
- extensions: the ServerHello extensions contain but are not limited to the following:
- key_share or key_share_pqc: the server without quantum-safe capabilities ignores the client's key_share_pqc extension and the supported_pqc extension, and specifies a single NamedGroup value, after selecting a conventional classical algorithm. Otherwise, the server ignores the extensions corresponding to classical algorithms and selects a quantum-safe algorithm, specifying the NamedGroup value. The server also includes a list of cryptographic parameters corresponding to the selected primitive.
- server_random: in the case that a KEM algorithm is selected, server_random is a secret encrypted using the client's public key. Note that in the case of KEM, the client generates a new key pair for every connection and securely stores the secret key. Otherwise, server_random has a conventional interpretation as in the traditional TLS protocol.

5.2 Authentication

We discuss hybrid authentication in the framework of TLS 1.3 in *HybridB* mode. This work considers the scenario where only the server is authenticated. The ClientHello message in the framework of the *HybridB* mode must include the following:

- extension:

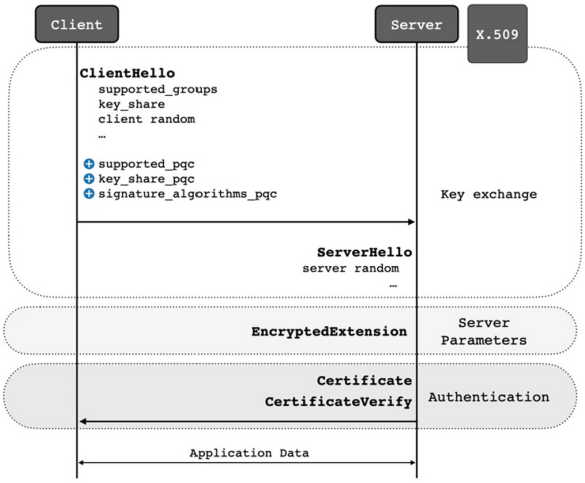


Fig. 2. A scheme of a TLS Handshake in the framework of the *HybridB* mode, given that the client supports post-quantum algorithms, but the server supports only classical algorithms.

- signature_algorithms_pqc: this extension includes quantum-safe digital signature algorithms supported by the client.

If the server does not support quantum-safe digital signature algorithms, then it ignores the client's signature_algorithms_pqc extension. Moreover, no modifications will be made to the existing portion of the X.509 certificate corresponding to authentication. The server's Certificate and CertificateVerify messages are not modified, and the rest of the protocol is carried out as usual.

If the server supports quantum-safe algorithms, then the X.509 certificate is modified to include public keys and signatures, corresponding to the quantum-safe algorithms. The server's Certificate message includes a quantum-safe certificate. The server's CertificateVerify message includes the quantum-safe algorithm used and the corresponding signature.

We provide the reader with Fig. 1 and Fig. 2 to facilitate a better understanding of the proposed solution.

5.3 Alternative Solution

An alternative to the proposed hybrid mode *HybridB* is another hybrid mode, which we refer to as *HybridA*. The *HybridA* mode has been a topic of research in multiple works [9–11, 13]. Note that its use-cases are different from that of the *HybridB* mode. *HybridA* mode is designed to allow for multiple algorithms to be used together for key exchange and/or authentication within the same protocol. Although strengthening the protocol's security, this design introduces a noticeable amount of design challenges and modifications to the TLS protocol. For instance, the components algorithms have to be negotiated in a "this and that" fashion instead of the "this or that" fashion. These component algorithms then need to be combined, that is, more than a single public

key needs to be conveyed. A decision needs to be made about how the keys must be combined. Authentication brings its own challenges, for instance, the CertificateVerify message does not have a built-in way of being extended and so some modifications to the protocol logic must be made. In fact, *HybridA* involves more substantial changes to the protocol's logic than *HybridB*.

We mentioned that *HybridB* does not offer the same security as *HybridA* but is intended for use-cases when entities are not required to be secured with different algorithms. Moreover, *HybridB* is an excellent candidate for the (not-so-distant) future, when quantum-safe algorithms are standardized, but some endpoints have not yet migrated to quantum-safe, as we have seen with ECC [4–6]. We also stated that *HybridB* does not involve as many design considerations or modifications to the TLS protocol as *HybridA*.

5.4 Design Challenges

The *HybridB* mode brings about a few design challenges. Recall, that during the key exchange procedure, the client creates new extensions, namely supported_pqc, signature_algorithms_pqc, and key_share_pqc. Adding new extensions requires modification to the logic of the TLS protocol. In addition, the NamedGroup enum must be updated to include quantum-safe algorithms. The certificate must also be updated to include quantum-safe algorithms, public keys, and signatures. This modification requires special attention since these changes should not affect backward compatibility. Moreover, the size limit of the X.509 certificate must be considered.

There is more than one way to design the *HybridA* mode, as discussed in [9, 11, 13]. We will consider a design that is closest to the *HybridB* mode. We assume that the two algorithms for key exchange and digital signature are negotiated individually by adding new extensions, and the NamedGroup enum is updated to include quantum-safe algorithms. We also assume that new key share extensions are created to convey keys for quantum-safe algorithms. These steps are equivalent to the corresponding steps in the framework of *HybridB and* require new logic to be included in the protocol. In the framework of the *HybridA* mode, the challenge now becomes deciding how the keys must be used. For one, each party might concatenate the keys established by each component algorithm and use the result in the protocol's key schedule. This requires new logic. Another challenge that must be considered in the design of the *HybridA* mode is that only the ClientHello and ServerHello messages support extensions. The certificate in the framework of *HybridA* is modified much like in the *HybridB* mode. However, the CertificateVerify message does not have any built-in way of being extended, so in the case of *HybridA*, some changes must be made to the protocol's logic. We provide the reader with Fig. 3 and Fig. 4 to illustrate high-level summary of the design challenges that both hybrid modes bring about.

Note, that much depends on the design of the hybrid modes. There are multiple ways to design *HybridA* mode, shown in [9–11, 13]. There are other ways to design the *HybridB* mode. However, in most cases, there aren't as many modifications done to the protocol in the framework of *HybridB*, regardless of the design, compared to the *HybridA* mode.

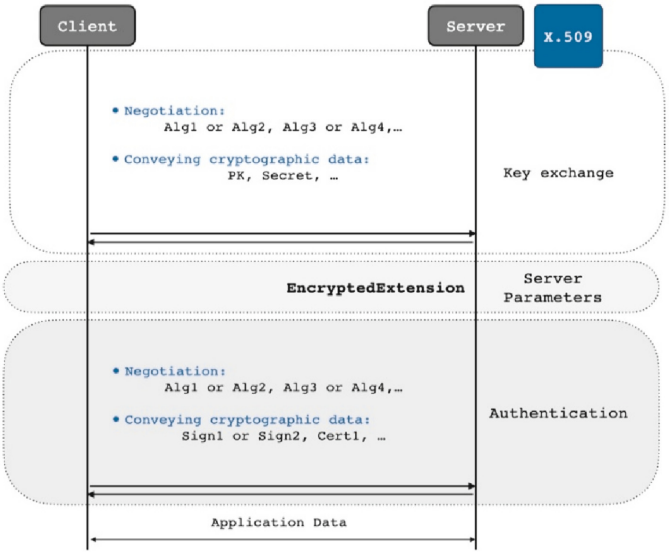


Fig. 3. High-level summary of the engineering decision considerations needed in the framework of the *HybridB* mode.

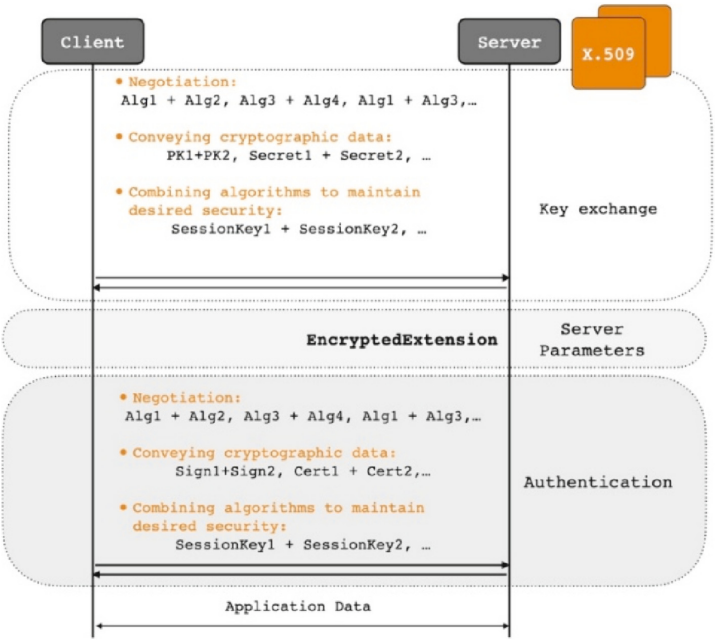


Fig. 4. High-level summary of the engineering decision considerations needed in the framework of the *HybridA* mode.

6 Conclusions and Future Work

In this work we considered the subject of cryptographic agility in network protocols and proposed a novel hybrid mode in the framework of the TLS1.3 protocol. The proposed hybrid mode differs from the conventional hybrid mode in the sense that it does not combine multiple cryptographic algorithms to be used together. Therefore, the proposed hybrid solution does not aim to secure data in the event of one or more algorithms being broken. The goal of the proposed solution is to allow for seamless negotiation between classical and quantum-safe protocols. This solution evades some of the engineering challenges brought by the conventional hybrid mode and does not require extensive modifications to the protocol itself.

Providing a safe and low-impact technology to prepare the IT infrastructure for the post-quantum challenges is necessary to engage companies and executives as soon as possible. This paper shows how we can answer this challenge by leveraging a hybrid approach based on software.

The next steps are to demonstrate the hybrid approach's viability and amend the strategy with practical steps:

We apply the hybrid approach with a network of manufacturers in real industrial scenarios.

We provide feedback to the research to further enhance the post-quantum strategy and the hybrid technology.

As mentioned in this paper the industry is moving to automation and autonomy of the entire supply chain. The short-term objective is to establish resiliency and sustainability, for humans living on Earth and keep costs under control or reduce costs. However, the long-term objective is to solve the problem that we will run out of resources on our planet in the foreseeable future. The solution is to conquer our solar system and begin asteroid mining and space manufacturing. Given the vast distances in space and the extremely hostile environment for humans, automation and autonomy are the only way to achieve this goal.

References

1. Gambetta, J.: Expanding the IBM Quantum roadmap to anticipate the future of quantum-centric supercomputing, 10 May 2022. https://research.ibm.com/blog/ibm-quantum-roadmap-2025. Accessed 24 Jan 2023
2. Gilbert, W., Tanttu, T., Lim, W.H., et al.: On-demand electrical control of spin qubits. Nat. Nanotechnol. (2023)
3. Moody, D.: NIST PQC: Looking into the Future. NIST (2022). https://csrc.nist.gov/csrc/media/Presentations/2022/nist-pqc-looking-into-the-future/images-media/session-1-moody-looking-into-future-pqc2022.pdf. Accessed 24 Jan 2023
4. Langley, A.: Forward secrecy for Google HTTPS, December 2011. https://www.imperialviolet.org/2011/11/22/forwardsecret.html. Accessed 24 Jan 2023
5. Moeller, B., Bolyard, N., Gupta, V., Blake-Wilson, S.: Elliptic curve cryptography (ECC) cipher suites for Transport Layer Security (TLS). RFC 4492, May 2006. https://rfc-editor.org/rfc/rfc4492.txt. https://doi.org/10.17487/RFC4492. Accessed 24 Jan 2023

6. National Institute of Standards and Technology. Specification for the Digital Signature Standard (DSS). Federal Information Professing Standards (FIPS) 186-2, January 2000. https://csrc.nist.gov/CSRC/media/Publications/fips/186/2/archive/2001-10-05/documents/fips186-2-change1.pdf. Accessed 24 Jan 2023
7. Stern, M.: Transitioning National Security Systems to a Post Quantum Future, 30 November 2022. https://csrc.nist.gov/csrc/media/Presentations/2022/transitioning-national-security-systems-to-a-post/images-media/session3-stern-transitioning-national-security-systems-pqc2022.pdf. Accessed 24 Jan 2023
8. Barker, E., Chen, L., Davis, R.: August 2020. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-56Cr2.pdf. Accessed 24 Jan 2023
9. Stebila, D., Fluhrer, S., Gueron, S.: Design issues for hybrid key exchange in TLS 1.3, 08 July 2019. https://datatracker.ietf.org/doc/html/draft-stebila-tls-hybrid-design-01#page-10. Accessed 24 Jan 2023
10. Kiefer, F., Kwiatkowski, K.: Hybrid ECDHE-SIDH key exchange for TLS. Internet-Draft draft- kiefer-tls-ecdhe-sidh-00, Internet Engineering Task Force (2018). https://datatracker.ietf.org/doc/html/draft-kiefer-tls-ecdhe-sidh-00. Accessed 24 Jan 2023
11. Schanck, J.M., Stebila, D.: A Transport Layer Security (TLS) Extension For Establishing An Additional Shared Secret, 17 April 2017. https://datatracker.ietf.org/doc/html/draft-schanck-tls-additional-keyshare-00. Accessed 24 Jan 2023
12. Hoffman, P.E.: The transition from classical to post-quantum cryptography. Internet-Draft draft-hoffman- c2pq-05, Internet Engineering Task Force, May 2019. https://datatracker.ietf.org/doc/html/draft-hoffman-c2pq-05. Accessed 24 Jan 2023
13. Crockett, E., Paquin, C., Stebila, D.: Prototyping post-quantum and hybrid key exchange and authentication in TLS and SSH. Cryptology ePrint Archive, Paper 2019/858 (2019)
14. Whyte, W., Zhang, Z., Fluhrer, S., Garcia-Morc, O.: Quantum-safe hybrid (QSH) key exchange for Transport Layer Security (TLS) version 1.3. Internet-Draft draft-whyte-qsh-tls13–06, Internet Engineering Task Force (2017). https://datatracker.ietf.org/doc/html/draft-whyte-qsh-tls13-06. Accessed 24 Jan 2023
15. Rescorla, E.: The Transport Layer Security (TLS) Protocol Version 1.3, August 2018. https://www.rfc-editor.org/rfc/rfc8446. Accessed 24 Jan 2023

Author Index

S. Femmam and P. Lorenz (Eds.): ICCNT 2022, LNDECT 205, p. 171, 2024.
https://doi.org/10.1007/978-3-031-59619-3