MODERN SYSTEMS 2025

International Conference of Modern Systems Engineering Solutions

ISBN: 978-1-68558-316-3

October 26th - 30th, 2025

Barcelona, Spain

MODERN SYSTEMS 2025 Editors

Alan Martin Redmond, Sophia Antiplois (CSTB), France

MODERN SYSTEMS 2025

Forward

The International Conference of Modern Systems Engineering Solutions (MODERN SYSTEMS 2025) continues a series of events focusing on systems development considering the variety of combination between requirements, technologies, and the application domains. The conference was held on October 26-30, 2025 in Barcelona, Spain.

We are witnessing a paradigm shift in systems engineering approaches caused by several facets of society and technology evolution. On one side, the mobility, the increase in processing power and the large storage capacity created the capacity needed to deliver services to everybody, everywhere, anytime. On the other side, new computation approaches, data gathering, and storage combined with advances in intelligence-based learning and decision-making, allowed a new perspective for systems engineering.

The advanced pace of technological achievements is supported by Cloud/Edge/Fog-based computing, High Performance Computing (HPC), Internet of Things (IoT), Big Data, Deep Learning, Machine Learning, along with 5G/6G communications (integration of terrestrial/special systems) and mobility. As such, deployment, operation and technologies, integration, maintenance became a cornerstone for developing systems complying with functional and non-functional requirements.

We take this opportunity to thank all the members of the MODERN SYSTEMS 2025 Technical Program Committee as well as the numerous reviewers. The creation of such a broad and high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to the MODERN SYSTEMS 2025. We truly believe that, thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the MODERN SYSTEMS 2025 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope the MODERN SYSTEMS 2025 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress with respect to modern systems. We also hope that Barcelona provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city

MODERN SYSTEMS 2025 Chairs

MODERN SYSTEMS Steering Committee

Alan Martin Redmond, Freelancer, France

Mo Mansouri, Stevens Institute of Technology, USA / University of South-Eastern Norway, Norway

Evangelos Manthos, Aristotle University of Thessaloniki, Greece

Hany Moustapha, École de Technologie Supérieure, Canada

Sylvain Cadic, AXONE – Paris, France

Adriana Alexandru, National Institute for Research & Development in Informatics - ICI Bucharest,

Romania
Oliver Michler, Technische Universität Dresden, Germany

MODERN SYSTEMS 2025 Liaison for Certification on Systems
Paul Martin, SE – Scholar, USA

MODERN SYSTEMS 2025 Publicity Chair
Lorena Parra Boronat, Universidad Politecnica de Madrid, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

MODERN SYSTEMS 2025

Committee

MODERN SYSTEMS 2025 Steering Committee

Alan Martin Redmond, Freelancer, France
Mo Mansouri, Stevens Institute of Technology, USA / University of South-Eastern Norway, Norway
Evangelos Manthos, Aristotle University of Thessaloniki, Greece
Hany Moustapha, École de Technologie Supérieure, Canada
Sylvain Cadic, AXONE – Paris, France
Adriana Alexandru, National Institute for Research & Development in Informatics - ICI Bucharest, Romania
Oliver Michler, Technische Universität Dresden, Germany

MODERN SYSTEMS 2025 Liaison for Certification on Systems

Paul Martin, SE – Scholar, USA

MODERN SYSTEMS 2025 Publicity Chair

Lorena Parra Boronat, Universidad Politecnica de Madrid, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

MODERN SYSTEMS 2025 Technical Program Committee

Mohd Helmy Abd Wahab, Universiti Tun Hussein Onn, Malaysia
George H. Abdou, New Jersey Institute of Technology, USA
Mohd Ashraf Ahmad, Universiti Malaysia Pahang, Malaysia
Abdel Ghani Aissaoui, University of Tahri Mohamed of Bechar, Algeria
Abdullah Al-Alaj, Virginia Wesleyan University, USA
Basheer Mohamad Al-Maqaleh, Thamar University, Yemen
Adriana Alexandru, National Institute for Research & Development in Informatics - ICI Bucharest, Romania
Hamid Ali Abed AL-Asadi,Iraq University College, Iraq
Abdullah Al-Mamun, University of Nevada, Reno, USA
Siddeeq Yousif Ameen, Duhok Polytechnic University, Iraq
Federico Aromolo, Scuola Superiore Sant'Anna, Pisa, Italy
Taha Bennani, University of Tunis el Manar, Tunisia
Valerio Besozzi, University of Pisa, Italy
Sofiane Bououden, University Abbes Laghrour Khenchela, Algeria
Antonio Brogi, University of Pisa, Italy
Sylvain Cadic, AXONE - Paris, France
Xiaoqiang Cai, The Chinese University of Hong Kong, Shenzhen, China
Luigi Maxmilian Caligiuri, Foundation of Physics Research Center (FoPRC), Italy

Copyright Information

Table of Contents

A Framework for Adaptability, re-use and Deconstruction of Buildings, Aligned with the Principles of Circular Economy

Alan Martin Redmond
Auto-Entrepreneur
Montigny-le-Bretonneux, France
e-mail: a.redmond@numeriquetech.eu

Abstract— **This article presents a framework that will bring a new perspective to circular economy products and processes within the construction industry. The framework will utilize Federated Enterprise Architecture approach that is traditionally used in the aerospace, automotive, and oil and gas industries, for built environment needs. The framework will address 5 main objectives : i) to analyze and develop a decentralized Federated Framework for construction and renovation processes; ii) to optimize the re-usability and recycling of building materials and components; iii) to investigate the validation process for framework solutions across large-scale pilots in diverse contexts ; iv) to promote and implement innovative tools for stakeholder collaboration and green finance integration; and v) to provide guidance for policy, standardization, and stakeholder adoption. This article acts as prerequisite to prepare the construction industry for transition to Federated Enterprise Architecture practices.**

Keywords-Federated framework; interoperability; adaptability; re-use and deconstruction of buildings, life cycle assessment; construction and renovation; circular economy; pilot studies

I. Introduction

It is estimated that half of the total CO2 emissions of a building arise during the construction process and the production of the material components. Furthermore to reduce the construction industry's contribution to global warming, its immediate decarbonisation is necessary and cannot only depend on incremental CO2 savings over the life cycle of a building but must include the planning and construction process. There is growing acknowledgement for circular construction as a means to develop, use, and re-use buildings, sites and infrastructure without unnecessarily exploiting natural ressources, polluting the environment and damaging ecosystems [1]. Moreover, The European Environment Agency recognizes that "almost 75% of the building stock is currently energy inefficient and more than 85% of today's buildings are likely to still be in use in 2050. Energy renovation of buildings is ongoing but at a very slow rate"[2]. In addition, 'Ecochain' has identified that in many industries, the supply chain accounts for more than 80% of the environmental impact and in addition the supply chain managers that source from different suppliers contribute to a massive impact on their product footprint [3]. In 2018, 'The European Commission' led an intiative to focus on supporting regions and EU countries to develop national bioeconomy strategies that will enhance knowledge on biodiversity and ecosystems, monitor progress towards a sustainable bioeconomy, promote good practices to operate the bioeconomy and enhance the benefits of biodiversity. In order to unlock investments and markets, the commission mobilised stakeholders in developing sustainable biobased solutions while launching a €100 million circular bioeconomy thematic investment platform [4]. According to the World Economic Forum 'Bioeconomy is emerging as a transformative force for sustainable development, leveraging biological resources and innovative technologies to address global environmental challenges [5]. Building upon these challenges and initiatives there is a growing need for solutions that extend the service life of buildings, support material reuse and recycling, and improve stakeholder collaboration through shared data and digital tools. This article outlines the development of a Federated Enterprise Architechure framework for developing solutions, methods and processes to meet the outlined objectives and align with the European Comission's Built4People Partnership [6] contributing to : (1) Increased adaptability of buildings, (2) Reduction of Waste, (3) Support for local and regional economic development, and (4) Policy evolution. The article has three main sections covering the design of the framework (objectives etc.) in Section II, the methodlogy in Section III and Section IV highlights the potential demonstrations. The acknowledgement and conclusions close the article.

II. Designing A Framework

The main goal of the framework is to deliver a comprehensive and sustainable framework for circular construction and renovation, ensuring adaptability, reuse, and deconstruction of building components while minimizing environmental impact and maximizing stakeholder value. The outcome is an innovative approach that integrates digital technologies, advanced processes, and novel materials into an interoperable and decentralized Federated Framework for (de)construction and renovation. This framework shifts away from monolithic, proprietary systems and creates an interconnected ecosystem where tools, platforms, and processes maintain autonomy while contributing to a unified, efficient workflow. The Federated Framework is designed to support decision-making across all stages of a building's life cycle from design and construction to reuse and deconstruction.

A. Objectives and Ambition

The overall objective is to deliver a more sustainable framework for circular construction and renovation, ensuring adaptability, reuse, and deconstruction of building components while minimizing environmental impact and maximizing stakeholder value. The project proposal leverages cutting-edge digital technologies, advanced methodologies, and innovative materials to align with the European Union's goals for a sustainable, people-centric built environment. The framework adopts a holistic and decentralized approach, integrating federated architectures, life-cycle-based methods, and participatory design processes to address the challenges of resource efficiency, carbon reduction, and material circularity. This vision is realized through a suite of interoperable tools, validated across diverse geographical and climatic contexts. The Strategic Objectives (SO) that will enable the achievement of the framework are described below:

a) To analyze and develop a decentralized Federated Framework for construction and renovation processes (objective 1): This objective focuses on creating an interoperable architecture that integrates advanced digital solutions, such as graph technology, Digital Twins (DTs), and blockchain-based systems. The federated framework facilitates interoperability and information sharing between semi-autonomous de-centrally organized Line of Businesses (LOBs), in particular reference to AI & Agents.

b) To optimize the re-usability and recycling of building materials and component (objective 2): The framework will enhance the traceability and performance of sustainable materials, prioritizing bio-based, CO2-storing, and modular solutions such as, the use of waste wood, bio-based insulation, Supervisory Control and Data Acquisition (SCADA) controls, Heating, Ventilation, and Air Conditioning (HVAC) for energy management systems, and product inventories for supply chains. Through advanced Life Cycle Analysis (LCA) and predictive maintenance tools, the framework will establish best practices for disassembly and reuse, enabling a shift towards a circular construction model.

c) To investigate the validation process of the Framework solutions across large-scale pilots in diverse context (objective 3): Two pilot sites across Portugal and Slovenia, and one virtual pilot study (referring to test simulations for example interoperability issues of shared models) between Romania and Denmark, will perform the adaptability and performance analysis of the Federated Framework under real-life and close to real life conditions. These pilots will focus on delivering test solutions in residential and non-residential settings, enhancing adaptability, resource efficiency, and scalability.

d) To promote and implement innovative tools for stakeholder collaboration and green finance integration (objective 4): This objective involves creating a virtual living lab and green finance platform to connect investors with sustainable building initiatives. It integrates digital building logbooks and financial models to enhance the scalability and economic viability of circular practices. The outcome is to provide support and knowledge on investment opportunities, such as ‘Growth and Income Fund’ and ‘Feeder Fund’ [7] (one of many smaller investment funds that pool investor money, which is then aggregated under a single centralized fund, allowing for reduced operation and trading costs).

e) To provide guidance for policy, standardization, and stakeholder adoption (objective 5): The framework will deliver actionable recommendations for regulatory bodies, certification authorities, and industry stakeholders to support the standardization and scaling of circular construction practices. Moreover, the framework will engage in 6 key stewardship of activities such as: standard identification and monitoring; collaboration with standards and policy bodies; gap analysis and recommendations development; advisory policy framework working group; workshops and stakeholder engagement; and policy briefs and contributions, that will increase awareness on best practices for design, adaptability, reuse and deconstruction.

Table 1 explains the various challenges that will be applied to each objective and its evaluation defined through measurement.

TABLE I. REQUIREMENTS & MEASUREMENTS

Evaluation		
Objectives	***Requirements***	***Measurable Key Performance Indicators (KPIs)***
1: Framework	The Enterprise Architecture incorporates methods and processes that focus on operational analysis – what the stakeholders need to accomplish and system analysis – what the system has to accomplish for the stakeholders.	KPIs include increased reuse rates of construction materials by 30% and a 25% reduction in embodied carbon across pilot projects.
2: Reusability & Recycling	The framework will demonstrate how to optimize Building Management Systems (BMS) by using Agentic AI to minimise GHG emissions during the full building operational life cycle that are essential for maintaining high levels of user comfort and well-being, which directly translate into high User QoE (Quality of Experience) KPIs.	The challenges that exist particularly for building systems is that electronic systems or products such as consumer products become obsolete long before the device wears out or fails and are simply discarded and sent to landfill. KPIs include a 40% increase in material recovery rates and a 20% cost reduction in renovation projects.
3: Validation	After the demonstrations the framework will create Impact Assessment Methodology (IAM) that will score the demonstrations based on the solutions validation requirements for both embodied and operational	KPIs include successful deployment of solutions in a shopping mall in Portugal with a total area of ~200,000 m2 that hosts over 150 stores on 5 floors and a small site 40 m² of built space targeting quantified reductions in

Evaluation		
Objectives	***Requirements***	***Measurable Key Performance Indicators (KPIs)***
	carbon.	embodied carbon by 40% and operational carbon by 50 – 75% respectively.
4: Green Finance	The overall objective for framework is to provide a one-stop shop [8] of information and support for SMEs in the local value chain but to also encourage larger companies to invest in disruptive technologies.	KPIs include engaging 50+ SMEs in local value chains and connecting them with funds/Venture Capitalists (VCs) to secure green investments.
5: Policy Guidance	The framework will contribute to the activities of the Built4People partners and to the Built4People network on innovation clusters through the achievements of the demonstration.	KPIs include the publication of 5 policy briefs and engagement with over 300 stakeholders.

B. The scope of the Framework

The Federated Framework consists of 16 interconnected modules that encompass key enabling tools, processes, and methods. These include advanced IT solutions such as graph technologies, digital twins, distributed ledger systems (blockchains), and Common Data Environments (CDEs), as well as methodologies like LCA, Model-Based Systems Engineering (MBSE), and business process mapping. Each module addresses specific aspects of circular construction, such as material traceability, predictive maintenance, and user-centric design.

TABLE II. FRAMEWORK: INNOVATIVE TOOLS, PRODUCTS & TECHNIQUES, PROCESS & METHODS

Framework		
Innovative Tools (INNT)	***Products and Techniques (P&T)***	***Process and Methods (P&M)***
INNT1) AI, IoT and Agents for BMS	P&T1) Digital Building Logbooks including DPP, BRP and MP	P&M1) Buinsess Process Mapping
INNT2) Graph Technology	P&T2) Security Transactions including Data Encryption, and Payments Certs (eIDAS) etc.	P&M2) Life Cycle Analysis including SCBA
INNT3) Large Language Models including LangChain - GenAI, and Chat models	P&T3) Open Source: Open LCA and Open API	P&M3) Federated Architectures including Linked Data CDE
INNT)4 Digital Twins, BIM and GIS Platform	P&T4) Semantic Analysis Techniques, RAG and Indexes	P&M4) MBSE
INNT)5 Distributed Ledger Technologies / Blockchains		P&M5) ETL Data Integration
INNT6) Preventative and Predictive Maintenance - RAM		P&M6) Real-Time Linked Data Space

a. Legends (Digital Product Passport – DPP; Building Renovation Passport – BRP; Material Passport – MP)

The semi-autonomous federated systems architecture will support State of the Art LOBs that uniquely provide solutions and by integrating them with products, techniques, processes and methods they will pioneer co-creation (see table 2). Furthermore, the pilot studies will transition the Use Cases (see table 4), which are bundle of selected technology advancements, products, techniques and methods & processes from Table 2 at the demonstrations initial conceptual stage to an advanced operational development stage based on constant evolution and learning. Moreover, using the Federated Architecture will present simulations of the decentralized platforms' abilities to connect with each other in a "common data-space" of open collaboration pooling of information. At pilot level, the framework will encapsulate three very different pilot studies, for which two of them are very much real-life scenarios representing residential (Slovenia) and non-residential projects (Portugal). The other Pilot study comprises of two countries (Denmark and Romania) working on virtual models to test their CDE platforms and interoperability. The impact of these Pilot studies will be evaluated at city council level in the Ukraine to provide added value to the project circular economy results.

III. METHODOLOGY

The methdology section comprises of four sections addressing an exploratory research stage; i) the framework vision, ii) the process, iii) the phases, and iv) the use cases.

A. The frameworks vision

The framework's vision is the development of products/materials/services including those that contribute to disassembled and reused, and CO2-storing materials etc. and also the cyphering of materials via Graph Technology. Moreover Graph Technology (ISO/IEC 39075:2024) Information Technology, Database Languages and GQL defines data structures [9]. The framework will provide structured and unstructured data from existing Relational Database and web services such as ECO Building Materials Suppliers Catalogs, deconstruction – reuse warehouse of materials, certified environmental product declaration catalogs and product environmental profiles in compliance with ISO 14025 standard [10], community engagement platforms, European circular economy stakeholder platforms, and environmental monitoring and IoT platforms.

The integrating advances of bio-based materials manufacturing technology for example Ceramics and Glass and the use of digital solutions (AI, property and Knowledge Graphs (KG), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), LangChain) with economy principles i.e. targeting VCs, Angel funds to provide investment opportunities to enterprises that develop and reuse, deconstruction materials in a life-cycle optimization and circular economy perspective, will offer solutions that not only mitigate environmental impacts, but also drive economic growth and societal well-being.

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-316-3

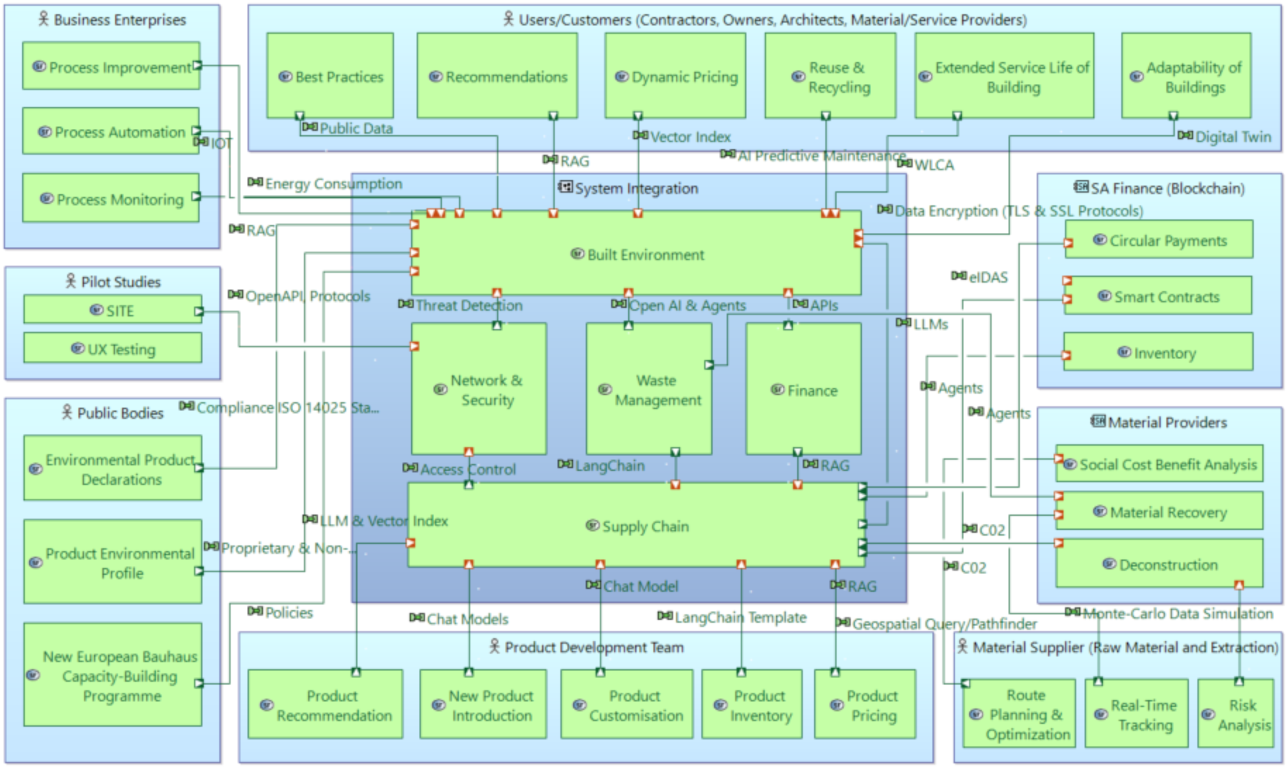

Figure 1. The Framework MBSE for Federated Enterprise Architecture.

B. *The Process*

The goal of the framework is to bring a new perspective to circular economy products and processes. However, fostering strategies and roadmaps is not enough and the framework will utilize Federated Enterprise Architecture approach that is traditionally used in the aerospace, automotive, and oil gas industries, for built environment needs.

Figure 1 illustrates the use of MBSE, which is a process offered in the framework to define the solution design. The pilot studies will transition from the initial conceptual stage (pilot study demonstrators original assumptions) to an advanced operational development stage based on constant evolution and learning. In the context of the framework Figure 1 defines the solutions need analysis, it is an integrated systems open architecture identifying components (e.g., material providers), functions (e.g., supply chain), and exchange items (CO2 carbon content, policies, predictive & preventive maintenance, etc.) pooling of information.

In comparison to traditional architectures that highlight connections of how various components such as ontologies are used and interchanged with semantic platforms and where over ambitious proposals identify lots of connected applications, the framework project builds on a system design approach. It will promote iteration to accommodate the pilot studies' ever changing needs.

In addition, new knowledge through the project's evolving research techniques can be implemented and tested to provide better practices and recursion to reach a level to finalize market readiness. Such analysis captures opportunities for disruptive and innovative solutions, processes and methods.

Furthermore, the Market Key Results Investment Platform, New Business Models, and Policy Recommendations will build upon private investments initiatives such as SMEs involvement in the project. Their presence shall extend contribution measures to local and regional value chain approaches, in order to increase innovation buy-in from users. This initiative has driven the proposal to focus on leveraging green finance investment and creating Scale & De-Risk Accelerators for bio-based materials and products (financial simulations). The impact is to not only provide sustainability, standardization, and governance but to also deliver a shared digital connected infrastructure for supporting decision makers with real monetary opportunities 'Bankability' reflecting and helping the circular economy.

C. *The Phases*

To successfully demonstrate this concept and enable a posterior replication of the results obtained within such a project, a clear and well-defined methodology has been defined, which consist in the following 5 phases: Operational Analysis – Define Stakeholder Needs and

 ISBN: 978-1-68558-316-3

Environment; Solution Analysis – formalize solutions requirements; logical architecture – develop solutions logical architecture; physical architecture – develop solutions physical architecture; and End Product Breakdown Structure (EPBS) – formalize solutions component requirements.

a) Phase 1: Identify specific needs for reuse & recycling of building elements and products (Operational Analysis) - this phase involves establishing the project's base-line while capturing and consolidating operational needs from the stakeholders for correct development and later deployment of the solutions. Moreover, it can be related to "Concept operations" where assumptions on the characteristics of the pilot studies have been made.

b) Phase 2: Model and support the circular economies supply to extend service life of buildings (Solution Analysis) - the data from P1 will contribute to P2 development of a Federated Architecture framework enabling the adoption/adaption of the existing methods and processes by identifying the boundary of the solutions, processes and consolidate the requirements. This phase can be referred to as the "operational concept" stage. The progress is more precise for individual pilots such as defining the business case, what the solutions must accomplish for the users, while also modeling functional data flows and dynamic behavior for integration approaches aligned with federated CDE and innovative tools. Whereas the previous phase investigates the overall intent of the pilot studies, OpsCon identifies what they will do (ISO/IEC/IEEE29148) [11].

c) Phase 3: Design and implement inclusive, accessible, sustainable, resilience, performant (energy, cost, etc) solutions for the built environment (Logical Architecture) - P3 represents a "white box" where models, methods are designed to be transparent, allowing the pilot study demonstrators to understand the internal workings of solutions. This phase built upon the information of P1 and the process and architecture of P2, will provide visibility into the pilot studies decision-making process, making it easier to identify the key features and rules contributing to their development. Moreover, P3 will define how the individual architectures will work to fulfill expectations such as the successful deployment of solutions to prove their capability.

d) Phase 4: Configure and integrate solutions in local and regional value chains (Physical Architecture) - P4 concentrates on using the knowledge of how the solutions will be developed to actually testing them in the 3 pilot studies. These 3 pilot studies consist of collecting real data from the field and deploying the practices to be replicated solutions, products, techniques, methods and processes based on platform assessment to the Ukraine (similar to the DAREED platform [12]). All specifications of interfaces deployment configurations, trade-off analysis of the integrated solutions is tested and evaluated.

e) Phase 5: Increase awareness on best practices for design for adaptability, reuse and deconstruction – the managing of the industrial criteria and integration strategy based on the impact outcomes of the pilot studies are assessed for market segment and commercial success. In addition, the outreach of solutions are analysed and considered to propose business models and recommendations to legislators.

D. Use Cases

The defined set of innovative solutions developed during the framework will be tested via 8 use cases addressing different target groups. Use cases will represent an already proven concept for the viability of combining these solutions, products & techniques, and process & methods towards achieving increased availability, access, and management of lifecycle data in the built environment. These use cases are one of the cornerstones of the project, as its conception fosters replication, bringing framework to a larger public and set of users.

TABLE III. SET OF INNOVATIVE SOLUTIONS DEVELOPED & TESTED

Use Cases		Innovation Tools						Products and Techniques				Process and Methods						Markets		
		1	2	3	4	5	6	P&T1	P&T2	P&T3	P&T4	P&M1	P&M2	P&M3	P&M4	P&M5	P&M6	M1	M2	M3
UC1	Circular Construction and Reuse Framework		X		X						X	X	X						X	X
UC2	Digital Platform		X		X				X					X		X	X			
UC3	Real-Time Linked Data-Space		X		X			X									X			
UC4	RAM Knowledge Platform	X					X	X					X							
UC5	Cybersecurity and Supply Chain Transparency					X			X											
UC6	LCA Models for Renovation Planning and Design		X	X	X		X	X	X	X			X	X	X		X			
UC7	Operational Carbon Management Systems	X	X		X		X	X		X				X	X		X			
UC8	Accelerator for Green Finance Investments		X		X	X		X	X		X	X						X	X	X

Table 3 maps and aligns with Table 2 Framework: Innovative Tools, Products and Techniques, Process and Methods, including Markets where the 8 Use Cases provide a combination of the frameworks modules for selection during the pilot studies, i.e., several modules can be applied to each UC.

TABLE IV. EIGHT USE CASES

A Combination of Framework Modules	
Use Cases	***Proposed Solutions, Products & Techniques, and Process & Methods***
UC1	*Circular Construction and Reuse Framework*: demonstrates different circular and sustainable building solutions to make building and infrastructure better, e.g., following circular and sustainability requirements such as design for adaptability, re-use, and durability. Using Al and graph technology (tracking) UC1 will help source materials locally that travel shorter distances, consuming less fuels and fewer carbon emissions. The preference will be to source reused materials or alternatively materials that are bio-based. UC1 will advise supporting local businesses through local sourcing which, can lead to economic growth and job creation, and social benefit. In addition, the sourcing of financing mechanisms, green insurance and micro-credit for sustainable development via Carbon Platform will provide enterprises with incentive to adapt to greener solutions connected to green finance thus providing a win-win situation.
UC2	*Digital Platform of Solutions*: aligned with CDE to improve collaboration, planning, management and automation within construction projects. The concept of Federated CDE is connected to the project's overall methodology of Federated Architecture Approach. The platform functionality for the framework requires solutions and technologies for

A Combination of Framework Modules	
Use Cases	***Proposed Solutions, Products & Techniques, and Process & Methods***
	development to include: i) folders with documentation, ii) a platform providing a list of materials and tasks, while also mapping to standards where all materials and tasks are awaiting for implementation in the early phases such as selected bills of materials and structures by assemblies. In addition, functions that can provide business intelligence reporting of sales and inventory while also connecting the quantified data with processes that have been identified as a bottleneck.
UC3	*Real-Time Linked Data-Spac***e** [13] [14]: use case focuses on integrating supply chain monitoring data analysis, such as data from sensors and IoT devices in existing or similar buildings, into an interoperable digital twin knowledge graph. This integration supports real-time visualization of embodied carbon, indoor climate metrics, and the adaptability of building systems. The focus of this approach is on optimizing supply chain processes to align with circular economy objectives. Knowledge graphs further enhance this system by linking data on materials, supply chains, environmental performance, and stakeholder roles, thereby enabling informed decision-making. Financial tools, such as agents for circular payments, product pricing, and commercial contracts, ensure that supply chain operations align with circular economy principles.
UC4	*RAM Knowledge Platform* [15]: will demonstrate the potential of preventative and predictive maintenance algorithms and systems to enable calculations on mechanical systems incurred by wear from the moment they are activated. The RAM knowledge platform identifies the useful service life of a system, product, or service by applying real-time monitoring against the preventative and predictive models, extended by proactive, Just-in-Time (JIT) sequence of preventive and corrective maintenance actions and upgrades. The RAM knowledge platform supports information/calculations on system configuration identification elements such as existing, internally developed, reusable components that may consist of Commercial Off-the-Shelf (COTS) products, and Non-Development Items (NDP – hardware and software configuration). Such processes are aligned with early defects detection "Poka-Yoke" [16] and reduce electronic and mechanical materials sent to landfills.
UC5	*Cybersecurity and Supply Chain Transparency*: use case reflects how blockchains will trace the sourcing of CO2-storing materials such as sustainably sourced long-lived bio-based materials and products and innovative lower emission materials/aggregates. In fact each transaction or exchange of information is recorded in a "block", which is then validated by network members before being added to the existing chain. Once validated, information becomes immutable and traceable. This technology offers unique guarantees in terms of security, transparency and traceability of exchanges, without the need for a centralized trusted intermediary. In UC5 Blockchains will create digital "product passports", containing all the information on a product's composition, manufacture and use. Furthermore according to [17] the concept of "decentralized AI" (DeAI) envisions open source, transparent AI through several blockchain technologies. Decentralized storage and distributed computing networks enhance data integrity, while smart contracts ensure transparent model access and tracking.
UC6	*LCA models for Renovation Planning and Design*: use case will demonstrate how tools developed under the framework will empower building stakeholders to streamline processes, enhance efficiency, and drive sustainable transformations. CO2 studies can be directly imported into openLCA [18] and standard LCA repositories, enabling UC6 users to assess flows and processes in their impact evaluations. UC6 will store all relevant data on a platform incorporating Digital Building Logbooks, which include Building Renovation Passports, BIM-based information, DPP, MP and GIS data. This integrated approach facilitates informed decision-making while ensuring compliance with the Corporate Sustainability Reporting Directive (CSRD). To advance beyond standard LCA principles, UC6 will adopt MBSE. The Carbon platform will generate life cycle inventories and evaluate the environmental impacts of material choices through comprehensive life cycle assessments.
UC7	*Operational Carbon Management Systems*: demonstrates the power of AI agents for energy monitoring and optimization including dynamic Pricing. The operational carbon management systems access real-time feeds and data from captured sensors (IoT) related to energy consumption of building elements and products. The introduction of AI agents and graph technology enhances IoT capabilities and across diverse energy systems. The approach applied to smart buildings is reliant on AI agents for controlling lighting and optimizing energy consumption as they are programmed to learn from their environment and improve over time. The framework will analyse patterns of electricity usage and optimize it such as turning off lights in unoccupied rooms or adjust the HVAC systems based on current occupancy, thus providing Occupant well-being. Furthermore, it will facilitate environmental sustainability by tracking energy performance, carbon emissions, and environmental impact.
UC8	*Accelerator for Green Finance Investments*: Green finance appears to be one of the leading technology solutions which, will further promote the increase in regulating environmental impact activities [19]. Therefore, the asset holders, bondholders, and issuers, among others, will have to refocus their efforts to guarantee that green finance is more useful, significant, inclusive, and environmentally protection oriented. UC8 will focus on Getting Buy-In to advance Green Infrastructure – Creative solutions for green infrastructure are only as viable as those who back them.

IV. DEMONSTRATIONS

To test these innovations and use cases, the demonstrations will cover a spectrum of many items included in the EU policy and market trends regarding data management in the built environment. This broad coverage aims to ensure that the proposed solutions offer high replication potential thanks to a demonstration plan. In this sense, the framework UCs will be tested by scenarios of different building typologies, energy grids and data architectures, via the involvement of a living lab (TRL6) and 2 large-scale pilots (TRL7-8) covering a variety of use cases and target users and 1 virtual pilot that will act as 'Development, Test and Evaluation (DT&E)' before the 2 large-pilots regarding 'Operational, Test & Evaluation' (OT&E) [20]. In addition, ISO 31000 Risk Management Plan will also be used to monitor the project progress [21].

A. Virtual Pilot Study: Denmark

The Danish virtual pilot will focus on a fully digital demonstration, leveraging advanced DT technology and BIM tools to address challenges in interoperability, circular construction, and supply chain integrity. This pilot aligns with the proposals objectives by simulating circular economy principles in construction and renovation processes. The virtual demonstrator will integrate open standards (Industry Foundation Classes ISO 16739-1:2024) [22], Information Container for linked Document Delivery under ISO 21597-2:2020 [23], Cybersecurity measures, and lifecycle optimization methodologies to showcase adaptability, reuse, and practices focused on carbon emission reduction. The primary aim of this pilot is to establish effective systems for sourcing materials locally, enabling businesses to develop robust recycling and reuse frameworks. This approach reduces reliance on resource-intensive production processes, minimizes waste generation, and fosters sustainable practices. Local sourcing is encouraged by transparency and decision-making capabilities enabled by the platform. The Carbon platform, equipped with graph technology conforming to ISO/IEC 39075:2024, will enable precise tracking and sourcing of materials, prioritizing reused materials and bio-based alternatives.

B. *Pilot Study Slovenia*

The demonstration planned for Cirkulane (Slovenia), is a prefabricated, residential wooden house, addressing the challenges of Build4People topics, including circularity and sustainability. The demonstrator ‘GORSKO’ aims to focus on:

- Development of ecological modular walls designed for sound and thermal insulation.
- Reduce the carbon footprint and resources by using wood as a main structural material, including the use of waste wood, bio-based (wood, wood fibres, wood chips, straw, clay, sheep wool) insulation, also insulation based on waste textile, and focusing on providing locally produced materials and products.
- Extend the prefabrication processes: In addition to the wooden elements, there is designs to implement a BIM connection to automated production with CNC machines of insulation panels (wood fibre-based).
- Source products inventory and identification to optimize transport and installation.
- Optimisation of preparation and installation process: 3D scanning, e-site, AR/VR use for installation/supervision.
- Developing a smart building (IoT, Digital twin) and setting up smart management and maintenance as a service.

The Expected outcomes planned are related to improve adaptability of building design and solutions for different uses, increased reuse of buildings components and increasing the end-of-life value, extend services life by smart maintenance services, and increase awareness and deploy best services in the demonstrators supply chain.

C. *Pilot Study: Portugal*

The Portuguese demonstration will concentrate on the Mechanical and Electrical (M&E) components that contribute to operation costs. The challenges relate to current SCADA systems that collect and manage data from a building's Command & Control infrastructure, which oversees major energy systems such as lighting, HVAC, and power distribution. These systems operate in an event-based manner, meaning that human operators manually analyses data, respond to alarms, and make real-time decisions based on detected anomalies, incidents, or diverse operational needs. While this approach allows for direct human oversight and control, it also is reactive, and carries low efficiency, leading to potential delays in optimizing energy usage and system performance. In the To-Be scenario, new energy analytics services will be implemented, such as load forecasting, energy sourcing classifications, anomaly detection, and virtual consumption dis-aggregation to generate meaningful data & insights to be fed into the vertically embedded Agentic AI that will orchestrate building management system. The AI-driven system will utilize endogenous building information extracted from the internal active building management systems, as well as exogenous energy grid information incoming from the local energy Transmission System Operator/Distribution System Operator.

V. Conclusion and Future Work

This paper focused on a Federated Enterprise Architecture approach to integrate digital technologies, via advanced processes, with novel materials into a decentralised Framework for (de)construction and renovation. The existing challenges of non-interconnected frameworks associated with monolithic, proprietary system provided the initiative to design an interoperable ecosystem for tools, platforms, and processes that maintained autonomy. The article outlines 5 objectives and their potential requirements and measurable KPIs. The methodology examines in detail the use of MBSE as a process offered in the framework to define the solution design. Furthermore, the methodology describes the evaluation of research methods and their philosophical assumptions in five phases, where each is aligned to MBSE.

The concept of the pilot studies is to facilitate the transition of the selected UCs bundle from the initial conceptual stage to an advanced operational development stage based on constant evolution and learning. The framework’s adaptive process will be measurable based on the outcomes of the demonstrations. Furthermore, the anticipated results of the process will provide a long lasting impact on our understanding of how an efficient design for adaptability, re-use and deconstruction of buildings should be approached, as well as to support EU regulation on those issues. Within the context of the paper’s output it has identified how to increase adaptability of buildings and the reduction of waste via the Federated Enterprise Architecture, while analyzing supporting structures such as leveraging green finance investment for local and regional economic development, and moreover providing sustainability, standardization, and governance towards the circular economy policy evolution.

Acknowledgment

The author would like to take this opportunity to acknowledge the contribution of F6S (Ireland), InoSens (Serbia), VISABEIRA (Portugal), UCN (Denmark), SGG

(Slovenia), & Neo-Eco (France) to the Call: HORIZON-CL5-2024-D4-02 (Efficient, sustainable and inclusive energy use) for which, this article has been based on.

REFERENCES

[1] M. Niemeyer, and N. Kozub, "New European Bauhaus Circular Housing in Ukraine (Project B)," New European Bauhaus, beautiful | sustainable | together, https://www.acash.org.pk/wp-content/uploads/2024/07/221207_NEB_circular_housing_compressed.pdf, July 2024

[2] European Environment Agency, "Building and construction," https://www.eea.europa.eu/en/topics/in-depth/buildings-and-construction?activeTab=07e50b68-8bf2-4641-ba6b-eda1afd544be, modified 16 December 2024

[3] Z. Quist, "Life Cycle Assessment (LCA) – Everything you need to know", Ecochain, https://ecochain.com/blog/life-cycle-assessment-lca-guide/ 18 June 2024.

[4] European Commission: Directorate-General for Research and Innovation, Bioeconomy – The European way to use our natural resources – Action plan 2018, Publications Office, 2018, https://data.europa.eu/doi/10.2777/79401

[5] World Economic Forum, Emerging Technologies, "What is the bioeconomy and how can it drive sustainable development? https://www.weforum.org/stories/2024/07/bioeconomy-sustainable-development/, 12 July 2024

[6] European Construction, Built Environment and Energy Efficient Building Technology Platform, 'Innovative Built Environment', and Europe Regional Network, 'World Green Building Council', "Built4People Partnership – Strategic Research & Innovation Agenda 2021-2027, European Partnership under Horizon Europe, Built4People – SRIA/Roadmap,"https://built4people.eu/wp-content/uploads/2023/10/Built4People-Partnership-SRIA-March-2022.pdf, March 2022.

[7] J. Chen, "Growth and Income Fund Definition, Investment Mix, Examples", Investopedia (part of the Dotdash Meredith publishing family), https://www.investopedia.com/terms/g/growth-income-fund.asp, last updated 05 August 2022.

[8] T. Segal, "One-Stop Shop: Definition, Example, History, Pros and Cons",Investopedia (part of the Dotdash Meredith publishing family), https://www.investopedia.com/terms/o/onestopshop.asp, last updated 26 June 2024.

[9] ISO/IEC 39075:2024, Information Technology – Database Language – GQL, Technical Committee: ISO/IEC JTC 1/SC 32 & ICS: 35.060, https://www.iso.org/standard/76120.html, Published (Edition 1, March 2024).

[10] ISO 14025:2006, Environmental Labels and Declarations – Type III environmental declarations – Principles and procedures, Technical Committee: ISO/TC 207/SC 3 & ICS: 13.020.50, https://www.iso.org/standard/38131.html, Published, July 2006.

[11] ISO/IEC/IEEE 29148:2018, Systems and Software Engineering – Life Cycle Processes – Requirements Engineering, Technical Committee: ISO/IEC JTC 1/SC 7 & ICS: 35.080, https://www.iso.org/standard/72089.html, Published November 2018.

[12] H. Wicaksono, B. Yuce, K. McGlinn, and O. Calli, " "Linked Data for Smart Energy System", 'Buildings and Semantics, Data Models and Web Technologies for the Built Environment,' edited by P. Pauwels and K. McGlinn, Chapter 2. Smart Cities and Buildings,, pp.254 – 256, doi: 10.1201/9781003204381-13, 2023.

[13] M. Page, and G. Cecconi, "European Data Spaces and the Role of data.europa.eu", Publications Office of the European Union; 2023, available from: doi/10.2830/1603.

[14] E. Curry, "Real-time Linked Dataspaces, Enabling Data Ecosystems for Intelligent Systems, Chapter 1. Real-time Linked Dataspaces: A Data Platform for Intelligent Systems within Internet of Things-Based Smart Environments, pp. 3 – 14, doi: 10.1007/978-3-030-29665-0, published: 18 November 2019.

[15] C. S. Wasson, "System Engineering, Analysis, Design, and Development": Conept, Principles, and Practices, 2nd Edition, Wiley Series in Systems Engineering and Management, A.P. Sage, Series Editor, Chapter 34. System Reliability, Maintainability, and Availability (RMA) pp. 721 – 791, ISBN 978-1-118-44226-5, published December 2015.

[16] H. Obinson, *"Case study: Poka Yoke tool in Bohemian Brew Peru, a small business",* editor Using Poka-Yoke techniques for early defect detection. Sixth International Conference on Software Testing Analysis and Review; 1997; United States. ISBN 978-1499122282 *Case study: Poka Yoke tool in Bohemian Brew Peru, a small business*. Available from: https://www.researchgate.net/publication/373397768_Case_study_Poka_Yoke_tool_in_Bohemian_Brew_Peru_a_small_business [accessed 04 September 2025].

[17] M. Li, "Watch decentralized AI in 2025: The convergence of AI and crypto," Forbes, Nov. 12, 2024. Available: https://www.forbes.com/sites/digital-assets/2024/11/12/watchdecentralized-ai-in-2025-the-convergence-of-ai-and-crypto, cited in N. Kshetri, "Building Trust in AI: How Blockchain Enhances Data Integrity, Security, and Privacy, Digital Object Identifier 10.1109/MC.2024.3505012, Date of current version: 29 January 2025.

[18] A. Ciroth, "openLCA nexus quick explanation Web-based Life Cycle Assessment data exchange and web shop", version 1.0, May 2013.

[19] HM Government, "Greening Finance: A Roadmap to Sustainable Investing, published 18 October 2021, https://www.gov.uk/government/publications/greening-finance-a-roadmap-to-sustainable-investing accessed 04 September 2025.

[20] Defense Acquisition University, 'Developmental Test and Evaluation (DT&E)', AETM 011, DAU Glossary Definition, https://www.dau.edu/acquipedia-article/developmental-test-and-evaluation-dte# accessed 04 September 2025.

[21] ISO 31000:2018, Risk Management – Guidelines, edition 2, International Standard to be Revised [90.92], Technical Committee: ISO/TC 262, ICS: 03.100.01, https://www.iso.org/standard/65694.html#lifecycle, Published February 2018.

[22] ISO 16739-1:2024, Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries, Part 1: Data Schema, Technical Committee: ISO/TC 59/SC 13 & ICS : 25.040.40,https://www.iso.org/standard/84123.html, Published March 2024.

[23] ISO 21597-2:2020, Industry Container for Linked Document Delivery – Exchange Specification, Part 2: Linked Types, Technical Committee: *ISO/TC 59/SC 13 &* ICS : 35.240.67 91.010.01, https://www.iso.org/standard/74390.html, Published November 2020.

Rethinking the Role of Department of Defense Architecture Framework in System-of-Systems Architecture Design

Zhemei Fang and Yuxuan Liu
School of Artificial Intelligence and Automation
Huazhong University of Science and Technology
Wuhan, Hubei 430074 China
email: zmfang2018@hust.edu.cn

Jianbo Wang
China Ship Development and Design Center
Wuhan, Hubei 430064 China
email: jbwcn@hotmail.com

Abstract—**While Department of Defense Architecture Framework (DoDAF) remains widely adopted for architecture modeling, its application to System-of-Systems (SoS) design still faces significant challenges according to feedbacks from practitioners in industry and academia. Existing research often focuses on model creation or tool support but lacks a comprehensive examination of the issues behind the unsuccessful applications. Thus this paper analyzes the root causes of unsuccessful DoDAF applications, including the perspectives of common misconceptions, inherent shortcomings, methodological inadequacies, limitations of modeling tools, and cultural and organizational barriers. Based on the challenges observed, we further explore how the Unified Architecture Framework (UAF) and SysML 2.0 could alleviate some of these limitations. Based on this analysis, we propose three improvement directions: iterative, process-driven architecture modeling, AI-assisted model generation and evolution, and domain-specific meta-model customization with consistency assurance. The study concludes that treating architecture models as evolving decision-support tools, rather than static documentation, significantly enhances their value in SoS design and provides actionable guidance for improving DoDAF and other architecture frameworks in practice.**

Keywords-architecture design; department of defense architecture framework; system-of-systems; misconceptions.

I. INTRODUCTION

Architecting is increasingly being adopted by organizations to manage the growing complexity of human-made systems, particularly large-scale SoS such as those in defense and air transportation. The latest ISO/IEC/IEEE 42010:2022 standard (Software, systems and enterprise - architecture description) [1] defines architecture as "fundamental concepts or properties of an entity in its environment and governing principles for the realization and evolution of this entity and its related life cycle processes". Meanwhile, the standard introduces the term Architecture Description Framework (ADF) (replacing architecture framework in the 2011 version) to formalize the conventions and common practices of architecture description—a tangible work product that communicates the otherwise intangible and abstract concept of architecture [1].

The ADF has evolved from the C4ISR architecture framework to DoDAF, then to the Unified Profile for DoDAF/MODAF (UPDM), and most recently to the UAF. Despite this evolution, DoDAF remains the predominant ADF in the defense sector [2]. At the same time, most commercial modeling tools have gradually aligned their underlying meta-models with the UAF meta-model, enhancing tool interoperability while still maintaining support for DoDAF-based practices. Current DoDAF models [3][4] are compatible with UAF meta-models.

However, concerns about DoDAF have been raised over the years, including inconsistencies across architectural views [5], challenges in effectively utilizing architecture models for downstream applications [6], difficulties in accommodating new technologies, such as cloud computing and big data [7]. Although UAF was introduced to address some of these challenges, it inherits many of the same weaknesses. This critique is frequently acknowledged within the Model-based Systems Engineering (MBSE) community as well [8]. Interestingly, these issues are more commonly acknowledged in informal exchanges [8] than systematically addressed in published research. This gap highlights a critical need for more rigorous investigation into the practical barriers that hinder the effective application of ADFs in real-world SoS contexts.

This paper aims to uncover the reasons behind unsuccessful application of DoDAF, as a representative ADF, in supporting SoS architecture design. The perspectives include prevalent misconceptions about DoDAF's intended role, limitations in existing modeling tool support, methodological gaps in modeling approaches, and organizational and cultural barriers to model adoption. Building on this analysis, we propose several potential directions to achieve an enhanced use of DoDAF as well as other ADFs.

The paper is organized as follows. Section II reviews related work on architecture frameworks. Section III analyzes the key challenges of applying DoDAF to SoS design. Section IV discusses improvement opportunities. Section V concludes the study and suggests future research.

II. RELATED WORK

The importance of architecture, along with the supporting ADFs that guide its formal representation, has been increasingly acknowledged across both academic and industrial domains in recent years.

Early research by Wagenhals and Levis [9] pioneered a structured methodology for developing DoDAF models

ISBN: 978-1-68558-316-3

using IDEF0. Subsequently, numerous studies have adopted and extended this approach for DoDAF models development (e.g., [10]-[12]). In DoDAF model development, the Systems Modeling Language (SysML) has progressively superseded IDEF0 as the preferred modeling approach [11]. Current research and practice continue to demonstrate the framework's relevance, with active applications documented in recent works [3][4].

The U.S. Department of Defense (DoD) concluded its development of the DoDAF framework with the 2009 release (DoDAF 2.02). This transitioned to the UPDM, developed by the Object Management Group (OMG), as an interim solution. OMG subsequently established the UAF as the current standard [13]. Hause [14] indicates that the UAF was developed to address interoperability challenges by reducing disparities among architecture frameworks, modeling tools, standards, processes, data exchange formats, and domain terminology in ADF implementations.

From the 31st to 34th Annual INCOSE International Symposium proceedings, numerous implementation case studies of the UAF have been documented. For example, Martin [15] proposed an aspect-oriented approach aimed at harmonizing architectural frameworks to enhance interoperability and better support MBSE practices. Later, Martin [16] demonstrated how MBSE enhances an organization's ability to plan for capability deployments, and manage portfolios of systems, services, people, technologies, processes, and facilities critical to fielded capabilities. Carroll et al. [17] successfully implemented UAF in modeling the global copper market enterprise, noting its efficacy in fostering systems thinking beyond traditional engineering roles. Hause et al. [18] specifically addressed enterprise software architecture challenges through UAF modeling. Most recently, Martin et al. [19] and Gagliardi et al. [20] extended UAF's utility to Mission Engineering (ME), showcasing its adaptability to complex defense and aerospace applications, and the resultant modeling process and models are standardized in the U.S. DoD's Mission Architecture Style Guide (MASG) [21].

Alongside these applications of UAF, significant legacy challenges persist. Gagliardi et al. [20] highlight that "even a relatively simple Resource Architecture model requires significant time and effort to develop", emphasizing the need for careful upfront planning. Their findings suggest three critical prerequisites for effective UAF adoption: 1) scoping the modeling effort, 2) assessing modeling risks, and 3) establishing a model federation plan—all of which should be addressed prior to commencing development. Similarly, Fang et al. [22] pointed out that the relationship between DoDAF description models and architecting decisions is ambiguous—a limitation that also persists in UAF.

Modeling languages and tools also present challenges. Trujillo and Madni [23] highlight that modeling languages—particularly SysML—pose a high entry barrier, primarily due to the extensive training required to interpret increasingly complex models. In response, Morkevicius et al. [24] advocate for implementing UAF within the SysML v2 environment, anticipating that the updated specification may mitigate some inherent limitations of current SysML implementations. Regarding tooling considerations, Maier [25] indicates that a good modeling tool should manage significant redundancy in representations by using referencing instead of duplication and employing automated checks; nevertheless, there remains a clear risk of model proliferation beyond practical usefulness.

In summary, while the evolution from DoDAF to UAF has led to improved standardization and broader applicability in both defense and enterprise contexts, practical challenges remain prevalent across modeling frameworks, languages, and tools. The literature reveals a persistent tension between the theoretical promise of ADFs and their real-world implementation barriers—many of which stem from complexity, tool limitations, and organizational constraints. These gaps underscore the necessity for a deeper investigation into the root causes hindering effective ADF application, particularly in complex SoS environments. Building upon these insights, this study aims to critically examine the key obstacles to DoDAF adoption and propose actionable strategies for enhancing its practical utility.

III. Practical Challenges and Inherent Shortcomings of DoDAF in SoS Architecture Design

The unsuccessful applications of DoDAF in supporting SoS architecture design stem from a fundamental misunderstanding of its intended role, limited support from modeling tools, inadequate methodological guidance, and practical and cultural barriers to model adoption, as shown in Fig. 1. This section examines these four aspects in detail.

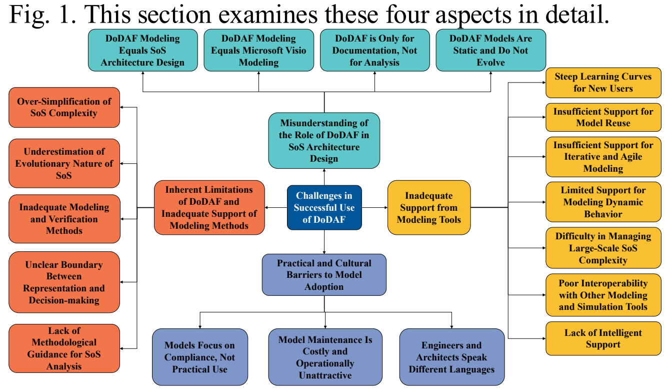

Figure 1 . Practical challenges of DoDAF in SoS architecture design.

A. *Misunderstanding of the Role of DoDAF in SoS Architecture Design*

Based on our practical modeling experiences and interviewing with modeling experts in industry, we summarize four common misunderstandings of the DoDAF's role in SoS architecture design.

1) Misunderstanding I: DoDAF Modeling Equals SoS Architecture Design

This misunderstanding often arises among outsiders who have unrealistically high expectations of DoDAF. They mistakenly assume that creating DoDAF models is equivalent to completing SoS architecture design.

In fact, DoDAF provides a structured set of standardized views and establishes a formal framework for representing

 ISBN: 978-1-68558-316-3

SoS architecture. However, a practical and effective SoS architecture design involves not only representation but also decision-making and evaluation—aspects that DoDAF alone does not fully address. Therefore, additional methodologies, such as operational simulation, trade-space analysis, and optimization, are essential to complement DoDAF for achieving a comprehensive SoS architecture design.

These inflated expectations often lead to significant disappointment, ultimately causing them to overlook the actual value of DoDAF models.

2) Misunderstanding II: DoDAF Modeling Equals Microsoft Visio Modeling

This misunderstanding often arises among practitioners who have some experiences with DoDAF modeling but have not delved into the underlying theories. They assume that creating DoDAF views is simply about drawing static diagrams, like flowcharts, without considering the underlying semantic relationships, constraints, and traceability.

In fact, DoDAF is model-based, not merely diagram-based. While it employs visual representations, it is fundamentally a structured architecture framework, not just a collection of disconnected drawings. Tools like Visio and similar diagramming software allow freeform visualization but do not enforce architectural consistency or data integrity. In contrast, DoDAF models should be developed using structured modeling tools (e.g., Cameo Enterprise Architecture, Sparx EA, IBM Rhapsody) that enforce rules and ensure consistency between capabilities, systems, and services across multiple views.

This misunderstanding can lead to superficial architecture modeling that lacks architectural rigor. Organizations may create visually appealing but structurally meaningless diagrams that fail to support real system development. Without architectural rigor, inconsistencies and logical errors may go unnoticed, ultimately undermining the effectiveness of the architecture.

3) Misunderstanding III: DoDAF is Only for Documentation, Not for Analysis

DoDAF is often misperceived as merely a documentation framework, rather than a foundation for architectural analysis and informed decision-making. This misunderstanding stems in part from the limitations of current practices and tools, which often fail to deliver on the promise of model-based analysis. Despite many tools claiming to support analytical functions, the actual use of DoDAF models for quantitative or qualitative analysis remains challenging in practice.

Several factors contribute to this gap. First, many DoDAF-compliant tools focus heavily on model visualization and reporting, offering limited support for integrated simulations, trade-off analysis, or impact assessments. Second, users may lack clear methodological guidance on how to leverage architectural description models for analytical purposes, especially in complex SoS contexts. Lastly, architecture models are often developed in isolation from operational or technical data, limiting their usefulness for real-time or predictive analysis.

As a result, DoDAF models are frequently underutilized in decision-making processes, reducing their value to stakeholders and reinforcing the notion that they are static deliverables rather than dynamic decision-support artifacts.

4) Misunderstanding IV: DoDAF Models Are Static and Do Not Evolve

Some organizations mistakenly treat DoDAF models as static, one-time deliverables rather than as evolving artifacts that must be continuously updated as the system evolves. This misconception largely arises from the inadequate support current modeling tools provide for iterative development and model maintenance.

SoS architectures are dynamic, requiring continuous updates to DoDAF models to reflect new requirements, emerging threats, and evolving technologies. Architecture models should support versioning, impact analysis, and iterative refinements throughout the SoS lifecycle.

When this need for evolution is overlooked, DoDAF models quickly become outdated and disconnected from the actual SoSs they are intended to represent, resulting in misalignment between architectural intent and operational reality.

5) Summary

The misunderstandings stem not only from a general lack of familiarity with DoDAF, but also from widespread disappointment with its practical applications. These challenges arise from inherent limitations within DoDAF and supporting methods, inadequate support from current modeling tools, and cultural resistance to adopting model-driven approaches.

B. Inadequate Support from Modeling Tools

From the perspective of modeling tools, the issues can be categorized into the following aspects.

1) Steep Learning Curves for New Users

Existing DoDAF tools often present steep learning curves, particularly for multidisciplinary teams involving architects, engineers, and operators. This hinders effective collaboration, especially when stakeholders have varying levels of modeling expertise.

2) Insufficient Support for Model Reuse

Model reuse is a fundamental benefit of architecture description modeling [23]. However, in practice, the tightly coupled nature of elements within DoDAF-based architecture models often impedes effective reuse. This rigidity limits the adaptability of existing models to new systems or evolving contexts. While some of these issues stem from tool implementations, the underlying challenges are also rooted in the structural constraints and design philosophy embedded in the DoDAF metamodel itself.

3) Insufficient Support for Iterative and Agile Modeling

SoS architecture design is typically an iterative process, yet most DoDAF tools do not effectively support version control, impact analysis, or automatic updates. Furthermore, the weak integration between different design phases (e.g., from capability planning to system design) makes it difficult to transition seamlessly from conceptual models to executable or detailed design artifacts.

4) Limited Support for Modeling Dynamic Behavior

Most DoDAF tools are primarily designed to represent static structures and relationships. While activity and sequence models offer some capability to model and analyze dynamic behaviors, they lack the flexibility needed to handle a wide range of scenarios. This limitation makes it challenging to perform simulations or visualizations that accurately reflect the operation of SoS under varying conditions, thus reducing the practical utility of architecture models in operational analysis and decision-making.

5) Difficulty in Managing Large-Scale SoS Complexity

When dealing with complex SoS architectures, comprising a large number of activities, systems, and interfaces, many tools exhibit performance bottlenecks. This includes slow user interface responsiveness and delays in rendering large diagrams. Moreover, as the interconnections between elements grow more intricate, users often find it difficult to trace dependencies, leading to confusion and decreased confidence in the models.

6) Poor Interoperability with Other Tools

Despite the growing emphasis on integrated modeling environments, current DoDAF tools often operate in silos. They lack interoperability with executable modeling tools, such as Modelica, Simulink, or AnyLogic. Data format inconsistencies and the absence of standardized exchange mechanisms hinder seamless integration, resulting in duplicated efforts and inconsistencies between architectural models and executable simulations.

7) Lack of Intelligent Support

The modeling process can be cumbersome, adding to the already heavy workload of architects and SoS engineers, who are responsible for many other tasks. Current modeling tools offer limited intelligent assistance, such as automated reasoning, consistency checking, or even model auto-generation. The integration of advanced technologies, such as large language models (LLM), holds significant potential to improve these processes by offering smarter support.

C. Inherent Limitations of DoDAF and Inadequate Support of Modeling Methods

From the perspective of inherent limitations and inadequate methodological support, five key issues can be identified: the first two stem from the intrinsic limitations of DoDAF itself, while the latter three arise from shortcomings in existing modeling methods.

1) Over-Simplification of SoS Complexity

While the goal of ADFs is to develop stable blueprints, expressed through various views, for complex SoS—similar to blueprints for building architecture—the boundaries of an SoS are far more intricate than those of a building. The diversity of stakeholders, unclear boundaries (and sometimes even objectives), varying development timelines for constituent systems, and the occurrence of complex, unexpected emergent behaviors all contribute to the difficulty of representing an SoS. As a result, ADFs tend to oversimplify the inherent complexity of SoS, making the choice of appropriate abstraction critically important.

2) Underestimation of Evolutionary Nature of SoS

SoSs are inherently dynamic, evolving continuously in response to changing requirements, constituent system upgrades, and unforeseen operational conditions. However, DoDAF often treats architecture models as static snapshots rather than living artifacts that demand iterative validation and continuous adaptation. While views such as CV-3 (Capability Phasing) and SV-8 (Systems Evolution Description) attempt to address system evolution, they largely depict it as a predefined, static process. Furthermore, many types of changes are overlooked—for example, frequent updates to OV-5b (Operational Activity Model) and OV-4 (Organizational Relationship Chart) are seldom adequately captured or represented.

3) Inadequate Modeling and Verification Methods

Although many modeling methods have been proposed over the years, some fundamental issues remain, primarily stemming from the inherent subjectivity of the modeling process. A typical example is the lack of a systematic understanding of granularity levels, which leads to inconsistent model granularity—some levels are overly detailed while others are too vague, resulting in a disorganized hierarchy. These seemingly minor issues can hinder the development of effective and reliable models.

In terms of verification, most existing methods rely on syntactic checks and rule-based reasoning [5], which are insufficient for detecting complex logical errors. This limitation undermines the reliability of the models and erodes user confidence in their correctness and utility.

4) Unclear Boundary Between Representation and Decision-making

DoDAF models are designed to structure vague or incomplete information, define and formulate decision-making problems, and guide architectural decisions [22]. However, these decision-making issues often remain obscured within the architecture models. This ambiguity creates confusion, leading to uncertainty about whether the models are flawed due to insufficient modeling experience or a lack of adequate decision analysis.

5) Lack of Methodological Guidance for SoS Analysis

While DoDAF defines a set of views, it offers limited guidance on how to use these views to conduct architecture evaluations, trade-space exploration, or impact analysis. Users are often left to interpret the views without a clear methodological framework, leading to inconsistent and ineffective practices. More critically, in many real-world applications, users struggle to identify latent deficiencies or potential shortcomings in the architecture design as represented by the models.

D. Practical and Cultural Barriers to Model Adoption

Beyond the structural limitations of DoDAF and the constraints of current modeling tools, the successful adoption of architecture models in real-world SoS projects also faces practical and cultural challenges. These issues reflect broader organizational behaviors and workflow mismatches that hinder the integration of DoDAF-based modeling into engineering practice.

1) Models Focus on Compliance, Not Practical Use

In many defense projects, DoDAF models are developed primarily to satisfy contractual or regulatory requirements rather than to support real-world design decisions. This compliance-driven mindset turns modeling into a box-checking exercise, where deliverables are created to pass reviews but rarely maintained or reused afterward. Even when the importance of architecture modeling is acknowledged, organizations often lack incentives or processes to keep these models up to date throughout the system's lifecycle. Once initial approvals are secured, model updates are deprioritized, reinforcing the perception that architecture models are static documents rather than evolving, decision-support tools. As a result, the long-term value of model-based systems engineering is significantly diminished.

2) Model Maintenance is Costly and Operationally Unattractive

The effort required to keep architecture models aligned with rapidly changing systems often outweighs the perceived benefits. Teams may prefer to directly update prototypes or source code, bypassing the architecture layer entirely. As a result, models quickly become outdated and are abandoned, viewed as an unsustainable overhead rather than a valuable asset for ongoing development

3) Engineers and Architects Speak Different Languages

A cultural gap exists between architects, who work within frameworks like DoDAF, and engineers, who focus on building and testing systems using simulation environments or programming languages. Engineers often find that DoDAF models are too high-level to support executable behavior or real system implementation in tools like Python or Simulink. This disconnect hampers collaboration and limits the effectiveness of architecture-driven development, leaving the architecture models isolated from actual system implementation.

IV. Opportunities for Improvement

Based on the identified challenges, we first evaluate whether UAF and SysML 2.0 can address some of these issues, and then propose several directions to enhance the practical application of DoDAF—applicable to UAF as well—in supporting SoS architecture design.

A. UAF's Capability to Address the Issues

As discussed in Section II, the UAF consolidates multiple architecture frameworks and offers more comprehensive views and dimensions compared to DoDAF. At its core, UAF establishes an integrated meta-model that enhances the semantic consistency and structural rigor of architecture representations. This unified meta-model also enables improved traceability from architectural elements to capability objectives by systematically linking functions, resources, and operational activities to capability definitions and performance measures.

Importantly, the OMG provides extensive support for UAF adoption, including the UAF Domain MetaModel (DMM), the UAF Modeling Language (UAFML), and a practical guide for enterprise architecture development. These resources offer more structured methodological guidance and clearer modeling practices than DoDAF, contributing to improved usability and standardization in SoS architecture design. Furthermore, UAF aligns more closely with MBSE principles and SysML [26], facilitating tighter integration between SoS architecture modeling and system lifecycle management.

Nevertheless, despite addressing fragmentation and enhancing semantic clarity, UAF still faces practical adoption challenges—particularly in terms of modeling methodology, tool maturity, and organizational constraints—as discussed in Section III.

B. SysML 2.0's Capability to Address the Issues

The current modeling language, SysML, is undergoing a significant transformation with the development of SysML 2.0. The SysML 2.0 standard focuses on three core elements, the underlying Kernal metaModeL (KerML), modeling semantics and syntax in the SysML, and the Application Programming Interface (API) and services [27]. It integrates graphical and textual modeling approaches, bridging the language gap between system architects and domain engineers. At the same time, it enhances modeling flexibility and efficiency, while supporting model sharing and automation. This revision aims to improve usability for systems engineering practitioners by introducing these more intuitive language constructs, enhanced expressiveness, and better model organization.

SysML 2.0 also defines standardized APIs that enable seamless integration with simulation engines and verification tools, significantly enhancing interoperability across the system development lifecycle. Moreover, it offers improved composability, allowing for more coherent and scalable representations of hierarchical structures—from SoSs to individual systems and components.

Moreover, its support for a formal textual syntax makes it naturally compatible with LLMs (e.g., ChatGPT, DeepSeek), enabling more interactive model manipulation, streamlined workflows, and reduced modeling complexity [28].

SysML 2.0 holds strong potential to address many of the challenges outlined in Section III; however, most of these anticipated benefits have yet to be validated in practice, and realizing them would require significant retooling of existing tools and workflows.

C. Improvement Suggestions

1) Architecture Description Models Reflect Architecting Process more than Architecture Outcomes

Rather than building complete DoDAF models upfront, development teams should focus on creating evolving, minimal viable models. Fig. 2 illustrates an iterative architecture modeling process that encompasses architecture modeling, analysis, evaluation, and decision-making. Simultaneously, enabling different stakeholders to contribute at varying levels of detail promotes better collaboration and aligns with agile development principles.

Fig. 3 demonstrates an example of iterative architecture modeling process that integrates DoDAF models, executable models (e.g., ExtendSim, Anylogic), and decision models. The decision models include qualitative decisions that help collect constraints/rules and clarify the information for architecture models, and quantitative decision-making and evaluations based on executable simulation results. Compared to the traditional paradigm [9], the key emphasis is placed on an iterative modeling process rather than delivering a complete set of architecture models all at once. Our core argument is that architecture models should serve as a means to guide and evolve with the architecting process, rather than simply capture its final products.

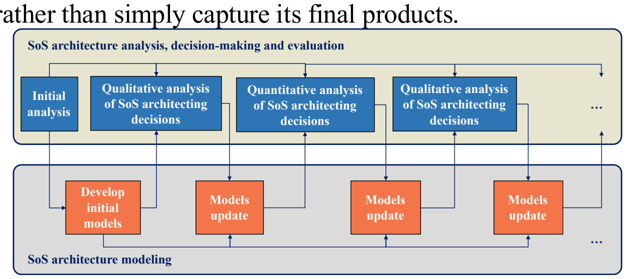

Figure 2 . Iterative architecture modeling process.

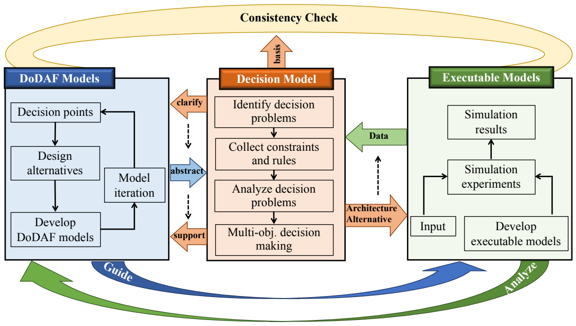

Figure 3 . An example for iterative architecture modeling process that integrates DoDAF models, executable models, and decision models.

2) AI-Assisted Architecture Modeling and Design

Recent artificial intelligence (AI) technologies offer significant potential for supporting SoS architecture design. As listed in Fig. 4, AI can support this process in four key areas: AI-assisted architecture modeling, AI-assisted architecture selection, AI-assisted architecture verification, and AI-assisted architecture evolution.

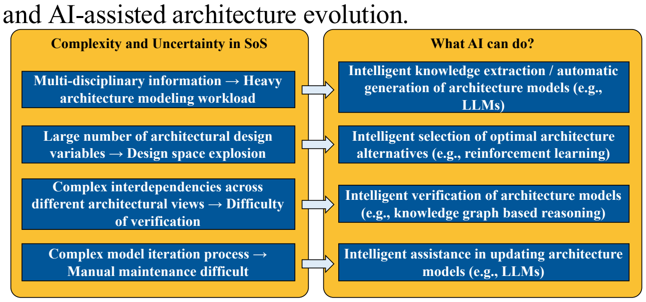

Figure 4 . Issues requiring AI assistance and potential solutions.

Among these areas, AI-assisted architecture modeling and evolution have attracted significant attention in the past two years, primarily due to the challenges associated with manual model development and maintenance, which are both labor-intensive and error-prone. Fig. 5 illustrates the generation process of architecture models (e.g., SysML or DoDAF models) using LLMs, which support the automatic generation of functional/component decompositions, activity models, and other artifacts in standard XML format. These standard XML models can then be transformed into XML structures compatible with SysML or DoDAF specifications.

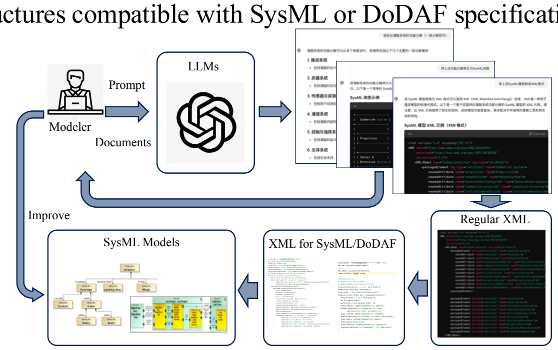

Figure 5 . Architecture model generation framework based on LLMs.

AI-driven techniques, when integrated with model version control, also show strong potential for automatically detecting inconsistencies, recommending updates, and managing complex dependencies. Furthermore, the ability to synchronize SysML/DoDAF/UAF models with real-time operational data could greatly enhance the timeliness and accuracy of model updates throughout the design lifecycle.

3) Customized Metamodel Development and Underlying Consistency Assurance

To better support domain-specific needs, organizations can develop customized meta-models that extend or specialize existing frameworks (e.g., DoDAF, UAF). These tailored meta-models allow for more precision in addressing specific requirements of a given system or domain. An integrated process of SoS architecture development and meta-model development is illustrated in Fig. 6.

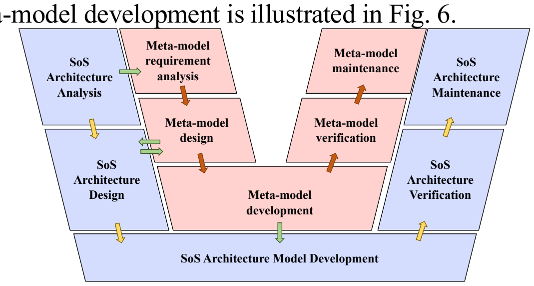

Figure 6 . SoS architecture design process with meta-model development.

It is important to note that developing customized meta-models introduces the challenge of maintaining consistency across different modeling views and with other frameworks used by different organizations. To address this, consistency assurance mechanisms must be integrated into the meta-model development process. This includes defining clear consistency rules and validation methods to ensure that models derived from the customized meta-model align with the intended system structure and behavior, while also ensuring better compliance with existing meta-models.

V. Conclusion and Future Work

This paper has analyzed the key challenges facing DoDAF in the SoS architecture design, including misconceptions, method limitations, inadequate tool support, and organizational barriers. Our findings indicate that the core issue lies in treating DoDAF as a static documentation tool rather than a dynamic decision-support asset that must evolve throughout the lifecycle. Several key lessons emerged from this investigation. We observed that organizational and technical barriers are deeply intertwined, each exacerbating the other. A recurring difficulty was distinguishing whether problems originated from DoDAF's inherent limitations, tooling deficiencies, or methodological misapplication. While newer frameworks like UAF offer improved semantic consistency, our findings temper expectations regarding their immediate utility, as they still face challenges in method and tool maturity. The integration of AI-assisted modeling presents a promising yet challenging path forward.

Future work will focus on three directions: developing a lightweight iterative modeling plugin to integrate architectural models with decision-support tools; creating a specialized prompt engineering framework for LLMs tailored to SoS architecture tasks; and establishing quantitative metrics to empirically validate improvements in model maintenance efficiency and decision-support capability. Eventually, transforming DoDAF from a documentation exercise into an evolving intelligent decision-support process represents quite a promising direction for enhancing its practical value in complex SoS environments.

References

[1] ISO/IEC/IEEE, "ISO/IEC/IEEE 42010:2022(E) Systems, Software and Enterprise - Architecture Description," 2022.

[2] Joint Staff, "Charter of the Joint Requirements Oversight Council and Implementation of the Joint Capabilities Integration and Development System," 2021.

[3] G. Bangjun, C. Yunfeng, W. Xinyao, and Y. Wangwang, "Modeling the Anti-UAV Swarm System Architecture Based on DoDAF," in 2023 9th Int. Conf. Big Data Inf. Anal., Haikou, China, Dec. 2023, pp. 409–412.

[4] A. Aghamohammadpour, E. Mahdipour, and I. Attarzadeh, "Architecting Threat Hunting System Based on the DoDAF Framework," J. Supercomput., vol. 79, pp. 4215–4242, 2023.

[5] M. Vinarcik, "The Problem with DoDAF Models," in 22nd Annu. Syst. Mission Eng. Conf., Tampa, FL, 2019.

[6] S. W. Mitchell, "Transitioning the SWFTS Program Combat System Product Family from Traditional Document-Centric to Model-based Systems Engineering," Syst. Eng., vol.17, no.3, pp.313-329, 2014.

[7] M. Zaman, "Dynamic Resilient Enterprise Architecture Model (DREAM) Adoption in Defense Digital Architecture Management to Mitigate Disconnected Systems and Processes," Ph.D. dissertation, George Mason Univ., 2025.

[8] [Reddit], "Change My View: Model-Based Systems Engineering in 2024 is Still More Hype Than Value," r/systems_engineering, 2024. [Online]. Available: https://www.reddit.com/r/systems_engineering/comments/1bpavpi/change_my_view_model_based_systems_engineering_in/?rdt=57765 [Accessed: Sept, 21st, 2025].

[9] L. W. Wagenhals, I. Shin, D. Kim, and A. H. Levis, "C4ISR Architectures II. A Structured Analysis Approach for Architecture Design," Syst. Eng., vol. 3, no. 4, pp. 248–287, 2000.

[10] C. Piaszczyk, "Model Based Systems Engineering With Department of Defense Architectural Framework," Syst. Eng., vol. 14, no. 3, pp. 305-326, 2011.

[11] R. Wang and C. H. Dagli, "Executable System Architecting Using Systems Modeling Language in Conjunction With Colored Petri Nets in a Model-Driven Systems Development Process," Syst. Eng., vol. 14, no. 4, pp. 383-409, 2011.

[12] M. Amissah and H. A. H. Handley, "A Process for DoDAF-Based Systems Architecting," in 2016 Annu. IEEE Syst. Conf., Orlando, FL, USA, 2016, pp.1-7.

[13] J. N. Martin and D. P. O'Neil, "Enterprise Architecture Guide for the Unified Architecture Framework (UAF) ," in Proc. INCOSE Int. Symp., vol. 31, pp. 242-263, 2021.

[14] H. M. Hause, "Rebuilding the Tower of Babel: The Case for a Unified Architecture Framework," in Proc. INCOSE Int. Symp., vol. 23, no. 1, pp. 1460–1474, 2013.

[15] J. N. Martin, "Aspect-Oriented Architecting Using Architecture Frameworks," in Proc. INCOSE Int. Symp., vol. 31, no. 1, pp. 210–226, 2021.

[16] J. N. Martin, "Extending UAF for Model-Based Capability Planning and Enterprise Portfolio Management," in Proc. INCOSE Int. Symp., vol. 32, no. 1, pp. 15-35, 2022.

[17] K. Carroll, A. Lyle, R. Lewark, C. Medina, and A. Morkevičius, "SoS - Global Solutions to Global Problems Using UAF," in Proc. INCOSE Int. Symp., vol. 34, pp. 2400-2412, 2024.

[18] M. Hause and L. O. Kihlström, "Modeling Enterprise Software with UAF," in Proc. INCOSE Int. Symp., vol. 34, pp. 2452-2475, 2024.

[19] J. N. Martin and K. E. Alvarez, "Using the Unified Architecture Framework in Support of Mission Engineering Activities," in Proc. INCOSE Int. Symp., vol. 33, pp. 1156–1172, 2023.

[20] M. Gagliardi, M. C. Hause, J. N. Martin, and M. A. Phillips, "Darth Vader's Secret Weapon: Implementing Mission Engineering With UAF," in Proc. INCOSE Int. Symp., vol. 34, pp. 1719-1747, 2024.

[21] Office of the Under Secretary of Defense for Research and Engineering, Department of Defense Mission Architecture Style Guide. Washington, DC, USA: U.S. Department of Defense, 2025.

[22] Z. Fang, X. Zhao, and F. Li, "Architecture Design Space Generation via Decision Pattern-Guided Department of Defense Architecture Framework Modeling," Systems, vol. 12, no. 2, p. 336, 2024.

[23] A. E. Trujillo and A. M. Madni, "MBSE Methods for Inheritance and Design Reuse," in Handbook of Model-Based Systems Engineering, A. M. Madni, N. Augustine, and M. Sievers, Eds. Cham, Switzerland: Springer, 2023.

[24] A. Morkevicius and G. Krisciuniene, "Towards UAF Implementation in SysML V2," in Proc. INCOSE Int. Symp., vol. 34, pp. 2452-2475, 2024.

[25] M. W. Maier, "Adapting the Hatley-Pirbhai Method for the Era of SysML and Digital Engineering," 2022 IEEE Aerosp. Conf., Big Sky, MT, USA, 2022, pp. 1-12.

[26] J. N. Martin and D. Brookshier, "Linking UAF and SysML Models: Achieving Alignment Between Enterprise and System Architectures," in Proc. INCOSE Int. Symp., vol. 33, pp. 1132–1155, 2023.

[27] Object Management Group, "OMG Systems Modeling Language (SysML) Version 2.0 Beta 2 (Revision 2024-02), Part 2: SysML v1 to SysML v2 Transformation," OMG Doc. No. ptc/2024-02-01, Feb. 2024.

[28] J. K. DeHart, "Leveraging Large Language Models for Direct Interaction With SysML v2," in Proc. INCOSE Int. Symp., vol. 34, pp. 2168–2185, 2024.

An Interference-Aware vCPU Scheduling Framework for Paravirtualized Real-Time Industrial Control Systems

Jessica Müller*, Steven Dietrich*, Michael Massoth†
*Bosch Rexroth AG; †Hochschule Darmstadt
*Lohr a. Main, Germany; †Darmstadt, Germany
email: {jessica.mueller4, steven.dietrich}@boschrexroth.de
email: michael.massoth@h-da.de

***Abstract*—Virtualization enables flexible, software-defined architectures in industrial automation, but introduces new challenges, such as resource contention and unpredictable latencies. This paper presents an interference-aware scheduling approach based on paravirtualized VM profiling. By dynamically classifying virtual CPUs (vCPUs) considering dominant I/O usage and preventing simultaneous execution of tasks with overlapping I/O demands, the method improves determinism and responsiveness. Simulated under realistic workloads, the scheduler significantly reduces utilization peaks, eliminates overload conditions, and stabilizes workload distribution. These results demonstrate the potential of the approach to enhance the predictability and efficiency of virtualized industrial control systems.**

Keywords-virtualization; industrial automation; real-time systems; VM scheduling; interference mitigation; vCPU classification; hypervisor scheduling; industry 4.0.

I. Introduction

The ongoing shift towards flexible production systems is a defining feature of Industry 4.0 (I4.0), where adaptability, modularity, and responsiveness are critical design goals [1]. To support this transformation, industrial control systems are deployed increasingly as Virtual Machines (VMs) hosted on centralized hypervisor platforms. This virtualization enables software- defined control, efficient resource utilization, and dynamic system reconfiguration without modifying physical hardware. However, the consolidation of time-sensitive applications onto shared virtualized infrastructures introduces new challenges. In particular, resource contention at the I/O or CPU level can lead to unintended temporal interference between virtual machines [2]. Such interference may impact the timing behavior of control applications and thus affect the predictability and reliability required in industrial automation environments. We contribute a new interference-aware scheduling approach that explicitly accounts for cross-VM interference at scheduling time rather than relying on conservative worst-case scheduling.

The remainder of this paper is organized as follows. Section II introduces Multi-Virtual-Machine (Multi-VM) environments, outlining industrial use cases, the state of VM scheduling in practice, and the shortcomings that motivate our work. Section III details the proposed interference-aware scheduling approach based on paravirtualized VM profiling, covering its design rationale, architectural components, and integration workflow. Section IV describes the experimental testbed and simulation scenarios used to evaluate the scheduler under realistic industrial conditions. Section V presents and interprets the results, with a focus on latency, interference mitigation, and their implications for industrial automation. Section VI concludes the paper and sketches avenues for future research.

II. Multi-Virtual-Machine Environments

To understand the challenges and design requirements of interference-aware scheduling, it is first necessary to analyze how industrial multi-VM environments are structured, how scheduling is currently implemented, and where existing limitations arise.

A. Industrial Use Cases and Requirements for Multi-VM Systems

In the context of I4.0, industrial control systems are deployed increasingly as VMs hosted on centralized computing platforms. Instead of being distributed across multiple embedded devices, control logic, HMIs, and edge analytics are consolidated into a single physical system running multiple VMs concurrently [3]. This architectural shift enables streamlined system integration, centralized updates, and flexible resource allocation in modular and reconfigurable production environments. To ensure strong isolation and low overhead, these virtualized control systems typically are managed by a Type 1 hypervisor [4].

A central requirement for such deployments is deterministic behavior for time-critical control loops. In particular, short and stable control cycle times – typically in the range of 1–10 ms – are essential for guaranteeing timely responses to sensor inputs and actuator commands [5][6]. Any temporal deviations caused by VM scheduling delays or resource contention at the hypervisor layer must therefore be minimized to maintain the overall system's functional integrity and reliability. An overview of this architecture is illustrated in Figure 1.

Within a Systems-of-Systems (SoS) setup, a hybrid control architecture is feasible: autonomous local real-time loops handle fast dynamics, while a lightweight supervisory layer coordinates setpoints and resource constraints across VMs.

B. Technical Overview: Current VM Scheduling in Industry

In modern industrial environments, Type-1 hypervisors play a critical role in consolidating control systems, HMIs, and edge computing workloads into virtualized infrastructures. These bare-metal hypervisors, such as VMware ESXi, Microsoft Hyper-V, or open-source solutions like Xen and KVM (with

 ISBN: 978-1-68558-316-3

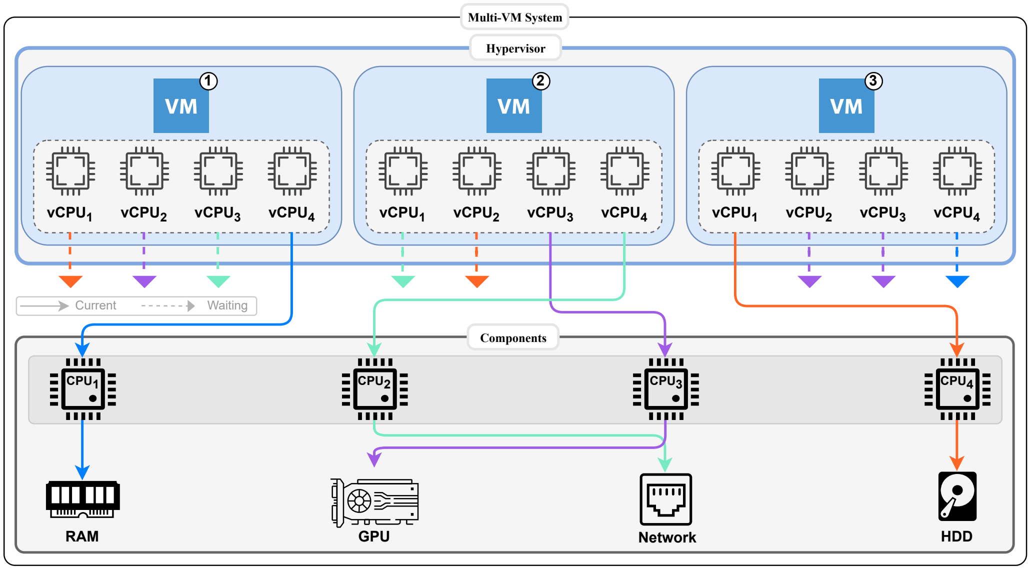

Figure 1. Multi-VM architecture with shared resources under a Type 1 hypervisor.

real-time extensions), operate directly on the host hardware and manage the allocation of physical CPU resources to VMs.

The core responsibility of the hypervisor's scheduler is to map vCPUs of the guest VMs to physical CPUs (pCPUs) on the host system [7]. Current scheduling strategies predominantly rely on variants of fair-share, priority-based, or real-time aware algorithms:

- **Fair-Share Schedulers**, such as the default Credit Scheduler in Xen or the Completely Fair Scheduler in KVM, aim to distribute CPU time evenly across VMs, based on configurable weights or credits. These are designed for general-purpose workloads and maximize overall utilization [7].
- **Priority-Based Scheduling** is commonly used to assign static or dynamic priorities to VMs or individual vCPUs. High-priority tasks receive preferential CPU access, which is suitable for scenarios with mixed workloads where certain VMs are more critical than others [8].
- **Real-Time Extensions** are offered in hypervisors, such as VMware ESXi with *Latency Sensitivity Mode* or KVM with the *PREEMPT_RT* patch. These mechanisms allow for stricter control over scheduling behavior, including CPU pinning (affinity), isolation from non-real-time workloads, and reservation of exclusive resources [9].

In the context of industrial automation, schedulers often leverage CPU affinity and isolation techniques to bind critical control VMs to dedicated cores, thereby reducing variability introduced by co-located workloads. Additionally, reservation mechanisms allow guaranteeing a minimum share of CPU time to latency-sensitive VMs [7].

Hypervisors may also employ I/O-aware scheduling policies, attempting to balance compute and I/O workloads across VMs. However, in standard configurations, CPU and I/O scheduling remain decoupled, which can introduce indirect effects on determinism – especially under high system load [10].

Overall, current hypervisor scheduling mechanisms are designed to ensure fair, efficient, and scalable CPU usage across virtual machines. While real-time features exist, their practical integration into industrial VM setups often requires careful tuning and architectural planning.

C. *Identified Shortcomings and Interference Issues*

Despite the availability of real-time extensions and resource isolation features, current hypervisor scheduling mechanisms remain susceptible to temporal interference – particularly in I/O-intensive scenarios [4]. In virtualized industrial environments, where deterministic control loops must operate within strict cycle times of 1–10 ms, even minor deviations in execution timing can compromise system integrity.

A key source of such deviation lies in the interaction between vCPU scheduling and I/O operations. Although CPU time may be reserved or pinned for a control VM, I/O subsystems (e.g., disk, network, or fieldbus interfaces) are typically shared among multiple VMs and rely on asynchronous handling through interrupt-driven mechanisms or hypervisor-level emulation [11][7]. These operations introduce latency that is neither fully visible nor fully controllable by the guest operating system, leading to non-deterministic delays in input acquisition or actuator response.

Moreover, hypervisors often decouple I/O scheduling from CPU scheduling, which makes it difficult to coordinate compute and communication timing holistically [12]. For instance, when multiple VMs compete for access to shared I/O resources – such as a virtual NIC or storage backend – context switches, interrupt storms, or emulation delays may disrupt the timing guarantees required by control applications [11][12]. These effects further are amplified under system load, where best-effort workloads or background processes inadvertently interfere with time-critical VMs, despite configured priorities or affinity.

As a result, cycle-time violations and jitter become increasingly probable in consolidated setups, particularly when industrial controllers, HMIs, and monitoring tools coexist on the same host [12]. Without holistic temporal coordination across all relevant subsystems – including CPU, memory, and I/O paths – the promise of determinism in virtualized control architectures remains difficult to fulfill under real-world conditions.

III. Interference-Aware Scheduling via Paravirtualized VM Profiling

To address the timing deviations and interference issues identified in multi-VM environments, a novel scheduling approach is introduced that explicitly considers the I/O behavior of virtual machines and their interactions at runtime.

A. Design Motivation and Objectives

In industrial environments increasingly shaped by digitalization and I4.0, conventional scheduling mechanisms are reaching their limits. These mechanisms were not typically designed to meet the specific demands of virtualized control systems [13]. A major issue in this context is I/O interference, which leads to unpredictable latencies and violations of strict cycle times. This undermines the reliability of industrial control applications, where, for instance, a guaranteed 1 ms cycle time is critical – but in practice, often only a worst-case latency of around 100 ms can be assured [6].

The aim of the newly conceived scheduling approach is therefore to proactively mitigate such interference through deliberate planning. This enables more reliable system availability, as the state of the I/Os is known at all times. It not only facilitates dynamic load balancing but also allows for foresighted resource allocation for potential emergency scenarios, such as interrupt-driven, I/O-intensive operations. In addition to the classical objective of optimal process and vCPU distribution, this approach strengthens overall system stability under real-time conditions.

B. Architectural Overview of the Profiling Scheduler

The proposed scheduler architecture consists of two tightly integrated components: A classification unit and a scheduling unit. As soon as a vCPU becomes eligible for execution, it is passed to the classification unit, which determines the dominant I/O resource it is expected to interact with. This classification is based on a lightweight analysis of the task characteristics within the vCPU and assigns it to an I/O category, such as GPU-bound, RAM-bound, cache-bound, network-bound, or disk-bound. The process is performed immediately before each scheduling decision, ensuring that classification always reflects the current system context without relying on historical profiling data. An example for the classification is shown in Figure 2.

Once classified, the vCPU is passed to the scheduling unit, which maps it to an appropriate pCPU core. The central policy

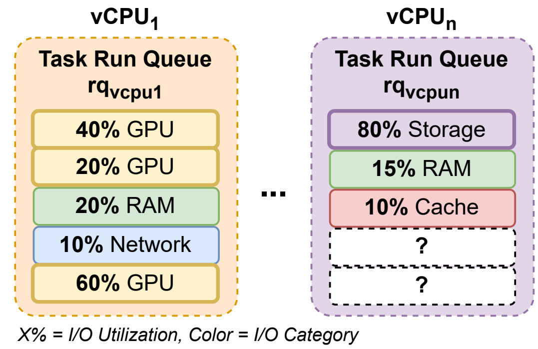

Figure 2. Classification of vCPUs based on their run queue.

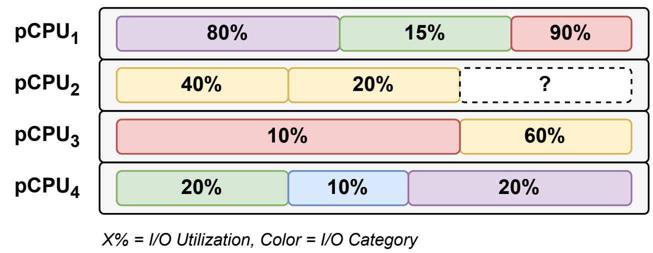

Figure 3. Scheduling Timelines.

enforced by the scheduler is to avoid concurrent execution of vCPUs from the same I/O category on different physical cores. This interference-aware constraint ensures that no two vCPUs with similar I/O access patterns simultaneously contend for the same shared hardware resource. By isolating I/O categories across cores within a given time window, the system prevents unpredictable latency spikes caused by overlapping access to memory buses, storage devices, or network interfaces. An example for the scheduling is shown in Figure 3.

The entire process is executed synchronously and on-demand: Every time a vCPU enters the ready queue, the classification and scheduling decisions are computed in a single step. This approach maintains high responsiveness while avoiding background profiling overhead.

C. Integration into Virtualized Environments

Practical deployment of the interference-aware scheduler requires integration at the hypervisor's kernel scheduling layer. On Linux-based hypervisors, such as KVM, this can be realized via the `sched_ext` framework, which permits external schedulers to be loaded without modifying the core kernel [14]. Hypervisors lacking comparable extensibility – such as Xen, VMware ESXi, or Microsoft Hyper-V – necessitate direct modification of the scheduler code, although the required changes remain confined to the scheduling path and do not affect device drivers or memory management [15].

The logic supports two operating modes. First, it can function as a standalone scheduler that assigns vCPUs solely on the basis of I/O classification. Second, it can act as a refinement stage atop an existing real-time scheduler (e.g., Earliest Deadline

First), reordering the run queue to prevent concurrent execution of vCPUs with matching I/O profiles and thus minimizing interference while preserving deadline guarantees.

Effective classification depends on visibility into each VM's internal run queue. To provide this information, every guest transmits a compact summary of its runnable tasks to the hypervisor via a dedicated hypercall or paravirtual channel. Implemented as a small guest-kernel module, this mechanism imposes no changes on user-space applications and can be shipped alongside standard paravirtualization drivers [16].

Because the classification and mapping occur only when a vCPU becomes ready, the additional computational burden is negligible, making the approach suitable for resource-constrained industrial hosts where deterministic timing and minimal overhead are paramount.

IV. Experimental Setup and Simulation

To evaluate the effectiveness and timing behavior of the proposed interference-aware scheduling concept, a controlled simulation environment was developed that allows systematic analysis under reproducible conditions.

A. Simulation Environment

The simulation was implemented in a Python-based Jupyter Notebook environment. The scheduler was developed as a custom computation that calculates the run queue assignments for all virtual CPUs based on predefined workload scenarios. These run queues represent the scheduling decisions over time and were subsequently used as input for a discrete-event simulation implemented with the SimPy framework. In this setup, SimPy emulates the execution of the virtual CPUs according to the generated schedule and enables measurement of timing-related metrics, such as waiting times and utilization. All experiments were conducted offline without deploying an actual hypervisor or virtual machines, allowing controlled and repeatable evaluation of the scheduling logic under synthetic conditions.

B. Scenarios and Assumptions

The simulation comprised a set of predefined scenarios with varying workload intensities, resource utilization patterns, and virtual machine configurations. For each scenario, synthetic datasets were generated to represent categorized vCPUs, including their expected resource demands and arrival times. It was assumed that all vCPUs were pre-classified and that the system operated under ideal conditions without allocation delays or interference between components. Resources were modeled deterministically, with fixed maximum capacities and no variability due to physical hardware behavior or contention effects. The main objective of this simulation was to validate the feasibility of the proposed scheduling approach and to provide initial performance insights under controlled conditions. Due to these simplifications, results should be interpreted as indicative rather than fully representative of complex real-world environments.

V. Results and Interpretation

The subsequent analysis summarizes the outcomes of the conducted simulation experiments, highlighting key behavioral differences between the baseline and the optimized scheduling strategies.

A. Scenario Overview and Scheduling Behavior

Figure 4 illustrates the execution timeline of the baseline scheduling strategy, where vCPUs are assigned to the shortest available run queue without considering their expected runtimes or I/O dependencies. In this scenario, all physical CPU cores initially process tasks in a balanced manner, resulting in nearly synchronous task completions across the cores. However, during execution, a pronounced idle period occurs on a single CPU core that must wait for a shared I/O operation to complete before further processing can continue. This blocking leads to an extended idle phase on that core and increases the total processing time for the workload.

Additionally, Figure 5 illustrates the utilization of the I/O components observed during the simulation of the same baseline execution. The diagram highlights a specific time interval between 13 and 16 time units, where the GPU utilization temporarily exceeds 100% due to concurrent access from multiple tasks. This overcommitment results in contention for the shared GPU resource, causing blocking delays that propagate back to the scheduling timeline and extend the overall processing time. The example demonstrates that purely queue-length-based scheduling not only produces unpredictable idle periods but also leads to excessive I/O load peaks that further degrade system performance and determinism.

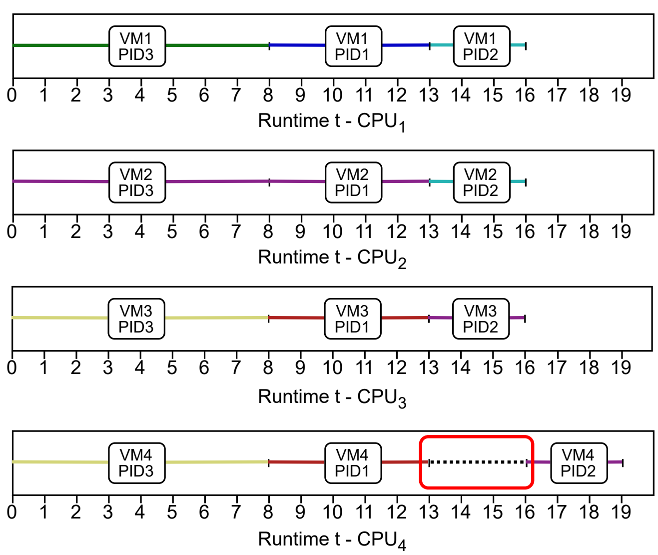

Figure 4. Scheduling Timelines without Optimization.

Figure 6 shows the execution timeline obtained with the proposed scheduling approach, where overlapping execution of equally categorized tasks is explicitly avoided. In this configuration, the scheduler assigns vCPUs so that tasks of

 ISBN: 978-1-68558-316-3

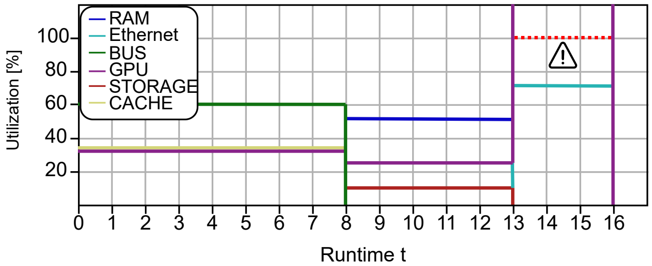

Figure 5. I/O Component Utilization after Simulation without Optimization.

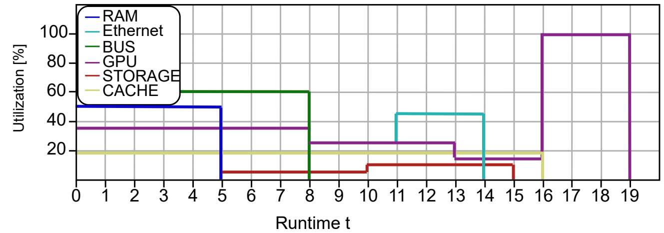

Figure 7. I/O Component Utilization after Simulation with 0% Overlapping Optimization.

the same category do not run concurrently on different cores, thereby preventing the I/O blocking effects observed in the baseline scenario. As a result, no idle periods occur during execution, and the overall processing time is reduced. However, this strict separation also leads to a less uniform workload distribution across cores, as visible in the timeline. While this setup demonstrates the feasibility of deterministic, non-overlapping scheduling, the approach can be relaxed to allow controlled overlap between task categories, providing additional flexibility to balance I/O and CPU utilization more evenly if required.

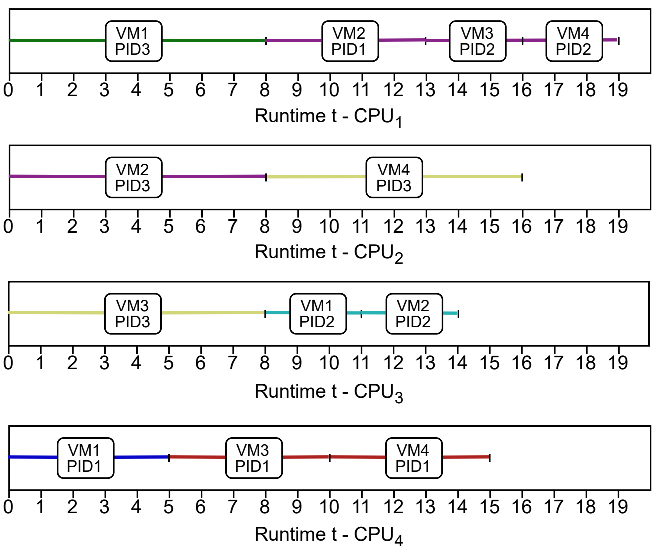

Figure 6. Scheduling Timelines with 0% Overlapping Optimization.

Additionally, Figure 7 illustrates the I/O utilization observed during the simulation of the optimized scheduling scenario. In contrast to the baseline case, no overcommitment beyond 100% occurs, confirming that the separation of I/O categories effectively reduces contention and stabilizes resource usage over time.

B. Quantitative Metrics and Performance Comparison

To objectively evaluate the effectiveness of the proposed vCPU scheduling optimization, a set of quantitative utilization and CPU load metrics was collected before and after the optimization. The components were categorized as RAM, ETHERNET, BUS, GPU, STORAGE, and CACHE. The results demonstrate several significant improvements.

First, the optimization led to a more constant utilization of critical I/O components over time. For example, the average utilization of ETHERNET decreased from 12.35% to 10.5%, while CACHE utilization was reduced from 16.0% to 13.6%. This more even distribution of load helps prevent unpredictable fluctuations and enables better planning of system resources.

Second, the optimization effectively reduced utilization peaks. The maximum utilization of ETHERNET dropped from 70% to 45%, representing a reduction of more than one third, while CACHE gets its maximum utilization cut by half, from 34% to 17%. For STORAGE, the maximum utilization was also reduced by approximately 33%. By lowering these peaks, the system achieves a smoother and more predictable workload profile, which is particularly important for time-sensitive applications.

Third, the optimization ensured that no component exceeded 100% utilization at any time. Before the optimization, GPU occasionally reached utilization peaks of up to 113%, indicating that tasks temporarily demanded more I/O capacity than was available, which resulted in waiting times and delays. After the optimization, all components remained consistently below 100% utilization, preventing overload conditions and eliminating unnecessary queuing of I/O operations.

In addition to improvements in I/O utilization, the distribution of CPU load across cores became more balanced. While the CPU loads were initially nearly identical across cores, but after optimization, the loads were more differentiated. Although this led to slightly differing completion times for the individual CPU cores in this synthetic example, this effect is not critical in real-world applications. In practical scenarios, there is a continuous inflow of new tasks, so the timing of core idle phases becomes irrelevant. The system benefits far more from the improved predictability and absence of overload situations than it is impacted by minor variations in per-core runtime.

Overall, the results clearly show that the optimization keeps component utilization more constant, reduces peak loads, prevents overload conditions, and distributes CPU workloads more evenly. This combination significantly increases the stability and responsiveness of the system without introducing adverse side effects for unaffected components, such as RAM or BUS.

C. Implications for Industrial Deployments

The presented optimization is highly relevant for industrial control environments, where virtualized systems must deliver consistent performance and comply with strict timing requirements.

By ensuring that critical I/O resources, such as Ethernet interfaces, storage subsystems, and GPU accelerators remain reliably below full utilization, the approach effectively prevents situations where tasks are forced to wait due to resource contention. This directly supports predictable cycle times, which are essential for machine control and safety-related processes.

The increased stability of resource usage also simplifies planning and verification against industrial standards, reducing the need for oversized hardware reserves and enabling more efficient system designs.

In real-world deployments, minor differences in CPU completion times, as observed in synthetic tests, have no practical impact, since industrial workloads are typically characterized by continuous streams of tasks. Under these conditions, the advantages of smoother utilization profiles and the elimination of overload situations clearly outweigh any variations in per-core timing, resulting in higher system availability and more robust operation under changing load conditions.

Moreover, the more balanced distribution of CPU load contributes to improved thermal behavior and can help extend the lifespan of hardware components, which is an important factor in embedded and industrial-grade platforms. Overall, the optimization provides a practical means of enhancing determinism, efficiency, and resilience in virtualized industrial environments.

VI. Conclusion and Future Work

This paper presented an interference-aware scheduling approach based on paravirtualized VM profiling, designed to improve the determinism and predictability of virtualized industrial control systems. By classifying vCPUs according to their dominant I/O resource usage and preventing the concurrent execution of equally categorized tasks, the proposed method effectively reduced utilization peaks and eliminated overload conditions that often lead to unpredictable latencies.

Experimental evaluation under synthetic conditions demonstrated that the optimization can maintain consistently lower maximum utilization across critical components, such as Ethernet, storage, and GPU, while achieving a smoother distribution of workload over time. Although slight variations in per-core completion times were observed, these effects are negligible in real-world industrial environments where continuous task streams are common.

Future work will focus on extending the approach beyond offline simulation and integrating the scheduler into production-grade hypervisors to validate its effectiveness under real workloads and mixed I/O patterns. Additionally, further research will explore adaptive scheduling strategies that dynamically adjust the degree of task separation based on system load and application criticality. Investigating the impact of the approach on power consumption, thermal behavior, and long-term hardware reliability in embedded industrial platforms also represents an important direction for future studies.

References

[1] R. Stark, *Virtual Product Creation in Industry*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2022, ISBN: 978-3-662-64299-3. DOI: 10.1007/978-3-662-64301-3.

[2] V. Meyer, U. L. Ludwig, M. G. Xavier, D. F. Kirchoff, and C. A. F. De Rose, “Towards interference-aware dynamic scheduling in virtualized environments”, in *Job Scheduling Strategies for Parallel Processing*, D. Klusáček, W. Cirne, and N. Desai, Eds., Cham: Springer International Publishing, 2020, pp. 1–24.

[3] C. Serôdio, P. Mestre, J. Cabral, M. Gomes, and F. Branco, “Software and architecture orchestration for process control in industry 4.0 enabled by cyber-physical systems technologies”, *Applied Sciences*, vol. 14, no. 5, 2024, ISSN: 2076-3417. DOI: 10.3390/app14052160.

[4] J. Müller, M. Giani, D. Deubert, and M. Massoth, “Virtualization in industrial production - a survey focusing on virtual and virtualized industrial controls”, in *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2024, pp. 1–7. DOI: 10.1109/ETFA61755.2024.10710668.

[5] M. Gundall, C. Glas, and H. D. Schotten, “Feasibility study on virtual process controllers as basis for future industrial automation systems”, in *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, vol. 1, 2021, pp. 1080–1087. DOI: 10.1109/ICIT46573.2021.9453651.

[6] S. Dietrich, G. May, O. Wetter, H. Heeren, and G. Fohler, “Performance indicators and use case analysis for wireless networks in factory automation”, in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017, pp. 1–8. DOI: 10.1109/ETFA.2017.8247605.

[7] S. Xi, J. Wilson, C. Lu, and C. Gill, “Rt-xen: Towards real-time hypervisor scheduling in xen”, in *Proceedings of the Ninth ACM International Conference on Embedded Software*, ser. EMSOFT ’11, Taipei, Taiwan: Association for Computing Machinery, 2011, pp. 39–48, ISBN: 9781450307147. DOI: 10.1145/2038642.2038651.

[8] V. Roy, “A context-aware internet of things (iot) founded approach to scheming an operative priority-based scheduling algorithms”, *Journal of Cybersecurity and Information Management*, vol. 13, pp. 28–35, Jan. 2024. DOI: 10.54216/JCIM.130103.

[9] M. A. Altahat, K. Mhaidat, and O. Al-Khaleel, “Quantitative analysis of hypervisor efficiency and energy consumption in heterogeneous multi-vm environments with varied server workloads”, *Simulation Modelling Practice and Theory*, vol. 141, p. 103 102, 2025, ISSN: 1569-190X. DOI: https://doi.org/10.1016/j.simpat.2025.103102.

[10] E. Jeannot, G. Pallez, and N. Vidal, “Io-aware job-scheduling: Exploiting the impacts of workload characterizations to select the mapping strategy”, *The International Journal of High Performance Computing Applications*, vol. 37, no. 3-4, pp. 213–228, 2023. DOI: 10.1177/10943420231175854. eprint: https://doi.org/10.1177/10943420231175854.

[11] J. Peixoto, J. Martins, D. Cerdeira, and S. Pinto, “Birtio: Virtio for real-time network interface sharing on the bao hypervisor”, *IEEE Access*, vol. 12, pp. 185 434–185 447, 2024. DOI: 10.1109/ACCESS.2024.3512777.

[12] N. Borgioli *et al.*, “An i/o virtualization framework with i/o-related memory contention control for real-time systems”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 11, pp. 4469–4480, 2022. DOI: 10.1109/TCAD.2022.3202434.

[13] U. Bilgen, E. Y. Tat, and M. ErelÖzçevik, “A novel digital twin enabled weighted task optimization framework for software defined data centers in industry 4.0”, in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, 2025, pp. 1–6. DOI: 10.1109/WCNC61545.2025.10978632.

[14] T. L. K. Community, *Sched_ext – bpf-based scheduler class*, https://docs.kernel.org/scheduler/sched-ext.html, Accessed: 2025-09-12, 2024.

[15] N. Kraljevic, B. Djordjevic, and V. Timcenko, “File system performance comparison in full hardware virtualization with esxi, kvm, hyper-v and xen hypervisors”, *Advances in Electrical and Computer Engineering*, vol. 21, no. 1, pp. 7–14, 2021. DOI: 10.4316/AECE.2021.01002.

[16] K. T. Raghavendra, S. Vaddagiri, N. Dadhania, and J. Fitzhardinge, “Paravirtualization for scalable kernel-based virtual machine (kvm)”, in *2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, 2012, pp. 1–5. DOI: 10.1109/CCEM.2012.6354619.

Q-Learning Performance on the CartPole Environment Under Observation Noise and Reward Variants

Steven Ren
Department of Computer Science
University of Massachusetts Amherst
Amherst, Massachusetts, USA
email: skren@umass.edu

Taieba Tasnim
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
email: ttasnim6386@tuskegee.edu

Berkeley Wu
Auburn City School
Auburn, Alabama, USA
email: tulipfan002@hotmail.com

Mohammad Rahman
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
email: mrahman@tuskegee.edu

Fan Wu
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
email: fwu@tuskegee.edu

***Abstract*—This paper evaluates Q-learning performance in the CartPole reinforcement learning environment under varying levels of observation noise and two distinct reward functions, in the broader context of designing robust learning-based controllers for cyber-physical systems. Specifically, we compare the standard step-based reward with a cosine-based reward designed to encourage upright pole balance. Observation noise is modeled as Gaussian noise, with standard deviations scaled to the range of each observation variable. Through multiple training runs at different noise levels, we evaluated convergence behavior, pole angle stability, and cumulative rewards. Our results show that observation noise significantly impairs learning under standard reward, whereas cosine-based reward improves robustness and promotes more stable policies. By linking reinforcement learning with noise-robust control design, this work directly contributes to the understanding of Q-learning under noisy environments and represents a step toward applying reinforcement learning to real-world cyber-physical systems, where noise and variability are inherent.**

***Keywords-reinforcement learning; q-learning; noise; reward; cyber-physical systems*.**

I. Introduction

The CartPole system is a classic reinforcement learning environment that is commonly used to benchmark various algorithms. In this research, the open-source Python library Gymnasium, developed by the Farama Foundation, is used to implement the CartPole environment [1]. The CartPole system—also known as the inverted pendulum—is a fundamental control problem used to test reinforcement learning algorithms. While much prior work has demonstrated successful applications of reinforcement learning algorithms to CartPole [2], real-world factors such as sensor noise and the design of reward functions have been less explored. This paper studies the impact of additive observation noise and shaped reward functions on Q-learning convergence and policy behavior.

The main contributions of this work are:

- Application of a Q-learning algorithm to CartPole under noisy observation inputs.
- Comparison of a standard reward function with a cosine-based reward function shaped by the pole angle.
- Evaluation of convergence episodes, pole angle statistics, and performance variance across noise levels.

The remainder of the paper is organized as follows. Section II provides a background on Q-learning, the epsilon-greedy policy, and CartPole. Section III outlines the methodology, including observation boundaries, noise modeling, reward functions, and training steps. Section IV presents and discusses the experimental results. Section V concludes the research with a summary and directions for future work.

II. Literature review

Q-learning, a concept first introduced by Watkins more than three decades ago, values delayed rewards in reinforcement learning [3]. It operates by estimating the optimal action-value function and aims for long-term reward maximization without requiring a model of the environment. The epsilon-greedy algorithm is a simple and commonly used method in reinforcement learning that attempts to balance exploration and exploitation [4]. In recent years, reinforcement learning has found increasing application in control problems, particularly in robotics and other cyber-physical systems where adaptive behavior is essential [5]. Q-learning, due to its simplicity and ability to handle discrete actions, has been successfully applied in robotic navigation and control [6]. Additional advancements have been made on top of the original Q-learning function, such as Efficient Q-learning, which improves computation through newly defined state and action spaces, a new reward function, and an optimized selection strategy [7]. The Deep Q-learning algorithm also extends from Q-learning by using a deep neural network to approximate the action-value function. With certain modifications, it has been applied to efficiently solve two-player zero-sum Markov games [8], in addition, it performs with good stability and optimality [9].

Above are previous studies of Q-learning and their various applications. More recently, researchers have also turned their attention to the robustness of reinforcement learning methods in noisy and uncertain environments, particularly in cyber-physical systems where safety is critical. A study by Krish et al. on observation noise robustness utilizes a tree-based algorithm for neural network control systems to identify the smallest amount of observation noise that can cause the neural network-based controller to violate safety constraints. They apply the algorithm on several systems such as Gymnasium's CartPole and LunarLander, along with two aircraft systems [10]. In another study, Nazrul demonstrates how reinforcement learning can be applied to optimize sampling frequency in cloud-based cyber-physical systems, enabling dynamic adjustment based on network conditions and system state. In a vehicle cruise control case, this approach outperformed fixed sampling strategies by balancing control performance with network efficiency [11].

In context of exploring how Q-learning performs under noisy environments, this paper will also briefly introduce another popular reinforcement learning algorithm, SARSA (State-Action-Reward-State-Action), which will serve as a baseline for comparison. The key difference between the two algorithms is that while Q-learning learns the value of the optimal policy, SARSA learns the value of the current policy being followed [12][13]. Details of the key terms mentioned here are explained in the following background section.

SUMMARY OF NOTATION

θ Pole angle in radians.
$\overline{\theta}$ Mean of the pole angle over an episode.
$\mathrm{Var}(\theta)$ Variance of the pole angle over an episode.
r Reward given to the agent.
γ Discount factor applied to future rewards.
ϵ Probability of taking a random action in ϵ-greedy policy.
$Q(s,a)$ Estimated value of taking action a in state s.
α Learning rate for Q-value updates.
n Number of steps within an episode.
$\tilde{o}_i$ Noisy observation values for observation i.
o_i True observation values for observation i.
σ_i^2 Variance of the Gaussian noise added to o_i.

III. SYSTEM AND PROBLEM FORMULATION

The CartPole system is a simulation used to solve the cart-pole problem, described as: “A pole is attached by an unactuated joint to a cart, which moves along a frictionless track. The pendulum is placed upright on the cart and the objective is to balance the pole by applying forces in the left and right directions on the cart” [2]. In Gymnasium's implementation, the agent is rewarded for each step taken while the pole remains upright. The environment terminates when the pole falls beyond a threshold or the cart moves out of bounds. An episode is defined as a sequence of actions that begins with a reset and ends with termination—either by failure or upon reaching the maximum of 500 steps. The maximum achievable reward per episode is 500, which serves as the convergence threshold.

The goal of this work is to train a reinforcement learning model using Q-learning to solve the CartPole system of balancing a pole in the presence of observation noise and then analyze the impact of noise and reward choice on performance. The system receives continuous observation values for cart position, cart velocity, pole angle, and pole angular velocity, which are subject to additive Gaussian noise to simulate real-world inaccuracies. These noisy observations are then discretized to define a finite set of states. At each step, an action is selected to maximize the cumulative reward for an episode. Two reward functions are used along with different levels of noise, and the convergence behavior and pole stability are assessed to understand the impact of noise and reward on the learning process. Figure 1 is a block diagram showing the overall Q-learning CartPole system with noise.

Q-learning is a model-free reinforcement learning algorithm that learns action-value functions based on observed transitions [3]. The Q-function describes the Q-table, which holds all action-value pairs and their corresponding Q-values (a 1×2 array where index 0 represents the reward for the action "left" and index 1 represents the reward for the action "right"):

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\Big[r_{t+1} + \gamma \max_{a'} Q(s_{t+1},a') - Q(s_t,a_t)\Big] \quad (1)$$

Here, $Q(s_t,a_t)$ denotes the current estimate of the action-value function, the expected return of taking action a_t in state s_t at time step t. The parameter α is the learning rate, r_{t+1} is the reward received after taking action a_t, and γ is the discount factor. The term $\max_{a'} Q(s_{t+1},a')$ denotes the maximum predicted future reward obtainable from the next state s_{t+1} over all possible actions a'.

The Q-learning update rule can be interpreted as follows: take the current Q-value for this state-action pair and update it using the immediate reward just received, plus the best Q-value expected from the next state [11]. The learning rate α determines how strongly this new estimate influences the update, while the discount factor γ controls the importance given to future rewards.

To balance the trade-off between exploration (trying new or less-used actions) and exploitation (choosing the best-known action), we apply the epsilon-greedy policy, which helps choose the action based on current state [14], defined as:

$$a(s) = \begin{cases} \text{random action}, & \text{if } \varepsilon > \text{rand}() \\ \arg\max_a Q(s,a), & \text{otherwise} \end{cases} \quad (2)$$

Here, ε is the epsilon value, a probability between 0 and 1 that determines the chance of choosing a random action, and it gradually decays over time toward a small constant. The function rand() represents a randomly sampled float from a uniform distribution over the interval $[0,1]$. The expression $\arg\max_a Q(s,a)$ denotes the action that currently has the highest Q-value for the state s. This exploration policy ensures sufficient exploration during early training episodes, while gradually favoring the exploitation of the learned Q-values as training progresses [15].

IV. METHODOLOGY

This section outlines the approach used to evaluate Q-learning performance, including the environment setup, state and action representations, and implementation details of the learning process.

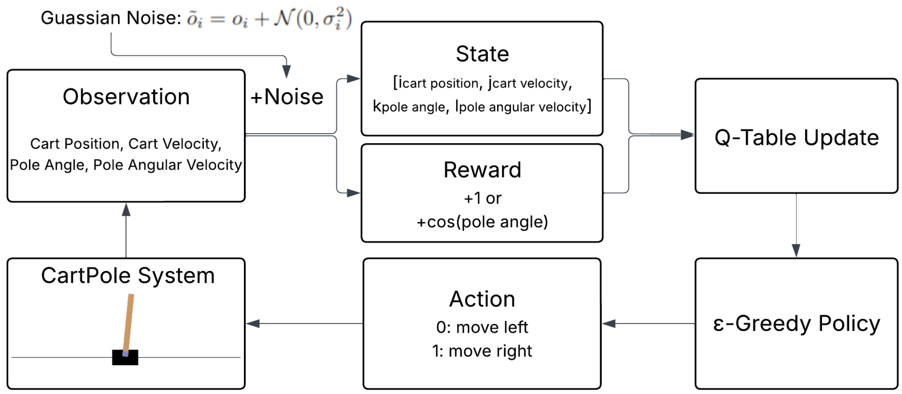

Figure 1. Block diagram of Q-learning CartPole system with noise.

A. Environment and Observations

The CartPole environment is implemented using the Gymnasium library developed by the Farama Foundation [1]. The system simulates a cart moving along a one-dimensional, frictionless track with a pole attached to it via an unactuated hinge joint. The agent receives observations in the form of a four-dimensional state vector: cart position x, cart velocity v, pole angle θ, and pole angular velocity ω.

Variables are continuous and bounded within defined limits:

TABLE I. OBSERVATION SPACE RANGES OF CARTPOLE ENVIRONMENT

Observation	Range
Cart Position (x)	$[-4.8, 4.8]$
Cart Velocity (v)	$[-5.0, 5.0]$
Pole Angle (θ)	$[-0.418, 0.418]$ radians
Pole Angular Velocity (ω)	$[-10.0, 10.0]$

Note that the observation space here differs from Gymnasium's original infinite range for Cart Velocity and Pole Angular Velocity. A limitation of this environment is its discrete action space, restricted to two binary actions: 0 for moving left and 1 for moving right. The velocity affected by the force applied to the cart is not fixed and depends on the pole's angle. We cannot directly specify a particular amount of force as an action [16].

B. Observation Noise

To simulate imperfect sensor measurements encountered in real-world systems, additive Gaussian noise is applied to each component of the observation vector:

$$\tilde{o}_i = o_i + \mathcal{N}(0, \sigma_i^2) \quad (3)$$

Here, o_i represents the true observation, and σ_i is the standard deviation of the noise applied to observation i, proportional to the variable's range. This noise is injected before state discretization, meaning it may cause the agent to misclassify its current state. Several noise levels are tested—specifically, 0.0 (no noise), 0.01, 0.05, and 0.1—to evaluate their effect on learning performance and control stability.

C. State Discretization

Since Q-learning operates on a discrete state space, each continuous observation variable is divided into a fixed number of bins. These bins are uniformly spaced within each variable's range. The state is encoded as a tuple of discretized indices corresponding to the binned values of cart position, cart velocity, pole angle, and pole angular velocity. The combination of these indices uniquely identifies a state in the Q-table. In this paper, we use 8 bins for cart position, 8 bins for cart velocity, 20 bins for pole angle, and 20 bins for pole angular velocity. Note that a larger number of bins sharply increases computational complexity [17]. This discretization reduces the infinite continuous observation space to a manageable number of discrete states, at the cost of precision. Observation noise can cause transitions between neighboring bins, introducing non-determinism into state transitions.

D. Reward Functions

In this work, two reward functions are evaluated:

1) Default Reward: A constant reward of +1 is given at each step as long as the pole remains upright and the cart stays within bounds. This is the default reward under the gymnasium environment.

2) Cosine-Based Reward: The reward is defined as:

$$r = \cos(\theta) \quad (4)$$

This function rewards the agent more when the pole angle θ is near vertical ($\theta = 0$) and penalizes deviations from the position. Since $\cos(0) = 1$, this function shapes the agent's behavior toward learning actions that minimize pole deviation, instead of just surviving.

E. Training Details

All reward function and noise level combinations are trained over 10,000 episodes, with each episode capped at 500 steps. The Q-learning hyperparameters used are:

- Learning rate (α) = 0.1
- Discount factor (γ) = 0.95
- Epsilon (ε) starts at 1.0 and decays exponentially to a minimum of 0.001

These parameters were found to perform well in the local environment: a Windows laptop with modern CPU and GPU, though they can be adjusted based on performance goals.

For each episode, the following statistics are recorded:

- Total reward: Sum of rewards per episode
- Pole angle mean and variance: Metrics that show how well Q-learning stabilizes the pole

Pole angle mean formula:

$$\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i \tag{5}$$

Pole angle variance formula:

$$\mathrm{Var}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \overline{\theta})^2 \tag{6}$$

Here, n is the number of steps in the episode, and θ_i is the pole angle at step i.

F. SARSA Baseline

A baseline comparison using the SARSA algorithm is run under the same settings as the Q-learning CartPole system, with the function being:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\Big] \tag{7}$$

At each time step, the agent updates its action-value estimate $Q(s_t, a_t)$ based on the actual reward received, the next observed state s_{t+1}, and the next action a_{t+1} selected by the current policy. The SARSA update is policy-dependent, as the learned values directly reflect the behavior policy being followed, including any exploration strategy. The same data as Q-learning is collected, allowing for a direct baseline comparison. This enables an assessment of how the off-policy approach of Q-learning influences learning performance relative to the on-policy nature of SARSA under noisy observations.

V. Results and Discussion

The bar graph (Figure 2) shows the number of episodes required to reach convergence—defined here as achieving a total reward of 500—in the CartPole environment across different combinations of reward functions and observation noise levels. A maximum of 10,000 episodes was allowed, with bars reaching that value indicating non-convergence within the limit. We can observe that, for a default reward, only the training with no noise successfully converges within the 10,000 episode limit. However, the cosine reward function, which penalizes larger pole angles, shows the ability to converge at noise levels up to 0.01. This suggests that the cosine reward function can offer an improved Q-learning experience and encourage more stable control behavior, allowing for the CartPole system to stay upright.

The two box plots (Figure 3 and Figure 4) show the mean and variance of pole angles across episodes for different combinations of noise levels and rewards. For example, in the mean pole angle plot for cosine reward, a single dot represents the mean pole angle over all steps taken within one episode.

For the default reward function, the mean pole angle remains close to zero when there is no noise, indicating that the pole stays centered. However, as the noise level increases to 0.01 and beyond, the mean pole angle shifts and becomes more

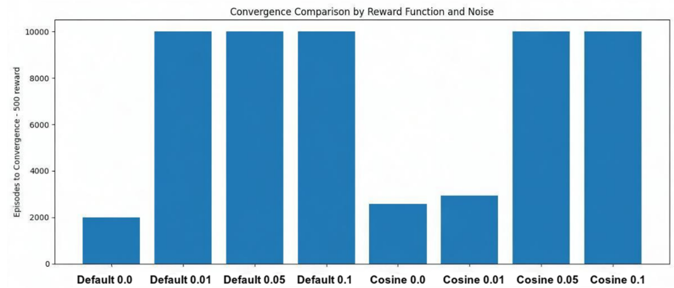

Figure 2. Q-Learning bar plot comparing episodes to convergence of different reward function and noise level combinations.

spread out, which is an expected behavior. This suggests that the training struggles to maintain balance under noisy conditions. The corresponding variance plots further reinforce this observation, showing a notable increase in pole angle variance with rising noise levels. Specifically, the median variance increases and the spread widens, indicating more frequent and extreme pole oscillations during training.

In contrast, the cosine reward function exhibits much better performance. Less outliers are observed at 0.0 noise level demonstrating the cosine reward's ability to promote steadier control even in uncertain environments.

As for the SARSA algorithm, the bar graph (Figure 5) shows the number of episodes required to reach convergence and the two box plots (Figure 6 and Figure 7) show the mean and variance of poles angles across episodes for different combinations of noise levels and rewards just like the Q-learning figures. The SARSA bar graph can be seen to have a similar points of convergences as the Q-learning bar graph. Also, similar to Q-learning, it can be seen that under cosine reward, the variance is more consistent across noise levels.

Overall, these plots show that for Q-learning the default reward function leads to unstable learning in the presence of noise, while the cosine reward function encourages more stable and consistent control. This aligns well with the convergence analysis, where the cosine reward enabled convergence at the 0.01 noise level, in contrast to the lack of convergence when noise was added under the default reward training. The CartPole system can be seen to behave similarly under the SARSA algorithm. These results demonstrate that careful reward design—such as using a cosine-based function that penalizes large pole angles—can improve robustness in reinforcement learning for the CartPole environment.

VI. Conclusion and Future Work

This paper explored the impact of observation noise and reward function design on the performance of Q-learning in the CartPole reinforcement learning environment and its relevance to cyber-physical systems. Our results demonstrate that observation noise significantly affects the stability and reliability of convergence. When the default reward function was used, even small amounts of noise impaired learning and control performance. In contrast, the cosine reward function

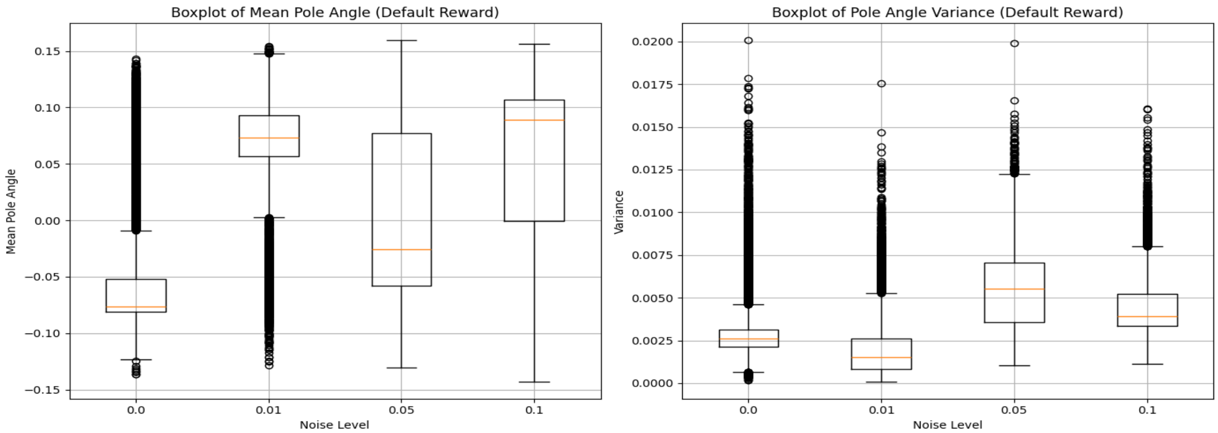

Figure 3. Q-learning box plot of default reward pole angle mean and variance.

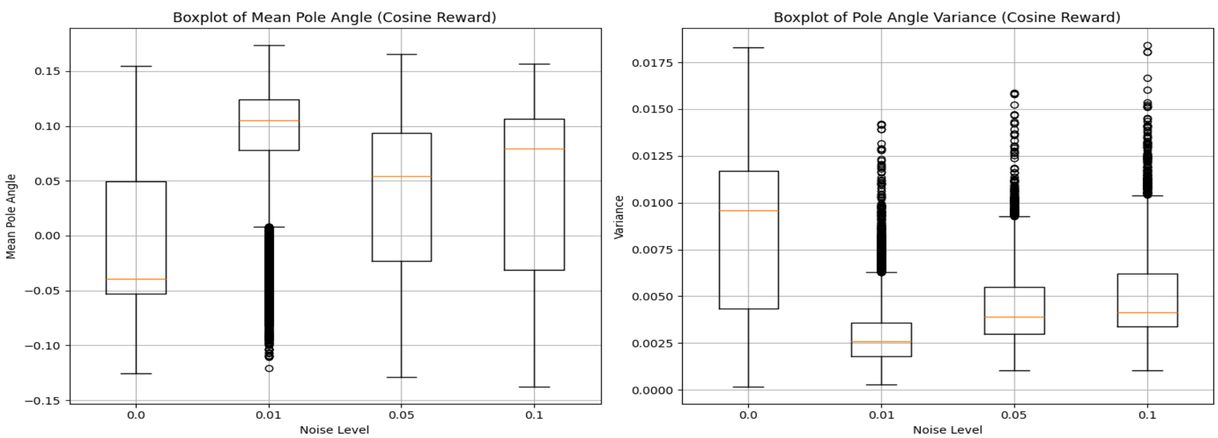

Figure 4. Q-learning box plot of cosine reward pole angle mean and variance.

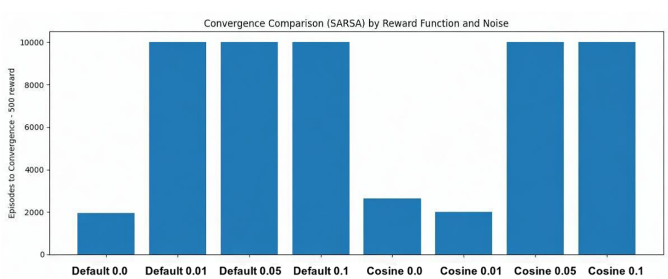

Figure 5. SARSA bar plot comparing episodes to convergence of different reward function and noise level combinations.

showed improvement in robustness, guiding the system to learn more stable policies under the noisy conditions.

Future work should extend this investigation beyond simulation by applying the experimental setup to a real-world physical system, where noise and variability are inherent and unavoidable. This would validate whether the observed benefits of different rewards translate into performance on real hardware.

Additionally, since this work used tabular Q-learning with discretized state spaces, a future direction is to examine how such methods can generalize to more complex or continuous environments. Although discretization provides interpretability and simplicity, it is often limited in scalability. Extending this framework using neural networks could bridge the tabular approach and deep reinforcement learning, enabling policies learned in idealized environments such as CartPole to generalize more effectively to higher-dimensional control tasks.

Finally, another promising direction is to develop or integrate noise detection and filtering techniques to help the system adapt its reinforcement learning process under uncertainty. Exploring combinations of reward functions, noise adaptation, and learning strategies can offer new insights into building intelligent, robust, and fault-tolerant cyber-physical systems capable of operating effectively in complex real-world environments.

Acknowledgment

The work is supported in partial by the National Science Foundation under NSF grants #2417608, and #2234911, #2209637, #2131228, and #2100134, Any opinions, findings, recommendations, expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

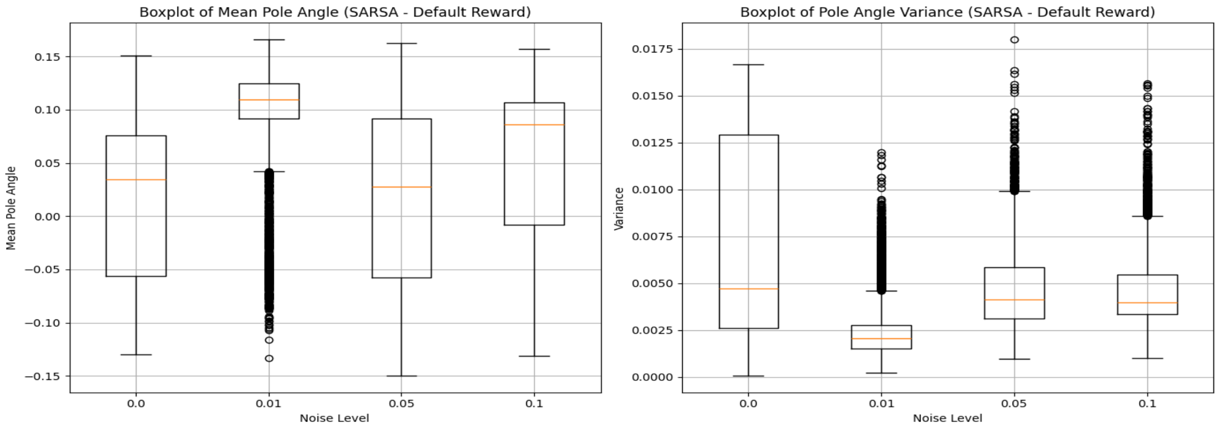

Figure 6. SARSA box plot of default reward pole angle mean and variance.

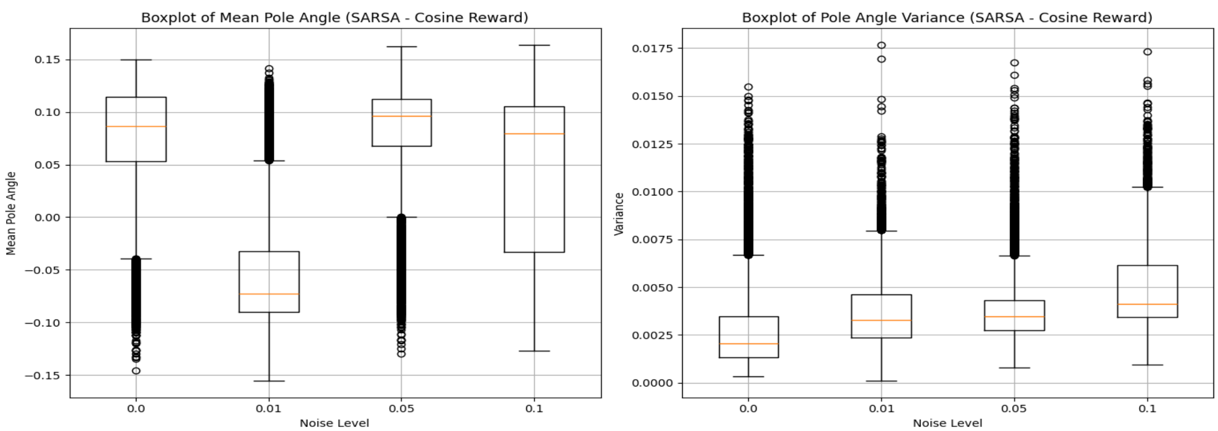

Figure 7. SARSA box plot of cosine reward pole angle mean and variance.

REFERENCES

[1] M. Towers *et al.*, “Gymnasium: A Standard Interface for Reinforcement Learning Environments,” *arXiv*, 2024. Available: https://arxiv.org/abs/2407.17032.

[2] S. Kumar, “Balancing a CartPole System with Reinforcement Learning - A Tutorial,” *arXiv*, 2020. Available: https://arxiv.org/abs/2006.04938.

[3] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, pp. 279–292, 1992. doi: 10.1007/BF00992698.

[4] Q. Nguyen, N. Teku, and T. Bose, “Epsilon Greedy Strategy for Hyper Parameters Tuning of A Neural Network Equalizer,” in *Proc. 2021 12th Int. Symp. on Image and Signal Processing and Analysis (ISPA)*, Zagreb, Croatia, 2021, pp. 209–212. doi: 10.1109/ISPA52656.2021.9552055.

[5] B. Singh, R. Kumar, and V. P. Singh, “Reinforcement learning in robotic applications: a comprehensive survey,” *Artificial Intelligence Review*, vol. 55, pp. 945–990, 2022. doi: 10.1007/s10462-021-09997-9.

[6] K. Zhu and T. Zhang, "Deep reinforcement learning based mobile robot navigation: A review," in Tsinghua Science and Technology, vol. 26, no. 5, pp. 674-691, Oct. 2021, doi: 10.26599/TST.2021.9010012.

[7] A. Maoudj and A. Hentout, "Optimal path planning approach based on Q-learning algorithm for mobile robots," Optimal path planning approach based on Q-learning algorithm for mobile robots, Applied Soft Computing, Volume 97, Part A,2020, doi: 10.1016/j.asoc.2020.106796.

[8] J. Fan, Z. Wang, Y. Xie, and Z. Yang, "A Theoretical Analysis of Deep Q-Learning," *arXiv*, 2020, Available: https://arxiv.org/abs/1901.00137.

[9] N. Gholizadeh, N. Kazemi, and P. Musilek, "A Comparative Study of Reinforcement Learning Algorithms for Distribution Network Reconfiguration With Deep Q-Learning-Based Action Sampling," in IEEE Access, vol. 11, pp. 13714-13723, 2023, doi: 10.1109/ACCESS.2023.3243549.

[10] V. Krish, A. Mata, S. Bak, K. Hobbs, and A. Rahmati, “Provable observation noise robustness for neural network control systems,” Research Directions: Cyber-Physical Systems, vol. 2, p. e1, 2024. doi:10.1017/cbp.2023.5.

[11] S. Nazrul, M. Rahman, and F. Wu, "Reinforcement Learning for Adaptive Sampling: Managing Network Resources in Cloud-Based Cyber-Physical Systems," SoutheastCon 2025, Concord, NC, USA, 2025, pp. 1136-1141, doi: 10.1109/SoutheastCon56624.2025.10971596.

[12] F. AlMahamid and K. Grolinger, "Reinforcement Learning Algorithms: An Overview and Classification," 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), ON, Canada, 2021, pp. 1-7, doi: 10.1109/CCECE53047.2021.9569056.

[13] G. A. Rummery and M. Niranjan, “On-line Q-learning Using Connectionist systems,” Technical Report CUED/F-INFENG/TR 166, Nov. 1994. Available: https://www.researchgate.net/publication/2500611_On-Line_Q-Learning_Using_Connectionist_Systems. Accessed: September, 2025.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[15] J. Lin, Y. Ye, S. Li, H. Zhang, and P. Zhao, "Improving Exploration in Deep Reinforcement Learning for Incomplete Information Competition Environments," in IEEE Transactions on Emerging Topics in Computational Intelligence, doi: 10.1109/TETCI.2025.3555250.

[16] “CartPole Environment,” Gymnasium Documentation. [Online]. Available: https://gymnasium.farama.org/environments/classic_control/cart_pole. Accessed: September, 2025.

[17] D. Cao *et al.*, "Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review," in Journal of Modern Power Systems and Clean Energy, vol. 8, no. 6, pp. 1029-1042, November 2020, doi: 10.35833/MPCE.2020.000552.

Development of a WebRTC-Based Video Calling Application to Predict Quality of Service Patterns Using an Artificial Neural Network

Carlos Moreno, Ezequiel Frias
Communications Department
Central University of Venezuela
Caracas, Venezuela
email: carlos.m.moreno@ucv.ve,
ezequieljfrias20@gmail.com

Vinod Kumar Verma
Computer Science and Engineering
Sant Longowal Institute of
Engineering and Technology
Punjab, India
email: vinod5881@gmail.com

Eric Gamess
MCIS Department
Jacksonville State University
Jacksonville, Alabama, USA
email: egamess@jsu.edu

***Abstract*—The development of video-calling applications using Web Real-Time Communication (WebRTC) represents an efficient and modern solution for real-time communications, enabling the direct transmission of audio, video, and data between browsers with no need for additional plugins. This research aimed to design and develop a WebRTC-based video-calling application capable of predicting Quality of Service (QoS) patterns through the implementation of an Artificial Neural Network (ANN). The proposal focused on analyzing key indicators (e.g., latency, jitter, and packet loss) that play a critical role in shaping user-perceived quality. The development of the predictive model was performed by using a Recurrent Neural Network (RNN) of the Long Short-Term Memory (LSTM) type. To validate the solution, four representative scenarios were established: acceptable quality, moderate degradation, critical quality, and extreme conditions. The results demonstrated that the LSTM model successfully captured the temporal behavior of QoS metrics and generated predictions within acceptable ranges according to standards defined by specialized organizations and industry leaders. It is concluded that the integration of LSTM neural networks into WebRTC applications constitutes a viable and effective strategy to enhance proactive QoS management and optimize the end-user experience.**

Keywords-Quality of Service; WebRTC; Video-Calling; Neural Networks; Prediction.

I. INTRODUCTION

Web Real-Time Communication [1]-[6] (WebRTC) is a set of open-source emerging technologies and APIs that enable real-time, peer-to-peer communications (audio, video, and data) directly between web browsers and mobile applications. It does not require intermediaries, plugins, or external software, making it a cornerstone of modern, decentralized communication systems. Due to its low latency, WebRTC has permitted the development of many new applications, revolutionizing how people interact online. It is now present in the majority of video conferencing systems (e.g., Google Meet), live streaming platforms, VoIP services, collaborative workspaces, online education platforms, file sharing, and multiplayer gaming. The use of WebRTC in browser-to-browser applications is expanding significantly as demand for real-time communication on the web grows, due to its standardized APIs [1] (e.g., `getUserMedia`, `RTCPeerConnection`, and `RTCDataChannel`), versatility, cross-platform compatibility, mandatory encryption for all media and data, and native integration on modern web browsers (e.g., Chrome, Firefox, Edge, and Safari).

This work proposes to develop an intuitive user interface for a WebRTC-based video call application and to analyze Quality of Service (QoS) parameters extracted from packets collected. In addition, the study seeks to design and implement a neural network model capable of predicting QoS patterns based on collected data, followed by a rigorous evaluation of its predictive performance. By combining user interface development, protocol-level traffic analysis, and advanced deep learning techniques, this research provides a systematic framework for addressing QoS prediction in real-time communication systems. The proposed approach intends to enhance both the accuracy and reliability of service quality estimation, thereby contributing to the optimization of WebRTC-based video call applications.

Recurrent Neural Networks (RNNs) of type Long Short-Term Memory (LSTM) with multiple outputs were used in this work since they are designed to handle sequential or time-series data. Unlike traditional networks, RNNs have an internal memory allowing them to use information from previous inputs to influence current outputs. Multiple outputs are used because the QoS output variables are correlated.

The rest of this paper is organized as follows. Section II discusses several peer-reviewed literature works conducted within this area of research and the problem addressed in this work. Section III describes the methodology employed, while Section IV presents and analyzes the results. Finally, Section V concludes the paper and discusses possible future work.

II. RELATED WORK

The study of real-time communication systems and QoS prediction has been widely addressed in the last two decades. Several studies have investigated the likelihood of network underperformance, anomalies, and failures, as well as the possibility of improving the QoS by applying artificial intelligence techniques.

Since WebRTC is an emerging technology, it is not considered in most of the work done in this area so far. For example, the study in [7] performed anomaly detections in network traffic using different models such as Isolation Forest, Naïve Bayes, XGBoost, LightGBM, and SVM classification. The results revealed that some of these models exhibit impressive performance and accuracy, highlighting the strengths and limitations of each one. The authors suggested integrating deep learning techniques, such as convolutional and RNNs. Another significant contribution in the area came from Garcia and Salcedo [8], who developed a model for failure prediction in IP networks using Artificial Neural Networks (ANNs). The study focused on detecting LAN failures, such as timeouts and connection rejections, demonstrating that ANNs can significantly improve the accuracy of fault diagnosis. In [9], the authors proposed a QoS prediction model called Topology-Aware QoS-GRNN (TAQ-GRNN), which incorporates gated RNNs of LSTM type. Even if their model could be integrated into WebRTC, the authors did not consider this possibility. The authors of [10] chose six specific QoS/QoE metrics and extracted the associated values from a VoIP measurement campaign in an LTE-A environment, before employing a set of recurrent neural networks (simple RNN, LSTM, and GRU) to predict the behavior of the selected QoS/QoE metrics. Aziz, Ioannou, Lestas, Qureshi, Iqbal, and Vassiliou [11] proposed a prediction model for QoS by using an RNN to integrate a Bidirectional Long Short-Term Memory (BLSTM). It can predict the QoS-aware network traffic for over 13 hours with high accuracy. They compared the RNN-BLSTM with other algorithms (i.e., LSTM, ARIMA, SVM). Their architecture is suitable for 5G and 6G mobile networks. The work in [12] did another relevant investigation within the field of QoS and Deep Learning, with the classification of multimedia traffic by using Convolutional Neural Networks. The authors of [13] developed a model for Service QoS prediction based on feature Mapping and Inference. In [14], Gerard, Bonilla, Bentaleb, and Céspedes proposed a Machine Learning (ML) model to enhance Forward Error Correction (FEC) efficiency. According to their findings, it corrects up to 60% of errors and achieves 2.5 times better energy efficiency than standard WebRTC.

Some work has been done in the area with the use of WebRTC. For example, Google [15] has deployed ML-based Bandwidth Estimation (BWE) systems within WebRTC that utilize a combination of LSTM and dense layer architecture to process real-time statistics (e.g., RTT and packet loss). This architecture enables superior proactive congestion prediction, significantly reducing parameters such as video freezes and connection drop rates. Sakakibara, Ohzahata, and Yamamoto [16] validated the creation of highly accurate No-Reference (NR) Quality of Experience (QoE) models solely based on WebRTC client statistics (jitter and bandwidth). Their models offer computationally efficient Deep Neural Networks (DNNs) or Temporal Convolutional Networks (TCNs) suitable for client-side monitoring. A doctoral thesis from Bingol [17] studied the convergence of AI techniques and WebRTC to predict QoE indicators, as they are more representative of user satisfaction than QoS.

This work differs from other state-of-the-art approaches in several key aspects. First, our architecture is proposed to predict QoS in interactive video calls specifically, and not for streaming or other applications. Second, we initially establish a robust comparative methodology by evaluating four distinct RNNs (GRU single output, GRU multiple outputs, LSTM single output, and LSTM multiple outputs) against three crucial performance indicators (Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error) to select the optimal model for implementation. Third, both the training and the subsequent operational deployment of the application rely exclusively on real-world measurements captured under diverse and varying network congestion conditions. Fourth, by utilizing a NoSQL Firebase Firestore [18] database for WebRTC metrics, this architecture provides superior scalability and high throughput with optimized performance and low latency.

Given all the aspects discussed previously, in the state-of-the-art, it is evident that in networks, QoS has become a crucial aspect to ensure an optimal user experience. This implies that the services and applications in use operate constantly. To achieve the best quality, it is necessary to invest in high-quality network infrastructure and carry out network monitoring. However, it is also important to develop applications with advanced capabilities that enable the prediction of QoS patterns. These applications might include artificial neural networks.

Based on the findings presented in the state of the art, two research questions arise:

- Question 1: Which QoS metrics can be considered to measure, analyze, and predict QoS patterns?
- Question 2: Which specific type of neural network predicts better QoS in a WebRTC-based video call?

III. METHODOLOGY

In this section, the implemented methodology is described, which includes the definition and characterization of the four study scenarios and the three indicators, the development of the WebRTC-based application and its final recurrent neural network architecture, after evaluating four possible alternatives, as well as the operation of the predictive model.

A. *Establishment of Scenarios and Quality of Service Parameters*

In this subsection, the QoS parameters considered in WebRTC were identified and defined, establishing criteria and metrics for the evaluation. The parameters selected were (1) latency, (2) jitter, and (3) packet loss rate. Four scenarios were chosen according to Rec. ITU-T G.1010 [19] as specified in Table I.

TABLE I. SCENARIOS SELECTED FOR STUDY

Scenario	Bandwidth	Latency	PLR	Description
(1) Acceptable Quality	50 Mbps	20 ms	0%	Ideal
(2) Moderate Degradation	2 Mbps	100 ms	3%	Congestion
(3) Critical Quality	0.8 Mbps	200 ms	10%	Deficient
(4) Extreme Conditions	0.3 Mbps	500 ms	20%	Degraded

B. *Development of the User Interface*

The user interface was developed using JavaScript along with the React framework, which allowed for the creation of a dynamic, modular, and scalable web application. For the implementation of real-time video calls, the PeerJS [20] library was used. PeerJS is a JavaScript library built on top of WebRTC that simplifies Peer-to-Peer (P2P) data, audio, and video communication in web browsers.

C. *Extraction, Data Processing, and Pattern Analysis*

For the collection of real-time metrics related to QoS during video calls, the `getStats` [21] function provided by the WebRTC API was used. In order to evaluate the application's performance across various connectivity contexts, the Network Link Conditioner tool [22], available on macOS, was used.

Figure 1 depicts the testbed for measurements. The client with the Network Link Conditioner tool is connected to the Internet via the SimpleFibra provider, using a 400 Mbps fiber optic WAN access link. Internally, the WLAN connection is established through a Wi-Fi 5 (IEEE 802.11ac) network, operating on the 5 GHz band, channel 161, with an 80 MHz channel width. On the other hand, the remote client is connected to the Internet via the NetUno provider, also through a fiber optic link, with a bandwidth of 200 Mbps. In its WLAN, Wi-Fi 5 is also used on the 5 GHz band, channel 153, with an 80 MHz channel width.

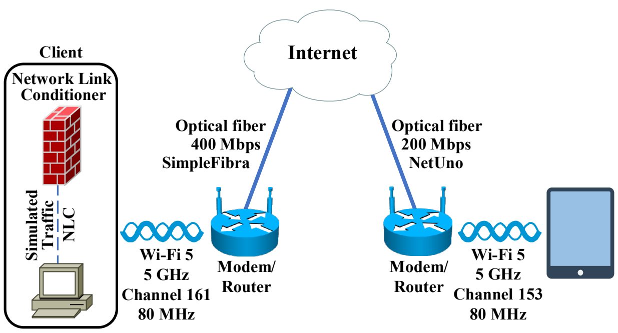

Figure 1. Testbed for Measurements

A mechanism was implemented to request a statistical report at 2-second intervals, in order to capture sudden variations in connection quality that might be overlooked with a longer interval. Then, based on the data obtained from each sample, structured JSON objects were built. Metrics collected for the object were: (1) timestamp, (2) jitterVideo, (3) jitterAudio, (4) roundTripTimeVideo, (5) roundTrip-TimeAudio, (6) packetsLostVideo, (7) PacketsLostAudio, (8) PacketsReceivedVideo, and (9) PacketsReceivedAudio. Using Formulas 1 and 2, the delay and packet loss rate were computed from the collected values.

$$delay = \frac{RoundTripTime}{2} \quad (1)$$

$$PacketLossRate = \frac{packetsLost}{packetsReceived + packetsLost} \times 100 \quad (2)$$

After constructing the metrics object, data were transmitted and stored in a Firebase Firestore [18] database (a cloud-based NoSQL database). To normalize the selected variables, the Python `MinMaxScaler` method from the `sklearn.preprocessing` [23] library was used.

D. *Development of the Neural Network*

For the analysis of patterns and the prediction of network conditions based on the collected metrics, it was decided to implement an RNN formed by four layers: one RNN input layer (receiving the six QoS metric values), one RNN layer (cell), one dense layer, and one output reshape that re-dimensioned the dense layer (outputting the six predicted QoS metric values). For the purpose of identifying the most suitable neural network model for predicting the QoS metrics, four experimental configurations were designed and evaluated. These configurations are based on the recurrent cell (first 2 layers): Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM). For each type of architecture, two output approaches were explored: one focused on predicting a single variable at a time (single output) and another capable of estimating multiple metrics simultaneously (multiple outputs). The four experimental models evaluated were: recurrent cell GRU single output (GRU-1), recurrent cell GRU multiple outputs (GRU-M), recurrent cell LSTM single output (LSTM-1), and recurrent cell LSTM multiple outputs (LSTM-M). The hyperparameters selected are shown in Table II.

TABLE II. HYPERPARAMETERS PER NEURAL NETWORK MODELS

Hyperparameter	GRU-1	GRU-M	LSTM-1	LSTM-M
Output (steps)	1	30	1	30
LSTM Layers	2	2	2	2
Neurons per Layer	128	128	128	128
Optimizer	Adam	Adam	Adam	Adam
Learning Rate	0.001	0.001	0.001	0.001
Epochs	12	11	32	25
Batch Size	16	16	16	16

The training of the four configurations was conducted using a general single dataset comprising values obtained under the four network conditions defined in Table I, during one hour.

The initial 80% of this general dataset was used exclusively for model training (training set), while the remaining 20% (corresponding to the most recent data) was reserved for testing (testing set). Each model was trained individually, respecting its specific architecture. During the training process, the `EarlyStopping` technique was applied. In each scenario, the model that achieved the best performance during validation was saved, in order to be formally evaluated later on another test set.

E. *Evaluation of the Neural Network*

To compare the performance of the different models, three evaluation metrics were defined and applied to the test set: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Single-output models (for both GRU and LSTM) achieved lower error metrics. For example, LSTM achieved errors of 0.0601 and 0.0994 for MAE and RMSE, respectively, demonstrating remarkable accuracy in predicting the next immediate point. However, these models presented significant limitations for long-term predictions, such as high computational inefficiency, cumulative error propagation, and underutilization of temporal relationships. In contrast, multiple-step (multiple-output) models were designed to address these challenges. Although their error metrics were slightly higher on average, the multi-output LSTM (MAE=0.1205, RMSE=0.2044) showed better capability for predicting extended series in a stable and coherent manner, mitigating the negative effects of error accumulation, and reducing computational cost per inference. Finally, the LSTM-M architecture was selected, consisting of one input layer (input 60 time steps and 6 features), one hidden LSTM layer (output 128 nodes), one dense layer (output 180 nodes), and one output reshape layer (output 30 time steps and 6 features), to perform predictions across the four scenarios (see Table I) without requiring retraining.

F. Measurements and Prediction of QoS per Scenario

Each call generated approximately 150 sets of samples (one every 2 seconds), capturing the following QoS parameters: audio and video jitter, audio and video round-trip time, packet loss rate, and number of packets received per channel. The model started operating from the first minute of the call, as sufficient data history was available at that point. The prediction model operated with a sliding window of historical values (60 steps) and predicted values (30 steps ahead).

IV. RESULTS AND ANALYSIS

The results of the four scenarios studied with the selected neural network are presented in the following sections, using embedded Python code within the general application built with PeerJS for WebRTC.

A. Acceptable Quality

Figure 2 shows that the model successfully estimated the video latency, closely following the actual signal trend. No significant offsets or error accumulation were observed, demonstrating the model's ability to adapt to stable network conditions.

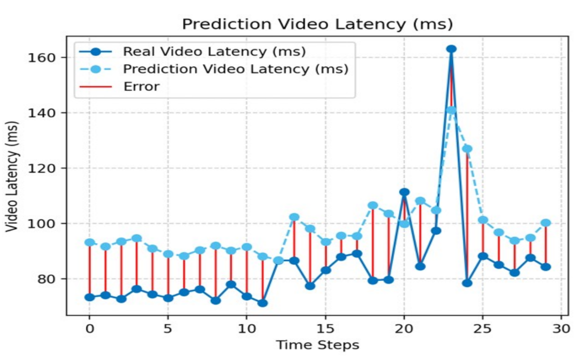

Figure 2. Actual Test Data vs. Model Prediction for Video Latency in Last-Minute Scenario 1

Figure 3 illustrates that the audio latency predictions exhibited a high level of agreement with the actual data. The model was able to maintain the trend without notable deviations, validating its ability to model this metric properly in low-variability environments.

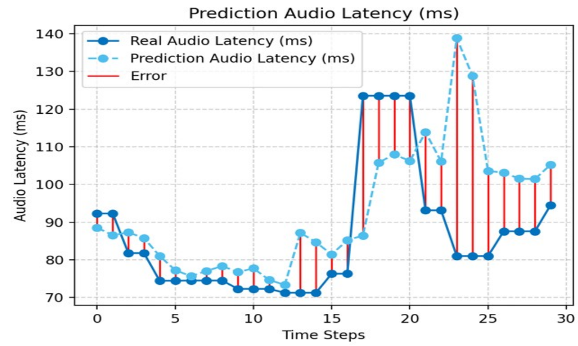

Figure 3. Actual Test Data vs. Model Prediction for Audio Latency in Last-Minute Scenario 1

Figure 4 shows that, although the model accurately predicted most video jitter values, an outlier was detected near 500 ms, indicating an anomaly in an otherwise stable network.

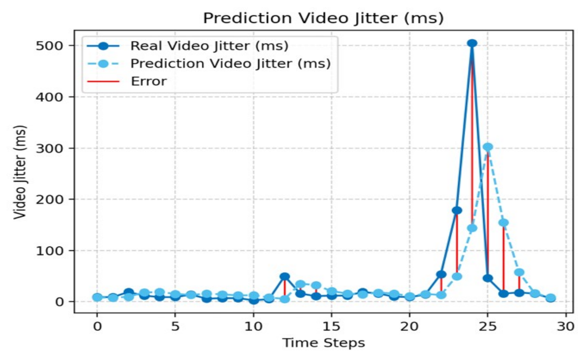

Figure 4. Actual Test Data vs. Model Prediction for Video Jitter in Last-Minute Scenario 1

As depicted in Figure 5, the audio jitter was predicted with minimal errors, showing highly stable behavior. This reinforces the idea that under ideal conditions, the model was capable of accurately capturing slight fluctuations in audio quality.

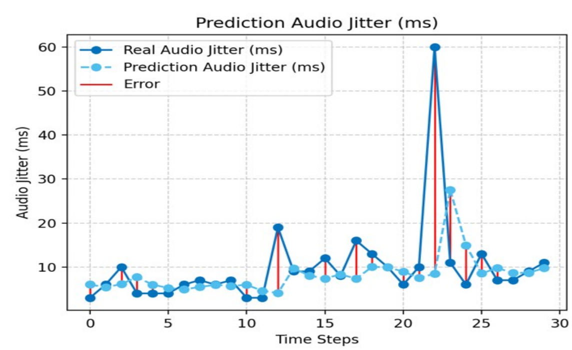

Figure 5. Actual Test Data vs. Model Prediction for Audio Jitter in Last-Minute Scenario 1

In Figure 6, it can be noted that the video packet loss rate was practically zero throughout the whole experiment, with the model predicting values close to zero.

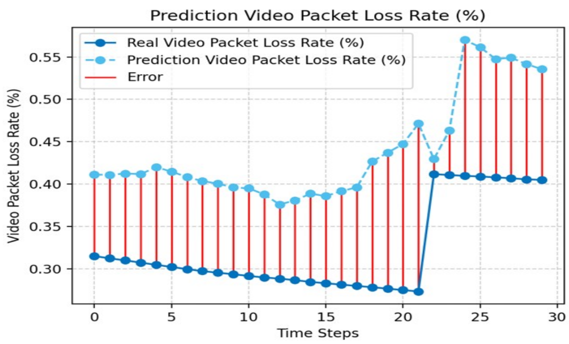

Figure 6. Actual Test Data vs. Model Prediction for Video Packet Loss Rate in Last-Minute Scenario 1

As shown in Figure 7, the audio packet loss rate prediction remained near zero, close to the measured data.

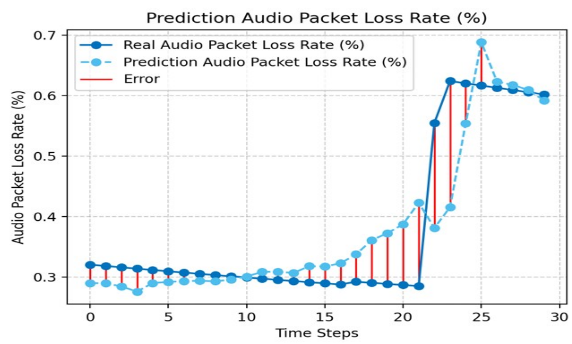

Figure 7. Actual Test Data vs. Model Prediction for Audio Packet Loss Rate in Last-Minute Scenario 1

B. *Moderate Degradation*

In Figure 8, it can be seen that the video latency showed an increase in variability compared to the acceptable quality network scenario (see Figure 2). While the model adapted well to average values, it exhibited limitations in predicting sudden latency spikes, which is expected given the less stable nature of the network in this scenario.

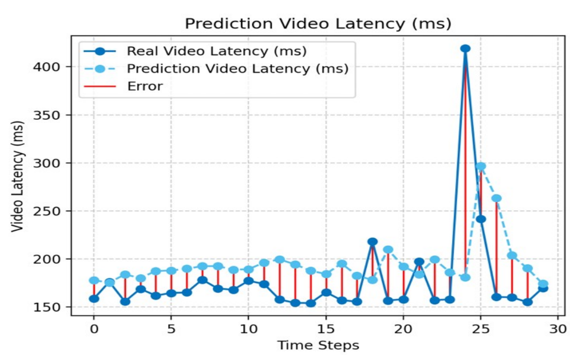

Figure 8. Actual Test Data vs. Model Prediction for Video Latency in Last-Minute Scenario 2

Figure 9 illustrates that the audio latency model effectively followed the general trend of the data, although, as with the video latency (see Figure 3), discrepancies arose when estimating extreme values. The performance is considered acceptable.

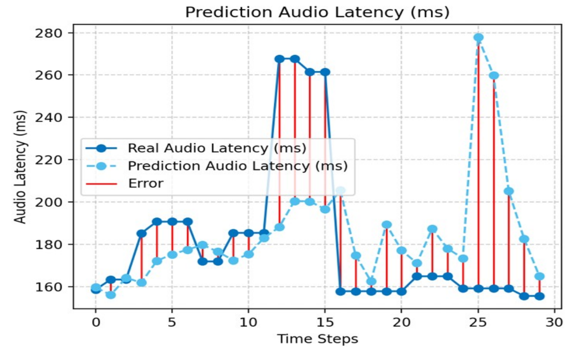

Figure 9. Actual Test Data vs. Model Prediction for Audio Latency in Last-Minute Scenario 2

According to Figure 10, the video jitter showed greater dispersion than the first scenario (see Figure 4). Nevertheless, the model could follow the overall trend, though with slightly reduced accuracy. This suggests that it can adapt to more dynamic conditions, yet with an increasing margin of error.

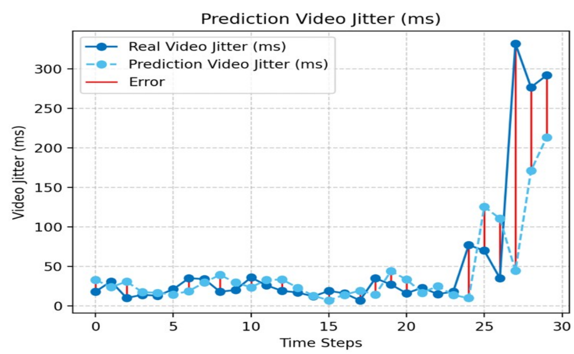

Figure 10. Actual Test Data vs. Model Prediction for Video Jitter in Last-Minute Scenario 2

As depicted in Figure 11, the behavior of the audio jitter showed wider fluctuations than in the first scenario (see Figure 5). The model maintained an acceptable ability to reflect the direction of changes compared to a stable environment (see Figure 5).

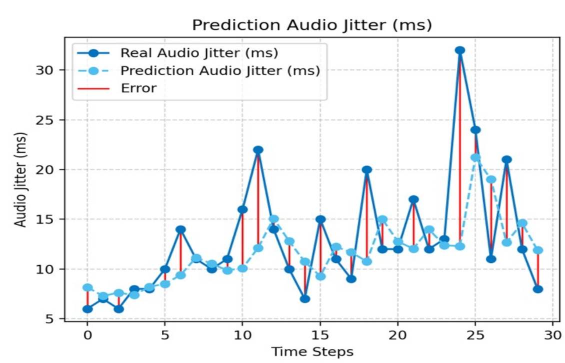

Figure 11. Actual Test Data vs. Model Prediction for Audio Jitter in Last-Minute Scenario 2

Regarding the video packet loss rate, Figure 12 indicates that the model faced greater difficulties in anticipating the actual pattern due to the intermittent and unpredictable nature of this type of traffic on a congested network. Even so, it managed to represent the overall trend of the fluctuations correctly.

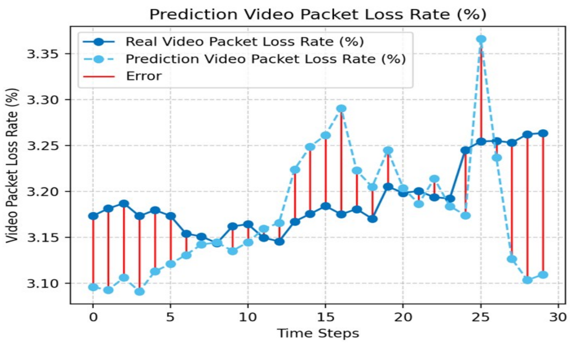

Figure 12. Actual Test Data vs. Model Prediction for Video Packet Loss Rate in Last-Minute Scenario 2

Figure 13 reveals that the audio packet loss rate exhibited variability similar to that of video (see Figure 12), although with lower intensity. The model captured the overall trend adequately, despite occasional discrepancies, demonstrating its adaptability while highlighting limitations in scenarios with irregular loss.

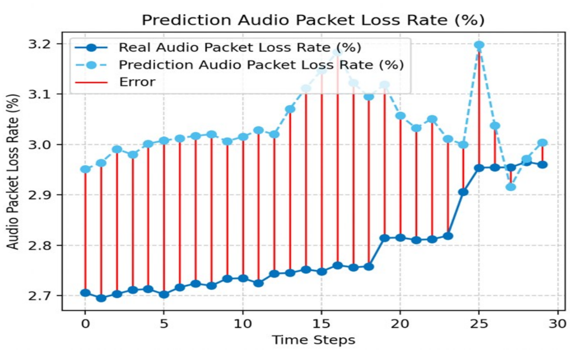

Figure 13. Actual Test Data vs. Model Prediction for Audio Packet Loss Rate in Last-Minute Scenario 2

C. *Critical Quality*

Figure 14 shows a pronounced deviation between the actual video latency values and the model's predictions. Although the model was generally able to follow the trend, differences in absolute values were evident, especially during periods of higher delay. This lack of precision can be attributed to the high baseline latency and the significant network instability.

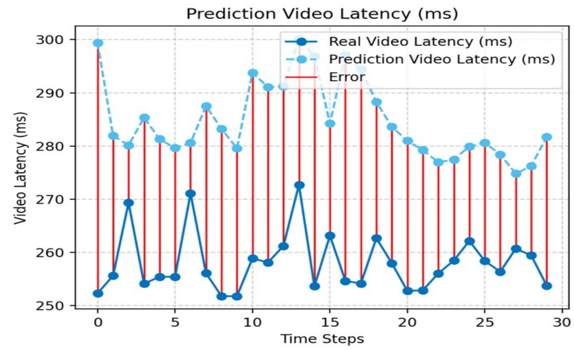

Figure 14. Actual Test Data vs. Model Prediction for Video Latency in Last-Minute Scenario 3

Similar to the video latency (see Figure 14), the audio latency also suffered discrepancies as shown in Figure 15. Although the model reasonably followed the trend, significant deviations were noted at the highest delay peaks.

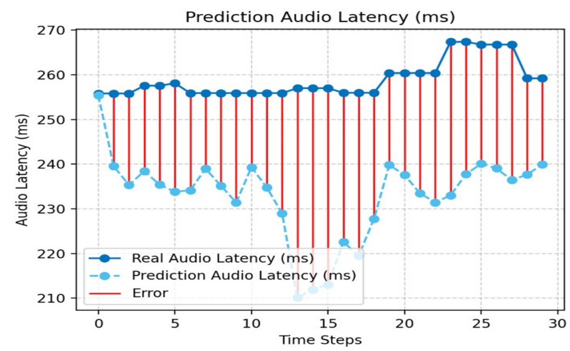

Figure 15. Actual Test Data vs. Model Prediction for Audio Latency in Last-Minute Scenario 3

In Figure 16, it can be seen that in the case of the video jitter, the model showed relatively stable performance. However, it struggled to replicate certain abrupt peaks present in the actual data. Despite this, the predictions reasonably captured the overall jitter dynamics, demonstrating the model's partial ability to adapt to rapid delay variations under critical conditions.

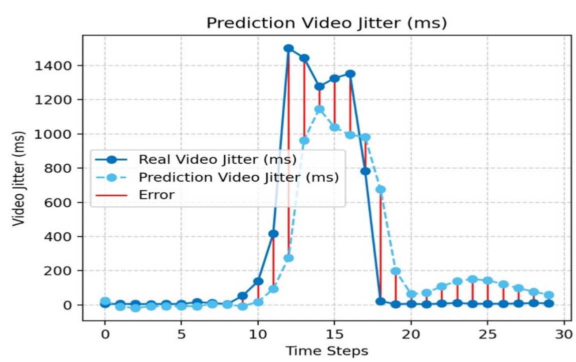

Figure 16. Actual Test Data vs. Model Prediction for Video Jitter in Last-Minute Scenario 3

As depicted in Figure 17, the audio jitter also exhibited behavior similar to that of video (see Figure 16). The model managed to follow the overall trend but faced notable difficulties during sudden changes. This illustrates that while the model can adapt to moderate fluctuations, it has limitations when faced with highly unstable events.

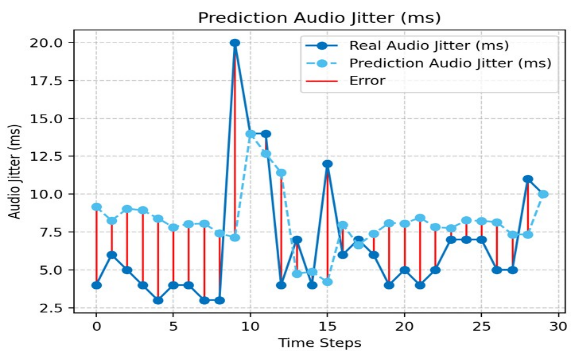

Figure 17. Actual Test Data vs. Model Prediction for Audio Jitter in Last-Minute Scenario 3

Regarding the video packet loss rate, Figure 18 indicates that the predictions generally remained close to the actual values. However, fluctuations were observed that the model was unable to predict accurately.

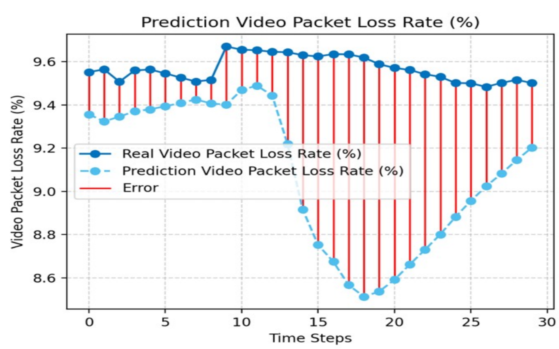

Figure 18. Actual Test Data vs. Model Prediction for Video Packet Loss Rate in Last-Minute Scenario 3

As shown in Figure 19, the audio packet loss rate exhibited patterns similar to those of video (see Figure 18). That is, while the model's predictions generally tracked the actual values, discrepancies emerged, reflecting its difficulty in anticipating abrupt changes.

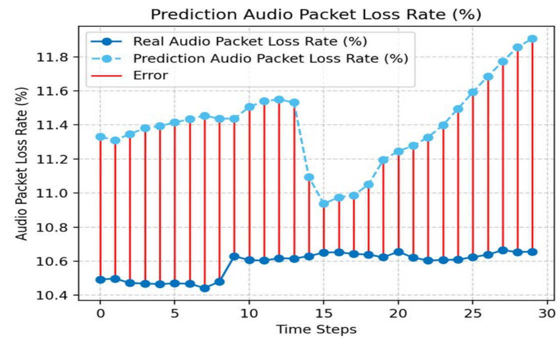

Figure 19. Actual Test Data vs. Model Prediction for Audio Packet Loss Rate in Last-Minute Scenario 3

D. Extreme Conditions

As can be seen in Figure 20, the model was able to reasonably follow the behavior of the video latency, adequately reproducing the most significant peaks present in the actual data. Although there are some discrepancies between the real and predicted values, the overall trend was effectively captured.

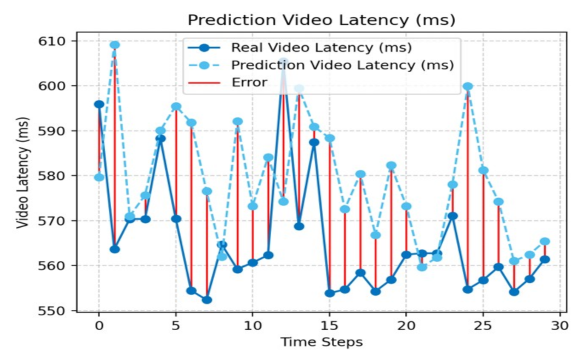

Figure 20. Actual Test Data vs. Model Prediction for Video Latency in Last-Minute Scenario 4

In contrast with the video latency (see Figure 20), the audio latency predictions exhibited greater deviations from the actual values, as shown in Figure 21. Increased dispersion and variability were observed, suggesting that the model has more difficulty adapting to rapid and erratic changes for this metric. Nevertheless, the overall trend was partially maintained, indicating that the model still achieved a coherent structural response.

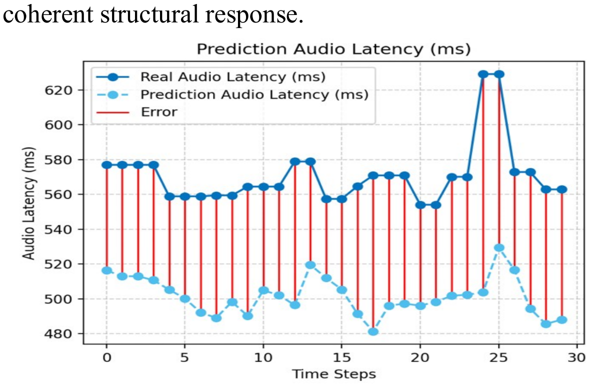

Figure 21. Actual Test Data vs. Model Prediction for Audio Latency in Last-Minute Scenario 4

In Figure 22, it can be seen that in the case of the video jitter, the model showed a reasonable ability to follow the general signal dynamics, although with specific differences in the maximum values. The predictions were consistent with the variation patterns, reflecting the model's ability to capture changes in delay instability, even if it did not achieve millimeter-level accuracy.

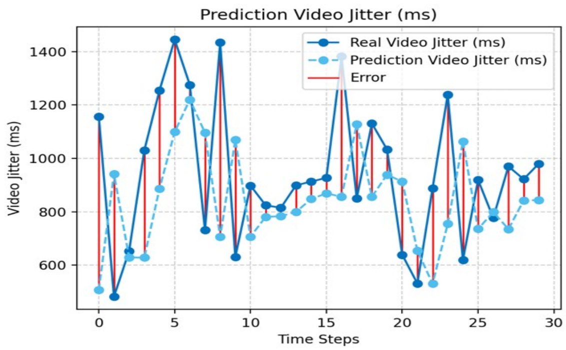

Figure 22. Actual Test Data vs. Model Prediction for Video Jitter in Last-Minute Scenario 4

As with the video jitter (see Figure 22), Figure 23 reveals that the audio jitter predictions provided an acceptable representation of the signal variations. Although discrepancies occurred at specific moments, especially during the most abrupt peaks, the model managed to represent the underlying behavior of the metric, reaffirming its partial ability to adapt to extreme fluctuations.

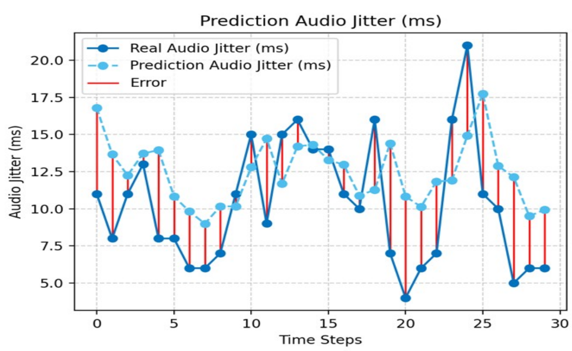

Figure 23. Actual Test Data vs. Model Prediction for Audio Jitter in Last-Minute Scenario 4

Regarding the video packet loss rate, Figure 24 indicates that the model's predictions showed an average difference of around 2 percentage points compared to the actual data.

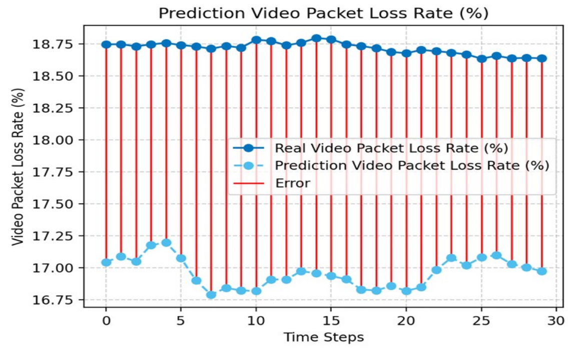

Figure 24. Actual Test Data vs. Model Prediction for Video Packet Loss Rate in Last-Minute Scenario 4

Similar to the video packet loss rate (see Figure 24), Figure 25 shows that the audio packet loss rate predictions reproduced the general structure of the actual signal, albeit with slight offsets at certain points. While an exact match was not achieved for all values, the model maintained acceptable coherence in terms of dynamics, correctly capturing the variation pattern in adverse environments.

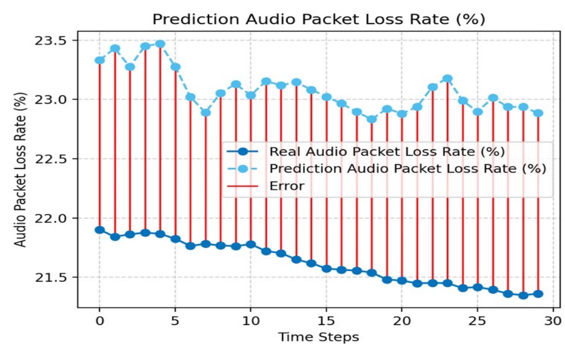

Figure 25. Actual Test Data vs. Model Prediction for Audio Packet Loss Rate in Last-Minute Scenario 4

V. CONCLUSIONS AND FUTURE WORK

After completing each of the phases outlined in this research project, it can be concluded that the integration of technologies such as WebRTC and RNNs represents a viable and modern alternative for addressing the problem of QoS prediction in video-call applications.

The research demonstrated that WebRTC, as a base technology, facilitates the creation of real-time communication environments with measurement and adaptation capabilities, removing previous technological barriers. The versatility of WebRTC, combined with a robust simulation infrastructure, made it possible to collect real metrics under different network conditions, thereby enriching the training of the predictive models.

During the system development, it was evidenced that LSTM-type neural networks are capable of capturing the temporal behavior of the evaluated metrics (latency, jitter, and packet loss rate), allowing the anticipation of their future evolution with an acceptable level of accuracy, especially under stable or moderately degraded conditions. In more extreme scenarios, with packet losses of up to 80% or abrupt variations in delay, the model showed limitations in absolute accuracy, although it was still able to reflect the general trends of network behavior. This characteristic is particularly useful for implementing early warning mechanisms or dynamic adaptation that can be activated before communication quality noticeably degrades for the user.

One of the most significant contributions of this work was demonstrating that a deep-learning-based model can be fed with the first few minutes of a call to generate reliable predictions of its future behavior.

The adopted predictive approach demonstrated robustness when trained across multiple network scenarios, which allowed the neural network to learn diverse patterns and therefore generalize better under new conditions.

The following recommendations are proposed to strengthen the developed solution and encourage future research: expansion of the dataset, inclusion of new QoS and QoE metrics, implementation of the model in real production environments, exploration of more complex architectures, and design of autonomous network management systems.

ACKNOWLEDGMENT

We acknowledge the Central University of Venezuela (UCV), Sant Longowal Institute of Engineering and Technology (SLIET), and Jacksonville State University (JSU) for partially funding this project.

REFERENCES

[1] World Wide Web Consortium (W3C), “WebRTC 1.0: Real-Time Communication Between Browsers”, March 2025, https://www.w3.org/TR/webrtc/

[2] H. Alvestrand, Overview: Real-Time Protocols for Browser-Based Applications, RFC 8825, January 2021, doi: 10.17487/RFC8825, https://www.rfc-editor. org/info/rfc8825

[3] E. Rescorla, “Security Considerations for WebRTC”, RFC 8826, January 2021, doi: 10.17487/RFC8826, https://www. rfc-editor.org/info/rfc8826

[4] E. Rescorla, “WebRTC Security Architecture”, RFC 8827, January 2021, doi: 10.17487/RFC8827, https://www.rfc-editor.org/info/rfc8827

[5] C. Perkins, M. Westerlund, and J. Ott, “Media Transport and Use of RTP in WebRTC”, RFC 8834, January 2021, doi: 10.17487/RFC8834, https://www.rfc-editor.org/info/rfc8834

[6] H. Alvestrand, “Transports for WebRTC”, RFC 8835, January 2021, doi: 10.17487/RFC8835, https://www.rfc-editor.org/info/rfc8835

[7] S. Ness, V. Eswarakrishnan, H. Sridharan, V. Shinde, N. Venkata Prasad Janapareddy, and V. Dhanawat, “Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques”, IEEE Access, vol. 13, pp. 16133–16145, August 2025, doi: 10.1109/ACCESS.2025.3526988.

[8] G. Garcia and O. Salcedo, “Predicción de Fallos en Redes IP Empleando Redes Neuronales Artificiales”, Trabajo de Grado para Magister en Ciencias de la Información y las Comunicaciones, Universidad Distrital Francisco José de Caldas, Colombia, 2017.

[9] Y. Wang, Z. Jia, X. Zhang, B. Shao, H. Wang, and X. Xing, “TAQ-GRNN: A Topology-Aware QoS Prediction Model Based on Gated Recurrent Neural Networks”, 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS 2024), August. 2024, pp. 303-308, doi:10.1109/ DDCLS61622.2024.10606915.

[10] M. Di Mauro, G. Galatro, F. Postiglione, W. Song, and A. Liotta, “Evaluating Recurrent Neural Networks for Prediction of Multi-Variate Time Series VoIP Metrics”, 2024 22nd Mediterranean Communication and Computer Networking Conference (MedComNet 2024), Nice, France, 2024, pp. 1-8, doi: 10.1109/MedComNet62012.2024.10578296.

[11] W. A. Aziz, I. I. Ioannou, M. Lestas, H. K. Qureshi, A. Iqbal, and V. Vassiliou, “Content-Aware Network Traffic Prediction Framework for Quality of Service-Aware Dynamic Network Resource Management”, IEEE Access, vol. 11, pp. 99716–99733, August 2023, doi: 10.1109/ACCESS.2023.3309002.

[12] Z. Wu, Y.-N. Dong, X. Qiu, and J. Jin, “Online Multimedia Traffic Classification from the QoS Perspective Using Deep Learning”, Computer Network, vol. 204, pp. 1–13, Elsevier, January 2022, doi: 10.1016/j.comnet.2021.108716.

[13] P. Zhang, J. Ren, W. Huang, Y. Chen, Q. Zhao, and H. Zhu, “A Deep-Learning Model for Service QoS Prediction Based on Feature Mapping and Inference”, IEEE Transactions on Services Computing, vol. 17, no. 4, pp. 1311–1325, August 2024, doi: 10.1109/TSC.2023.3326208.

[14] J. Gerard, D. C. Bonilla, A. Bentaleb, and S. Céspedes, “Optimizing Quality and Energy Efficiency in WebRTC with ML-Powered Adaptative FEC”, 2024 IEEE International Conference on Multimedia and Expo Worlshops (ICMEW 2024), July 2024, pp. 57-64, doi: 10.1109/ICMEW63481. 2024.10645390.

[15] “Optimizing RTC Bandwidth Estimation with Machine Learning”, https://engineering.fb.com/2024/03/20/networking-traffic/optimizing-rtc-bandwidth-estimation-machine-learning/

[16] K. Sakakibara, S. Ohzahata, and R. Yamamoto, “Deep Learning-Based No-Reference Video Streaming QoE Estimation Using WebRTC Statistics”, 2024 IEEE International Conference on Artificial Intelligence in Information and Communication (ICAIIC 2024), February 2024, pp.1-7, doi: 10.1109/ICAIIC60209.2024.10463278

[17] G. Bingol, “Advancing Video Communication: From WebRTC Quality Prediction to Green Appplications”, Ph.D. Dissertation, Department of Electrical and Electronical Engineering, University of Cagliari, Cagliari, Italia, 2025.

[18] Google Firebase Firestore, “Cloud Firestore”, https://firebase. google.com/docs/firestore

[19] ITU-International Telecommunications Union, “G.1010: End-User Multimedia QoS Categories”, https://www.itu.int/rec/T-REC-G.1010-200111-I

[20] PeerJS, “PeerJS Simplifies WebRTC Peer-to-Peer Data, Video, and Audio Calls”, https://peerjs.com

[21] WebRTC for Developers, “Breaking Changes in getStats”, https://www.webrtc-developers.com/breaking-changes-in-getstats/

[22] Apple, “Network Link Conditioner”, https://nshipster.com/network-link-conditioner

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python”, Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.