Enabling Intelligence Inclusiveness in Edge to Cloud Continuum: Challenges and Opportunities

Javier Palomares*, Estefanía Coronado*†, Cristina Cervelló-Pastor‡, and Shuaib Siddiqui*
*i2CAT Foundation, Barcelona, Spain;
Email: {javier.palomares, estefania.coronado, shuaib.siddiqui}@i2cat.net
†High-Performance Networks and Architectures, Universidad de Castilla-La Mancha, Albacete, Spain;
Email: estefania.coronado@uclm.es
‡Department of Network Engineering, Universitat Politècnica de Catalunya, Castelldefels, Barcelona, Spain;
Email: cristina.cervello@upc.edu

***Abstract*—Edge to Cloud Continuum is a concept that integrates cloud computing and cellular networks that has been gaining popularity due to its potential to provide a seamless user experience and address the challenges of managing complex multi-domain networks involving massive IoT devices. Enabling intelligence in the Edge to Cloud Continuum can further enhance its capabilities, offering benefits such as reduced latency, improved scalability, enhanced resource utilization, and increased context awareness. This paper provides insights into the opportunities and challenges of enabling intelligence in Edge to Cloud Continuum, highlighting the potential of this technology. This study presents a comprehensive review of the existing literature on enabling intelligence in Edge to Cloud Continuum, to reach the research questions that will construct the PhD. Various tools and technologies that can be used to integrate intelligence into the Edge to Cloud Continuum system were explored and analyzed. In addition, this study provides a detailed work plan for the upcoming months of the project.**

***Index Terms*—Edge to Cloud Continuum, Multi-domain, Enabling Intelligence, Distributed Systems, 6G.**

I. Introduction and Motivation

The fifth generation of cellular mobile communication systems (5G) is being implemented to offer low-latency and high-bandwidth communication services to mobile devices and industries with different service requirements. To achieve this, 5G is gradually being integrated into mobile operator infrastructures through the use of Software-Defined Networking (SDN) [1], Network Function Virtualization (NFV) [2], and Multi-Access Edge Computing (MEC) technologies [3]. Service-oriented architecture in cloud computing has significantly transformed software construction. The authors of the paper in [4] propose a service-oriented architecture that highlights the main challenges and opportunities in this area. They identify two main challenges, namely the need to maintain high service availability and provide end-to-end security solutions. On the other hand, they also present some opportunities, such as rapid service deployment and service discovery through federated clouds.

Meanwhile, the significance of Artificial Intelligence (AI)'s emergence in 5G and beyond lies in its potential to boost network and service efficiency and effectiveness [5] [6]. This is due to the capabilities of AI technologies such as machine learning and natural language processing, which can offer intelligent automation for network management, minimizing errors and reducing the need for human intervention. The emergence of a new branch of AI, known as Inclusive AI, is attributed to the requirement of AI systems to adhere to fundamental principles like transparency, accountability, accessibility, and fairness. It recognizes and addresses the diverse needs of all individuals and communities, requiring AI developers and users to program and use AI systems ethically and responsibly. In the study [7] the authors illustrate the process of developing a robot that utilizes AI and strives to achieve an inclusive society, while in [8] an inclusive and sustainable AI strategy for Europe based on human rights is presented. Inclusive AI must recognize and address potential biases in the systems to ensure that they do not perpetuate discrimination or exclusion. The work [9] highlights a key concern that the biases and limitations of AI systems may result from limited and incomplete datasets that do not accurately reflect the entire population or from biases inherent in the configuration of AI models by the scientific community. On the other hand, Microsoft has developed a guide [10] that outlines how to facilitate intelligence inclusiveness and how to recognize bias in Artificial Intelligence complex systems.

In this line, the authors of [11] present a thorough survey on the integration of AI and Edge Computing in future networks, in which they define intelligence, highlight various challenges phased and propose their approach for distributing it across network edges. Cloud computing has been a popular paradigm for the past decade, allowing organizations to utilize computing resources on demand, without investing in costly infrastructure. However, as organizations have become more reliant on cloud computing, a new challenge has emerged: how to manage resources across multiple cloud environments seamlessly. The surveys in [12], [13] provide an extensive analysis of the challenges faced by Cloud Computing. To address these challenges, the concept of the Edge to Cloud Continuum has emerged. The authors suggest that the Edge to Cloud Continuum vision seeks to offer several benefits, including decreasing service latency, improving context awareness, enhancing the quality of service, and increasing operational efficiency, among others.

The Edge to Cloud Continuum is a new paradigm that is defined by the authors of [14] as: *"a novel abstraction layer to express a continuous range of capacities between cloud and edge/fog"*. Multi-domain systems are the future of telecommunications because they offer several advantages over traditional single-domain systems. A seamless integration of multiple environments enables the seamless transfer of workloads and data across them. The Edge to Cloud Continuum can span public, private, and hybrid cloud environments, providing organizations with the flexibility to utilize the cloud services that best suit their needs. The surveys in [12], [13] provide an extensive analysis of the challenges faced by Cloud Computing. The Edge to Cloud Continuum vision, according to the authors, aims to provide various advantages, beyond those already stated. These include enhancing user experience, speeding up real-time processing, and allowing sensitive data to be stored near their origin. Moreover, it provides superior assistance for mobility and the utilization of mobile communication. The challenges described include but are not limited to heterogeneity in network components, ensuring a high-quality service, offloading computation, and addressing privacy and security concerns. Another big challenge presented is the creation of a unified framework that can automatically deploy and orchestrate a complete network to support expected service quality. The impact of enabling automated multi-domain orchestration of network services is analyzed in some works, as shown in [15]. This paper explores how to manage a service federation functionality in a user-transparent manner. The new modules proposed are necessary for coordination with peering domains, enabling descriptor orchestration, and enabling the exchange of pertinent resource information across multiple domains. The most recent paper in the related work about enabling intelligence in the Edge to Cloud Continuum [16] presents a comprehensive system architecture of a multi-tier orchestration platform for Connected, Cooperative, and Automated Mobility (CCAM), which discusses the benefits and challenges of its integration. The paper highlights that the proposed architecture has the potential to improve scalability, cost-effectiveness, flexibility, and resource utilization. However, there are concerns regarding how to enable and manage inclusive intelligence, from a multi-domain perspective to handle topics such as data privacy, security, regulatory compliance, and ethical implications.

The telecommunications sector is moving towards complex systems that comprise multiple domains and numerous IoT devices. These network resources will be spread out geographically and managed by different orchestrators, some of which may not even belong to the same operators. As a result, the network must be able to meet the needs of both the network itself and its users. Managing such a system is a challenging task. This is why the Ph.D. topic suggests the utilization of intelligent orchestrators to address this problem by collaborating to reach a consensus to agree on the best way to manage the network. Throughout the duration of the Ph.D., the focus will be on introducing a system framework that will allow for the incorporation of intelligence in an Edge to Cloud Continuum system.

This paper describes the approach taken in the first five months of the Ph.D. to identify the research topic. The initial contributions presented in this study include the following:

- A deep analysis of related work relevant to the topic.
- The evaluation and comparison of existing tools, devices, and technologies that can enable intelligence within the Edge to Cloud Continuum paradigm.
- The presentation of the work plan for the next years.

The remainder of this paper is organized as follows: Section II focuses on the research proposal, including the methodology used to formulate the research questions, the preliminary contributions, and the work plan. Conclusions and future lines of work are summarized in Section III.

II. Research Proposal

This PhD involves three main phases: the exploratory phase, the implementation phase, and the writing and dissemination phase. During the initial five months of the PhD, the primary focus is on conducting a thorough review of the relevant literature in order to identify areas where current knowledge is lacking. This will allow us to formulate research questions, identify challenges and opportunities, and develop scenarios that can benefit from and build on these identified gaps in the existing knowledge.

A. Identifying Key Research Questions

Based on the analysis of the pertinent literature, the research problems that will be investigated are the following:

- What are the benefits and challenges that arise from enabling inclusive intelligence in Edge to Cloud Continuum?
- How to enable Intelligence in a federation of orchestrators on the Highly Heterogeneous Edge-to-Edge to Cloud Continuum?
- How to efficiently deploy complex AI workflows on heterogeneous and distributed infrastructures to improve selected features?

B. Methodology

The procedure used to formulate the research questions was as follows: (i) selection of relevant surveys and papers for the Edge to Cloud Continuum topic published between 2020 and the present day; (ii) the selected works were classified based on the opportunities, open challenges, and use cases highlighted by the authors; (iii) conducted a comprehensive review of state-of-the-art tools and technologies to map the ecosystem according to the features they provide and the requirements from the selected works; (iv) identified scenarios that could benefit from the incorporation of AI into the Edge to Cloud Continuum; and (v) start designing an abstraction of the architecture and the necessary modules required for the Edge to Cloud Continuum scenario.

C. Preliminary Contributions

This section presents the preliminary contributions of the research, which include a set of opportunities and challenges that arise from the inclusive AI integration in the Edge to Cloud Continuum, a set of scenarios identified as potential beneficiaries of intelligence in the Edge to Cloud Continuum, as well as an analysis of relevant existing technologies and tools that could help start enabling this intelligence.

1) Main challenges and opportunities: Existing literature indicates that incorporating intelligence into the Edge to Cloud Continuum has the potential to: (i) improve the performance and accelerate data processing by optimizing the use of network resources; (ii) enhance reliability by using intelligent decision-making mechanisms to reduce network downtime; (iii) improve scalability by adapting to changing network demands as needed; and (iv) reduce operator costs by optimizing resource usage, reducing energy consumption and minimizing equipment maintenance.

However, to identify the areas where the Ph.D. can contribute, the emphasis will be placed on the difficulties encountered to detect the gaps that need to be addressed:

- Lack of variety of trustful inclusive and unbiased data sets: A significant concern is that the imperfections and prejudices of artificial intelligence systems may arise from inadequate and incomplete data sets that don't faithfully represent the entire population, or from the biases that are built into the design of AI models by the scientific community.
- Lack of a unified Edge to Cloud Continuum framework: The main challenge is the creation of a unified framework that can automatically deploy and orchestrate a complete complex multi-domain network that supports the assurance of high-quality services, the offloading of computation, addressing privacy and security concerns and supports expected service quality, among others.

2) Identified Scenarios: Through an examination of the current literature, some scenarios that could potentially benefit from the incorporation of intelligence into the Edge to Cloud Continuum have been identified:

- The healthcare sector: this could provide access to real-time data analysis and insights. For instance, in a pandemic situation, where healthcare worldwide organizations generate vast amounts of data from various sources, including Electronic Health Records (EHRs), medical imaging, and patient monitoring devices. Hospitals could use the shared generated knowledge to derive meaningful insights for better decision-making, patient care, and further research.
- The transportation sector: enabling intelligence in the Edge to Cloud Continuum could improve traffic management, logistics, and safety. For example, a smart traffic management system can use real-time data to optimize traffic flow and prevent accidents. The automotive industry is a promising consumer due to the high mobility and service demand with stringent QoS requirements.

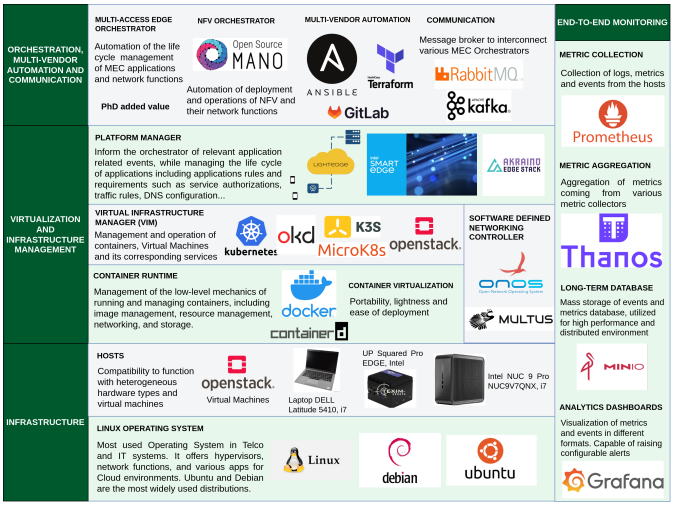

Fig. 1. Canvas of the tools and technologies analysed

- The military sector: this could improve the outcome of armed conflicts, by improving the coordination and communication among different armies of allies participating in the same conflict. Typically, each country sets up its own private ad-hoc network for communication at their base. However, managing and sharing the knowledge obtained from the alliance could improve coordination and ultimately lead to better outcomes in the situation.

3) Tools and Technologies Analysis: As was commented previously, the Edge to Cloud Continuum paradigm integrates multiple technologies to provide efficient cloud services across different domains. However, since the development of 5G Core, NFV, and MEC technologies are being handled by different standardization bodies, the deployment, integration, and interaction of these solutions to support the desired features and end-to-end performance of the 5G ecosystem is not well-coordinated. To address this, a comprehensive review of existing tools, devices, and technologies was conducted to map the ecosystem according to their features. This would aid in identifying the additional modules that this Ph.D. can develop to meet the stakeholders' requirements for integrating inclusive AI in the Cloud Continuum paradigm.

Fig. 1 illustrates the aforementioned review, which adds value by identifying where intelligence should be located to enhance network management. Two types of intelligence are distinguished: intelligence orchestration and orchestration of intelligence. Intelligence orchestration involves managing the lifecycle of an application in the most optimal and efficient manner. In contrast, the orchestration of intelligence involves orchestrating intelligent modules or functions, without necessarily providing intelligent actions themselves. The canvas is structured into four primary sections, which include: (i) the orchestration layer, (ii) the management and virtualization layer, (iii) the underlying infrastructure, and (iv) the end-to-end monitoring tools.

D. Work Plan

This section provides an overview of the stages of the work plan that will direct the research activities for this Ph.D.:

- *Phase 1*. Further research will be conducted on papers that are aligned with or related to the main research questions and objectives. Some of the topics to be investigated include: (i) facilitating intelligence across multi-domain orchestration; (ii) examining how prior research proposes to incorporate end-to-end inclusive AI across a multi-domain platform; and (iii) identifying the primary challenges in enabling end-to-end intelligence. The duration of this phase is until half of the first year.
- *Phase 2*. After identifying gaps in the literature, this phase will focus on determining and designing the modules that could integrate inclusive AI into the system architecture. Defining at the same time how efficiently deploy complex AI workflows to improve selected features. As the topic of [16], which will be published on April 4th, 2023, is closely aligned with the research objectives of the Ph.D. and given the benefits highlighted by the authors, is going to be considered to be used as a starting point. The duration of this phase is until half of the second year.
- *Phase 3*. During this phase, the degree of compatibility and alignment between the proposed system architecture and established standardization research groups such as Experiential Networked Intelligence (ENI) and Zero touch network & Service Management (ZSM) will be evaluated. This phase will continue until the end of the second year.
- *Phase 4*. During the final phase, a thorough examination and assessment of the proposed intelligent framework's feasibility and performance will be conducted across different edge and cloud domains.

III. Conclusions

Given that Edge to Cloud Continuum has been gaining popularity due to its potential to provide a seamless user experience and address the challenges of managing complex multi-domain networks, an in-depth analysis of the related work was conducted to reach the research topics that will construct this Ph.D. This paper presents a comprehensive research plan on this topic, together with the tasks that were conducted within the first five months of the Ph.D.

This research included a comparison of the challenges and opportunities that different authors have on the subject of the inclusion of intelligence in the Edge to Cloud Continuum paradigm. This study has formulated the following research questions: (i) what advantages and obstacles emerge from implementing comprehensive intelligence in Edge to Cloud Continuum?; (ii) how can Intelligence be enabled in a union of orchestrators on the Highly Heterogeneous Edge-to-Edge to Cloud Continuum?; and (iii) how can complicated AI workflows be deployed effectively on diverse and distributed infrastructures to enhance specific features?

From the aforementioned research questions, this paper presents the initial findings in the form of an overview of key challenges and opportunities, an identification of potential use cases that can leverage this work, and an evaluation of various tools and technologies that can be employed for implementing the proposed system architecture. In addition to the main goals, a work plan for the rest of the Ph.D.'s timeline is proposed.

Acknowledgment

This work has received funding from the EU's H2020 projects Zero-SWARM (101057083) and AI@EDGE (101015922). The authors would also like to acknowledge CERCA Programme / Generalitat de Catalunya for sponsoring this work. Furthermore, this work was also supported by the EU "NextGenerationEU/PRTR", MCIN, and AEI (Spain) under project IJC2020-043058-I.

References

[1] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," *Proc. of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[2] ETSI, "Network Functions Virtualisation (NFV); Architectural Framework," December 2014, ETSI GS NFV 002 V1.2.1.

[3] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[4] Y. Wei and M. B. Blake, "Service-Oriented Computing and Cloud Computing: Challenges and Opportunities," *IEEE Internet Computing*, vol. 14, no. 6, pp. 72–75, 2010.

[5] M. E. Morocho Cayamcela and W. Lim, "Artificial Intelligence in 5G Technology: A Survey," in *Proc. of IEEE ICTC*, 2018, pp. 860–865.

[6] Y. Arjoune and S. Faruque, "Artificial Intelligence for 5G Wireless Systems: Opportunities, Challenges, and Future Research Direction," in *Proc. of IEEE CCWC*, 2020, pp. 1023–1028.

[7] K. Okabe and Y. Hirata, "A Study of Management to Develop Adaptable AI-enabled Robots toward Inclusive Society," in *Proc. of IEEE SICE*, 2022, pp. 1369–1374.

[8] C. Fernández-Aller, A. F. de Velasco, Á. Manjarrés, D. Pastor-Escuredo, S. Pickin, J. S. Criado, and T. Ausín, "An Inclusive and Sustainable Artificial Intelligence Strategy for Europe Based on Human Rights," *IEEE Technology and Society Magazine*, vol. 40, no. 1, pp. 46–54, 2021.

[9] M. M. A. de Graaf, G. Perugia, E. Fosch-Villaronga, A. Lim, F. Broz, E. S. Short, and M. Neerincx, "Inclusive HRI: Equity and Diversity in Design, Application, Methods, and Community," in *Proc. of ACM/IEEE HRI*, 2022, pp. 1247–1249.

[10] Microsoft, "In Pursuit of Inclusive AI," https://www.microsoft.com/design/assets/inclusive/InclusiveDesign_InclusiveAI.pdf, Accessed on 26.03.2023.

[11] P. Ramachandran, S. Ranganath, M. Bhandaru, and S. Tibrewala, "A Survey of AI Enabled Edge Computing for Future Networks," in *Proc. of IEEE 5GWF*, 2021, pp. 459–463.

[12] Z. Mahmood, *Fog Computing: Concepts, Frameworks and Technologies*. Springer Cham, 01 2018.

[13] D. Rosendo, A. Costan, P. Valduriez, and G. Antoniu, "Distributed intelligence on the Edge-to-Cloud Continuum: A systematic literature review," *Journal of Parallel and Distributed Computing*, vol. 166, pp. 71–94, 2022.

[14] S. Moreschini, F. Pecorelli, X. Li, S. Naz, D. Hästbacka, and D. Taibi, "Cloud Continuum: The Definition," *IEEE Access*, vol. 10, pp. 131 876–131 886, 2022.

[15] J. Baranda Hortiguela, J. Mangues-Bafalluy, R. Martinez, L. Vettori, K. Antevski, C. J. Bernardos, and X. Li, "Realizing the Network Service Federation Vision: Enabling Automated Multidomain Orchestration of Network Services," *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 48–57, 2020.

[16] N. Slamnik-Kriještorac, G. M. Yilma, M. Liebsch, F. Z. Yousaf, and J. M. Marquez-Barja, "Collaborative Orchestration of Multi-Domain Edges From a Connected, Cooperative and Automated Mobility (CCAM) Perspective," *IEEE Transactions on Mobile Computing*, vol. 22, no. 4, pp. 2001–2020, 2023.