Infocommunications Journal

HTE 75 1949–2024

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

June 2024 Volume XVI Number 2 ISSN 2061-2079

Technically Co-Sponsored by

IEEE ComSoc IEEE Communications Society

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.
Infocommunications Journal is also included in the Thomson Reuters – Web of ScienceTM Core Collection, Emerging Sources Citation Index (ESCI)

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section
Supporters
FERENC VÁGUJHELYI – president, Scientific Association for Infocommunications (HTE)

The publication was produced with the support of the Hungarian Academy of Sciences and the NMHH

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027 • E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics
Phone: +36 1 463 4189 • E-mail: pvarga@tmit.bme.hu

Subscription rates for foreign subscribers: 4 issues 10.000 HUF + postage

Publisher: PÉTER NAGY

HU ISSN 2061-2079 • Layout: PLAZMA DS • Printed by: FOM Media

www.infocommunications.hu

Infocommunications Journal welcomes HTE 75

Pal Varga

Our publisher, HTE – the Scientific Association for Infocommunications – reaches its platinum age, turning 75. This ever-evolving, independent, professional organization was officially registered on January 29, 1949, and for 75 years, it has been providing objective representation for the entire infocommunications sector (ICT; telecommunications, information technology, media), facilitating the acceptance and development of technological advancements. The secret to its long-standing success lies in its ability to continuously renew itself: by creating modern scientific and professional platforms, it expands the knowledge base of those interested, and by utilizing the collective expertise of its members, it provides opinions on Hungarian and EU projects, compiles comprehensive technical reports, and offers customized training sessions.

Let's see what the summer issue of ICJ in 2024 has brought.

Simon Dahdal et al. present a novel MLOps framework to address the challenges of data scarcity and imbalance in smart manufacturing, particularly in Industry 5.0. The framework utilizes Generative Adversarial Networks (GANs), specifically a WGAN, to generate synthetic data that rebalances datasets, significantly improving the training and reliability of Machine Learning (ML) models. This approach was successfully applied to a real-life industrial scenario in Bonfiglioli's EVO plant, where it enhanced the accuracy of a binary classifier used for fault detection in gearboxes. Besides demonstrating the effectiveness of synthetic data in improving ML applications this study paves the way for future advancements in sustainable and zero-defect manufacturing within Industry 5.0.

The paper by G. Á. Németh and M. I. Lugosi presents a novel, open-source model-based testing framework specifically designed for finite state machine specifications. The framework's flexibility in model conversion and test generation makes it ideal for testing complex software systems across various domains, such as infocommunications. Notably, the framework includes a new heuristic test generation algorithm for the N-Switch Coverage Criteria, allowing for fine-tuned test criteria and coverage. Through analytical complexity assessments and empirical simulations, the framework's ability to balance test execution resources and fault coverage was demonstrated, offering valuable insights for test engineers to optimize their testing strategies.

M. Z. I. Nazir, M. Alqaradaghi and T. Kozsik introduce a new automated checker, "FindHidingMethod," that detects methodhiding issues in Java programs, a common programming weakness that can lead to unexpected results. Integrated into the SpotBugs static analysis tool, the checker identifies instances where static methods in subclasses inadvertently hide methods in their superclasses. The approach was rigorously tested on both custom test cases and real-world programs, demonstrating high precision in detecting these method-hiding issues. This tool significantly enhances the capabilities of static analysis in identifying programming errors in Java code.

T. Padmavathi, K. K. Cheepurupalli, and R. Madhu present a unique approach to channel estimation and interference cancellation in Filter Bank Multi-Carrier (FBMC) modulation, which is a strong candidate for 5G and beyond communications. Their paper introduces the use of multiple auxiliary symbols per pilot to neutralize imaginary interference, thereby enhancing throughput and channel capacity. The proposed Iterative Minimum Mean Squared Error (IMMSE) cancellation scheme significantly reduces interference at pilot and data positions. The authors demonstrate that FBMC, when paired with the proposed techniques, outperforms Orthogonal Frequency Division Multiplexing (OFDM) in terms of Bit Error Rate (BER) and system throughput, particularly at low transmission power. This advancement in channel estimation and interference cancellation promises improved performance for future communication systems.

The article by R. Praveen Kumar et al. introduces a novel hybridization of Machine Learning algorithms, specifically combining the Genetic Algorithm (GA) and Lion Algorithm (LA), to optimize Cluster Head (CH) selection in Wireless Sensor Networks (WSNs). The proposed model addresses key challenges in WSNs, such as node failures and unbalanced energy consumption, by implementing a non-uniform clustering strategy and multi-hop communication. The hybridized model enhances routing efficiency while considering energy, cost, time, network lifetime, and data accuracy. The performance of the model was validated through comparative studies, demonstrating significant improvements in throughput and packet loss ratios over existing models like WOA, GWO, LA, and GA. This innovative approach ensures an optimal balance between data transmission accuracy and energy efficiency, making it a robust solution for real-world WSN applications.

In their paper, B. Csóka, P. Fiala and P. Rucz explore the estimation of direction and distance of sound sources using microphone arrays. They specifically employ the MUSIC (Multiple Signal Classification) beamforming algorithm combined with Kalman filtering for tracking moving sources. While MUSIC proved effective and efficient for direction-ofarrival (DOA) estimations, particularly when enhanced with the Kalman filter for smoother tracking, distance estimation posed significant challenges, especially during practical measurements with unmanned aerial vehicles (UAVs). The study highlights the initial success in simulations but also underscores the difficulties encountered in real-world cases.

György Wersényi and Ádám Csapó studied and compared auditory and visual short-term memory using a serious game, in their paper. The participants of their experiments played a "finding pairs" game with either visual icons or auditory objects. Results showed no significant difference between visual and auditory memory overall, but familiar sounds were 2 recalled better than unfamiliar ones. They found that male and younger participants performed better, though further research with a larger sample size is suggested.

Pal Varga is the Head of Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics. His main research interests include communication systems, Cyber-Physical Systems and Industrial Io T, network traffic analysis, end-to-end QoS and SLA issues – for which he is keen to apply hardware acceleration and AI/ML techniques as well. Besides being a member of HTE, he is a senior member of IEEE (both in ComSoc and IES). He is Editorial Board member in many journals, Associate Editor in IEEE TNSM, and the Editor-in-Chief of the Infocommunications Journal.

An MLOps Framework for GAN-based Fault Detection in Bonfiglioli's EVO Plant

Simon Dahdal, *Student Member, IEEE*, Lorenzo Colombi, Matteo Brina, Alessandro Gilli, Mauro Tortonesi, *Member, IEEE*, Massimiliano Vignoli, and Cesare Stefanelli, *Member, IEEE*

***Abstract*—In Industry 5.0, the scarcity of data on defective components in smart manufacturing leads to imbalanced datasets. This imbalance poses a significant challenge to the development of robust Machine Learning (ML) models, which typically require a rich variety of data for effective training. The imbalance not only restricts the models' accuracy but also their applicability in diverse industrial scenarios. To tackle this issue, our research delves into the capabilities of Deep Generative Models, with a special focus on Generative Adversarial Networks, for the generation of synthetic data. This approach is aimed at rectifying dataset imbalances, thereby enhancing the training process of ML models. We demonstrate how synthetic data can substantially bolster the performance and reliability of ML models in industrial settings. Furthermore, the paper presents an innovative MLOps pipeline and architecture, meticulously designed to incorporate Deep Generative Models (DGMs) into the entire ML development cycle. This solution is automated and goes beyond mere automation; it is self-optimizing and capable of making necessary corrections, specifically engineered to address the dual challenges of data imbalance and scarcity, thus enabling more precise and dependable ML applications in smart manufacturing.**

***Index Terms*—Deep Generative Models, MLOps, Generative Adversarial Network, Industry 5.0, Synthetic Data Generation, Imbalanced Datasets.**

I. INTRODUCTION

THE integration of Machine Learning (ML) in industrial environments leads to increased operational efficiency and drives innovation in manufacturing processes and product development. This synergy of ML and big data is at the heart of Industry 5.0 [1]–[4], marking a shift towards more intelligent, adaptable, and sustainable manufacturing [5]. Maximizing the potential of ML in industry, however, necessitates addressing the unique challenges inherent in these settings.

In fact, *imbalanced datasets* – stemming from the scarcity of data on faulty components, due to the high yield of manufacturing processes [6] – *present a significant challenge for the effective adoption of ML in Industry 5.0 applications*. Most ML models assume the availability of extensive and varied datasets but the imbalance often results in models that are both limited and biased [7]. While Chaos Engineering, more and more used in IT, is starting to be proposed in industrial environments [8], manufacturing companies cannot be expected to address this issue by purposely inducing component failures or operational disruptions for the sake of data generation – an understandably costly and risky endeavor. Instead, a more promising solution lies in the augmentation of datasets through the generation of synthetic faulty component data with Deep Generative Models (DGMs) [9].

DGMs have seen remarkable advances in recent years, and have revolutionized the field of generative AI by effectively learning to capture complex data distributions and implementing high-fidelity sampling from them. In this context, most modern approaches leveraging diffusion-based and flow-based models have almost overnight become the most relevant approaches in many use cases – starting from image and video generation. However, *Generative Adversarial Networks (GANs)* arguably remain the best approach for the generation of tabular data [10]–[12], both in terms of sample fidelity and of performance at the training and at the inference levels.

In addition, since manufacturing processes exhibit characteristics that change over time, ML models need to be continuously re-evaluated and periodically re-trained. The deployment of a newly trained GAN typically results in enhanced performance capabilities, necessitating the retraining of the associated ML model with data generated by the improved fresh GAN. This requirement underscores the need for a more effective process management approach. By ensuring that the ML model is trained with the most recent data, its accuracy and applicability to the current manufacturing context are maintained, thereby aligning with the evolving industrial demands. On top of that, the progression of any ML project from a proof of concept to production is often impeded by the absence of DevOps and MLOps expertise [13].

This paper investigates the adoption of GAN-based synthetic data generation and the realization of a robust MLOps platform in the context of a real and challenging industrial use case. We present a comprehensive solution specifically designed for the gearbox assembly and testing line of the Bonfiglioli EVO plant, but broadly applicable to any real-world industrial use case. Our solution improves Bonfiglioli's production line by introducing an ML-based pre-testing phase, that serves as an early filter to identify defects, thereby conserving resources and energy that would be otherwise used in the expensive testing phase for likely faulty gearboxes. By leveraging GAN-based synthetic generation of faulty gearbox data, our solution overcomes the low-performance issues typically exhibited by classifiers trained with imbalanced datasets. This would lead to significant cost savings other than improving time efficiency and throughput of the assembly line.

More specifically, the paper demonstrates the effectiveness of DGMs, and GANs in particular, in overcoming challenges associated with data scarcity and imbalance through synthetic data generation. We achieved significant improvements in

S. Dahdal, L. Colombi, M. Brina, A. Gilli, C. Stefanelli, and M. Tortonesi are with University of Ferrara, Italy (E-mail: {simon.dahdal, lorenzo.colombi, matteo.brina, alessandro.gilli, cesare.stefanelli, mauro.tortonesi}@unife.it)

M. Vignoli is with Bonfiglioli, Calderara di Reno, Bologna, Italy (E-mail: massimiliano.vignoli@bonfiglioli.com)

DOI: 10.36244/ICJ.2024.2.1

classifier performance using Wasserstein GANs (WGANs) and Conditional GANs (CTGANs). Compared to the baseline classifier trained on the original dataset, these approaches yielded a 6-point increase in F1-Score and a nearly 12-point improvement in F2-Score. The larger improvement in F2-Score is particularly important for our task of detecting broken gearboxes, as it reduces False Negatives. At the same time, our MLOps platform based on KubeFlow and KServe is capable of serving 200 requests per second thus satisfying – and significantly surpassing – the performance requirements of the use case for the performance of trained ML models.

II. Background and Related Work

In the last decade, data has been seen as the key driver of progress across various industries, emphasizing its role in sparking new ideas, insights, products, and better decision-making. This aligns with the vision of *Clive Humby's*, who stated in 2006 that "data is the new fuel," highlighting the need for vast amounts of high-quality information for progress and efficiency. Additionally, thanks to the insights gained from recent deep learning approaches, the extensive data collected from industrial plant sensors has emerged as the predominant driving force. However, as contemporary data-driven and ML-powered applications begin to emerge in various industries the reliance on substantial, high-quality data is essential. Challenges such as data incompleteness, poor quality, and inadequate quantity can pose significant obstacles [14]. To address this issue various solutions can be found in literature, including data-generation techniques [15].

A. Generative Adversarial Networks (GANs)

Generative AI has been demonstrated to be a revolutionary field, especially with the introduction of Deep Generative Models (DGM). These state-of-the-art models, which originate from the fusion of generative algorithms and deep learning, have the great capabilities to generate new, realistic data samples that replicate the features and patterns of the training data, called synthetic data. There are several architectural designs of DGMs, such as GANs and Variational Autoencoders, Energy-based models, Autoregressive models, and Diffusion models. They stand out with their ability to generate data that closely mirrors real-world data distributions. In essence, the objective of training DGM is to grasp an unknown probability distribution from a typically limited number of independent and identically distributed samples. Upon successful training, DGM can be employed to assess the likelihood of a given sample and generate new samples resembling those from the unknown distribution. [12]

In recent years, the popularity of Diffusion models has surged due to their ability to generate samples of excellent quality, especially in image generation. However, GANs still remain the reference solution for generating tabular data [10]–[12]. Initially introduced by Goodfellow et al. [16] leveraging game theory concepts, GANs are based on 2 networks the Generator and the Discriminator. The generator learns the data distribution through unsupervised learning to produce authentic adversarial samples. Simultaneously, the Discriminator distinguishes between real and synthetic (generated) samples. The learning process involves iterative updates to both the generator and the discriminator. The generator function produces samples from noise input, while the discriminator tries to distinguish real samples from the synthetic samples.

The peculiar architecture of GANs allows them to efficiently generate data that closely mimic real examples [17]. The use of a neural network to model the loss function allows GAN-based models to have a lower number of parameters compared to other DGMs, with significant advantages in terms of higher data sampling speed and low training time [9]. The GAN capability to generate data concurrently, rather than sequentially, also enhances their speed, making them more applicable and effective in real-world scenarios. The lower computational demand and high efficiency of GANs make them a powerful and efficient tool for creating high-quality data, that is fundamental in Industry 5.0.

Many types of GANs have been introduced in literature such as Deep Convolutional GAN, Conditional GAN, Pix2Pix GAN, Cycle GAN, and others [17]. One of the most interesting variants is the *Wasserstein GAN (WGAN)*, introduced by Arjovsky et al. in 2017 [18]. WGAN proposes a new cost function using the Wasserstein distance, also known as the Earth mover's distance, which is used to measure the distance between two probability distributions. This makes the training of WGANs generally much more stable than that of traditional GANs and significantly reduces the occurrence of the mode collapse phenomenon that usually makes GANs overfit on a specific class of data and therefore, preventing it from generating samples that belong to the targeted class. The Wasserstein loss also provides a more meaningful measure of convergence and thus more useful insights into generator performance. Structurally, the network remains largely the same, except that the activation function of the discriminator's output layer is replaced with a linear function instead of a Sigmoid function.

In addition to WGAN, another GAN-based generative model that has shown great potential is the *Conditional Tabular GAN* (CTGAN) [19], which is specifically designed to generate synthetic tabular data. CTGANs are also designed to reduce the mode collapse phenomena [20]. For this reason, CTGAN is particularly suited for use cases affected by data scarcity. CTGAN innovatively introduces a 'mode-specific normalization' technique for processing continuous features. This approach involves treating each feature independently. Firstly, a variational Gaussian mixture model is fitted to the feature. Subsequently, normalization of each value within the feature is performed using the mean and variance of the corresponding Gaussian component from the mixture. Furthermore, CTGAN modifies the GAN's loss function to enhance the generation of categorical features. This modification includes an additional term in the loss function: the cross-entropy between the one-hot encoding of the input and that of the generated data. This adjustment aids in conditioning the output of the categorical features, ensuring higher fidelity in the synthesized data. Another novelty introduced by the CTGAN is the Training-by-Sampling process, focusing on replicating the original dataset's feature distribution through strategic data sampling and conditional vector construction [20].

However, the presence of misclassified examples in the training set could impact the CTGAN generation performance. Since CTGANs are designed to condition the generator's output, their process can be adversely affected by biases and inaccuracies stemming from these misclassified data points in the original dataset. As a result, CTGANs may learn to generate an increased number of false positives and negatives. This highlights the importance of clean and accurately classified training data for the effective functioning of CTGANs.

B. Machine Learning Operations (MLOps)

MLOps is a relatively new discipline that emerged when machine learning models began to be deployed in production environments. It can be seen as a specialization of DevOps, which focuses on software development principles and practices. However, MLOps is specifically oriented towards the automation of ML workflows. [21]. A cardinal aspect is automation where possible because automating an end-to-end ML pipeline helps avoid manual tasks and reduces delays [22]. However, there also could be some manual steps and tasks to be done. For example, when the model performance goes down, below a certain threshold, a retraining process is triggered and as a result, a specialized ML engineer could decide whether to approve the update for the deployment. Another crucial MLOps aspect is the guarantee of reproducibility and repeatability of the experiments other than versioning of the models, code, and data [21].

In literature it is possible to find various works on the ML life-cycles [23], [24] but some activities are consistently mentioned. These common activities include data engineering, model engineering, operations, and support tools, as detailed by Faubel et al. [25]. Firstly, data engineering comprises activities that are exclusively related to data like data collection, analysis, and preparation. Secondly, model engineering refers to all the steps that serve to create a model. This step includes activities like model building, training, evaluation, selection, and packaging. All the activities mentioned before can be either performed manually or in an automated way. The operations part aims to maintain the quality of the model in a real-world scenario. Operations include continuous integration and continuous deployment (CI/CD), model testing, deployment, and monitoring. Lastly, support tools provide services like versioning, infrastructure management, and automation. Infrastructure management is a key component of an MLOps system and must address various challenges. These challenges include the heterogeneity of both hardware (CPUs, GPUs, etc.) and software (operating systems, ML modules, libraries, etc.), the lack of standards due to costly legacy machines and their inability to communicate with newer systems, resource management, which involves the efficient allocation and utilization of computational resources to optimize performance and cost, and scalability, as the infrastructure must expand to accommodate the increasing size and complexity of ML models [25].

In greater detail, an MLOps pipeline necessitates various technical components, as detailed in work by Kreuzberger et al. [21]. This comprehensive list includes certain ML-related components such as a feature store, a model registry, and a metadata store. Moreover, to adhere to CI/CD principles, a source code repository and a CI/CD component are required. Furthermore, to enable automated workflow (pipeline) an orchestration component is essential. Lastly, it is usually needed to include a model serving component and an infrastructure orchestrator. In particular, the infrastructure orchestrator plays a crucial role in providing the required computation resources. Depending on the level of automation of an MLOps pipeline application, it can be classified using MLOps maturity models. The most widely used are those developed by Google and Microsoft. Google's model comprises three levels. Starting from 0 denoting no automation, progressing to level 1 with ML pipeline automation, and culminating at level 2, with CI/CD Pipeline automation. [26]. In contrast, the Microsoft model consists of five levels and its structure includes both MLOps and DevOps aspects.

Today, MLOps in Industry 4.0 still is in the initial adoption phase, as highlighted by Faubel et al. in [25] but the development and use of AI in the various branches are increasing exponentially [27]. This is particularly evident when considering small and medium-sized enterprises (SMEs) that have recently embarked on their digital transition journey or find themselves in the middle of a transition phase [4]. Notably, SMEs often struggle with a limited IT workforce and insufficient expertise in the ML sector, as well as limited financial resources to effectively build ML-powered applications. Therefore, it is crucial to choose a suitable strategy. Adopting Google's MLOps level one may introduce unnecessary complexity and a steep learning curve for these businesses. Moreover, the physical nature of the production environment introduces additional challenges and constraints, such as safety concerns as illustrated in [13]. As a pragmatic alternative, SMEs may benefit more from a lower level of automation that prioritizes simplicity, aligning better with their operational capacity and limited skill sets. Considering that SMEs represent 99% of all businesses in the EU as refereed by the "Annual Report on European SMEs 2022/2023" published by the European Union, developing an approachable set of standards and best practices could have a significant societal impact in a relatively short time.

III. BONFIGLIOLI INDUSTRIAL USE CASE

Bonfiglioli (https://bonfiglioli.com) is a leading manufacturing company that designs and manufactures a wide range of gear motors, drive systems, planetary gear motors, reducers, and inverters with over 130 years of experience. Bonfiglioli is increasingly adhering to Industry 5.0 best practices, and implementing efficient and environmentally sustainable processes.

Within the wide range of manufacturing lines that Bonfiglioli is running, and constantly improving, the gearbox assembly and testing line located in the EVO Plant represents a particularly interesting one. This line utilizes sophisticated machinery for automated precision assembly and thorough testing processes. The assembly line is composed of three workstations (WS) as illustrated in Fig. 1. **WS1 - Differential Assembly** is responsible for the assembly of the differential

part of the gearbox. It collects data on insertion forces and tightening torque. **WS2 - Gearbox Assembly** similarly engages in the assembly process, here focusing on the entire gearbox. It gathers data on insertion forces and tightening torque, akin to the WS1 station. **WS3 - End-of-Line Testing** is the last station of the plant, which receives the assembled gearbox and implements the testing phase.

In turn, WS3 consists of two distinct machines. The first one is dedicated to generating cycle data through specific stress tests on the component under examination. This involves conducting two phases of static analysis at 800 revolutions per minute (RPM) and two phases of ramp analysis at 11,000 RPM. These tests are critical for assessing the component's performance under varying operational conditions. Concurrently, a second machine operates to monitor vibrational data. This machine is specifically tasked with tracking vibration levels as various RPM thresholds are surpassed. The data collected from this machine is crucial for understanding the vibrational characteristics of the component under different operational speeds, which is vital for ensuring its reliability and structural integrity.

To further improve the efficiency of the Bonfiglioli EVO gearbox assembly and testing line, the manufacturing process has been extended by integrating a pretesting phase, positioned between WS2 and WS3, that offers substantial benefits in enhanced sustainability and reduced costs. The pretesting phase serves as an early filter to identify defects, thereby conserving resources and energy that would be otherwise used in the more expensive final testing phase. It aids in minimizing waste by detecting quality issues at an earlier stage, allowing for rectification or recycling before extensive testing. Not only do the added early quality checks ensure higher overall product quality, aligning with sustainable manufacturing practices by reducing environmental impact, an aspect increasingly vital in today's market, but they also improve the overall good output of the assembly line. The realization of an effective pretesting phase requires addressing the issue of training an accurate classifier using an imbalanced dataset collected from the WS1 and WS2 machines. At the same time, there is the need to design and develop an MLOps platform capable of deploying, running, monitoring, and managing and updating ML models.

The dataset collected from WS1 and WS2 is centered around two processes: tightening torques and press-fit forces. The roughly 700 metrics contained by the dataset are critical as they directly impact the assembled gearbox's functionalities. An accurate and multi-step evaluation process, conducted in collaboration with domain experts, has led to the identification and prioritization of 66 key features deemed most influential on the gearbox's final performance and readiness. Based on these input features and the output of the tested gearboxes from the WS3 testing machine, we managed to construct the final dataset used to to train the ML pretest classifier.

However, naively training a classifier using the imbalanced dataset led to suboptimal performance. More specifically, using a logistic regression model we obtained a relatively high number of False Negatives, as discussed in Section IV-B. To address this issue, we designed a GAN-based synthetic generator for faulty component data.

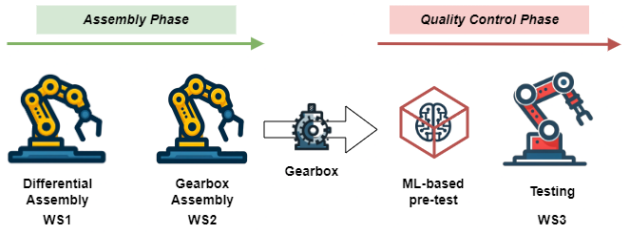


Fig. 1: Bonfiglioli gearbox assembly line at the EVO plant.

A. Design of GAN-based Synthetic Data Generation

Like with all artificial neural network-based solutions, the model architecture represents a critically important factor for the effectiveness of Generative Adversarial Networks (GANs). Model architecture design involves many decisions, starting from the number of layers and the number of neurons in each layer for both the Generator and Discriminator networks, as well as the activation function, the possible adoption of regularization constraints, the tuning of hyperparameters, etc. – all factors that can significantly affect a model's performance. Hence, employing an automated tool for the training process of a GAN can markedly improve its effectiveness, making such a tool an essential aid for practitioners.

In this specific case, Optuna (https://optuna.readthedocs.io/) was employed to systematically search for the optimal configuration of architecture and hyperparameters. For example, aspects such as the number of layers in each network, the number of neurons in each layer, the size of the input noise vector, the number of training epochs, and the size of the training batch are defined using the capabilities of the optimizer. Specifically, Optuna requires upper and lower bounds for the hyperparameter search space and, based on an objective function that needs to be minimized (or maximized), the search for the optimal combination of hyperparameters is done by selecting one of various heuristics given at the disposition of the users of Optuna. The parametric model of the GAN at hand has been specified and passed to Optuna as follows: The generator is composed of an input layer of the same size as the noise, which $\in [10, 500]$, N hidden layers with $N \in [1, 5]$, and many neurons in each layer that $\in [32, 512]$. The activation function chosen for each hidden layer is *LeakyReLU*, to avoid the dying ReLU phenomenon, where neurons become inactive and stop learning, effectively rendering them useless by causing the gradient to become zero. The output layer, on the other hand, has a size equal to the number of features in the dataset, as it is required to produce synthetic output data. The activation function for the output layer is *Tanh*, suitable as the dataset is initially normalized between -1 and 1. As for the optimizer, *RMSProp* has been selected, as often proposed in literature due its adaptive Learning Rate, Convergence Speed and stability even on very nonstationary problems [18]. Regarding the discriminator, the only changes are in the input layer, which has a size equal to the number of features in the data, as its function is to receive input data and decide whether it is synthetic or not. The output layer has a size of 1 and utilizes a linear activation function. Regarding the training,

a batch training methodology has been chosen, with the batch size determined by Optuna.

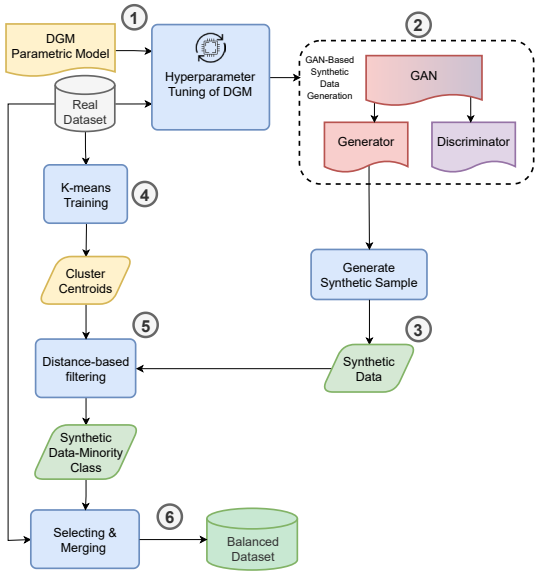


Fig. 2: WGAN Dataset Balancing algorithm.

Fig. 2 illustrates the data augmentation workflow using WGAN. The goal was to generate examples belonging to the minority class. For this reason, the workflow includes WGAN training, data generation and filtering steps. The latter is needed due to the presence of mislabeled examples in the training set and the WGAN's poor performance in generating categorical data. In detail, the steps are the following:

1) From the parametrically defined model, hyperparameter optimization is performed, minimizing the objective function.
2) The WGAN that achieves the minimum value of the objective function is stored in the Model Registry.
3) The best performing WGAN is used to generate a synthetic dataset much larger than the original one
4) A K-means model is trained on the original dataset to identify the centroids of two clusters, which presumably correspond to the two classes of gearboxes: 'good' and 'broken'.
5) Once the centroid of the minority class is identified, the synthetic data are filtered measuring the Euclidean distance between each example of the minority class centroid. Empirically, it has been found that even a simple approach such as the Euclidean distance has given good results. This process retains at least N example less distant than d_{max} from the centroid. N represents the difference between the number of examples in the majority and minority classes in the original dataset, and d_{max} is the empirically chosen maximum distance a point can have from the centroid to be classified as belonging to the minority class.
6) The synthetic filtered data is used to balance the initial dataset.

Moreover, for comparison purposes, we tested a second straightforward GAN-based model, called. CTGAN. It was configured specifically to generate examples belonging to the minority class. Specifically, were used the Synthetic Data Vault (SDV) [28] CTGAN implementation. SDV is a comprehensive Python library tailored for generating tabular synthetic data and offers various synthesizers including the CTGAN model.

B. Kubeflow

We implemented the MLOps platform on top of Kubeflow (https://kubeflow.org/), an MLOps framework developed and maintained by Google. Kubeflow aims to make it easier for organizations to develop, deploy, and manage ML workloads. Kubeflow includes Katib for hyper-parameters tuning, KServe for model serving, Jupyter Hub, and Kubeflow pipeline. It is possible to interact with these features through a web-based GUI. Due to Kubeflow's architecture, which is micro-services oriented and based on Kubernetes, it can be seamlessly integrated with other software components operating on top of Kubernetes. In addition, Kubeflow can be installed in every cloud or a local single-node Kubernetes cluster.

Using Kubeflow Pipelines (KFP) [29] it becomes possible to build and execute portable and scalable ML workflows. A pipeline is represented as a Directed Acyclical Graph, where each node is a component. At runtime, each component execution corresponds to a single container. Components can be exported for later use, in YAML format. This way, components are highly portable and facilitate code reuse. Pipelines are programmable using KFP SDK that lets easily convert a Python function into a container component using a simple Domain Specific Language. Another way to create a component is by using a custom Docker image, permitting the inclusion of different programming languages in the pipeline. From the pipeline SDK users can define, save, and execute pipelines. This allows triggering execution programmatically, enabling continuous training strategy. Kubeflow, also, provides a simple method to pass data between each step using its artifact store. An artifact is every object created during the execution. For example, an artifact could be a dataset or a set of metrics. For each artifact, metadata information is saved in a MySQL database, while the object is saved in a MinIO object-store. Therefore, the artifact store uses both MySQL and MinIO as backends.

C. MLOps Pipeline Implementation

1) Runtime Environment: For all of these reasons and the features stated above, Kubeflow has been chosen as the MLOps framework for testing deployment. The decision was made to deploy Kubeflow using raw manifests on a Kubernetes single-node cluster. Specifically, the Kubernetes environment chosen was MicroK8s, favored for its straightforward installation process and valuable add-ons. The Kubernetes node utilized was configured with 16 virtual CPUs, 32GB of RAM, and 50GB of storage memory.

2) Kubeflow Pipeline: The ML pipeline is implemented via KFP version 2, leveraging the Python KFP SDK. The pipeline is made of various components, as illustrated in Fig. 3. The

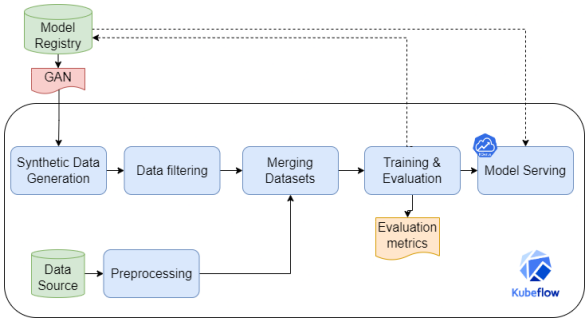
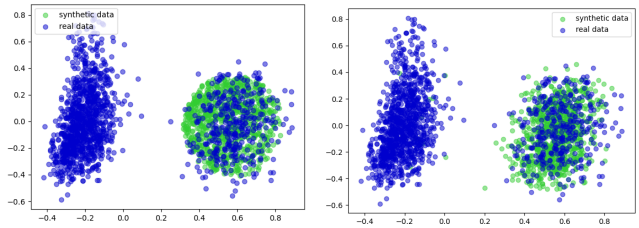


Fig. 3: Kubeflow MLOps pipeline.



(a) WGAN (b) CTGAN

Fig. 4: Original data and synthetic data plotted after the reduction to 2 dimensions with PCA.

initial component retrieves the latest version of the dataset stored in a CSV file from a Kubernetes volume and transfers it to Kubeflow's artifact store for subsequent utilization. This step is also responsible for acquiring the pre-trained GAN. The serialized GAN can be stored in the same volume as the dataset, or a specialized model registry, such as Mlflow. The adoption of a model registry offers many advantages, including decoupling the GAN development processes from the classifier training phase. Subsequently, the dataset undergoes preprocessing where the meaningless features are deleted and inputs are separated from outputs. Concurrently, the newest GAN available in the model registry is retrieved and used to generate a synthetic dataset. The synthetic dataset is filtered to keep only the examples belonging to the minority class and is used to balance the real dataset.

In the subsequent steps, the classifier, specifically a Logistic Regression model (LogReg), is trained on the enriched and balanced dataset. LogReg was specifically chosen for its inherently interpretable nature which is of great importance for manufacturing applications that require thoroughly evaluating and possibly certifying decision-making elements. However, other ML classifiers, including Decision Trees, Support Vector Machines, Gradient Boosting Machines, etc., could be used. After the training the model undergoes evaluation, and metrics are exported for comparative analysis through the Kubeflow dashboard. As a final step, the model is deployed into production using Kubeflow's built-in serving framework: *KServe* (https://kserve.github.io). KServe offers a Kubernetes Custom Resource Definition to enable out-of-the-box deployment of trained models onto various widely used serving runtimes, such as TFServing, TorchServe, Triton, and many others.

A scikit-learn inference service has been deployed using the KServe Python SDK. The predictor is configured with a minimum and a maximum number of replicas set at, respectively, 1 and 10. Additionally, constraints for the predictor pod resources have been defined, limiting it to use 0.5 CPU and 0.5 GB of memory. Finally, as shown in Fig. 3, the model can be saved in the model registry, to prepare it for subsequent utilization. This practice serves several purposes, including tracking the model and facilitating further evaluation. Additionally, the stored model can be deployed into production at a later stage.

IV. Experimental Evaluation

Let us demonstrate how WGAN and CTGAN adoption can improve the performance of the pre-testing phase in the Bonfiglioli assembly line.

A. GANs Performance

Starting with a visual comparison of the synthetic data generated by the two trained GANs. Fig. 4 illustrates this relationship. To create this visualization comparison, Dimensionality Reduction was employed. In detail, the Principal Component Analysis (PCA) algorithm was used to reduce the dimensionality of the dataset from 66 features which were heavily cleaned into 2 dimensions. The PCA has managed to create 2 distinct clusters, the left one is denser presenting the good gearboxes while the right cluster represents the broken gearboxes. For Both WGAN and CTGAN, We can observe the positioning of the **synthetic data** and the **original data**. We can observe that the CTGAN has successfully created a distribution that closely matches the original, dividing the data into two clusters that resemble the original clusters.

Secondly, for a more precise and detailed comparison of the two GANs performances, we employed a range of distance and similarity metrics that best describe data distributions. These metrics were selected due to their superior capability to elucidate the differences in the data distributions created by each GAN, offering a deep and thorough analysis of their respective performances. A brief description of each used metric: Starting with the *Wasserstein distance* (WD) indicating a close similarity between the original and synthetic datasets. A second metric is *Kolmogorov-Smirnov (KS) D statistic*, which represents the maximum difference between two empirical cumulative distribution functions. The larger the D value, the more significant the difference between the two distributions, the actual gearboxes and the synthetic gearboxes. The KS D statistic is based on the KS test which was used to compare the differences in probability distribution features data between the actual gearboxes and the generated synthetic gearboxes. The *KS-complement* (KS-C) was a supplementary index of the KS test, and its value equaled $1 - (D\ statistic)$. The *Correlation Similarity* (CS) has been computed. This metric measures the correlation between a pair of numerical columns and computes the similarity between the real and synthetic data. A score closer to 1.0 shows a perfect pairwise correlation. Finally, we

measured the *Jensen Shannon distance* (JS), an alternative to measure the distance between two probability distributions. All the previously presented distance metrics were applied feature-wise between original and synthetic datasets to then calculate the mean of the scores.

A summary table of compression metric results between the two GANs can be seen in Table I. The results demonstrate a marginally superior performance of the CTGAN compared to the WGAN, confirming the observations made in earlier visual comparison. An important evaluation of the inference time is crucial. The WGAN model has achieved 0.14 seconds to generate 1,000 samples and 0.32 seconds for 10,000 samples. Conversely, the CTGAN model was slower, requiring 0.24 seconds for 1,000 samples and 0.53 seconds for 10,000 samples, possibly due to its distinct architecture. However, both models exhibited swift inference performance.

TABLE I
GANs PERFORMANCE.

	KS-C	CS	WD	JS
WGAN	0.759	0.922	0.024	0.309
CTGAN	0.799	0.931	0.022	0.261

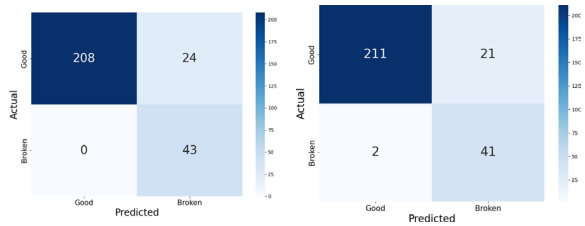
B. Classifier Prediction Performance

The last crucial test was to test how much the LogReg model improved by augmenting the original dataset with the synthetic data. The test was done as follows: Firstly, the cleaned original dataset was split into training and testing sets. Subsequently, the training dataset was augmented with the WGAN using the methodology described earlier (see section III-A). Three LogRegs were trained, one on just the original training set, a second with a WGAN-augmented training dataset, and a third one with the conditioned CTGAN-augmented training set. The confusion matrix of the results is shown in Fig. 5. The number of False Negatives has been reduced to zero in the case of WGAN and near zero in the case of CTGAN, this is due to the presence of mislabeled in the generated synthetic data as explained in the section II-A. More comprehensive and detailed metrics are reported in Table II.

Although CTGAN showed better performance in similarity metrics, WGAN, enhanced with unsupervised filtering, resulted in more effective classifier training. These results are attributed to the process of managing mislabeled data points in the dataset. A data augmentation by CTGAN failed to resolve and, in some cases, might worsen the situation by generating mislabeled data points introducing more biases and errors into the process. The final results showcase the achievement of the goal of reducing the number of incorrectly classified broken gearboxes (False Negative) and improving the performance of the classifier model in terms of key metrics such as Recall, F2 score, and the G-mean. This improvement underscores the enhanced ability of the model to accurately identify and classify gearbox conditions, marking a noteworthy advancement in predictive capabilities. In the light of zero waste tolerance, we can conclude that WGAN fits better in this specific use case, despite CTGAN performing similarly.

TABLE II
LOGREG MODEL PERFORMANCE METRICS [30].

	Original dataset	WGAN augmented	CTGAN augmented
Accuracy	0.90	0.91	**0.92**
Precision	0.65	0.64	**0.66**
Recall	0.81	**1.00**	0.95
F1 Score	0.72	**0.78**	**0.78**
F2 Score	0.77	**0.90**	0.88
G-Mean	0.86	**0.95**	0.93



(a) trained with WGAN balanced augmented dataset

(b) trained with CTGAN balanced augmented dataset

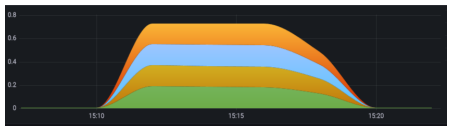
(c) trained with original balanced augmented dataset

Fig. 5: Comparison between confusion matrices of LogReg classifiers trained with different datasets.

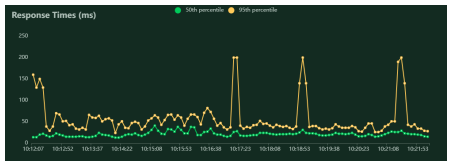
C. Classifier Serving Performance

To assess the performance of KServe, an observability stack has been deployed, specifically, the one integrated into MicroK8s has been used. This stack incorporates Prometheus for metrics scarping and Grafana for visualization. Additionally, an open-source load testing tool, Locust (https://locust.io/), was employed in the evaluation process. Key metrics, such as response time, number of pods, and pod resource consumption, were selected to provide meaningful insight into the performance characteristics of the KServe deployment.

Initially, a Locust test, with a peak of 20 users sending 10 requests per second (RPS), was started. This way, the inference service had to handle, approximately, 200 RPS. While in an idle state, only one pod is instantiated, as the RPS increased, the KServe autoscaler dynamically deployed new pods, ultimately reaching a total of 4 pods. Moreover, analyzing the resource usage revealed an equitable distribution of requests between each pod. Nonetheless, the Locust dashboard indicated that the 95th percentile of requests were served in less than 56ms. This test indicates how Kubernetes-backed KServe can easily scale in response to increasing load, maintaining efficient resource usage, as well as good performance under heavy load conditions.



(a) CPU usage of each Pod.



(b) Response time Locust chart.

Fig. 6: CPU usage and response time charts.

V. CONCLUSIONS & FUTURE WORK

Machine Learning (ML) presents a plethora of compelling and high-impact applications in Industry 5.0, although the end-to-end integration of ML-powered applications in the industrial scenario still presents many challenges. To address those issues, we realized a robust framework that automatically rebalances datasets by incorporating synthetic data generated by GANs, thus allowing to improve the performance of ML model training. Specifically, by employing a WGAN, we succeeded in training a binary classifier that can distinguish between good and broken gearboxes. This framework contributes meaningfully to Industry 5.0, particularly in terms of sustainability and Zero Defect Manufacturing. Furthermore, we realized a robust architectural pipeline for orchestrating the ML model deployment, serving, and update – thus ensuring levels of error resilience that are suited for the industrial environment. While our framework was designed for, and validated on, a real-life manufacturing line – namely the Bonfiglioli EVO gearbox assembly and testing line – it is of wide applicability. Looking ahead, we aim to measure the results of using synthetic data in Industry 5.0, such as estimating cost savings and throughput improvements in the assembly line and the acceleration that can be given to creating next-generation ML-based applications. Additionally, we plan to enhance the lifecycle management of the generated synthetic data, thereby extending the reproducibility of the ML models beyond on-the-fly generation. Finally, we also intend to explore more advanced unsupervised and semi-supervised anomaly detection techniques to identify broken gearboxes.

REFERENCES

[1] R. Venanzi *et al.*, "Enabling adaptive analytics at the edge with the Bi-Rex Big Data platform," *Computers in Industry*, vol. 147, p. 103 876, 2023. DOI: 10.1016/j.compind.2023.103876

[2] A. Frankó *et al.*, "Applied Machine Learning for IIoT and Smart Production Methods to Improve Production Quality, Safety and Sustainability," *Sensors*, vol. 22, no. 23, 2022. DOI: 10.3390/s22239148

[3] E. Jantunen *et al.*, "Maintenance 4.0 World of Integrated Information," in *Enterprise Interoperability VIII*. Springer International Publishing, 2019. DOI: 10.1007/978-3-030-13693-2_6

[4] A. E. Frankó *et al.*, "A Survey on Machine Learning based Smart Maintenance and Quality Control Solutions," *Infocommunications Journal*, vol. 13, no. 4, pp. 28–35, 2021. DOI: 10.36244/ICJ.2021.4.4

[5] P. Varga *et al.*, "Data-Driven Workflow Execution in Service Oriented IoT Architectures," in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 203–210, 2018. DOI: 10.1109/ETFA.2018.8502665

[6] L. Alzubaidi *et al.*, "A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, 2023. DOI: 10.1186/s40537-023-00727-2

[7] T. B. Nyíri *et al.*, "What can we learn from Small Data," *Infocommunications Journal*, vol. 15, no. SI, pp. 27–34, 2023. DOI: 10.36244/ICJ.2023.5.5

[8] M. Fogli *et al.*, "Chaos Engineering for Resilience Assessment of Digital Twins," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 2, pp. 1134–1143, 2024. DOI: 10.1109/TII.2023.3264101

[9] S. Bond-Taylor *et al.*, "Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7327–7347, Nov. 2022. DOI: 10.1109/tpami.2021.3116668

[10] S. De *et al.*, "Deep Generative Models in the Industrial Internet of Things: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 5728–5737, 2022. DOI: 10.1109/TII.2022.3155656

[11] D. Ezeh *et al.*, "An SDN Controller-Based Framework for Anomaly Detection Using a GAN Ensemble Algorithm," *Infocommunications Journal,* vol. XV, no. 2, pp. 29–36, June 2023. DOI: 10.36244/ICJ.2023.2.5

[12] L. Ruthotto *et al.*, "An introduction to deep generative modeling," 05 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/gamm.202100008, DOI: 10.1002/gamm.202100008

[13] C. Hegedűs *et al.*, "Tailoring MLOps Techniques for Industry 5.0 Needs," in *2023 19th International Conference on Network and Service Management (CNSM)*, 2023. pp. 1–7. DOI: 10.23919/CNSM59352.2023.10327814

[14] Y.-T. Chen *et al.*, "On the Private Data Synthesis Through Deep Generative Models for Data Scarcity of Industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 551–560, 2023. DOI: 10.1109/TII.2021.3133625

[15] Z. Pödör *et al.*, "A Practical Framework to Generate and Manage Synthetic Sensor Data," *Infocommunications Journal*, vol. XIV, no. 2, pp. 64–72, June 2022. DOI: 10.36244/ICJ.2022.2.7

[16] I. J. Goodfellow *et al.*, "*Generative Adversarial Networks*," 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

[17] Y. Li *et al.*, "The theoretical research of Generative Adversarial Networks: an overview," *Neurocomputing*, vol. 435, pp. 26–41, 2021. DOI: 10.1016/j.neucom.2020.12.114

[18] M. Arjovsky *et al.*, "Wasserstein GAN," 2017. [Online]. Available: https://arxiv.org/abs/1701.07875

[19] L. Xu *et al.*, "Modeling Tabular data using Conditional GAN," 2019. [Online]. Available: https://arxiv.org/abs/1907.00503

[20] O. Habibi *et al.*, "Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105 669, 2023. DOI: 10.1016/j.engappai.2022.105669

[21] D. Kreuzberger *et al.*, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," *IEEE Access*, vol. 11, pp. 31 866–31 879, 2023. DOI: 10.1109/ACCESS.2023.3262138

[22] L. Colombi *et al.*, "A machine learning operations platform for streamlined model serving in industry 5.0," in *NOMS 2024-2024 IEEE Network Operations and Management Symposium*, 2024. pp. 1–6. DOI: 10.1109/NOMS59830.2024.10575103

[23] "Practitioners guide to mlops: A framework for continuous delivery and automation of machine learning." [Online]. Available: https://services.google.com/fh/files/misc/practitioners_guide_to_mlops_whitepaper.pdf

[24] R. Subramanya *et al.*, "From DevOps to MLOps: Overview and Application to Electricity Market Forecasting," *Applied Sciences*, vol. 12, no. 19, 2022. DOI: 10.3390/app12199851

[25] L. Faubel *et al.*, "Mlops challenges in industry 4.0," *SN Computer Science*, vol. 4, no. 6, p. 828, 2023. DOI: 10.1007/s42979-023-02282-2

[26] G. Symeonidis *et al*., "MLOps - Definitions, Tools and Challenges," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022. pp. 0453–0460. DOI 10.1109/CCWC54503.2022.9720902

[27] D. Ficzere *et al*., "AI Toolbox Concept for the Arrowhead Framework," in *2023 19th International Conference on Network and Service Management (CNSM)*, 2023. pp. 1–7. DOI: 10.23919/CNSM59352.2023.10327821

[28] N. Patki *et al*., "The synthetic data vault," in *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct 2016. pp. 399–410. DOI: 10.1109/DSAA.2016.49

[29] E. Bisong, *Kubeflow and Kubeflow Pipelines*. Berkeley, CA: Apress, 2019, pp. 671–685. ISBN 978-1-4842-4470-8. [Online]. Available: DOI: 10.1007/978-1-4842-4470-8_46

[30] A. Géron, Hands-on Machine Learning with Scikit-Learn, *Keras, and TensorFlow*. O'Reilly, 8 2019.

Simon Dahdal (Student Member, IEEE) is a PhD Student at the Engineering Department of the University of Ferrara, in the Distributed Systems Research Group. He currently works on projects that focus on the application of Big Data and Machine Learning in High-Stakes Environments, in particular Industry 4.0/5.0 and Human Assistance & Disaster Recovery scenarios.

Lorenzo Colombi He is a research assistant at the University of Ferrara, in the Distributed Systems Research Group. He currently works on projects that focus on the application of Big Data, Machine Learning, and Machine Learning Operations in High-Stakes Environments, such as Industry 4.0/5.0 and Human Assistance & Disaster Recovery scenarios.

Matteo Brina He is a research assistant at the University of Ferrara, in the Distributed Systems Research Group. He currently works on projects that focus on the application of Big Data and Machine Learning in Industrial Environments.

Alessandro Gilli He is a research assistant at the University of Ferrara, in the Distributed Systems Research Group. He currently works on projects that focus on the application of Big Data and Machine Learning Operations in both Industry 4.0/5.0 and Human Assistance & Disaster Recovery scenarios.

Mauro Tortonesi (Member, IEEE) is an Associate Professor and the head of the Big Data and Compute Continuum research laboratory at the University of Ferrara, Italy. He received the Ph.D. degree in computer engineering from the University of Ferrara, in 2006. He was a Visiting Scientist with the Florida Institute for Human & Machine Cognition (IHMC), Pensacola, FL, USA, from 2004 to 2005 and with the United States Army Research Laboratory, Adelphi, MD, USA, in 2015. He participates / has participated with several roles in a wide number of research projects in the distributed systems area, with particular reference to Compute Continuum, Big Data, and IoT solutions in industrial and military environments. He has co-authored over 100 publications and has 4 international patents.

Massimiliano Vignoli He is the Chief Data Scientist at Bonfiglioli S.P.A., currently leading several predictive maintenance projects. He has extensive experience in managing the full life cycle of machine learning applications, from development to production.

Cesare Stefanelli (Member, IEEE) received the Ph.D. degree in computer science engineering from the University of Bologna, Italy, in 1996. He is currently a Full Professor of distributed systems with the Engineering Department, University of Ferrara, Italy. At the University of Ferrara he coordinates a Technopole Laboratory dealing with industrial research and technology transfer. He holds several patents, and coordinates industrial research projects carried on in collaboration with several companies. His research interests include distributed and mobile computing in wireless and ad hoc networks, network and systems management, and network security.

MTR Model-Based Testing Framework

Gábor Árpád Németh and Máté István Lugosi

***Abstract*—In this article we propose a novel, free and open-source model-based testing framework for finite state machine specifications. The various model handling and test generation options make the framework suitable for testing complex systems and provide a solid background for investigating different automated test design methodologies. The complexity and fault detection capabilities of the available algorithms are investigated through analytical analyses and simulations applying randomly injected faults.**

***Index Terms*—model-based testing, conformance testing, finite state machine, test generation algorithm**

I. INTRODUCTION

In software development, testing is a critical, but often resource-intensive process. Although test execution is automated in most big software companies, test design is typically done manually, which – considering the rapidly changing complex products – tends to be an ad-hoc, error-prone and time-consuming approach. Model-based testing (MBT) turns this costly and labour intensive task into an automated process. In MBT, the requirements of the product are described as a formal model and the test cases are derived automatically from this model.

This article focuses on the MBT of Finite State Machine (FSM) specifications [9], [18], [22], which have been extensively used in different problem domains such as telecommunication software and protocols [16], [17], pattern matching [3], hardware design [26], and embedded systems [8]. A number of academic and commercial tools are developed to support MBT [6], [19]. Commercial products for FSM-like specifications include Conformiq Designer[1] and Reactis Tester[2], but these are not open-source. GrapWalker (GW)[3], fMBT[4], and Modbat[5] are free and open-source FSM-based tools that are actively under development. GW has an easy-to-use graphical user interface (GUI), but test generation is mainly done by random traversals; it lacks efficient systematic routing algorithms [30]. fMBT generates test cases from converted Extended FSMs using random and other heuristics to fulfill a given coverage (such as permutations of consecutive elements). Modbat is specialized to testing the application programming interface (API) of a software [4], test cases can be generated by heuristic search.

In this article we present a new, free and open-source model-based testing framework – called $Model \gg Test \gg Relax$ (MTR) – for FSM specifications. With MTR the test engineer can import specification models from GW, apply different conversions on the model and generate tests. The variety of systematic test generation algorithms and their different settings provide the potential to the test engineer to balance between quality aspects and the resources required for testing.

The body of the article is organized as follows. Section II discusses related terms regarding FSMs and MBT. Section III overviews the main functionalities of the MTR framework, Section IV describes its working process. The different test generation strategies are summarized in Section V, our new algorithm created for N-Switch Coverage is also briefly discussed here. Section VI presents simulations for test generation algorithms investigating the complexity of automated test creation, the size and the fault coverage of the resulting test suites. Possible future directions are discussed in Section VII, the main results of the paper are concluded in Section VIII.

II. PRELIMINARIES

A. Finite State Machines

A Mealy Finite State Machine (FSM) M is a quadruple $M = (I, O, S, T)$ where I, O, S and T are the finite and non-empty sets of *input symbols*, *output symbols*, *states* and *transitions* between states, respectively. Each transition $t \in T$ is a quadruple $t = (s_j, i, o, s_k)$, where $s_j \in S$ is the start state, $i \in I$ is an input symbol, $o \in O$ is an output symbol and $s_k \in S$ is the next state. The number of states, inputs and transitions are denoted by n, p and m, respectively.

An FSM can be represented with a *state transition graph*, which is a directed labeled graph whose nodes and edges correspond to the states and transitions, respectively. Each edge is labeled with the input and the output, written as i/o, associated with the transition.

FSM M is *deterministic*, if for each (s_j, i) state-input pair there exists at most one transition in T, otherwise it is *non-deterministic*. If there is at least one transition $t \in T$ for all state-input pairs, the machine is said to be *completely specified*, otherwise it is *partially specified*. In case of deterministic FSMs the output and the next state of a transition can be given as a function of the start state and the input of a transition, where λ: $S \times I \rightarrow O$ denotes the *output function* and δ: $S \times I \rightarrow S$ denotes the *next state function*. Let us extend δ and λ from input symbols to finite *input sequences* I^* as follows: for a state s_1, an input sequence $x = i_1, \ldots, i_k$ takes the machine successively to states $s_{j+1} = \delta(s_j, i_j)$, $j = 1, \ldots, k$ with the final state $\delta(s_1, x) = s_{k+1}$, and produces an *output sequence* $\lambda(s_1, x) = o_1, \ldots, o_k$, where $o_j = \lambda(s_j, i_j)$, $j = 1, \ldots, k$. An FSM M is *strongly connected* iff for each pair of states $(s_j\ ,\ s_l)$, there exists an input sequence which takes M from s_j to s_l.

[1] Gábor Árpád Németh and Máté István Lugosi are with the Department of Computer Algebra, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary, (e-mail: nga@inf.elte.hu, mate.lugosi@gmail.com)

[1] Conformiq Designer, https://www.conformiq.com/products/

[2] Reactis Tester, https://www.reactive-systems.com/products.msp

[3] GraphWalker, https://graphwalker.github.io/

[4] fMBT, https://github.com/intel/fMBT

[5] Modbat, https://gitlab.com/cartho/modbat

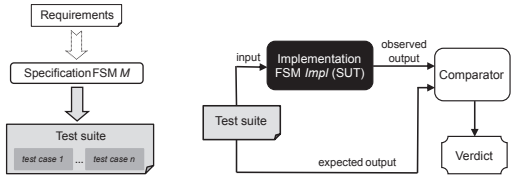
DOI: 10.36244/ICJ.2024.2.2

Two states, s_j and s_l of FSM M are *distinguishable*, iff there exists an $x \in I^*$ input sequence – called a *separating sequence* – that produces different output for these states, i.e.: $\lambda(s_j, x) \neq \lambda(s_l, x)$. Otherwise states s_j and s_l are *equivalent*. A machine is *reduced*, if no two states are equivalent.

An FSM M has a *reset message*, if there exists a special input symbol $r \in I$ that takes the machine from any state back to the s_0 initial state: $\exists r \in I : \forall s_j : \delta(s_j, r) = s_0$. The *reset is reliable* if it is guaranteed to work properly in any implementation machine $Impl$ of M; otherwise it is *unreliable*. A machine with reset capability is strongly connected, iff each state $s_j \in S$ is reachable from s_0.

The Extended Finite State Machine (EFSM) is an extension of the FSM formalism with variables, actions and guarding conditions over variables.

B. Model-based testing



(a) Test generation (b) Conformance testing

Figure 1. FSM model-based test generation and testing

The process of FSM model-based test generation is illustrated in Figure 1(a): From the requirements, an FSM M specification model is created. *Test cases* – which are the pairs of input sequences and expected output sequences of M – are generated automatically from this model. A set of test cases forms a *test suite*. The resulting test suite can be applied to the System Under Test (SUT) which can be considered as an $Impl$ implementation machine of specification M with an unknown internal structure, thus one can only observe its output responses upon a given input sequence – see Figure 1(b). *Conformance testing* checks if the *observed output sequences* of $Impl$ are equivalent to the *expected output sequences* derived from M – i.e. it determines if $Impl$ *conforms* to M.

Note that to connect the specification model to an actual SUT, a source code, called *adaptation code* needs to be created, that adapts the specification model to the SUT[6]. The adaptation code implements each element of the specification model as *keywords*. Such keywords are created for each transition of the specification model. Utilizing the adaptation code, one can transform abstract test cases into executable ones to effectively test the SUT.

C. FSM Fault Models

Fault models describe the assumptions of the test engineer about the implementation machine (s)he is about to test. For completely specified, deterministic FSMs the following three types of faults were proposed [10]:

I. Output fault: for a given state-input pair, FSM $Impl$ produces an output that differs from the one specified in machine M.
II. Transfer fault: for a given state-input pair, $Impl$ goes into a state that differs from the one specified in M.
III. Missing state or extra state

For non-deterministic and partially defined FSMs, the fault model above was extended with the following [7]:

IV. Missing or additional transitions

A usual assumption made in literature is that the faults do not increase the number of the states of the machine [18], thus the fault models of Chow [10] and Bochmann et al. [7] are typically restricted to output and transfer faults [18].

III. Overview of the Model ≫ Test ≫ Relax Framework

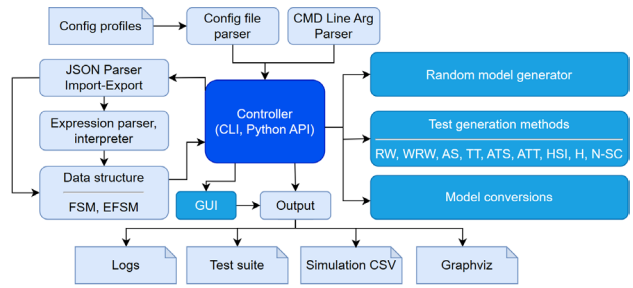


Figure 2. High level overview of $Model \gg Test \gg Relax$ framework

Figure 2 presents a high level overview of the architecture of the $Model \gg Test \gg Relax$ (MTR) model-based testing framework[7]. The user can import existing FSM or EFSM models or generate random ones, and can also make conversions on models (see Section IV-A). A wide range of algorithms can be utilized for test suite generation (see Section V) and an interface file can also be created that can be used as a skeleton for adaptation code creation (see Section IV-C). The parameters of the tool can be set by Command Line Interface (CLI) or by configuration profiles. Note that three different configuration profiles are delivered with the framework, optimized for testing, research and education purposes, respectively. Besides the generated test suite, the tool provides the following files to evaluate the results:

- *logs*: Six verbosity levels can be selected.
- *CSV file*: Comma-separated values summarize the main parameters of the model, in addition to the parameters and the results of the selected test generation algorithm.
- *Graphviz*[8] *file*: The models and the results of the applied test generation method can be visualized using this file.

The framework was implemented in C++ using the LEMON[9] library.

[6]Sometimes it is also referred as "glue code" as it glues the model and SUT together. In some cases – based on the abstraction level of the specification model and the applied testing tools – this adaptation code can be partially or completely generated.

[7] ***Model*** ***≫*** ***Test*** ***≫*** ***Relax***. https://modeltestrelax.org/, https://gitlab.inf.elte.hu/nga/ModelTestRelax

[8] Graphviz. Graph visualization software. https://graphviz.org/

[9] Library for Efficient Modeling and Optimization in Networks (LEMON), http://lemon.cs.elte.hu

IV. WORKING PROCESS OF THE MODEL ≫ TEST ≫ RELAX FRAMEWORK

A. Model making and manipulations

1) Model import: The specification model is defined in a JSON[10] format that is similar[11] to the one used by GraphWalker (GW). Thus, the user is able to create a model in the GUI of GW Studio and import it to our framework.

2) Model generation: It is possible to generate random FSMs with different parameters for simulation purposes.

3) Model conversions and manipulations: MTR provides the following conversion options to manipulate models:

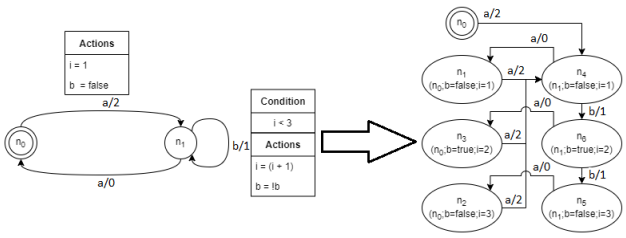


Figure 3. EFSM → FSM model conversion

- *EFSM → FSM model conversion:* For each possible state-variable value combination (that can be reached within the EFSM from the initial state considering the actions and guarding conditions of transitions[12]) a distinct state will be created in the converted FSM – see Figure 3. The conversion results in the well-known state explosion problem [18], but one can limit the range of variables. FSM test generation methods can be applied on the converted model and the adaptation keywords need to be implemented only once for each transition of the EFSM specification (parameterized by variables)[13].
- *Partially specified → completely specified conversion:* For each undefined state and input symbol pair a loop transition is added without an output symbol.
- *State minimization:* Helps the design phase of the formal specification by converting non-reduced machines into reduced ones merging equivalent states.
- *Add/remove reset*: Add/remove reliable or unreliable reset transitions to the model in one step.
- *Error injection:* Model-based mutation testing (MBMT) [5] can be applied by injecting given number of random transfer, output, missing or additional transition faults to the model. With this functionality one can investigate the fault detection capabilities of different test generation methods with simulations that apply the test suites generated from unmodified models to modified models.

[10] JSON. https://www.json.org/

[11] There are some differences in the model handling of GW and MTR as GW does not follow the (E)FSM formalism completely. Thus, some conversion is required if one would like to import the model of GW into MTR, but this is described in detail in the ”5.1.2. Editing models using GraphWalker Studio” section of MTR User Guide.

[12] If some states cannot be reached – thus they are not added to the converted model – then MTR displays a warning message.

[13] Note that an application example (OpenIddict) is delivered with MTR that tests the main functionalities of the Oauth 2.0 [2] protocol using an EFSM model and shows how the EFSM → FSM conversion and the testing on the converted model works.

B. Test generation

The framework contains numerous algorithms for test suite derivation with varying complexities and fault coverages enabling the test engineer to find an appropriate trade-off between resources allocated for the generation and execution of tests and the quality of the SUT. The available methods are described in Section V.

C. Test execution

MTR generates an interface file that contains the elements of the model and can be used when writing the adaptation code. The adaptation code should contain the following steps:

- STEP 1: Parse the next element of the test suite – that consists of an input/output list – generated by MTR.
- STEP 2: Execute actions corresponding to the given element of the input list.
- STEP 3: Check the result with assert if it corresponds to the one that can be expected from the given element of the output list.

Note that sample test projects are also delivered with the framework[14] which can be utilized as a base to understand the modelling and adaptation code creation processes before one creates an own test project. Each project contains the following parts:

- An FSM or EFSM model that describes the specification of the SUT.
- An adaptation code which implements each element of the specification model.
- A SUT that is provided by an external link.

Also note that MTR provides an option to export the generated test suites into GW and thus write the corresponding adaptation code there.

V. TEST GENERATION ALGORITHMS

Table I summarizes the available test generation algorithms in MTR and their main properties. A brief description is given for all algorithms in the following subsections and simulation results are presented in Section VI.

A. Random Walk (RW)

Random Walk (RW) starts from the initial state of the machine and in each step a transition leading from the current state is selected randomly and traversed entering a new state. The former step is executed until a given stop condition – the preset percentage of states or transitions have been visited – is fulfilled. MTR also provides an option to set selection probabilities for each transition of the model.

Although RW is unsuitable for the functional testing of large-scale software (as the length of the test sequence is not bounded and thus can be much longer than the optimal solution) and for regression testing (due to the randomness of transition traversals), it can be useful for exploratory testing of a new functionality.

[14] These projects can be accessed in folder *sample_models / applications*

TABLE I
THE MAIN ASPECTS OF TEST GENERATION ALGORITHMS

Algo.	Model	Complexity of test generation	Complexity of test suite	Structure of test suite	Coverage and other notes
RW	FSM / EFSM	Not bounded	Not bounded	1 test sequence	Given percentage of state/transition coverage (based on stop condition).
AS	FSM	$O(n^2)$	$O(m)$	1 test sequence	100% state coverage
TT	FSM	$O(n^3 + m)$	$O(m)$	1 test sequence	100% state and transition coverage. Guarantees to find output faults.
ATS	FSM	ATS0 (standard): $O(n^3 + m)$, ATSa/x (iterative): $O(\eta(n^3 + m))$, $\eta < 2 \cdot n$	ATS0: $O(m)$, ATSa/x: $O(\eta \cdot m)$, $\eta < 2 \cdot n$	1 test sequence (with subparts)	100% state and transition coverage. Guarantees to find output faults.
ATT	FSM	$O(m(n^3 + m))$	$O(m^2)$	1 test sequence (with subparts)	100% state and transition coverage. Guarantees to find output faults.
HSI	FSM	$O(p^{\theta+1} \cdot n^3)$	$O(p^{\theta+1} \cdot n^3)$	Structured test suite with multiple test sequences	Guarantees to discover output and transfer faults and θ extra states.
H	FSM	$O(p^{\theta+1} \cdot n^3)$	$O(p^{\theta+1} \cdot n^3)$	Structured test suite with multiple test sequences	Guarantees to discover output and transfer faults and θ extra states. Improvement of the HSI-method
N-SC	FSM	$O((N+1) \cdot m^{(k+1)(N+1)}, k = 0..N$	$O((N+1) \cdot m^{N+1})$	1 test sequence	Covers all topologically possible, consecutive $N+1$ transitions.

B. All-State (AS)

The All-State (AS) test generation method produces a test sequence that visits every state of a deterministic, strongly connected[15] FSM at least once. It applies the Nearest Neighbour (NN) heuristic [15] which searches in each step for the closest unvisited state until such state exists.

The AS test generation has $O(n^2)$ time complexity, the length of the generated test sequence is $O(m)$.

C. Transition Tour (TT)

The Transition Tour (TT) [23] algorithm generates the shortest test sequence that visits all transitions of a deterministic, strongly connected[15] FSM at least once, then returns to the initial state.

The problem of creating the test sequence above was reduced to the Directed Chinese Postman Problem (DCPP) [12] with unit costs for the edges of graph G (where G corresponds to specification machine M). The related algorithm [12], [20], [25] consists of the following two parts:

I. Augmenting the original graph G by duplicating some edges to make it Eulerian graph G_E.
II. Finding an Euler tour over G_E.

The time complexity of TT test generation and the length of the resulting test sequence is $O(n^3 + m)$ and $O(m)$, respectively. The generated test sequence guarantees to discover all output faults, but does not guarantee to find transfer faults.

D. All-Transition-State (ATS)

The All-Transition-State (ATS) algorithm [31] creates a test suite for deterministic, strongly connected[15] FSMs that fulfills the first two formal conditions of the ATS criteria [13]:

I. For all t transitions: The test suite should cover at least one walk that contains t and then reaches all states of M.
II. There has to be at least one walk to all states which does not include transition t (if feasible).

The ATS algorithm uses a preamble part responsible for traversing all transitions of FSM M first, then a postamble part responsible for traversing all states of M to fulfill both conditions, but on different graphs. These different graphs include the state transition graph of the specification FSM M and its subgraphs, where some t transitions are filtered out. The preamble part is realized using the TT algorithm without returning to the initial state at the end and the postamble part applies the NN heuristic [15] which searches in each step for the closest unvisited state until such state exists.

There are 3 different versions of the ATS algorithm:

1. Standard version (ATS0),
2. Iterative version without iteration limit (ATSa),
3. Iterative version with iteration limit (ATSx).

The three versions of the algorithm differ in how condition II can be fulfilled. The user has the choice to find a trade-off between coverage and the overall length of the test suite. ATS0 has a total complexity of $O(n^3 + m)$ and an $O(m)$ overall length for the test suite [31]. ATSa and ATSx have a total complexity of $O(\eta(n^3 + m))$, where $\eta < 2 \cdot n$ and the total length of the resulting test suite is $O(\eta \cdot m)$ [31]. The generated test sequence guarantees to discover all output faults and to find most of transfer faults [31].

E. All-Transition-Transition (ATT)

This algorithm is the naive implementation of the first two conditions of the All-Transition-Transition (ATT) criteria [13]:

I. For all t transitions: The test suite should cover at least one walk that contains t and then reaches all transitions of the FSM.
II. There has to be at least one walk to all transitions which does not include transition t (if feasible).

For condition I, the ATT algorithm uses a preamble part that traverses all transitions of the FSM, then a postamble part that traverses all transitions of the machine again. Condition II can be fulfilled in a similar way, but the preamble part is applied on different filtered graphs of the specification.

The complexity of ATT test generation and the length of the test sequence is $O(m \cdot (n^3 + m))$ and $O(m^2)$, respectively.

F. Harmonized State Identifiers (HSI)

In this algorithm, the Harmonized State Identifiers (HSI) state verification method [21], [27] is applied to create a structured test suite for reduced, deterministic, strongly connected FSMs with reliable reset[16] capability [29]. The algorithm contains the following main parts:

- A *state cover set (SCS)* $Q = \{q_0, \ldots, q_{n-1}\}$ that is used for reaching all states; the problem was reduced to create a spanning tree from the s_0 initial state.

[15]If reset transitions exist, MTR applies them in the test suite only if the strongly connected assumption cannot be fullfilled without them.

[16]If the model has unreliable resets, then MTR generates a distinct test suite first, that checks if all reset transitions are implemented in the SUT properly.

- A *separating family of sequences* of Z responsible for verifying end states. The Z set is a collection of sets $Z_j, j = 0, \ldots, n-1$ of sequences (one set for each state) where for every non-identical pair of states s_j, s_l there exists a separating sequence. In our implementation, the Z set is represented with a spanning forest over an auxiliary state pair graph, the edges of which are directed to state pairs that have a separating input.

Based on the parts discussed above, the algorithm consists of two stages. The first stage identifies all states of the FSM and the second stage checks all remaining transitions. The resulting algorithm is the generalization of the W [10] and Wp [14] methods and it guarantees to find all output and transfer faults of FSM $Impl$.

The total length of the resulting test suite and the complexity of test generation is $O(p \cdot n^3)$ [29]. By extending the method above it will also guarantee to find θ given number of extra states in the implementation, resulting $O(p^{\theta+1} \cdot n^3)$ complexity.

G. H-method

The H-method [11] creates a test suite for reduced, deterministic, strongly connected FSMs with reliable reset[16]. The resulting test suite guarantees to discover all output and transfer faults and preset θ number of extra states in $Impl$.

Just like the HSI-method, the H-method also uses a Q SCS to travel to states that need to be verified. It also uses Harmonized State Identifiers for state identification and transition checking, but instead of using predetermined state identifiers, it selects the appropriate ones on-the-fly, thus shortening the test suite.

The algorithm consists of 4 stages:

- STAGE 1: Let the test suite be the SCS sequences, extended by every possible $\theta + 1$ long permutation of the input symbols.
- STAGE 2: For each two sequences u and v of the SCS Q, check if the test suite has sequences uw and vw such that w distinguishes the states $\delta(s_0, u)$ and $\delta(s_0, v)$. If there are no such sequences, select an appropriate w and add uw and vw to the test suite.
- STAGE 3: For each sequence $u\alpha$ where u is in the SCS Q, and α is a sequence of the input symbols with a length up to $\theta+1$, and a v sequence which is also in the SCS Q, check if the test suite has sequences $u\alpha w$ and vw such that w distinguishes the states $\delta(s_0, u\alpha)$ and $\delta(s_0, v)$. If there are no such sequences, select an appropriate w and add $u\alpha w$ and vw to the test suite.
- STAGE 4 (if $\theta > 0$): For each sequence $u\alpha$ where u is in the SCS Q, and α is a sequence of the input symbols with a length up to $\theta + 1$, and for each $u\beta$ where β is a prefix of α, check if the test suite has sequences $u\alpha w$ and $u\beta w$ such that w distinguishes the states $\delta(s_0, u\alpha)$ and $\delta(s_0, u\beta)$. If there are no such sequences, select an appropriate w and add $u\alpha w$ and $u\beta w$ to the test suite.

The complexity of the test generation and the resulting test suite is $O(p^{\theta+1} \cdot n^3)$ [28].

Note that although the original paper [11] mentioned that the length of the test suite depends on the order in which transitions are checked, no corresponding method is described for this. MTR proposes different strategies for processing transitions. We found that the most effective solution is that when the transitions are sorted by input symbols that produce the most diverse output symbols (i.e. the algorithm prefers input symbols in processing that are able to separate more states), the results presented in Section VI apply this approach.

H. N-Switch Coverage (N-SC)

N-Switch Coverage (N-SC) [10] covers all topologically possible, consecutive $N + 1$ transitions of reduced, deterministic, strongly connected FSMs. Note that article [10] only introduced the criteria that need to be fulfilled, but no corresponding algorithm had been given and to the best of our knowledge there still hasn't been published any.

Thus, we created a new heuristic algorithm for N-SC, that takes N and a k iteration limit (scaling from 0 to N) as input parameters and briefly works as follows:

STEP 1: Initialization:

- Set test sequence ts as empty: $ts := \{\}$.
- Set current state to the initial state: $s_c := s_0$.
- Create an ordered list $L = \zeta_1 ... \zeta_K$ for all possible, consecutive $N+1$ transitions (including loop transitions) in FSM M. Mark all $\zeta_1 ... \zeta_K \in L$ elements as *uncovered*.
- Initialize the next element of L to be covered: $\zeta^n := \{\}$.

STEP 2: Covering $N + 1$ transitions: Repeat until all elements of L are marked as *covered*:

- STEP 2.1: Select an ζ_x element of L that is marked as *uncovered* and for which its s_y start state is the nearest[17] from the s_c current state: $\zeta^n := \zeta_x$.
  - If $s_c \neq s_y$: Add the $s_c \rightarrow s_y$ path into ts. $s_c := s_y$.
- STEP 2.2: Add the next transition $t = (s_j, i, o, s_l)$ of ζ^n to ts. Set s_c to the end state of t: $s_c := s_l$.
- STEP 2.3: Check if ζ^n is covered:
  - If yes:
    - STEP 2.3.1: Mark the element in L corresponding to ζ^n as *covered*.
    - STEP 2.3.2: Check with k iteration limit if a ζ_i element of L marked as *uncovered* is partially covered with the last k steps, i.e. the first k elements of ζ_i is covered with the last k transitions in ts:
      - If yes: $\zeta^n := \zeta_i$ and continue with STEP 2.2.
      - Otherwise: continue with STEP 2.
  - If no:
    - Continue with STEP 2.2

Note that by changing iteration limit k, one can create a trade-off between the length of the resulting test sequence and the test generation time.

List L contains maximum $O(m^{N+1})$ elements each one with a length of $N + 1$, thus the length of the resulting test sequence is $O((N + 1) \cdot m^{N+1})$. STEP 2 iterates over all elements of L and for each element, STEP 2.2 adds test sequences with $N + 1$ length. STEP 2.3.2 checks partially covered elements with $O(m^{k \cdot (N+1)})$ worst case complexity,

[17] This can be found by breadth-first search from s_c.

resulting in $O((N+1)\cdot m^{(k+1)\cdot(N+1)})$ test generation complexity[18].

VI. SIMULATION RESULTS

The simulations were executed on servers running an Ubuntu 22.04.2 LTS operating system with 4 GB memory and one core of a shared AMD EPYC 7763 64-core CPU with 2445 MHz clock frequency.

Strongly connected, reduced, completely specified, deterministic random FSMs with reliable reset capability were generated with MTR in different Scenarios to investigate the performance of the algorithms – see Table II.

TABLE II
INVESTIGATED SCENARIOS

ID	Number of states			Density / $\lvert I\rvert$	$\lvert O\rvert$	simulation goal
	min.	max.	size of step			
Scenario 1	5	2000	5	5	10	complexity
Scenario 2	5	100	5	5	2	fault cov.
Scenario 3	5	100	5	5	10	fault cov.

A. Complexity investigations

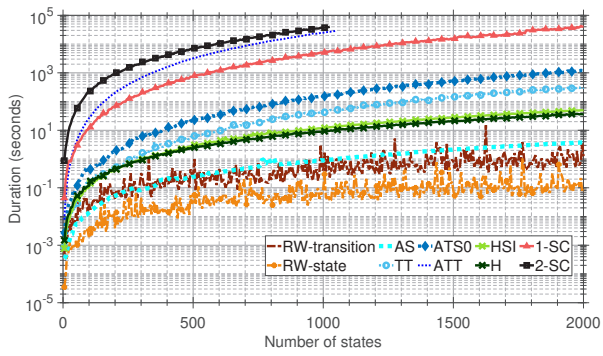


Figure 4. Scenario 1: Test generation time

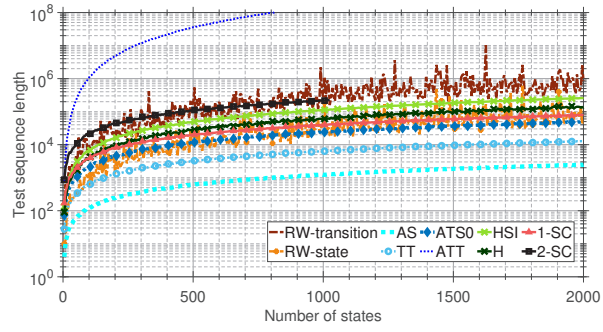


Figure 5. Scenario 1: Test sequence length

[18]In case of completely specified, deterministic FSMs, L contains $O(p^{N+1})$ elements, thus the complexity of the test sequence and test generation is $O((N+1)\cdot p^{N+1})$ and $O((N+1)\cdot p^{(k+1)\cdot(N+1)})$, respectively.

Scenario 1 investigates the time required for test generation and the overall length of the test sequences in function of the number of states in the specification machine.

Figure 4 shows the test generation time for RW (both with 100% state and 100% transition coverage), AS, TT, ATS (with ATS0 standard version), ATT, HSI (with $\theta = 0$), H (with $\theta = 0$), 1-SC (with $k = 1$) and 2-SC (with $k = 2$) algorithms[19]. Figure 5 shows the overall length of the resulting test sequences for the same settings.

As expected, the test generation time of 1-SC, 2-SC and ATT is the longest. The complexity of TT and ATS0 test generation is around the cubic function of the number of the states. The test generation complexity is less than the theoretic cubic upper limit in case of the HSI and the H method because each member of the separating family of sequences typically consists of a test sequence with a length of 1 or 2 instead of the theoretical worst case of $n-1$ length. The H performs better as it is an improvement of the HSI. AS solves a much easier problem to visit all states with NN that is reflected in its complexity. The test generation time of RW is the least as it only selects a transition randomly and checks the stop condition at each step.

The length of the test suite is the shortest in case of AS, that only visit states with NN. The size of the test suite is a linear function of the number of states in case of TT and ATS0. The ones generated by the HSI and H are significantly bigger as in this case the test suite systematically checks all states and verifies the end states of the remaining transitions, although they are much shorter than the theoretic upper limit due to the reason discussed previously. Its size is between 1-SC and 2-SC and the improvement of the H over HSI can be clearly seen. The ATT performs the worst as in each step it tries to create a transition adjacent walk before visiting all transitions.

B. Fault coverage investigations

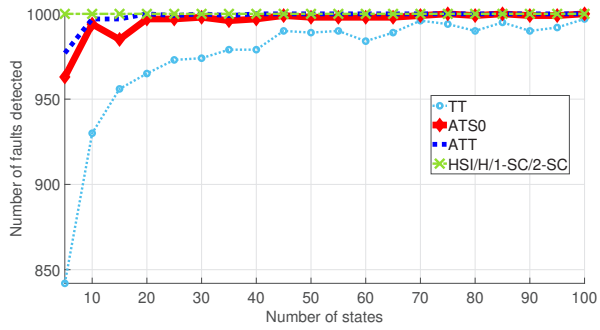


Figure 6. Scenario 2: Number of discovered faults

In Scenarios 2 and 3 the fault coverage of test suites generated by different algorithms is investigated with randomly injected transfer faults[20]. Each data point in these scenarios

[19]ATT runs out memory above 1050 states, 2-SC is investigated only up to 1025 states as its execution time grows rapidly.

[20]Transfer faults are selected for investigations because of the reasons described in Section II-C and the fact that output faults are guaranteed to be found by algorithms that traverse all transitions of the specification FSM.

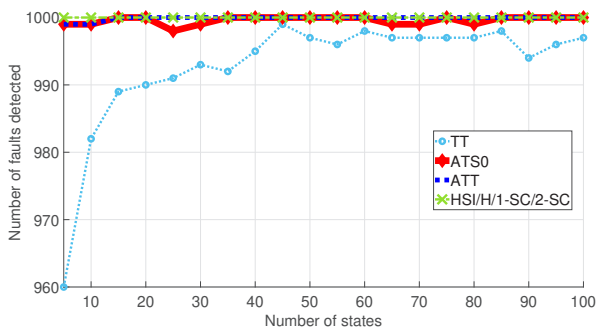


Figure 7. Scenario 3: Number of discovered faults

is obtained by 1000 simulation runs; in each simulation a single transition fault is injected into a distinct random FSM with given parameters and it is observed how many times from these 1000 distinct cases do the algorithms discover the fault. The results for FSMs with 2 and 10 output symbols are presented in Figures 6 and 7, respectively[21]. As expected H, HSI, 1-SC and 2-SC discovered all 1000 transfer faults regardless of the number of states. The ATS0 and ATT algorithms perform just a little worse and TT gives results that can still be acceptable depending on the application domain. In Scenario 3 there are more possible output symbols than in Scenario 2, thus the fault coverage of the test generation algorithms increases, but the trends are similar.

C. SIP UAC registration example

Simulations were also performed to investigate the overall length and the fault coverage of the generated test suites for the specification machine presented in Figure 8 which describes the SIP (Session Initiation Protocol) [1] registration process of the User Agent Client[22].

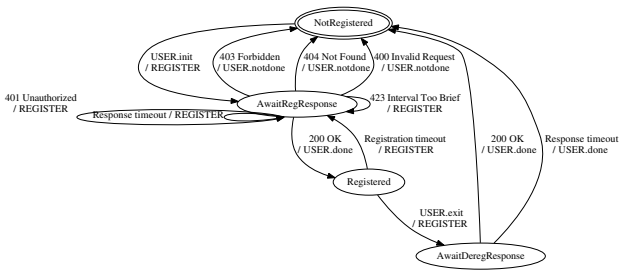


Figure 8. FSM for the registration process of the SIP user agent client

The overall length of the test sequence generated by AS, TT, ATS0, ATT, HSI (with $\theta = 0$), H (with $\theta = 0$), 1-SC (with $k = 1$) and 2-SC (with $k = 2$) is 3, 19, 47, 65, 32, 32, 76 and 372 transitions, respectively.

[21]Note that the results of AS algorithm are not presented in the figures to be able to discover the fault coverage of more robust algorithms precisely. In Scenario 2 AS discovers only 85-184 faults at and below 20 states and 206-256 faults at and above 25 states. In Scenario 3 AS finds 136-222 faults at and below 20 states and 202-261 faults at and above 25 states.

[22]Here only the signaling level was considered; a description about how this FSM can be constructed from the related call-flows is presented in [24].

As the machine contains 4 states and 12 transitions, $12 \cdot (4-1) = 36$ diverse transition faults are possible. Thus, 36 faulty models were created and the fault coverage of the different test suites was investigated. The AS and TT were able to find 6 and 33 faults, respectively. The ATS0, ATT, HSI, H, 1-SC and 2-SC were able to discover all 36 possible faults.

VII. FUTURE WORKS

We have the following plans for future MTR enhancements.

First, we would like to introduce incremental test generation algorithms which identify the effects of changes in the test suite derived from a previous system specification and only update those parts that are necessary. Thus, test generation time can be significantly reduced and different testing goals (such as regression testing, sanity testing) can be clearly isolated from each other.

We also plan to extend our framework to handle Communicating Finite State Machine models and timers which are essential in reliable communication protocols.

As a long term goal, we would like to apply some upper level logic which based on input data – the structure of the specification model, the problem domain, the testing goals and the resources allocated for testing – can automatically propose a test suite that best suits the needs of the test engineer.

VIII. CONCLUSION

In the current article, we introduced a novel model-based testing framework that can be used in the systematic testing of complex software in diverse problem domains such as infocommunications. The framework offers a wide range of model conversion and test generation options.

The test criteria and test coverage can be fine-tuned by selecting a given test generation algorithm and its parameters. The related algorithms were summarized, and a new heuristic test generation algorithm for the N-Switch Coverage Criteria [10] has also been presented. The complexity of test generation and the size of the resulting test suite for the implemented test generation algorithms were investigated via analytical worst case complexity calculations and by empirical analyses. The fault coverage of the generated tests was also observed by simulations. The results let the test engineer find an appropriate trade-off between sources allocated for test execution and the coverage of tests depending on testing goals.

ACKNOWLEDGEMENTS

The first author was supported by the project "Software and Data-Intensive Services" Nr. 2019-1.3.1-KK-2019-00011 financed by the Hungarian National Institute of Science and Innovation.

The authors would like to thank the students who took part in implementation of the following part of the framework: Zsolt Csáky for EFSM handling and EFSM → FSM transformation, Tódor Dávid Nyeste for H and HSI-methods, Tomás Varga for N-SC test generation. The authors would also like to express their gratitude to Levente Hunyadi for his valuable technological advice.

REFERENCES

[1] RFC 3261: SIP: Session Initiation Protocol, 2002. https://tools.ietf.org/html/rfc3261 Accessed: 2023-07-04.

[2] RFC 6749: The OAuth 2.0 Authorization Framework, 2012. https://datatracker.ietf.org/doc/html/rfc6749 Accessed: 2024-01-09.

[3] Paul Ammann and Jeff Offutt. *Introduction to Software Testing*. Cambridge University Press, New York, NY, USA, 2nd edition, 2016. DOI: 10.1017/9781316771273.

[4] Cyrille Artho, Martina Seidl, Quentin Gros, Eun-Hye Choi, Takashi Kitamura, Akira Mori, Rudolf Ramler, and Yoriyuki Yamagata. Model-based testing of stateful APIs with Modbat. In *Proc. 30th Int. Conf. on Automated Software Engineering (ASE 2015)*, pages 858–863, Lincoln, USA, 2015. IEEE. DOI: 10.1109/ASE.2015.95.

[5] Fevzi Belli, Christof J. Budnik, Axel Hollmann, Tugkan Tuglular, and W. Eric Wong. Model-based mutation testing—approach and case studies. *Science of Computer Programming*, 120:25–48, 2016. DOI: 10.1016/j.scico.2016.01.003.

[6] Maicon Bernardino, Elder M. Rodrigues, Avelino F. Zorzo, and Luciano Marchezan. Systematic mapping study on MBT: tools and models. *IET Software*, 11(4):141–155, 2017. DOI: 10.1049/iet-sen.2015.0154.

[7] Gregor von Bochmann, Anindya Das, Rachida Dssouli, Martin Dubuc, Abderrazak Ghedamsi, and Gang Luo. Fault Models in Testing. In *Proceedings of the IFIP TC6/WG6.1 Fourth International Workshop on Protocol Test Systems IV*, pages 17–30, Amsterdam, The Netherlands, 1991. North-Holland Publishing Co.

[8] Eckard Bringmann and Andreas Krämer. Model-based testing of automotive systems. In *Proceedings of the 2008 International Conference on Software Testing, Verification, and Validation, ICST* '08, pages 485–493, Washington, DC, USA, 2008. IEEE Computer Society. DOI: 10.1109/ICST.2008.45.

[9] Manfred Broy, Bengt Jonsson, Joost-Pieter Katoen, Martin Leucker, and Alexander Pretschner (Eds.). *Model-Based Testing of Reactive Systems*. Springer, 2005. DOI: 10.1007/b137241.

[10] T. Chow. Testing software design modelled by finite-state machines. *IEEE Transactions on Software Engineering*, 4(3):178–187, May 1978. DOI: 10.1109/TSE.1978.231496.

[11] Rita Dorofeeva, Khaled El-Fakih, and Nina Yevtushenko. An Improved Conformance Testing Method. In Farn Wang, editor, *Formal Techniques for Networked and Distributed Systems – FORTE 2005*, volume 3731 of Lecture Notes in Computer Science, pages 204–218. Springer, Berlin, Heidelberg, 2005. DOI: 10.1007/11562436_16.

[12] Jack Edmonds and Ellis L. Johnson. Matching, Euler tours and the Chinese postman. *Mathematical Programming,* 5(1):88–124, 1973. DOI: 10.1007/BF01580113.

[13] István Forgács and Attila Kovács. *Practical Test Design*. BCS, The Chartered Institute for IT, 2019.

[14] S. Fujiwara, G. v. Bochmann, F. Khendec, M. Amalou, and A. Ghedamsi. Test selection based on finite state model. *IEEE Transactions on Software Engineering*, 17(6):591–603, 1991. DOI: 10.1109/32.87284.

[15] Gregory Z. Gutin, Anders Yeo, and Alexey Zverovich. Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discret. Appl. Math.*, 117(1-3):81–86, 2002. DOI: 10.1016/S0166-218X(01)00195-0.

[16] Drago Hercog. *Protocol Specification and Design. In Communication Protocols*. Springer, Cham, 2020. DOI: 10.1007/978-3-030-50405-2_2.

[17] Gerard J. Holzmann. *Design and Validation of Protocols*. Prentice-Hall, 1990.

[18] David Lee and Mihalis Yannakakis. Principles and Methods of Testing Finite State Machines – A Survey. *Proceedings of the IEEE*, 84(8):1090–1123, 1996. DOI: 10.1109/5.533956.

[19] Wenbin Li, Franck Le Gall, and Naum Spaseski. A survey on model-based testing tools for test case generation. In *Tools and Methods of Program Analysis*, pages 77–89. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-71734-0_7.

[20] Y. Lin and Y. C. Zhao. A new algorithm for the directed chinese postman problem. *Computers and Operations Research*, 15(6):577–584, 1988. DOI: 10.1016/0305-0548(88)90053-6.

[21] Gang Luo, Alexandre Petrenko, and Gregor V. Bochmann. Selecting Test Sequences for Partially-Specified Nondeterministic *Finite State Machines. In Proceedings of the IFIP WG6.1 7th International Workshop on Protocol Test systems VI*, pages 91–106. Springer, 1995. DOI: 10.1007/978-0-387-34883-4_6.

[22] Matheus Monteiro Mariano, Érica Ferreira de Souza, André Takeshi Endo, and Nandamudi Lankalapalli Vijaykumar. *Comparing graph-based algorithms to generate test cases from finite state machines. Journal of Electronic Testing*, 35(11–12):867–885, December 2019. DOI: 10.1007/s10836-019-05844-6.

[23] S. Naito and M. Tsunoyama. Fault detection for sequential machines by transition-tours. In *Proceedings of the 11th IEEE Fault-Tolerant Computing Conference (FTCS 1981)*, pages 238–243. IEEE Computer Society Press, 1981.

[24] Gábor Árpád Németh and Péter Sótér. Teaching performance testing. *Teaching Mathematics and Computer Science*, 19(1):17–33, 2021. DOI: 10.5485/TMCS.2021.0518.

[25] S. C. Orloff. A Fundamental Problem in Vehicle Routing. *Networks*, 4:35–64, 1974. DOI: 10.1002/net.3230040105.

[26] Volnei A. Pedroni. *Finite State Machines in Hardware. Theory and Design (with VHDL and SystemVerilog)*. The MIT Press, London, England, 2013. DOI: 10.7551/mitpress/9657.001.0001.

[27] Alexandre Petrenko, Nina Yevtushenko, Alexandre Lebedev, and Anindya Das. Nondeterministic state machines in protocol conformance testing. In *Proceedings of the IFIP TC6/WG6.1 Sixth International Workshop on Protocol Test Systems VI*, pages 363–378, NLD, 1993. North-Holland Publishing Co.

[28] M. Soucha and K. Bogdanov. SPYH-method: An improvement in testing of finite-state machines. *2018 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, pages 194–203, July 2018. DOI: 10.1109/ICSTW.2018.00050.

[29] Mihalis Yannakakis and David Lee. Testing finite state machines: Fault detection. *Journal of Computer and System Sciences*, 50(2):209–227, 1995. DOI: 10.1006/jcss.1995.1019.

[30] Muhammad Nouman Zafar, Wasif Afzal, Eduard Enoiu, Athanasios Stratis, Aitor Arrieta, and Goiuria Sagardui. Model-based testing in practice: An industrial case study using GraphWalker. In *14th Innovations in Software Engineering Conference*, ISEC 2021, New York, NY, USA, 2021. Association for Computing Machinery. DOI: 10.1145/3452383.3452388.

[31] Gábor Árpád Németh and Máté István Lugosi. Test generation algorithm for the All-Transition-State criteria of Finite State Machines. *Infocommunications Journal*, 13(3):56–65, 2021. DOI: 10.36244/ICJ.2021.3.6.

Gábor Árpád Németh obtained his MSc in Electrical Engineering and his PhD in Computer Science at the Budapest University of Technology and Economics (BME), Department of Telecommunication and Media Informatics (TMIT) in 2007 and 2015, respectively. He worked at Ericsson between 2011 and 2018 on a performance testing tool used in the telecommunication industry. Currently, he works at the Eötvös Loránd University (ELTE) on topics related to software testing.

Máté István Lugosi obtained his BSc in Computer Science at Eötvös Loránd University (ELTE) in 2021. Currently, he works at Ericsson on embedded software of microwave network devices. He studies in the MSc program of Computer Science at ELTE in the Cryptography specialization.

Automated checker for detecting method-hiding in Java programs

M. Z. I. Nazir, M. Alqaradaghi, and T. Kozsik

Abstract—**Method overriding is a valuable mechanism that happens when an instance method is defined in a subclass and has the same signature and return type as an instance method in the superclass. However, in Java, if those methods are static, then instead method hiding happens, which is a programming weakness and may produce unexpected results. Static analysis is an approach in software testing that examines code to identify various programming weaknesses throughout the software development process without running it.**

This paper addresses the detection of method-hiding problem in Java programs. We implemented a new automated checker under the SpotBugs static analysis tool that can detect the mentioned problem. According to our results, the checker precisely detected the related issues in both custom test cases and real-world programs.

Index Terms—**Java, method-hiding, precise automated checker, static analysis, SpotBugs tool**

I. INTRODUCTION

FREQUENT cyberattacks on IT infrastructures drive cybersecurity research [1]. It is crucial to keep software free of weaknesses. Method overriding (also called late binding, run-time polymorphism, and dynamic polymorphism) happens when an instance method is defined in a subclass and has the same signature (method's name, parameters' numbers, and parameters' types) and return type as an instance method in one of the superclasses. In this case, the method in the subclass will override the one in the superclass. This programming mechanism is valuable. It enables a class to derive from a superclass that exhibits similar behavior and subsequently customize and/or extend the behavior as required [2, 3]. While the compiler in Java does not require the *@Override* annotation[1] to be present for the overridden method. Doing so is advised for the following reasons:

1) The compiler will produce an error if the method is not present in one of the superclasses, informing the programmer that this is not actually overriding and that he must fix it.

2) If the overridden method is static, the compiler will generate another type of error [2] which will instruct about the necessity of removing that annotation because it is not possible to override a static method. Omitting the *@Override* annotation in the latter case will make the compiler ignore this issue, leading to the problem of method hiding.

More specifically, method hiding happens when a subclass defines a static method with the same signature and the return type as a static method in the superclass. The method in the superclass, in this case, hides the one in the subclass [2]. Overriding and hiding methods have distinct differences in determining which method is called from a specific location. In the case of overriding, the method called is determined during runtime based on the specific instance of the object being used. On the other hand, hiding determines the method called during compile time by considering the specific qualified name or method invocation expression used at the call [3].

Method hiding is neither considered an error nor a compilation failure. However, according to the SEI CERT Oracle Coding Standard for Java, method hiding should be avoided because it often leads to unexpected results, especially when programmers mistakenly expect method overriding. This has been clarified under Rule 06. Methods (MET) MET07-J [4]. Moreover, according to the same web page, no free automated static analysis tool can detect this issue in Java code. Static analysis approaches save time, effort, and money by identifying software flaws and security vulnerabilities early in the software development process [5, 6]. These techniques are capable of identifying a wide variety of security flaws and vulnerabilities, from simple programming errors to more complex concerns like access control difficulties [5].

The motivations of this paper are:

- According to the TIOBE index [7], Java is still one of the most widely used programming languages despite some decline in popularity.
- Java is used to create many long-lasting programs that we use on a daily basis. It is crucial to keep these applications up to date and fix any flaws.
- Static analysis techniques are useful for finding code flaws and security issues.

The contributions of the paper are:

- Design and implement an automated checker named *FindHidingMethod* under the SpotBugs static analysis tool (SB) [8], which raises an issue when finding method hiding bugs in Java programs.
- Assess our approach and report the assessment results using *recall*, *false alarm rate*, and *precision* metrics.

M. Z. I. Nazir was with the Department of Programming Languages and Compilers, ELTE, Eötvös Loránd University, Budapest, Hungary. He is now with the Technical University Munich (e-mail: bsvncs@inf.elte.hu).

M. Alqaradaghi is with the Department of Programming Languages and Compilers, ELTE, Eötvös Loránd University, Budapest, Hungary and Northern Technical University, Kirkuk, Iraq (e-mail: alqaradaghi.midya@inf.elte.hu).

T. Kozsik is with the Department of Programming Languages and Compilers, ELTE, Eötvös Loránd University, Budapest, Hungary
(e-mail: kto@inf.elte.hu).

[1] *@Override* annotation instructs the compiler that you intend to override a method in the superclass.

[2] error: static methods cannot be annotated with the *@Override*.

DOI: 10.36244/ICJ.2024.2.3

B. Problem Statement

The issue of defining a static method in a subclass that has the same signature as a static method in the superclass is known as method hiding. Here, the superclass method is hidden by the subclass method. It is important to avoid method hiding because it can lead to confusion and unexpected behavior, especially when programmers mistakenly expect method overriding. Listing 1 presents simple Java code for method overriding versus method hiding.

LISTING 1
METHOD OVERRIDING VERSUS METHOD HIDING

```
1 class SuperClass {
2       public static void methodHiding() {
3               System.out.println("methodHiding (SuperClass)");
4       }
5       public void methodOverriding() {
6               System.out.println("methodOverriding (SuperClass)");
7       }
8 }
9 class HidingVsOverriding extends SuperClass {
10      public static void methodHiding() {
11              System.out.println("Method Hiding (SubClass)");
12      }
13      public void methodOverriding() {
14              System.out.println("Method Overriding (SubClass)");
15      }
16 }
17 public class MainClass {
18      public static void main(String[] args) {
19              SuperClass bs3 = new HidingVsOverriding();
20              bs3.methodOverriding();
21              bs3.methodHiding();
22      }
23 }
```

```
Results:
Method Overriding (SubClass)
Method Hiding (SuperClass)
```

As we can see here, method hiding may create confusion. It may become a source of a programming error when a static method is called using an instance of the subclass (an object) because, in this case, an inexperienced or incautious programmer may expect dynamic binding of the call to a method implementation defined in the dynamic type of the object (the subclass). However, even though some of the IDEs today provide hints that an instance should not call static methods and attributes, rather it should be called by the class (because static methods and attributes belong to the class and not to an instance), we have to assume that some people have nothing more than their compiler and a simple text editor, which will not catch such issues.

II. RELATED WORKS

No automated static analysis tool available for free can find methods hiding problems within Java programs [4]. However, some static analysis tools target various issues regarding the method overriding mechanism, which is very similar to the method hiding's problem. In this section, we present them.

PMD source code analyzer [9] targets the problem of useless overriding methods [10]. The related checker only raises an issue when an overridden method does not do any more than the method it overrides, marking it as useless. SB targets various issues related to method overriding in Java [11]. We will go through them in detail. The first one is when there is a call for an overridable method that performs a security check. This is considered an issue because the overridden method may compromise it and omit the checks. We have implemented this checker, in our previous work [12].

Due to the similarity of the upcoming SB's rules, we will explain them by grouping them into two groups. The first is related explicitly to how *equals*, *compareTo*, and toString methods are overridden. SB raises an issue when **a)** the *hashCode* method is not being overridden by the class overriding the *equals* method. This is an issue because, according to the contract of those two methods, equal objects should have equal hashcodes (i.e., calling the *hashCode* method on each of the two objects must produce the same integer result). **b)** a class defines a covariant version of the *equals* method. **c)** a class defines a covariant version of *compareTo* method. The last two rules state that the parameters of *equals* and *compareTo* methods must have type *java.lang.Object* otherwise, it is considered a bug **d)** a class defines a *toString* method that is not actually the one in the *java.lang.Object* class. The latter is probably what the programmer intended. The second rules group targets different overriding related issues. Those rules will raise an issue **e)** when a method overrides a method included in an Adapter class that implements a listener defined in the *java.awt.event* or *javax.swing.event* package. SB considers this an issue because this method will not be called when an event occurs. **f)** when an overriding method changes the superclass contract related to the Liskov Substitution Principle defined in a superclass. This is an issue since a subclass instance can be cast and handled as an instance of the superclass. **g)** when a class overrides an *equals* method in a superclass, and both methods use the *instanceof* operator to decide whether two objects are equal. This is problematic since it is important to ensure those two equal methods are symmetrical, i.e., a.equals(b) == b.equals(a). If B is a subtype of A, then there is a good chance that this method's equivalence connection is not symmetric. A's equals method verifies that the argument is an instance of A, and B's equals method verifies that the argument is an instance of B.

SonarQube static analysis tool (Sonar) [13] targets some of the previously mentioned issues, as shown in Table 1. However, it also targets other method overriding related issues of Java, listing them as rules [14]. We present the most important ones. An issue will be raised when these rules are violated. **a)** while not mandatory, using the *@Override* annotation on compliant methods improves readability by making it explicit that methods are overridden. According to this rule, *@Override* should be used to override and implement methods; **b)** In JUnit testing, to make sure that the test cases are set up and cleaned up consistently, the overriding implementations of *setUp* and *tearDown* methods should call the parent implementations explicitly because those two methods provide some shared logic that is called before all test cases. This logic may change over the lifetime of your codebase; **c)** a record class has an array field and is not overriding *equals, hashcode*, or *toString* methods.

This is an issue because array fields are compared by their reference, and overriding *equals* is highly appreciated to achieve the deep equality check. The same strategy applies to *hashCode* and *toString* methods; **d)** although overriding the *clone* method without implementing the *Cloneable* interface can be helpful if a programmer wants to control how subclasses clone themselves, it's probably a mistake. So, this rule suggests that classes that override clone should implement *Cloneable* and call the *super.clone* method. **e)** a class implementing the interface *Cloneable* but does not override the *clone* method is considered an issue because *Cloneable* is a marker interface that defines the contract of the *clone* method, which is to create a consistent copy of the instance. Since the compiler cannot enforce the definitions of marker interfaces (because they have no own API), when a class implements Cloneable but does not override the clone method, it likely violates the contract for Cloneable. Finally, **f)** the *Object.finalize* method should not be overridden. Relying on overriding it to release resources or update the program's state is highly discouraged because there is no guarantee that this method will be called as soon as the last references to the object are removed, which may lead to many issues. Table 1 presents the summary of the previously explained related works.

TABLE I
RELATED WORKS SUMMARY

The issue/rule	PMD	SB	Sonar	Ours
Useless overriding methods	✓		✓	
The methods that perform security checks must be declared private or final		✓		
hashCode method is not being overridden by the class that is overriding the *equals* method		✓	✓	
A class defines a covariant version of the *compareTo* method		✓		
A class defines a covariant version of the *equals* method		✓	✓	
A class defines a *toString* method that is not actually the one in the *java.lang.Object* class		✓		
A class overrides a method implemented in the superclass Adapter wrongly		✓		
Do not use the *instanceof* operator to decide whether two objects are equal		✓	✓	
Method overrides should not change contracts		✓	✓	
@Override annotation should be used for overriding and implementing methods			✓	
Junit test cases should call super metthods			✓	
equals, *hashcode*, and *toString* methods should be overridden in records containing array field			✓	
Classes that override *clone* should implement *Cloneable* and call the *super.clone* method			✓	
"Cloneables" should implement clone			✓	
The *Object.finalize* method should not be overridden			✓	
Never declare a class method that hides a method declared in a superclass or super interface				✓

Method hiding is considered as a weakness [4]. Moreover, many state-of-the-art static analysis tools focus on various issues regarding method overriding, which is the basis of method hiding. Still, none of these tools focus on method hiding. Our work sheds light on this problem and implements a checker that raises an issue when finding one.

III. RESEARCH METHODOLOGY

This section thoroughly describes our checker design process and the steps involved in developing the custom test cases.

A. Checker design

Listing 2 presents the pseudocode of the implementation of our checker named *FindHidingMethod*, using SB version 4.7.3. SB is written mainly in Java, so we implemented our checker using Java. The process starts by visiting each class in the program, then getting a list of all its superclasses, i.e., the parent class, grandparent class, etc., until reaching the last superclass, which is always the Object class. For each of the visited classes, our checker will check each of the methods and raise an issue when it finds a hidden subclass. More specifically, an issue will be raised when there is a subclass-superclass pair that includes methods with the same name, both of which are static, non-private, and not main (because it is an odd case of a static method that may exist in a superclass-subclass pair). However, the checker also considers the possibility of the method being a constructor and some other odd cases where it will be excluded (not reporting as an issue).

The checker has been developed successfully and has passed our team's internal review and the SB tool's public reviews. For further information about the implementation coding, please refer to the public review of our checker implementation on the official website of the SB tool [15].

LISTING 2
THE PSEUDOCODE OF OUR CHECKER

```
1 procedure FindHidingMethod( aClass ) is
2  foreach method in declared methods of class aClass loop
3   if method is static and non-private and not SpecialCase then
4    foreach superClass in superclasses of aClass loop
5     foreach superMethod in declared methods of class superClass lo
6      if signature(method) = signature(superMethod) then
7       report
8      end if
9     end loop
10   end loop
11  end if
12 end loop
13 end

14 function SpecialCase( method ) is
15  return   method is "non-private void main ( String[] )"
16       or method is "non-private void main( )"
17       or method is a constructor
18       or method is static_initializer_block
19       or method is a generated method
20 end
```

Complexity Analysis: SB static analysis tool, which we built our checker under, inspects the Java byte code for different programming vulnerabilities and weaknesses. Since the bytecode of the subclass does not contain the bytecode of any methods of the superclass (i.e., inherited methods), we had to analyze all the superclasses. The latter leads to O(n) complexity (where n is the number of superclasses), then goes over each of the methods, leading to a second loop with O(m) complexity (where m is the number of methods in each class). Combined together, it gives O(n*m), i.e., a quadratic time complexity. However, we do not generate the bytecode for each method every time we check for the vulnerability of that method; rather, we call a built-in method of the SB tool's environment, called *visitClassContext* [15], only once for the class. It scans all the methods and information related to the class. This balances out the double loop, producing an efficient checker (i.e., with linear complexity).

An Exception to Rule 06. Methods MET07-J: According to the SEI CERT Oracle Coding Standard for Java web page, which includes a description of the targeted issue of this paper, there is a case that should not be considered a violation of this rule, i.e., it should not be counted as a method hiding issue. This exception only applies when an API's hidden methods are called; in this scenario, all calls to hidden methods make use of qualified names or method invocation expressions that clearly indicate which particular method is being called [3]. The previously mentioned exception case has not been considered through our checker's implementation for two reasons:

1) When there is a method hiding in a program, it is considered unsafe, regardless of whether the related static method is being called.
2) The produced Java byte codes for calling a static method using a class name and calling it using a class instance are the same. Since SB uses bytecode to inspect the flaws in programs, any checker built under SB is not able to differentiate these two as well. More specifically, when using an instance variable for calling a static method, the JVM smartly fixes it when producing the bytecode. It makes it seem like the calling was happening on the class name. See Listing 3, where both invocation types (from lines 10 and 15) produce the bytecode **invokestatic.**

Therefore, whenever there are two identical static methods in any superclass-subclass pair, our checker will raise an issue and report the second method as a bug, no matter if any of the methods are being called or not (or whether they are called on a fully qualified name (class name) or an instance). This, however, can be considered a limitation of our implementation, and it may be addressed by implementing a checker under a static analysis tool that inspects the program source code instead of the bytecode (for example, under PMD Source Code Analyzer).

Note: Checkers implemented under SB may interact with the SB framework and be able to use its properties using either the *OpcodeStackDetector* abstract class or the *Detector* Interface. However, the previously explained scenario also made us decide to choose the *Detector* interface, which is a lower level since using the *OpcodeStackDetector* will only add extra complexity to our program.

B. Custom test cases' design

To assess our generated checker and attain thorough coverage of the problem being studied. The following points are covered by the test cases that our team has built:

1) Flawed and unflawed: To assess our checker performance in both true positive (TP) and false positive (FP) aspects [3], we designed non-compliant (NC), flawed test cases, and compliant (C), unflawed test cases.
2) Unambiguous: the test cases are written in a clear and concise way, leaving no room for misinterpretation.
3) Validated expectations: every test case has a predetermined expected result.
4) Test objective: every test case has a distinct goal that identifies the precise component of the problem being investigated.
5) Independence: We created separate test cases to isolate and identify problems more effectively.

As a result, we produced 11 C and 9 NC test cases. Next, we go into detail on the design of these test cases.

NC test cases: These are the flawed constructs. These test scenarios are considered insecure and involve real issues, so our checker should report them. We could cover every scenario in which methods in Java applications might be hidden by creating nine NC test cases.

C test cases: These are the unflawed constructs, i.e., include safe coding scenarios to test if our checker successfully ignores them. To cover all possible scenarios, we have designed eleven test cases.

You can check the test cases from the public review of our checker implementation on the official website of the SB tool [15]. The word *good* is used in the file name and/or method name of the C test cases, while the word *bad* is used in the NC test cases.

[3] TP is the number of flawed constructs that are detected correctly by our checker, while FP is the number of unflawed constructs that are mistakenly reported by the checker.

LISTING 3
THE PRODUCED BYTECODE OF CALLING A STATIC METHOD USING A CLASS INSTANCE AND A CLASS-QUALIFIED NAME

```
1 Invocation.java
2
3 class Super {
4           public static void staticMethod() {
5           }
6 }
7 public class Invocation {
8           public void invocationOnInstance() {
9                       Super sup = new Super();
10                      sup.staticMethod();
11          }
12
13          public void invocationOnClass() {
14                      Super sup = new Super();
15                      Super.staticMethod();
16          }
17 }

Compiled from "Invocation.java"

 public class Invocation {
 public Invocation();
   Code:
     0: aload_0
     1: invokespecial #1        // Method java/lang/Object."<init>":()V
     4: return

 public void invocationOnInstance();
   Code:
     0: new           #7        // class Super
     3: dup
     4: invokespecial #9        // Method Super."<init>":()V
     7: astore_1
     8: aload_1
     9: pop
    10: invokestatic  #10       // Method Super.staticMethod:()V
    13: return

 public void invocationOnClass();
   Code:
     0: new           #7        // class Super
     3: dup
     4: invokespecial #9        // Method Super."<init>":()V
     7: astore_1
     8: invokestatic  #10       // Method Super.staticMethod:()V
    11: return
}
```

IV. RESULTS AND DISCUSSION

This section presents our checker's analysis results of both bug types. The first type is intentional bugs; the custom test cases that have been explained in the previous section, while the second are real-world bugs.

A. Analyzing the custom test cases

For evaluation purposes, our team designed custom test cases. They have been explained in detail in Section 3.2. We used three metrics to represent our checker's performance in identifying methods hiding issues in Java programs. Formulas 1 through 3 present these metrics respectively. To calculate the metrics values, we first ran our checker on the test cases and then computed the number of TP and FP.

$$Recall = TP/NC \quad (1)$$

$$False\ alarm\ rate = FP/C \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

Higher recall and precision values, while a lower value of false alarm rate indicates better performance. All values fall in the interval [0, 1].

Table 2 presents the results of running our checker on the custom test cases. It achieved optimal performance results.

TABLE II
ASSESSMENT RESULTS OF ANALYZING CUSTOM TEST CASES

Metrics	Values
NC test cases	9
C test cases	11
TP	9
FP	0
Recall	1.00
False alarm rate	0.00
Precision	1.00

B. Analyzing real-world software

We analyzed seven different pieces of software to determine how well our checker performed in identifying the target bug of this paper in real-world software. The software are: SB itself, maven-javadoc-plugin [17], mybatis-3 [18], spark [19], cayenne [20], Apache Hadoop [21], and Apache Dubbo [22] (In Table 3, they have been renamed to P1 through P7, respectively). Hence, we could only use the precision metric here because it is not straightforward to calculate the number of NC and C constructs when it comes to real-world software.

The analyzed software has been chosen arbitrarily from the GitHub web page. Out of 30 projects, the presented ones include method-hiding weaknesses. This indicates the popularity of this issue; it appears in 23% of the arbitrarily chosen projects. Table 3 presents the results, which revealed that our checker gave the highest possible precision for the analyzed programs. The numbers of TP and FP have been decided by manually reviewing the output report of the checker. You can check the second author's GitHub repository to find the reports of running the checker on the presented software [23].

TABLE III
ASSESSMENT RESULTS OF ANALYZING THE REAL-WORLD SOFTWARE

METRICS	P1	P2	P3	P4	P5	P6	P7
TP	4	2	6	38	2	35	6
FP	0	0	0	0	0	0	0
PRECISION	1.0	1.0	1.0	1.0	1.0	1.0	1.0

V. CONCLUSION

With the help of the SpotBugs static analysis tool, we have created and implemented a new checker called "FindHidingMethod" that can identify the issue of method hiding in Java programs. Our approach has been evaluated, and the results revealed that it was very precise when detecting related issues in the analyzed test cases and real-world programs.

REFERENCES

[1] M. Ufuk., Review of some recent European cybersecurity research and Innovation Projects, Infocommunications Journal, Vol. XIV, pp. 70–78, 2022, DOI: 10.36244/ICJ.2022.4.10.

[2] Java Documentation, Overriding and hiding methods, accessed on December 2023, https://docs.oracle.com/javase/tutorial/java/IandI/override.html

[3] Java Language Specification, Inheritance, Overriding, and Hiding, accessed on December 2023, https://docs.oracle.com/javase/specs/jls/se21/html/jls-8.html#jls-8.4.8

[4] SEI CERT Oracle Coding Standard for Java, MET07-J. Never declare a class method that hides a method declared in a superclass or super interface, accessed on December 2023, https://wiki.sei.cmu.edu/confluence/display/java/MET07-J.+Never+declare+a+class+method+that+hides+a+method+declared+in+a+superclass+or+superinterface

[5] B. Chess and J. West, Secure Programming with Static Analysis, Addison-Wesley, USA, 2007.

[6] M. Alqaradaghi, G. Morse and T. K., Detecting security vulnerabilities with static analysis – a case study, Pollack Periodica, Vol. 17, pp. 1–7, 2021, DOI: 10.1556/606.2021.00454.

[7] TIOBE Index for October, accessed on October 2023, https://www.tiobe.com/tiobe-index

[8] SpotBugs, Find Bugs in Java Programs, accessed on December 2023, https://spotbugs.github.io/

[9] PMD Source Code Analyzer., accessed on December 2023, https://pmd.github.io/

[10] Useless Overriding Method, accessed on December 2023, https://pmd.github.io/pmd/pmd_rules_java.html

[11] SpotBugs Bug Description, accessed on July 2024, https://spotbugs.readthedocs.io/en/latest/bugDescriptions.html

[12] M. Alqaradagh, M.Z.I. Nazir, and T. Kozsik, Design and Implement an Accurate Automated Static Analysis Checker to Detect Insecure Use of SecurityManager. Computers, 12(12), p. 247., Computers 2023, 12(12), 247; DOI: 10.3390/computers12120247

[13] SonarQube Static Code Analysis Tool., accessed on December 2023, https://www.sonarsource.com/products/sonarqube/

[14] Sonar rules, Java static analysis, accessed on January 2024, https://rules.sonarsource.com/java/

[15] M.Z.I. Nazir, Public Review of MET07-J Checker, accessed on December 2023, https://github.com/spotbugs/spotbugs/pull/2467

[16] visitClassContext method, accessed on July 2024, https://github.com/spotbugs/spotbugs/blob/cc2bad5559f662dd997059606abc9d7e659f2a45/spotbugs/src/main/java/edu/umd/cs/findbugs/Detector.java#L36

[17] maven-javadoc-plugin, Apache Maven Javadoc Plugin., accessed on December 2023, https://github.com/apache/maven-javadoc-plugin

[18] mybatis-3, MyBatis SQL mapper framework for Java, accessed on January 2024, https://github.com/mybatis/mybatis-3

[19] Apache Spark – A unified analytics engine for large-scale data processing, accessed on December 2023, https://github.com/apache/spark

[20] cayenne, Mirror of Apache Cayenne, accessed on January 2024, https://github.com/apache/cayenne

[21] Apache Hadoop, accessed on January 2024, https://github.com/apache/hadoop

[22] Apache Dubbo, The Java implementation of Apache Dubbo. An RPC and microservice framework, accessed on January 2024, https://github.com/apache/dubbo

[23] M. Alqaradaghi, Running FindHidingMethod checker on real-world software, accessed on January 2024, https://github.com/Midya-ELTE/Running-FindHidingMethod-checker-on-real-world-software/tree/main

M. Z. I. Nazir was born in Lahore, Pakistan, in 2001. He received the B.Sc. degree in informatics from ELTE, Eötvös Loránd University, Budapest, Hungary, in 2023. Currently, he is studying M.Sc. informatics at Technical University Munich. He worked previously at different companies in Hungary. On microservices at Ericsson, on networks at Thermofisher, and on file transfer at CERN.

M. Alqaradaghi was born in Baghdad, Iraq. She received a B.Sc. degree in Information Technology from the Middle Technical University, Baghdad, Iraq, in 2006 and an M.Sc. degree in Computer Science from Sam Higginbottom University, India, in 2015. She is now a Ph.D. candidate at the Department of Programming Languages and Compilers, ELTE, Eötvös Loránd University, Budapest, Hungary. Her research focuses on finding security vulnerabilities in Java code using static analysis tools. From 2015 to 2019, she worked as a Teaching Assistant at the Northern Technical University, Kirkuk, Iraq. Since 2019, she has been working as an instructor in the programming languages lab in the Department of Programming Languages and Compilers, ELTE.

T. Kozsik is currently an Associate Professor with the Department of Programming Languages and Compilers, Eötvös Loránd University (ELTE). His research interests include formal verification, programming paradigms (i.e., functional programming, concurrent programming, and quantum computing), static analysis, refactoring, and domain-specific programming languages.

Evaluation of FBMC Channel Estimation using multiple Auxiliary symbols for high throughput and low BER 5G and beyond communications

T. Padmavathi[1]*, Kusma Kumari Cheepurupalli[2], and R. Madhu[3]

***Abstract*—Filter Bank Multi-Carrier (FBMC) modulation has been recognized as a consistent contender and a possible successor for Orthogonal Frequency Division Multiplexing (OFDM) in 5G and beyond because of its outstanding spectral properties. The channel is assessed in FBMC using pilot-symbol aided channel estimation that provides robust estimates even for severe channel conditions. In the present work, neutralizing the imaginary interference at the pilot positions is focused while estimating the channel. To neutralize the imaginary interference, multiple auxiliary symbols have been proposed to enhance the throughput and channel capacity. The Iterative Minimum Mean Squared Error (IMMSE) cancellation scheme has been proposed to reduce the interference at the pilot and data positions. Transmission power, Bit Error Rate (BER) and throughput are computed for Filter Bank Multi Carrier (FBMC), OFDM and proven that better system performance is obtained for FBMC. The performance of channel estimation is evaluated through 5G standards and indicates that the usage of multiple auxiliary symbols per pilot leads to better throughput and low Bit Error Rate at low power transmission.**

***Index Terms*—Filter bank Multi Carrier, Auxiliary symbols, Imaginary interference, Bit Error Rate, Throughput, 5G and beyond Communications.**

I. INTRODUCTION

Cyclic Prefix - Orthogonal Frequency Division Multiplexing (CP-OFDM) is a noticeable transmission technique used in various wireless communication standards i.e 4G Long-Term Evolution (LTE) [1], Digital Audio and Video Broadcasting-terrestrial (DVB-T) [2], Wireless Local Area Networks (WLAN). In OFDM system, wideband frequency spectrum has divided into numerous parallel sub bands. One tap channel equalizer is adequate at the receiver because OFDM system is more robust to frequency selective channels. However, cyclic prefix insertion is required for CP-OFDM systems in order to get these benefits. However, cyclic prefix increases the redundancy in a time domain. Additionally, it must ensure the guard bands in the frequency domain to avoid excessive Out-of-Band (OOB) emission [3].

[1]* T. Padmavathi, [1]*Sr Asst professor, Department of Electronics and Communication Engineering, CVR College of Engineering, Ibrahimpatnam, India. (E-mail: padmatp41@gmail.com)

[2] Kusma Kumari Cheepurupalli, [2]Associate Professor, Department of Electronics and Communication Engineering, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, India.
(E-mail: chkusumasrinivas@gvpce.ac.in)

[3] R. Madhu, [3]Assistant Professor, Department of Electronics and Communication Engineering, University College of Engineering Kakinada (A), JNTUK Kakinada, India. (E-mail: madhu_ramarkula@rediffmail.com)

DOI: 10.36244/ICJ.2024.2.4

Due to these redundant resources in the time domain and frequency domain, the spectral efficiency of CP-OFDM systems is reduced.

New modulation techniques are required to overcome these constraints Future wireless systems must accommodate a wide range of communication needs, such as enhance Mobile Broadband (eMBB) and Ultra-Reliable Low Latency communication (URLLC) [4], [5] and so on. New modulation techniques are required to overcome these constraints. Filter Bank Multi Carrier (FBMC) is a new waveform technique that has many advantages over OFDM as a contender for use of 5G communication systems. The only notable difference is the substitution of OFDM for a multi-carrier system based on filter banks at the transmitter and receiver, which increases bandwidth utilization. FBMC technology uses a good time-frequency prototype filter, which has many advantages such as low spectral side lobes, high spectrum efficiency, and frequency offset robustness [6]. However, because the system is not strictly orthogonal, imaginary interference exists, but this necessitates the estimation of channel coefficients in the complex domain.

In recent years, FBMC-Orthogonal Quadrature Amplitude Modulation (FBMC-OQAM) has been noticed as a strong contender due to its spectral features [7-8]. However, there are some drawbacks, such as less compatibility with Multiple-Input and Multiple-Output (MIMO) [9]. The real orthogonality is maintained by FBMC-OQAM systems only, but it suffers from inherent interference in the complex domain. As a result, conventional channel estimation approaches of OFDM systems are incompatible with FBMC-OQAM systems [10-12]. To resolve this issue, Lee.et.al proposed a single prototype filter based FBMC-QAM.[13]. However, for ideal channels the FBMC-QAM system with a single prototype filter does not guarantee orthogonality between adjacent subcarriers and symbol intervals. As a result, the FBMC-QAM system with a filter-based design does not meet orthogonality constraints in the complex domain [14].

For FBMC-QAM systems, a conventional scattered pilot assisted channel estimation approach is used.[15] In this work, a prototype filter with a well-localized spectrum and a higher self-Signal to Interference Ratio (SIR) is used. After receiving a signal, a proto-type filtering is implemented, and the channel is estimated at the pilot position using Least Squares (LS) approach. However, the interference is not canceled at the pilot symbols using the conventional channel estimation scheme. As a result, the performance of channel estimation reduced for frequency selective channels.

Several preamble-based schemes for Channel Estimation (CE) have been proposed. The Interference Approximation Method (IAM) [16] and in the literature Interference Cancellation Method (ICM) [17] are two well-known interference mitigation algorithms. They either reduce or exploit imaginary interference to improve CE performance. In [18], a novel preamble structure for FBMC systems was proposed in conjunction with these methods.

To estimate time and frequency for frequency selective channels, a Minimum Mean Squared Error (MMSE) channel estimation method is applied [19-20]. By using this method, clustered pilots are used and only one OFDM symbol is taken into consideration at a time. In FBMC, while estimating the channel with MMSE channel estimation, time-varying characteristics of the wireless channels are not considered. Many authors investigated the inherent interference of pilot symbols and proposed channel estimation algorithms for auxiliary symbols [21-22]. Unfortunately, transmitting additional symbols necessitates giving up certain time-frequency blocks, which reduces bandwidth efficiency. Furthermore, the result of [23] reveals that the auxiliary symbol scheme has an unsatisfactory Bit Error Rate (BER), implying that channel estimate accuracy is low.[24] presents a compressed sensing-based channel estimation strategy that takes the channel's sparsity into account while still beginning with the auxiliary symbol approach. To design an Inter-carrier Interference (ICI)-free structure over a doubly selective channel using the auxiliary symbol approach, [25-26] time-frequency blocks must be sacrificed. As a result, this system exhibits low spectral efficiency and inaccurate channel estimation.

Pilot Aided Channel Estimation (PACE) is considered as one of the best techniques for tracking the wireless channel characteristics. A single auxiliary symbol is allotted to each pilot in order to apply PACE in FBMC [27-28]. But compared to the data symbol power, this approach needs higher auxiliary symbol power. Instead of one auxiliary symbol, two are used to solve this problem. One auxiliary symbol is utilized in this estimation technique to minimize the Peak to Average Power Ratio (PAPR) and achieve the maximum channel capacity for a specific range of Signal to Noise Ratio (SNR). The other symbol is utilized to prevent more power offset with the expense of computational complexity.

In this work, a different approach is proposed in choosing the auxiliary pilot symbols to reduce the power offset, at the same time to enhance channel capacity for low SNR. A randomly generated coding matrix is also included for the selection of the auxiliary pilot symbols.

Filter Bank Multi Carrier System model

The limitations of OFDM can be eliminated in FBMC by including pulse shaping filter in both time and frequency domain. Consequently, FBMC systems have more spectral containment signals and offer more effective use of the radio resources where no CP is required.

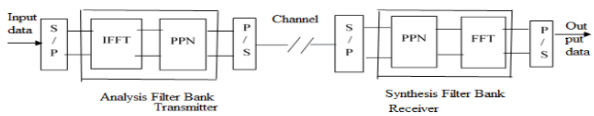


Figure 1. Transmitter and Receiver of FBMC System

In Filter Bank Multicarrier transmission shown in Figure.1. The transmitted signal from FBMC transmitter with data symbol $x_{l,k}$ and basis pulse $g_{l,k}(t)$ at l^{th} subcarrier position with time position k is given by

$$s(t) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} g_{l,k}(t) x_{l,k}(t) \quad (1)$$

where L is the number of subcarriers and K is the number of multicarrier symbols with basis pulse. The basis pulse $g_{l,k}(t)$ is expressed as

$$g_{l,k}(t) = p(t - kT) e^{j2\pi lF(t-kT)} e^{j\theta_{l,k}} \quad (2)$$

The prototype filter basis pulse $p(t)$ shifted in the time and frequency domain is given in the Eq.2. The throughput of FBMC can be maximized by low frequency spacing F and time spacing T. Multi carrier system orthogonality requires a condition $TF \geq 1$ and basis pulse localization in both time and frequency domain. The requirement of prototype filter $(\theta l, k = 0)\ p(t)$, frequency localization can be violated, because of CP in OFDM. In FBMC, Hermite polynomials $Hi(.)$ based basis pulse is employed. Because these polynomials produce great trade-off between time and frequency localization [29-30].

A different prototype filter $Rx(t)$ can be used at the receiver, so that the receiver basis pulses $q_{l,k(t)}$ can be expressed as in Eq.3

$$q_{l,k}(t) = P\ Rx(t - kT) e^{j2\pi lF(t-kT)} e^{j\theta_{l,k}} \quad (3)$$

To simplify the above analytical expression, the transmission system can be represented using a discrete time system model. The basis pulses are sampled with symbol rate $f_s = \frac{1}{\Delta t} = FN_{FFT}$ and obtains all the samples in large vector $g_{l,k} \in \mathbb{C}^{NX1}$. $Q = [g_{0,0------}g_{L-1,K-1}] \in \mathbb{C}^{NXLK}$ represent the receive basis pulse samples. In OFDM, the orthogonality denotes that $Q^H G = I_{L,K}$, whereas in FBMC, the real orthogonality condition maintains true if $\Re\{Q^H G\} = I_{L,K}$.

The received data symbol vector $Y \in \mathbb{C}^{LK\ X\ 1}$ over time and frequency selective channel is represented as in Eq.4

$$Y = diag\{h\}Dx + n \quad (4)$$

with $D = Q^H HG$

The channel can be represented by a vector $h \in \mathbb{C}^{LKX1}$ and complex Gaussian noise vector n is represented by a vector $n \sim N(0, P_n Q^H Q)$

II. EXISTING PILOT SYMBOL CHANNEL ESTIMATION APPROACH

The pilot symbols are known at the receiver before data transmission in pilot symbol channel estimation [31]. At the pilot position $(l,k) \in P$, the channel $h_{l,k}$ can be estimated using LS approach by dividing the received symbol $y_{l,k}$ with pilot symbol $x_{l,k}$. The estimate of channel at pilot position $\hat{h_P^{LS}} \in C^{|PX1|}$ using one -tap channel is given by Eq.5

$$\hat{h_p^{LS}} = diag(x_p)^{-1} y_p \quad (5)$$

Due to imaginary interference in FBMC, additional preprocessing becomes necessary. This can be done by precoding with help of Eq.6

$$X = cx \quad (6)$$

where the data symbol denoted by $\bar{x}$ and c is the precodingmatrix. The imaginary interference is cancelled at pilot position if the below Eq-7 is satisfied.

$$T\{q_p^H G\} Cx = 0 \quad (7)$$

Either Auxiliary Symbol Method [32] or Data spreading Approach can be used to frame precoding matrix C.

A. Proposed Auxiliary Pilot Symbol Approach

There are many symbols close to pilot symbols shown in Figure 2, resulting in imaginary interference. In the conventional technique, one of these symbols is used to cancel the imaginary interference [33].

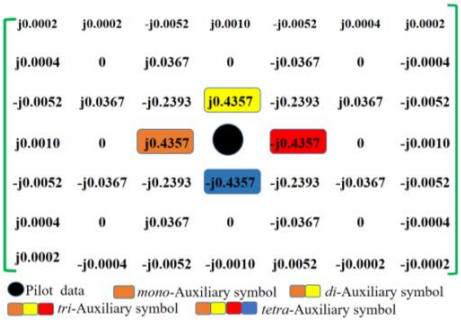


Figure 2: Selection of Auxiliary symbols close to the pilot symbol in the matrix

In order to cancel the imaginary interference, the neighbouring auxiliary pilot symbols are selected by using below Eq.8

$$X_p = [D_{P,P} \quad D_{P,D} \quad D_{P,A}] \begin{bmatrix} x_P \dots \dots \dots \\ x_D \dots \dots \dots \\ x_A \dots \dots \dots \end{bmatrix} \tag{8}$$

The vector $x_P \in R^{|P|X1}$represents all the elements of X at the pilot position. The same is applied to $x_D \in R^{|D|X1}$ at the data position & $x_A \in R^{|A|X1}$ at the auxiliary position. The row elements and column elements of D matrix forms the matrix $D_{P,D} \in C^{|P|X|D|}$ by taking pilot position and data position into account. The same is continued for $D_{P,P}$ and $D_{P,A}$.

Consider one auxiliary symbol per pilot, then $D_{P,A}$ represented by a diagonal matrix. From the elements of diagonal matrix shown in Figure. 2, the interference weight at auxiliary pilot position is 0.4357 is considered to compute the auxiliary pilot power offset K_A, and it is expressed using Eq.9[34].

$$K_A = \frac{P_A}{P_D} \tag{9}$$

where P_A is the power of auxiliary symbol, P_D is the power of data symbol. To cancel $N = 8$ closest interference symbols given in Figure.2. The auxiliary power offset given by

$$\frac{[4(0.2393)^2 + 3(0.4357)^2]}{(0.4357)^2} = 4.27$$

The auxiliary symbol must reimburse the imaginary interference from the adjacent symbols (without the auxiliary symbol), leading to an interference power of $(1 - 0.4357^2)$. Furthermore, the auxiliary symbol must be multiplied by $\frac{1}{0.4357}$ to compensate for the loss given by the interference weight. Thus, the auxiliary symbol power is $\frac{1-0.4357^2}{0.4357^2} = 4.21$ times higher than the data symbol power. In the proposed method it was observed that the auxiliary symbol power is reduced from 4.21 to 0.08 while increase from *mono* to *tetra* auxiliary symbol which is shown in Table.1.

TABLE I
AUXILIARY POWER OFFSET FOR INTERFERENCE WEIGHTS (ε=0.4357)

	mono-Auxiliary symbol	*di*-Auxiliary symbol	*tri*-Auxiliary symbol	*tetra*-Auxiliary symbol
$\frac{P_A}{P_D}$	$\frac{1-\varepsilon^2}{\varepsilon^2} = 4.21$	$\frac{1-2\varepsilon^2}{(2\varepsilon)^2} = 0.82$	$\frac{1-3\varepsilon^2}{(3\varepsilon)^2} = 0.25$	$\frac{1-4\varepsilon^2}{(4\varepsilon)^2} = 0.08$

With the increase in usage of auxiliary symbols, it is observed that a shortfall in data symbols. Even if there is a loss in data symbol, the throughput for certain range of Signal to Ratio (SNR) is not degraded. BER is one of the performance metrics which must be taken care of in channel estimation. While transmitting the data symbol and pilot symbol, channel estimation becomes complex due to the channel transfer function discontinuity at the edge of subcarriers. This is due to the estimation methods depending on an assumption that the channel delay taps are limited in time [35]. Even though the channel is estimated accurately, the resulting matrix multiplication is given in Eq.7 which is computationally complex.

The iterated MMSE Channel estimation method proposed in this paper works as follows:

Algorithm 1 Iterated MMSE Channel estimation

1. Compute transmission matrix with data given in Eq-2
2. Perform One-tap channel equalization followed by quantization then estimate
3. $x_{l,k}^{\hat{}(0)} = Q\{\frac{y_{l,k}^{(0)}}{h_{l,k}^{\hat{}(0)}}\}$, with $h^{\hat{}(0)} = diag\{D^{\hat{}(0)}\}$
4. Initialize the iteration with $i = 0$
5. Interference can be cancelled from the received signal.
6. $y^{(i+1)} = y - (D^{\sim(i)} - diag\left\{diag\{D^{\sim(i)}\}\right\})x^{\hat{}(i)}$
7. Obtain channel estimation data matrix $D^{\hat{}(i+1)}$ with less interference at pilot position.
8. Equalize the received signal with improved channel estimate and quantize, then obtain estimate of transmitted signal.
9. $x_{l,k}^{\hat{}(i+1)} = Q\{\frac{y_{l,k}^{(i+1)}}{h_{l,k}^{\hat{}(i+1)}}\}$
10. Consider $i = 0, 1, \dots \dots 4$ and repetition of step 4 to step 7.

where subscript $(.)^{(i)}$denotes i^{th}iteration step.

The modified MMSE channel estimation enhances accuracy by canceling the interference with the above method. The underlying correlation of channel matrix does not cancel the interference at the pilot positions because of nonlinearities in wireless channel. To overcome this problem, iteration step $i = 3\ and\ 4$ are considered and interference is perfectly cancelled.

III. EXPERIMENTAL RESULTS

FBMC has lower side lobes and smaller guard bands resulting in higher spectral efficiency than OFDM. In order to achieve the advantages of FBMC and to enhance the performance of the proposed channel estimation method, LTE standards are considered to implement the system. This improvement is also evaluated by implementing FBMC with $1.4MHz$ LTE spectrum. The subcarrier spacing of $F = 15KHz$ is considered in FBMC like OFDM signal. Cyclic prefix length

of $4.75\mu s$ and $K^{OFDM}(Number\ of\ symbols) = 14$, which results in the symbol duration as $1ms$ for OFDM signal, whereas FBMC allows to transmit $K^{OFDM}(Number\ of\ real\ symbols) = 30$ within the same symbol duration i.e $1ms$. Even though LTE occupies $1.4MHz$, OFDM only uses $L = 72$ subcarriers $(LF = 1.08MHz)$ whereas in FBMC, there are much lower sidelobes, the number of subcarriers is increased to $L = 87$ corresponds to $LF = 1.305MHz$.

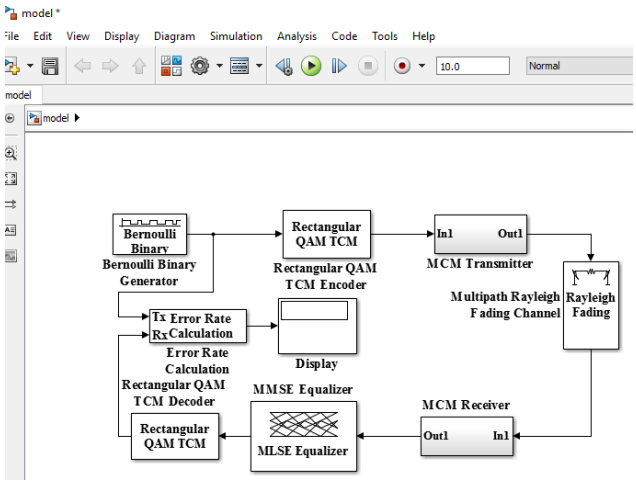


Figure. 3. Experimentation model for Throughput and BER

Transmission power

The pilot symbol pattern for FBMC is considered a diamond-shaped pattern, and the pilot density of |P| (KT LF) = 0:044 is chosen from the LTE standards. The data symbol power is the same as the pilot symbol power $P_P = P_D$. If the pilot power is increased by a factor of two, resulting in the same SNR for channel estimation (complex domain) and data transmission (real domain). The signal is transmitted with the 2.4955 GHz carrier through the Rayleigh fading channel to achieve several channel realizations. Modulation order, code rate, and turbo coding are taken from LTE standards to evaluate throughput. In turbo coding, while transmitting a signal the highest 15 Channel Quality Indicator (CQI) values, can be selected to get the highest throughput for all data bits.

To compare OFDM and FBMC, the same transmission power P_S is considered, and defined as:.

$$P_S = \frac{1}{KT}\int_{-\alpha}^{\alpha} E\{s(t)^2\} \tag{10}$$

The signal length in the time domain is given by KT and the transmitted signal $s(t)$ is given in Eq.1. For the same transmission power P_S, SNR of OFDM and FBMC is given by

$$SNR_{FBMC} = SNR_{OFDM}\frac{L^{OFDM}}{L^{FBMC}} \tag{11}$$

The SNR can be expressed as

$$SNR_{OFDM} = \frac{|D_{OFDM}|P_{D,OFDM}+|P_{OFDM}|P_{P,OFDM}}{L_{OFDM}K_{OFDM}P_n} \tag{12}$$

$$SNR_{FBMC} = \frac{|D_{FNA}|P_{D,FNA}+|P_{FNA}|P_{P,FNA}+|A_{FNA}|P_{A,FNA}}{L_{FBMC}K_{FBMC}\frac{P_n}{2}} \tag{13}$$

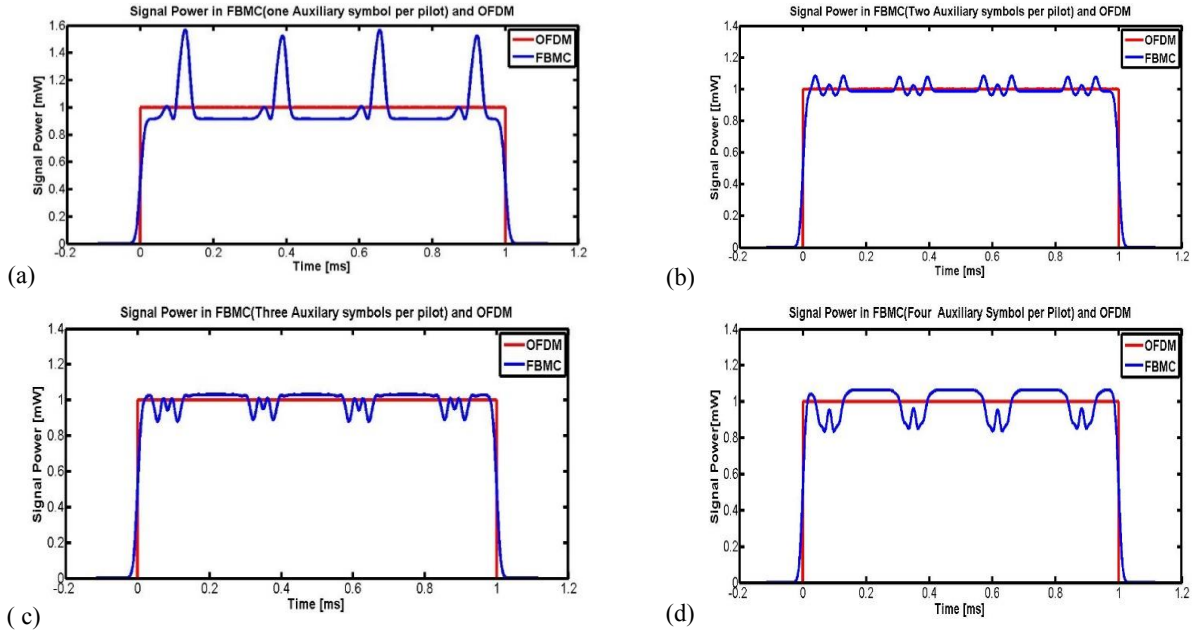


Figure 4. (a), (b), (c) and (d) are the signal power of FBMC and OFDM with number of Auxiliary symbols per pilot from *mono-* to *tetra*

Various performance metrics such as transmission power, throughput and BER are computed for Multi carrier modulations such as OFDM, FBMC using the computation model shown in Figure 3. While implementing this model higher order Quadrature Amplitude Modulation (QAM) like 4-where FNA in $P_{D,FNA}$ denotes FBMC based number of auxiliary symbols per pilot, and it varies from 1 to 4. P_n is the noise power. FBMC is operated in real domain, so the noise power P_n is reduced by a factor of 2.

If FBMC uses a greater number of subcarriers, the available transmission power must be dispersed over a wider bandwidth.

As a result, it reduces each symbol transmission power while keeping the noise power constant, resulting in a poor SNR.

The power required to transmit OFDM and FBMC are evaluated by varying the Auxiliary symbols per pilot from *mono-* to *tetra-* are shown in Figure.4.

If the number of auxiliary symbols is increased from mono- to tetra-the available power to transmit the data symbols also increases. Then the number of data symbols reduced, but the auxiliary symbol power can be significantly reduced, shown in Table 1.

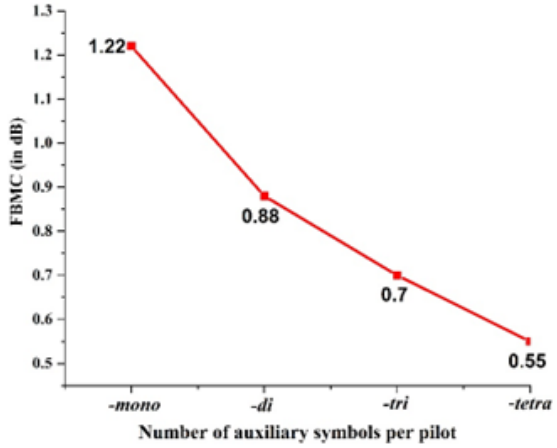


Figure 5: Shift of SNR for FBMC with auxiliary symbols per pilot from *mono-* to *tetra-*

Throughput computation

Throughput is computed by considering only one auxiliary symbol per pilot with various interpolation methods. In Moving block average-based interpolation, the average of all pilot estimates is considered within a time frequency range, whereas in linear interpolation the three closest pilot estimates are considered. The wireless channel is highly correlated in both frequency and time domain. Due to this, moving block average interpolation is better than the linear interpolation by nearly 1.7 *dB* SNR.

Based on the relationships given by Eq.11 and Eq.12, the SNR of FBMC and OFDM are calculated by varying auxiliary symbols per pilot from 1 to 4 and comparison with OFDM is shown in Fig. 5.

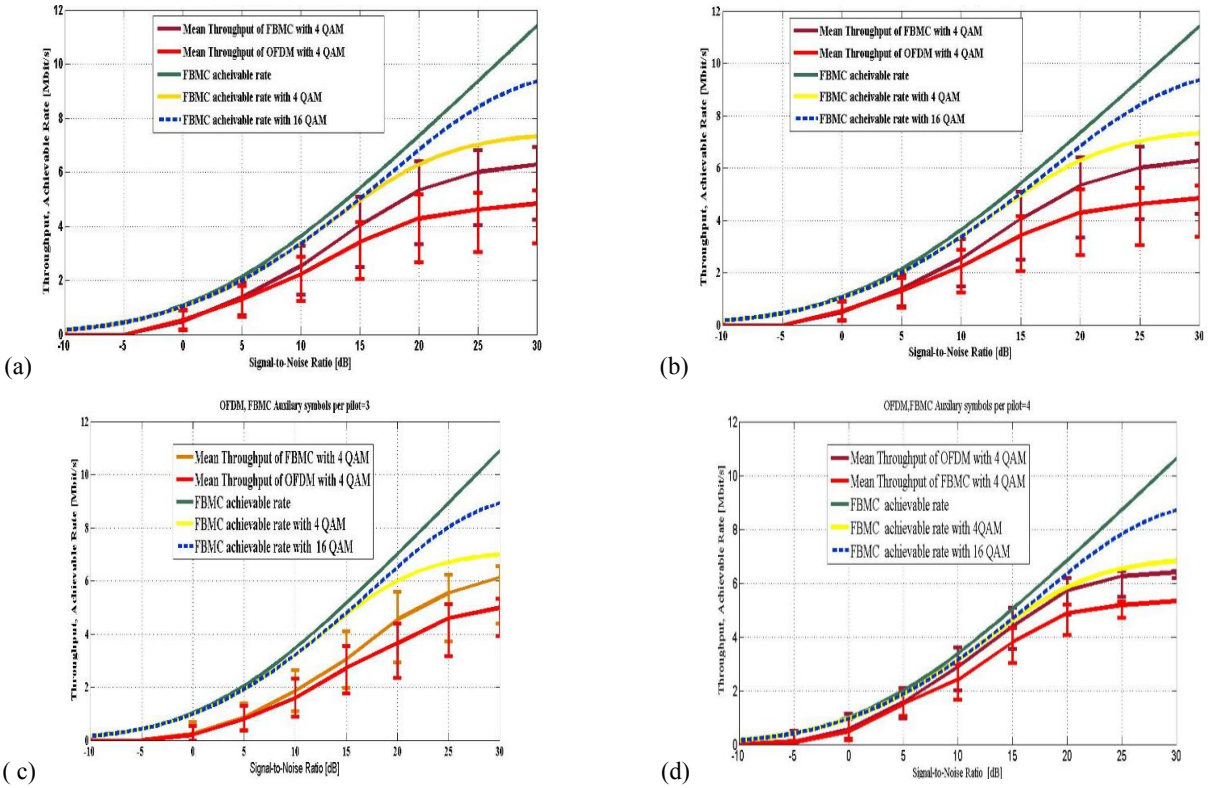


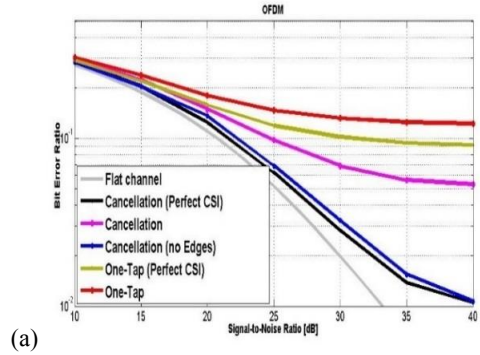
Figure 6: (a), (b), (c) and (d) are the throughput achievable rate for FBMC and OFDM with number of Auxiliary symbols per pilot from *mono-* to *tetra-*

From Figure 5 compared to OFDM, in FBMC an earlier shift of 1.22dB is observed for one auxiliary symbol per pilot. Similarly, an earlier shift of 0.88 dB is observed for two auxiliary symbols per pilot, an earlier shift of 0.7 dB is observed for three auxiliary symbols per pilot, and earlier shift of 0.55 dB is observed for four auxiliary symbols per pilot. Throughput is calculated by varying SNR and the effect of channel capacity for the measured throughput is shown in Figure. 6. For the computation of throughput such as mean throughput (average Throughput), achievable throughput (maximum throughput) the computational model is shown in Figure 3.

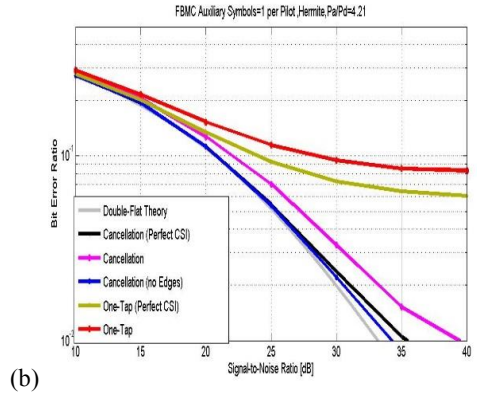
For a low SNR, the throughput is increased. The increased throughput of data symbols compensates the loss of throughput for few pilot symbols. Therefore, using multiple auxiliary symbols per pilot performs better than having one auxiliary symbol per pilot.

BER computation

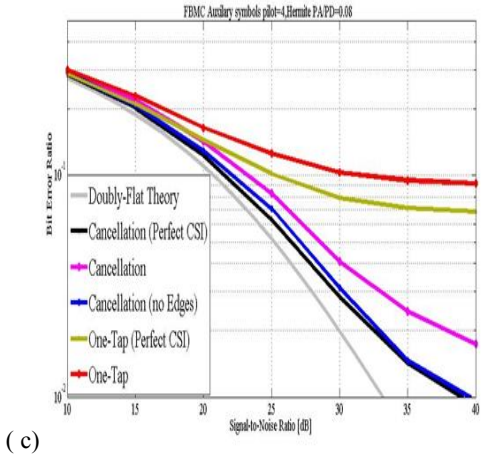
The BER performance of OFDM is evaluated with interference cancellation and shown in Figure 7. To improve the channel estimation accuracy pilot to data power offset of $P_A/P_D = 2$ is considered.



(a)



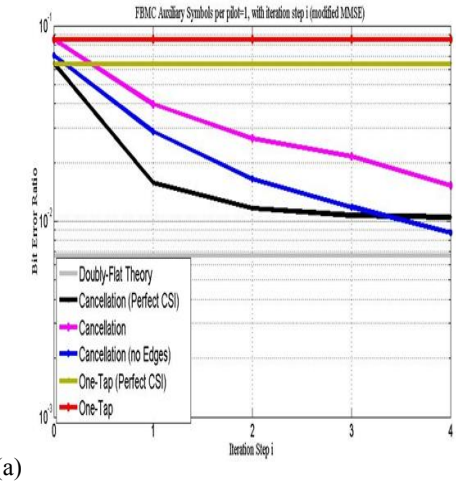
(b)



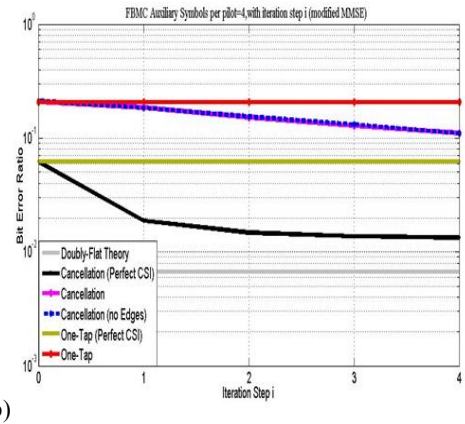
(c)

Figure 7:(a) BER performance of OFDM with interference cancellation; (b) BER performance of FBMC with interference cancellation with *mono*-Auxiliary symbols per pilot; (c) BER performance of FBMC with interference cancellation with *tetra*-Auxiliary symbols per pilot.

The performance of one Tap equalizer is poor if the interference is added to the noise. BER is computed, by excluding the time-frequency position close to the edge, i.e only points which are at the center of the frame are considered. Better BER performance (6dB) is obtained for interference cancellation with Perfect CSI (Channel State Information), and it is close to flat channel response. The proposed Auxiliary symbol approach for FBMC is compared with OFDM and plotted in Figure.7. Only locations in the middle of the frame are taken into consideration for computing BER, while time-frequency positions near the edge are excluded. Perfect CSI (Channel State Information) results in better BER performance (6dB) for interference cancellation, and the channel response is nearly flat for the FBMC with proposed auxiliary symbol approach. By varying the number of pilots from *mono to tetra* for FBMC, BER is computed and shown in Figure.8.



(a)



(b)

Figure 8: BER performance of FBMC with Modified MMSE(Iterated MMSE) and with *mono*-Auxiliary symbols per pilot (b) BER performance of FBMC with Modified MMSE(Iterated MMSE)and *tetra*- Auxiliary symbols per pilot

BER is also computed by performing a number of Montecarlo simulations and plotted and shown in Figure 8(a-b). From the Figure. 8. the BER performance of FBMC using the auxiliary symbol channel estimation method with auxiliary symbols per pilot from *mono*- to *tetra*-, it is observed that by employing *tetra*-auxiliary symbols per pilot, the interference

cancellation method obtains 3 dB less performance for the least $(\frac{P_A}{P_D})$ power offsets i.e only 0.08. FBMC shows better performance because of more power for data symbols and less channel-induced interference than OFDM.

By increasing the number of iteration steps from *mono-* to *tetra-* BER is reduced to 0.2 dB to 0.1 dB for interference cancellation scheme. If the number of iterations is increased the performance of BER is improved which is shown in Figure 7(a-b). Imaginary interference cancellation based on MMSE method achieves better performance compared to iteration $i = 1$ to $i = 4$. This is achieved only due to the cancellation of imaginary interference by increasing the auxiliary pilot symbols from mono to tetra.

IV. CONCLUSIONS

The proposed channel estimation and interference cancellation are suitable to any physical layer modulations such as OFDM and FBMC. To cancel the imaginary interference and to estimate the channel, auxiliary symbols are transmitted along with the pilot. When one auxiliary symbol is transmitted per pilot, the power offset, and transmission power enhancement are observed. The number of symbols allotted to data can be decreased by transmitting multiple Auxiliary symbols per pilot, while the power required to transmit the signal and power offsets can be reduced. When the SNR is low, the throughput gain of more data symbols compensates for the throughput loss of a fewer number of data symbols. As a result, using numerous symbols per pilot outperforms using only one auxiliary symbol. Due to the availability of more power for data symbols in FBMC, the BER performance is better than OFDM. The performance of the proposed MMSE channel estimation scheme is close to the ideal channel because it is designed to deal with large delay spread and Doppler spread channels. The auxiliary symbol-based channel estimation approach outperforms the conventional process and has the added benefit of higher throughput.

REFERENCES

[1] 3GPP TS 36.214: version10.1.0 Release 10. *Physical layer: Measurements*. 2011

[2] Ghosh, Sudipta, and Ankit Bass, "Implementation of digital video broadcasting-terrestrial (DVB-T) using orthogonal frequency division multiplexing (OFDM) on physical media dependent sub layer," *Int. J. Comput. Appl.*, vol. 44, no. 22, pp. 20–25, 2012.

[3] Pan, Lebing, Jun Ye, and Xiaobing Yuan, "Spectral precoding for out-of-band power reduction under condition number constraint in OFDM-based systems," *Wireless Personal Communications,* vol. 95, no. 2, pp. 1677–1691, 2017. **DOI**: 10.1007/s11277-016-3874-8

[4] Tanda, Mario. "Performance analysis of FBMC/OQAM systems in frequency-selective channels." *Digital Signal Processing 134* (2023): 103935.

[5] Pishvaei, Seyed Mahmoud, Behzad Mozaffari Tazehkand, and Jafar Pourrostam. "Design and performance evaluation of FBMC-based orthogonal time–frequency space (OTFS) system." *Physical Communication* 53 (2022): 101723.

[6] Gregoratti, David, and Xavier Mestre, "Uplink FBMC/OQAM-based multiple access channel: Distortion analysis under strong frequency selectivity," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4260–4272, 2016.

[7] Kim, Chanhong, Yeo Hun Yun, Kyeongyeon Kim, and Ji-Yun Seol, "Introduction to QAM-FBMC: From waveform optimization to system design," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 66–73, 2016. **DOI**: 10.1109/MCOM.2016.1600384CM

[8] Mahama, Sumaila, Yahya Jasim Harbi, Alister G. Burr, and David Grace, "Design and Convergence Analysis of an IIC-based BICM-ID Receiver for FBMC-QAM Systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 563–577, 2020. **DOI**: 10.1109/OJCOMS.2020.2992314

[8] Sim, Dongkyu, and Chungyong Lee, "A MIMO receiver with two-dimensional ordering for maximum likelihood detection in FBMC-QAM," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1465–1468, 2017. **DOI**: 10.1109/LCOMM.2017.2689003

[9] Kwon, Beom, Seonghyun Kim, and Sanghoon Lee, "Scattered reference symbol-based channel estimation and equalization for FBMC-QAM systems," I*EEE Transactions on Communications*, vol. 65, no. 8, pp. 3522–3537, 2017. **DOI**: 10.1109/TCOMM.2017.2710310

[10] Cheng, Xing, Dejun Liu, Chen Wang, Song Yan, and Zhengyu Zhu, "Deep learning-based channel estimation and equalization scheme for FBMC/OQAM systems," I*EEE Wireless Communications Letters*, vol. 8, no. 3, pp. 881–884, 2019. **DOI**: 10.1109/LWC.2019.2898437

[11] Fuhrwerk, Martin, Sanam Moghaddamnia, and Jürgen Peissig, "Scattered pilot-based channel estimation for channel adaptive FBMC-OQAM systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1687–1702, 2017. **DOI**: 10.1109/TWC.2017.2651806

[12] Lee, Taehyun, Dongkyu Sim, Bongsung Seo, and Chungyong Lee, "Channel estimation scheme in oversampled frequency domain for FBMC-QAM systems based on prototype filter set," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 728–739, 2018. **DOI**: 10.1109/TVT.2018.2885078

[13] Jeon, Donghyun, Seonghyun Kim, Beom Kwon, Hojae Lee, and Sanghoon Lee, "Prototype filter design for QAM-based filter bank multicarrier system," *Digital Signal Processing*, vol. 57, pp. 66–78, 2016. **DOI**: 10.1016/j.dsp.2016.05.002

[14] Bazzi, Jamal, Petra Weitkemper, and Katsutoshi Kusume, "Power efficient scattered pilot channel estimation for FBMC/OQAM," In *SCC 2015; 10th International ITG Conference on Systems, Communications and Coding*, VDE, pp. 1–6, 2015.

[15] J. F. Du and S. Signeel, "Novel preamble-based channel estimation for OFDM/OQAM systems," in *Proceedings 2009 IEEE International Conference on Communications*, Dresden, Germany, pp. 1–6, June 2009.

[16] E. Kofidis and D. Katselis, "Improved interference approximation method for preamble-based channel estimation in FBMC/OQAM," in *Proceedings IEEE European Signal Process.Conference*, Barcelona, Spain, pp. 1603–1607, September 2011.

[17] H. Wang, W. C. Du, and L. W. Xu, "Novel preamble design for channel estimation in FBMC/OQAM systems," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 8, pp. 3672–3688, 2016.

[18] Kwon, Beom, Seonghyun Kim, and Sanghoon Lee, "Scattered reference symbol-based channel estimation and equalization for FBMC-QAM systems," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3522–3537, 2017. **DOI**: 10.1109/TCOMM.2017.2710310

[19] Han, Hyungsik, Namshik Kim, and Hyuncheol Park, "Design of QAM- FBMC waveforms considering MMSE receiver," *IEEE Communications Letters*, vol. 24, no. 1, pp. 131–135, 2019. **DOI**: 10.1109/LCOMM.2019.2952375

[20] Xiaoli, Ma, G. B. Giannakis, and Shuichi Ohno. "Optimal training for block transmissions over doubly selective wireless fading channels." *IEEE Transactions on Signal Processing* 51.5 (2003): 1351–1366.

[21] Lee, Taehyun, et al. "Channel estimation scheme in oversampled frequency domain for FBMC-QAM systems based on prototype filter set." *IEEE Transactions on Vehicular Technology* 68.1 (2018): 728–739.

[22] Hassan, Beenish, et al. "Adaptive refined random orthogonal matching pursuit algorithm for FBMC/OQAM MIMO framework." *Alexandria Engineering Journal* 87 (2024): 319–328.

[23] Saideh, Michel, Marion Berbineau, and Iyad Dayoub, "On doubly selective channel estimation for FBMC-OQAM using the LMMSE filter for future railway communications," In *2018 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST)*, IEEE, pp. 1–6, 2018. **DOI**: 10.1109/ITST.2018.8566706

[24] Padmavathi, Thodupunoori, Kusma Kumari Cheepurupalli, and R. Madhu. "Analysis of Multi Carrier Modulation Techniques for 5G Physical Layer Communications Estimation of KPI." *Journal of Scientific & Industrial Research* 81.11 (2022): 1224–1232. http://nopr.niscpr.res.in/handle/123456789/60912

[25] Liu, Wenfeng, *et al*. "Preamble-based channel estimation for OQAM/FBMC systems with delay diversity." *IEEE Transactions on Wireless Communications* 19.11 (2020): 7169–7180.

[26] Chen, Da, Rui Wang, and Tao Jiang. "Channel estimation and pilot symbol optimization based on intrinsic interference utilization for OQAM/FBMC systems." *IEEE Transactions on Signal Processing 69* (2021): 4595–4606.

[27] Wang, Ying, *et al*. "Bi-orthogonality recovery and MIMO transmission for FBMC Systems based on non-sinusoidal orthogonal transformation." *Signal Processing* (2024): 109427.

[28] Rosenkilde, Johan, and Arne Storjohann, "Algorithms for simultaneous Hermite–Padé approximations," *Journal of Symbolic Computation*, vol. 102, pp. 279–303, 2021. **DOI**: 10.1016/j.jsc.2019.07.026

[29] Jeannerod, Claude-Pierre, Vincent Neiger, Éric Schost, and Gilles Villard, "Computing minimal interpolation bases," *Journal of Symbolic Computation*, vol. 83, pp. 272–314, 2017. **DOI**: 10.1016/j.jsc.2016.11.015

[30] Nissel, Ronald, and Markus Rupp, "On pilot-symbol aided channel estimation in FBMC-OQAM," In *2016 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, IEEE, pp. 3681–3685, 2016.

[31] Cui, Wenjia, Daiming Qu, Tao Jiang, and Behrouz Farhang-Boroujeny, "Coded auxiliary pilots for channel estimation in FBMC-OQAM systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no.5, pp. 2936–2946, 2015. **DOI**: 10.1109/TVT.2015.2448659

[32] Kong, Dejin, Xing Zheng, Yue Zhang, and Tao Jiang, "Frame repetition: A solution to imaginary interference cancellation in FBMC/OQAM systems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1259–1273, 2020. **DOI**: 10.1109/TSP.2020.2971185

[33] Choi, Jeong-Min, Youngho Oh, Hakju Lee, and Jong-Soo Seo, "Pilot-aided channel estimation utilizing intrinsic interference for FBMC/OQAM systems," *IEEE Transactions on Broadcasting*, vol. 63, no. 4, pp. 644–655, 2017. **DOI**: 10.1109/TBC.2017.2711143

[34] Singh, Prem, Himanshu B. Mishra, Aditya K. Jagannatham, and Kasturi Vasudevan, "Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4668–4682, 2019. **DOI**: 10.1109/TSP.2019.2925607

T. Padmavathi did M.Tech in Systems in signal Processing currently pursuing Ph.D in the field of Wireless Communications & Signal processing, from JNTU, Kakinada. She is currently working as Sr. Assistant Professor with the Department of Electronics and Communication Engineering, CVR College of Engineering, Ibrahimpatnam, Hyderabad, India. She has published papers in the field of MIMO, Channel Estimation in the reputed Journals

Kusma Kumari Cheepurupalli has done Ph.D in the field of Signal processing and Communications from Andhra University, Visakhapatnam. She did her M.Tech project in Naval Science Technological Laboratory, Vishakhapatnam. She is the Principal Investigator for the project titled 'Robust Signal Processing Techniques for RADAR/SONAR Communications", granted by DST under Women Scientist Scheme. She has got 4 best papers awards in the various Conferences. She has got more than 25 papers in the reputed journals. She has 3 patent grants and Co-author for the Text book "Semiconductor Physics and Devices", published by Deccan International Publisher. Currently working as Associate Professor in Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam.

Madhu Ramarakula was born in Warangal district, INDIA in 1980. He received the B.E (Bachelor of Engineering) degree in Electronics and Communication Engineering from Osmania University, Hyderabad, India, in 2003 and M.Tech degree in Communication Systems from Jawaharlal Nehru Technological University Hyderabad, India, in 2009. He received his Ph.D degree in Electronics and Communication Engineering from Andhra University, Visakhapatnam, India, in 2014. He is presently working as an Assistant Professor in the Department of Electronics and Communication Engineering, University College of Engineering Kakinada (A), JNTUK Kakinada, India. He has 10 years of teaching experience. He has published more than 40 research papers in various reputed national and international Journals and conferences. His research interests include Mobile and Cellular Communications, Antennas, Satellite communications and GNSS. He is a member of IEEE.

A Novel Hybridization of ML Algorithms for Cluster Head Selection in WSN

Praveen Kumar R., M. P. Prabakaran, Durai Arumugam, and J. Selvakumar

***Abstract*—Generally, Wireless Sensor Networks (WSNs) are infrastructure-less networks with thousands of sensor nodes that sense or monitor the physical and environmental changes and forward the collected data to a central node. Besides, WSN has become the most efficient technology for handling Internet of Things (IoT) devices. Still, challenges such as node failures, high traffic among the nodes, link failures, etc., limit the performance of WSNs. To solve the challenges in WSN, this paper aims to develop a novel non-uniform clustering model, where the Cluster Heads (CHs) are selected based on the candidate CH selection strategy that transfers the data. Moreover, unbalanced energy utilization and data redundancy are eliminated via multi-hop communication. For attaining the non-uniform clustering model, the routing among the data packets is done by the efficiency of the hybridization of the Machin Learning (ML) algorithms viz Genetic Algorithm (GA) and Lion Algorithm (LA) with the consideration of energy, cost, time, network lifetime, and data accuracy. Finally, the performance of the proposed model is verified and validated through a comparative study with the existing models.**

***Index Terms*—Wireless Sensor Networks, Genetic Algorithm, Lion Algorithm, Cluster head Selection, Internet of Things**

I. INTRODUCTION

Currently, IoT emerges as an incredible technology that enables anything that can be connected will be connected [1]. From a thing as tiny as a pill to a thing as large as an aircraft, it can be connected through IoT and collect and share data. With the innovation of digital intelligence, IoT connects various things associated with sensors and allows them to communicate without human interactions [2]. Besides, IoT virtually connects the world like a digital fabric that interlinks the digital and real universes. WSN generally consists of small sensor nodes managed and controlled through the base station (BS) [3]. Nevertheless, the battery capacity of the sensor nodes is limited by their size. Alternative power source deployment for every sensor node is practically infeasible since they are random and dynamic [4, 5]. In WSN, the entire network's performance relies on the lifetime of the sensor nodes, and the impairment resulting from dead nodes can drastically diminish the network's reliability. Accordingly, recent research is concentrating on the progress of innovative methodologies that minimize the nodes' vitality consumption [6-8].

Praveen Kumar R. is with the Dept. of ECE, Easwari Engineering College, Chennai, India. (E-mail: rpjcspraveen@gmail.com)

M. P. Prabakaran is with the Dept. of ECE, AKT Memorial College of Engineering and Technology, Kallakurichi, India.

Durai Arumugam is with the Dept. of IT, Easwari Engineering College, Chennai, India.

J. Selvakumar is with the Dept. of CSE, Sri Ramakrishna Engineering College, Coimbatore, India.

DOI: 10.36244/ICJ.2024.2.5

Also, WSN comprises many hubs, hubs management, and energy consumption strategies of every hub. Therefore, developing strategies to manage vitality accessibility is mandatory to ensure network lifetime [9, 10]. Only a few significant hubs are kept active through these topology control techniques to manage the availability and carry out the network capabilities [11]. The sensor nodes in the network are random and dynamic, as they may be either kept active or asleep. Hence, the hubs that possess maximal lasting energy are considered to be the significant ones [12, 13].

Rather than interconnecting every hub to sink a particular hub for transferring data, a pioneer node is nominated in every cluster concerning the node that possesses maximal lingering energy [14]. Other than the pioneer node/CH node, all other nodes are named as member nodes that transfer the data to the CH node [15]. At last, the gateway node establishes an appropriate connection among the lattices and transmits the data to the BS. Implementing clustering in WSN enables a typical communication bandwidth (BW), ensures an alleviated network, minimizes communication expenses, and eliminates data redundancy [16-19]. Previously, numerous types of research had been proposed to implement CH selection using ML and optimization concepts [19, 20]. Even though optimization strategies achieved better energy efficiency, network lifetime, and minimized delay, they need several parameters for delivering optimal solutions [21-24]. Moreover, choosing an appropriate algorithm to solve complex problems requires a better understanding and expertise in ML knowledge [25-27].

Although the rapid evolution of IoT and WSN offers unprecedented opportunities for network connectivity and data transmission, their potential is restricted by the limited lifespan of sensor nodes in the network. The ramifications of dead nodes causes reduced network performance, robustness and reliability. Hence, it is crucial to develop innovative technologies to reduce the energy consumption to ensure prolonged lifespan of sensor nodes. To achieve this objective, many existing techniques including Single-hop and Multi-hop CH Selection through GA [1], GA and Modified Particle Swarm Optimization (MPSO) [2], Fruitfly Optimization Algorithm (FFOA) and Glowworm Swarm Optimization (GSO) [3], New Individual Updating Strategies-based Hybrid Elephant Herding Optimization Algorithm [4], etc., are designed. These techniques aim to select the CH in the network to streamline communication, reduce energy consumption and eliminate data redundancy within the WSN. However, existing methodologies often

rely on complex optimization techniques or machine learning algorithms, which demand large computational resources, parameter tuning and knowledge expertise to implement effectively. Moreover, they face certain challenges like reduction of alive nodes, increased cost function, less scalability, limited adaptability, etc. To overcome these issues, we proposed a promising solution by integrating evolutionary algorithms like GA and LA. The objective of our study is to enhance the longevity and energy efficiency of sensor nodes in WSNs through optimal CH selection. By selecting CHs, we aim to extend the lifespan of sensor nodes and improve overall energy efficiency across the network. The developed mechanism integrates the exploration and exploitation capabilities of GA and LA to effectively identify and select CHs in the WSN. This approach not only ensures improved energy efficiency but also offers reliable communication throughout the network. By combining these two meta-heuristic optimization algorithms into a single algorithm, this study provides a comprehensive solution that effectively addresses the challenges faced by existing CH selection methods in WSNs. The vital implementation descriptions of this paper are specified as follows.

- Initially, the significant CHs for the WSN are chosen through a hybridized adaptive EA model.
- The pioneer node in each lattice is identified and selected using the hybridization of GA and LA.
- To nominate the proper CH, the significant parameters are chosen regarding the constraints, viz., distance, energy, and delay.
- The performance of the proposed model is validated through a comparative investigation with other traditional models.

The rest of the paper is arranged in a fashion as given below. Section II reviews the literature concerning various CH selection protocols using different approaches. Section III deliberates the proposed architecture and its objectives. Further, Section IV explains the conventional and proposed optimization concepts used. The simulation results and the achievements are discussed in Section V.

II. LITERATURE STUDY

In 2021, Praveen Kumar R *et al.* [1] developed a CH selection model for WSN using single-hop and multi-hop CH selection through GA (S/MHCH-GA). This work implemented a novel CH selection to attain significant performance by exploiting both single- and multi-hop criteria. Moreover, the S/MHCH-GA selected CH based on the energy constraint by estimating the initial and utilized energy. The empirical results proved that the S/MHCH-GA model attained an energy-efficient WSN model and outperformed the existing methods. However, this approach is limited adaptability to dynamic environments.

Praveen Kumar R et al. [2] have introduced a CH selection model for WSN using hybridization of optimization algorithms, namely GA and Modified Particle Swarm Optimization (MPSO). Herein, the main aim of this work was to attain an energy-efficient and network lifetime-enhanced WSN model by optimizing the parameters of energy utilization, delay, throughput, network lifespan, and energy efficacy. Finally, the performance of the proposed method was validated with a comparison of the conventional optimization algorithms. However, this framework requires fine-tuning of algorithm parameters leading to high time-consuming and resource-intensive.

In 2019, Kale Navnath Dattatraya and K. Raghava Rao [3] presented a fitness-based Glowworm swarm with the Fruitfly approach (FGF) for CH selection in WSN. The implementation used a hybridized strategy of the Fruitfly Optimization Algorithm (FFOA) and Glowworm Swarm Optimization (GSO). The significant parameters considered for optimization were network lifespan, energy, delay, and cost. The simulation investigation shows the performance of the proposed FGF method by outperforming the traditional optimization methods. However, this hybrid methodology faces issues like high complexity, and it is not scalable to handle large WSN.

In 2020, Sim Sze Yin and Yoni Danieli [4] developed a New Individual Updating Strategies-based Hybrid Elephant Herding Optimization Algorithm (NIUS-HEHOA) model for CH selection in WSN. It began with clustering, energy utilization estimation, and enhancing network lifespan. The obtained results showed the performance of the NIUS-HEHOA, which accomplished better throughput, alive nodes, and energy than other conventional methods. Although this framework achieved improved results than the conventional models, it lacks adaptability to the changing dynamic characteristics of WSN, leading to performance reduction in real-time application.

In 2018, A. Rajagopal *et al.* [5] addressed a CH selection model for WSN using the hybridization of Bacterial foraging Optimization (BFO) and bee swarm Optimization (BSO). However, the main objective was to minimize the packet delivery ratio and maximize energy efficiency. The performance of the hybrid BFO-BSO approach was validated in terms of delay, energy, and lattice count. However, the CH selection using this hybrid approach is time-consuming and resource-constraint, making it less applicable for real-world WSN scenarios.

In 2021, Umashankar ML *et al.* [6] adopted a hybrid Simulated Annealing (SA) approach to nominate efficient pioneer nodes for WSN. The aim was to attain energy efficiency by considering battery capabilities like size, rechargeable, and replaceable properties. The investigation analysis revealed better network lifespan and robustness

TABLE I
LITERATURE SURVEY

Authors	Methodology	Benefits	Drawbacks
Praveen Kumar R et al. [1]	Single-hop and Multi-hop CH Selection through GA	Energy-efficient CH selection, suitable for resource-constrained environment	Limited adaptability and scalability to dynamic environments
Praveen Kumar R et al. [2]	GA and Modified Particle Swarm Optimization (MPSO)	Optimizes energy utilization, throughput, network lifespan, and energy efficacy	Requires fine-tuning and time-consuming
Kale Navnath Dattatraya and K. Raghava Rao [3]	Fruitfly Optimization Algorithm (FFOA) and Glowworm Swarm Optimization (GSO)	Increases network lifespan, and minimizes delay and cost function	Complex and cannot handle large networks
Sim Sze Yin and Yoni Danieli [4]	New Individual Updating Strategies-based Hybrid Elephant Herding Optimization Algorithm	Energy optimization, improves network lifespan, increased throughput and alive nodes	Cannot adapt to the dynamic changes in WSN, making it inefficient for real-time application
A. Rajagopal et al. [5]	Bacterial foraging Optimization (BFO) and bee swarm Optimization (BSO)	Improves packet delivery ratio and maximizes energy efficiency	Time-consuming and resource-intensive
Umashankar ML et al. [6]	Simulated Annealing (SA) algorithm	Achieved improved energy efficiency, and better network lifespan than other conventional models	Increased computational overhead and limited reliability

through customized battery enhancements. However, this method faces challenges like high computational overhead, and limited reliability. Table 1 presents the literature survey.

In the related work section, we reviewed several existing algorithms for CH selection in WSN. Each technique offers certain advantages and disadvantages. The major drawbacks of the existing approaches include limited scalability, less adaptability, high computational overhead, demand of fine-tuning, large time consumption, etc. Moreover, some methods are resource-intensive, making it less applicable for real-world CH scenarios, since the WSN are resource-constrained environments. In addition, the existing works face difficulty in integrating the different optimization algorithms into a single approach, leading to increased system complexity. To resolve these issues, we proposed a novel CH selection protocol using the combination of Genetic Algorithm and Lion Algorithm. The developed methodology aims to improve the performance of WSN by enhancing the CH selection efficiency. This algorithm considers the constraints like distance, energy, delay, etc., for selecting the CH. Unlike the existing works discussed in the literature survey, which predominantly focus on single optimization algorithms or their hybrids, the designed approach strategically combines GA and LA to exploit their complementary strengths, enhancing the efficiency and effectiveness of CH selection. The improved exploration and exploitation phase of the developed algorithm enables to identify the optimal CH candidate in the solution space. In addition, the developed algorithm prioritizes CH selection nodes with high energy, offering robust and more reliable communication within the WS network. Moreover, the iterative CH selection process ensures prolonged network lifespan and increases the energy utilization, improving the WSN performances. Thus, the proposed algorithm addresses the challenges faced by the conventional models.

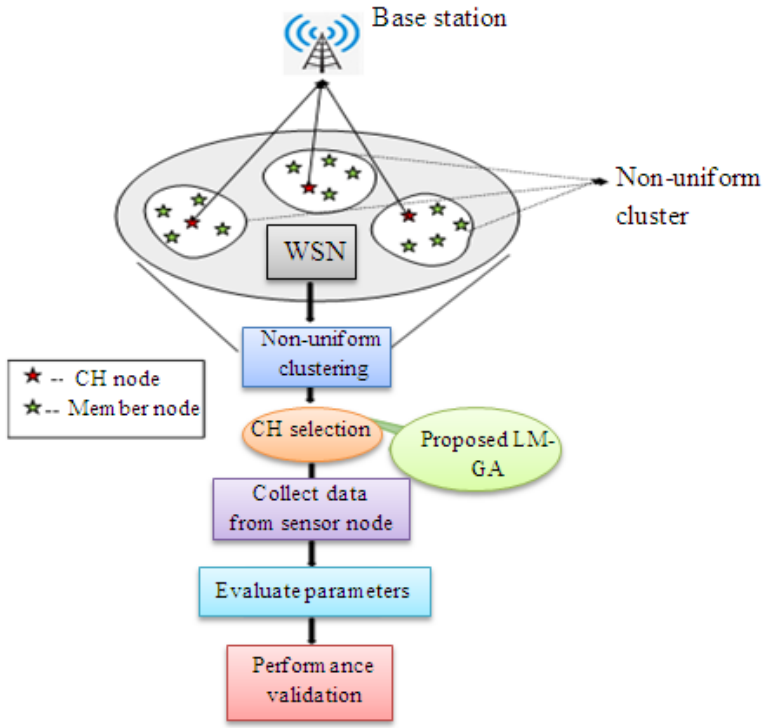


Fig. 1. Block Diagram of Proposed CH Selection Model for WSN

III. A NOVEL CH SELECTION MODEL FOR WSN

A. Proposed Architecture

This section discusses the novel CH selection protocol established to accomplish the desired WSN performance. Fig. 1 depicts the model of the proposed CH selection protocol. Initially, the significant CHs for the WSN are chosen through a hybridized adaptive EA model. The hybridized adaptive Evolutionary Algorithm (EA) model uses the combination of Genetic Algorithm (GA) and Lion Algorithm (LA) for optimal selection of CHs in WSN[32, 33]. The GA is an evolutionary optimization algorithm developed based on the concept of genetics. This approach iteratively evolves a population of candidate solutions to estimate the optimal solution by following the steps like selection, mutation, crossover, etc. Consequently, the LA model aims to refine solutions iteratively; thereby improving the models CH process. The pioneer node in each lattice is identified and selected using the hybridization of GA and LA. The significant parameters about the constraints, viz distance, energy, and delay, are chosen to nominate the proper CH. Here, a non-uniform clustering approach is employed to cluster the network, which also assists in minimizing network cost and battery costs. Thereby, a cost-effective system is achieved. Now, non-uniform clustered lattices are created based on the CH nodes' energy efficiency, and the cluster's other nodes are collectively termed member nodes. The CH nodes are selected through the proposed, and the election process iteratively takes place. According to the fitness of the proposed model, the CH nodes are evaluated based on the residual energy for each iteration. Each sensor node's energy drains through data transmission over the network by choosing an optimal CH that is reliable as far as efficient WSN is possible. The proposed model effectively carries out this process. The energy E estimation for every data transmission is stated in Eq. (1), where $E(Q)$ specifies the energy among the CH and member nodes, and the estimation of $E(Q)$ is given in Eq. (2). Besides, $U(i)$ is the energy between CH and BS in Eq. (3). Further, $E(P)$ is the vitality among two member nodes and the calculation is explained in Eq. (4), in which $K(N_j)$ is the j^{th} member node energy, $K(CH_i)$ is the i^{th} CH energy, n speaks out the CH count, and l points to the total number of nodes in the network.

$$E = \frac{E(Q)}{E(P)} \quad (1)$$

$$E(Q) = \sum_{i=1}^{n} U(i) \quad (2)$$

$$U(i) = \sum_{\substack{j=1 \\ j \in i}}^{l} (1 - K(N_j) \times K(CH_i) : 1 \leq i < n \quad (3)$$

$$E(P) = m \times max_{j=1}^{l} K(N_j) \times max_{i=1}^{n} K(CH_i) \quad (4)$$

The threshold distance after every iteration is calculated according to Eq. (5), which is the distance derivation among the nodes in the network. Here, BS is the base station, $D(Q)$ is the distance between the CH and member nodes as defined in Eq. (6), and $D(P)$ refers to the distance between two adjacent member nodes as presented in Eq. (7).

$$D = \frac{D(Q)}{D(P)} \quad (5)$$

$$D(Q) = \sum_{j=1}^{l}\sum_{i=1}^{n}\|N_j - CH_i\| + \|CH_i - BS\| \quad (6)$$

$$D(P) = \sum_{j=1}^{l}\sum_{i=1}^{n}\|N_j - N_i\| \quad (7)$$

Moreover, the network lifespan is directly proportional to the amount of data transferred. Reducing data transmission delay can improve the network lifespan. Eq. (8) shows the delay due to data transfer among the nodes, where $max_{i=1}^{n}(CH_i)$ is the total number of CHs.

$$D_t = \frac{max_{i=1}^{n}(CH_i)}{l} \quad (8)$$

These are the significant parameters identified to influence the WSN performance.

IV. TRADITIONAL AND PROPOSED GA AND LA MODELS

A. Traditional GA

Like all other optimization concepts, GA [29] begins with population initialization, fitness estimation, and fitness. The termination of GA is carried out by convergence like other optimizations as well. Nevertheless, the operations in between the GA varies with other optimization concepts. Generally, GA comprises S chromosomes with population R, and fitness is estimated for each chromosome. Following fitness, mate selection takes place to perform crossover and mutation. These are the main three steps in GA. The mathematical model of GA is expressed in Eq. (10), where $P(a_i)$ is the probability of selected individual S.

$$P(a_i) = \frac{f(a_i)}{\sum_{j=1}^{n} f(a_i)} \quad (10)$$

In many cases, the $P(a_i)$ chosen as unmodified from two-parent chromosomes N produces no new solutions. Moreover, the child's genes modify arbitrarily, and mutation is normally slow in GA. The steps in GA optimization are described below,

Population initialization: The GA optimization begins with the initialization of the population of potential solutions with fixed population size. The population size defines the number of solutions in each generation of the GA optimization.

Selection: Determine the fitness function for each individual in the population. The fitness function is evaluated based on objective function (optimization problem). After fitness evaluation, the individuals or solutions with higher fitness were selected for reproduction.

Crossover: In this phase, the selection solutions are combined to create new offspring through crossover operation. This phase involves exchanging the genetic materials between the selected individuals to create new solutions.

Mutation: After crossover step, the next process is mutation in which random changes are done to the created new solutions to prevent premature convergence to suboptimal solutions.

Termination: The final step is termination. The GA algorithm iterates through generations until a termination criterion is met. The termination conditions include maximum number of iterations, maximum convergence, etc.

B. Traditional LA

LA [28] is developed based on the biological life behavior of lions. Here, the male lion is R_{male}, the female lion is R_{female}, and the nomad lion is R_{nomad}. There are two main processes, namely crossover and mutation, and one auxiliary process, such as gender clustering, is concerned with mating. Hither, R_{male}, and R_{female} produce up to four cubs after computing crossover. Furthermore, cubs are the solutions attained by both $\square_{male}$s and $\square_{female}s$. Every cub is delivered through a different crossover mask C_m. In other words, m^{th} mask C_m is utilized for producing $R^{cubs}(m)$. Further, these cubs R^{cubs} are subjected to mutation, and produce another four cubs namely R^{new}. Subsequently, all these cubs are placed in the cubs' pool, and the gender clustering process takes place to decide R^{m_cubs} and R^{f_cubs}. The steps involved in LA are described below,

Initialization of Population: The LA optimization commences with the initialization of lions with fixed population size.

Position Initialization: *Assign and initialize the position of each lion in the population. The position indicates the potential candidate solution for solving the optimization problem.*

Fitness Evaluation: After initialization, determine the fitness of each lion in the population.

Update Leader: Select the lion with higher fitness value, and it is considered as leader. It guides the pride in the search process.

Pride Update: Then, based on the leader position, the positions of the pride are updated. This phase involves updating the lion's position to explore the search space around the leader's position.

Boundary Handling: In this step, the updated positions of the lions are verified whether it lies within the search space boundary. In case any lion deviated from the boundaries, its position was adjusted.

Hunting Behavior: Further, simulate the lion's hunting mechanism. Here, the lions explore the search space to find the potential prey.

Prey Capture and Sharing: On finding the prey, the lion shares the information with other lions in the population. The process aids in exchanging the information between the pride and leader, guiding the lions towards optimal solutions.

Termination Criteria: Check for termination conditions like maximum number of iterations or maximum convergence. If the termination is not met, repeat the steps 3 to 9.

V. SIMULATION RESULTS

A. Simulation Setup

The proposed CH selection for WSN using the hybrid approach was implemented in MATLAB 2018a on Intel core® core i3 processor 7020U@2.3 GHz, 8 GB RAM, 64-bit operating system. The efficiency and novelty of the implemented model were recorded via the simulation results. Hither, the network nodes spread over $100m \times 100m$ and amidst a BS. The initial vitality $V_{init} = 0.5$, free space vitality $V_{free\_s} = 10pj/bit/m^2$, power amplifier vitality $V_{power} = 0.0013pj/bit/m^2$, transmitter amplifier vitality $V_{trans} = 50nj/bit/m^2$, and data aggregation factor $V_{da} = 5nj/bit/signal$, and the total rounds count is 2000. The evaluation was implemented through various performance parameters such as energy, distance, delay, network lifetime, cost, and network accuracy. Furthermore, the performance of the proposed model is compared over various conventional models such as LA [28], GA [29], Grey Wolf Optimization (GWO) algorithm [30], and Whale Optimization Algorithm (WOA) [31].

B. Algorithmic Analysis

In this subsection, the performance of the proposed CH selection for WSN using the hybrid proposed model is discussed. The statistical analysis concerning mean, median, and standard deviation (SD) for proposed and traditional approaches concerning the number of alive nodes and normalized energy are given in Table II. The analysis clearly shows that the proposed model attained better energy efficiency. Figure 2 depicts the evaluation of the number of alive nodes over increasing iterations. This metric indicates the number of sensor nodes that are functioning within the network over the iterations. It enables to determine the network's resilience and robustness against node failures.

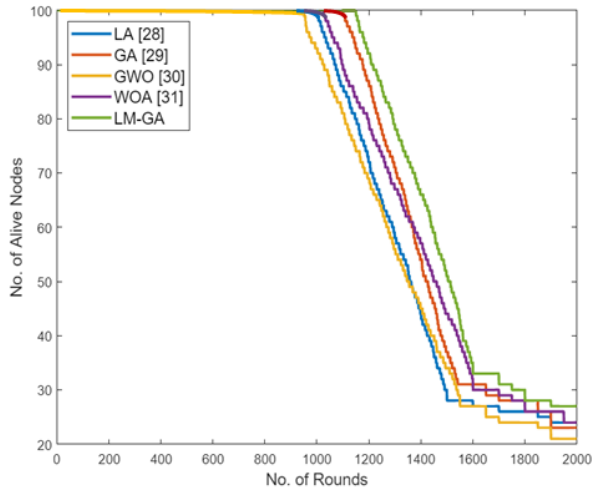


Figure 2: Evaluation of number of alive nodes

Here, the number of alive nodes are assessed over increasing iterations from 0 to 2000. The number of alive nodes achieved by the proposed method is compared and evaluated with the existing techniques such as GA, LA, GWO, and WOA. From the analysis, it is observed that the number of alive nodes in the existing techniques as well as the proposed model is constant that is 100 from 0 to 1000 iterations. After 1000 iterations, the number of alive nodes starts decreasing. At 2000 iterations, the number of alive nodes in conventional models like GA, LA, GWO, and WOA is 21, 25, 23, and 24, while the number of alive nodes in the proposed algorithm is 30. This analysis validates that the designed approach has the highest number of alive nodes than the existing models. This efficiency of the designed approach offers several advantages like improved network coverage, enhanced data delivery, and prolonged network lifetime.

TABLE II
THE STATISTICAL ANALYSIS CONCERNING MEAN, MEDIAN, AND SD FOR PROPOSED AND TRADITIONAL APPROACHES

Approaches		GA [29]	LA [28]	GWO [30]	WOA [31]	Proposed model
Mean	Alive Nodes	62.7	62.4	62.3	62.5	62.9
	Normalized Energy	0.20154	0.20187	0.20114	0.20199	0.20214
Median	Alive Nodes	75	77	77	73	79
	Normalized Energy	0.11547	0.11487	0.11211	0.11114	0.11747
SD	Alive Nodes	41.214	41.747	41.457	41.545	41.987
	Normalized Energy	0.20878	0.20854	0.20877	0.20477	0.20114

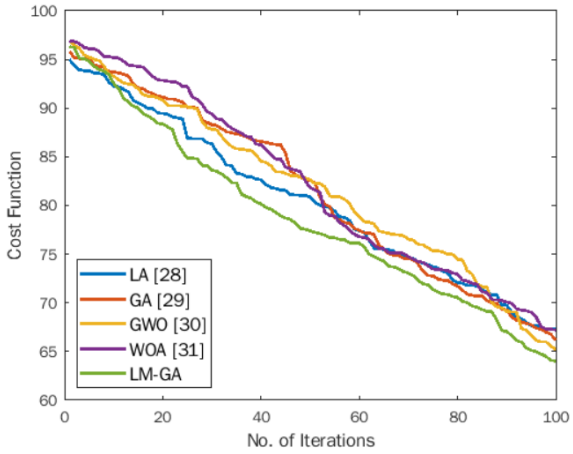


Figure 3: Evaluation of cost function

Figure 3 presents the comparison of cost function. The cost function is the metric, which measures the suitability of the sensor nodes for assuming the role of cluster heads. The cost function enables to assess multiple parameters like energy levels, communication range, energy consumption, communication overhead, and computational capabilities. Here, the cost function is compared with the existing techniques such as GA, LA, GWO, and WOA. The cost function is assessed over varying number of iterations from 0 to 100. From the evaluation, it is observed that the proposed algorithm obtained a minimum cost function compared to other techniques. At 100 iterations, the conventional models and proposed approach attained cost function of 68, 66, 70, 66, and 63, respectively. The minimal cost function indicates that the proposed algorithm contributes to reduced energy consumption, enhanced network coverage, efficient data aggregation, improved network performances, etc.

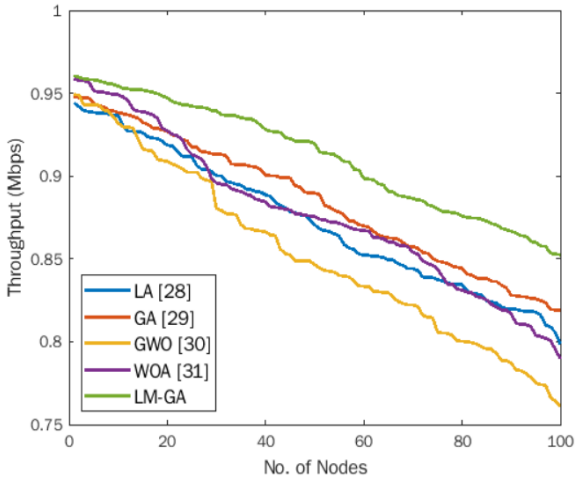


Figure 4: Throughput validation

Figure 4 depicts the validation of the throughput. The throughput metric measures the rate at which data packets are successfully transmitted from source to destination within the network. We measured the throughput in terms of megabits per second (Mbps). The higher throughput indicates improved network efficiency in transmitting data. Here, we evaluated the network throughput by varying the number of nodes from 0 to 100. The determined throughput of the developed algorithm is compared with the conventional models such as GA, LA, GWO, and WOA. From the analysis, it is observed that when the number of nodes increases, the network throughput decreases. When the number of nodes is 100, the network throughput of the above-stated existing techniques and proposed approach is 0.76, 0.78, 0.82, 0.8, and 0.86, respectively. This manifests that the developed algorithm improved the network throughput than the existing techniques. In addition, the comparative assessment highlights that the developed algorithm acts as a promising solution for improving network efficiency, enhancing data delivery, and facilitating reliable communication in wireless sensor networks.

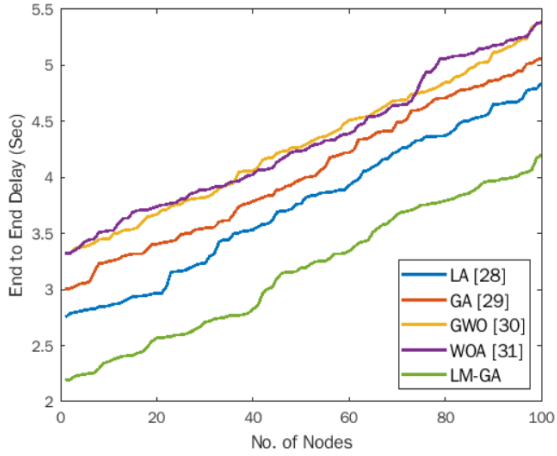


Figure 5: End-to-End delay

Figure 5 depicts the end-to-end delay. The end-to-end delay indicates the time taken for a data packet to travel from source to destination node. This metric includes all delays including transmission delay, propagation delay, queuing delay, and processing delay. To validate that the presented algorithm obtained minimum end-to-end delay, it is compared with existing techniques such as GA, LA, GWO, and WOA. Here, the end-to-end delay was assessed over increasing node count in the network. This evaluation states that the end-to-end delay increases on increasing the node count. When node count is 100, the above-mentioned conventional models and the proposed algorithm obtained end-to-end delay of 5.1s, 4.9s, 4.6s, 5.3s, and 3.8s, respectively. This lower delay implies that the developed algorithm delivers the data quickly with minimal latency in the network.

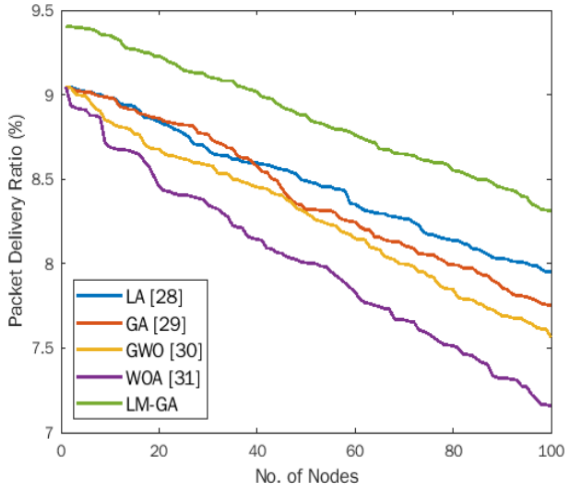


Figure 6: PDR validation

Figure 6 presents the validation of the packet delivery ratio (PDR) with the existing models. the PDR defines the proportion of the number of data packets successfully received at the destination to the total number of data packets sent by the source. It represents the effectiveness of data delivery in the network. The traditional models such as GA, LA, GWO, and WOA earned PDR of 7.2%, 8.2%, 7.6%, and 7.9%, while the developed methodology attained PDR of received at the destination to the total number of data packets sent by the source. It represents the effectiveness of data delivery in the network. The traditional models such as GA, LA, GWO, and WOA earned PDR of 7.2%, 8.2%, 7.6%, and 7.9%, while the developed methodology attained PDR of 8.5%. This illustrates that the developed algorithm offers better reliability and network performances than the existing techniques. Also, it indicates the data delivery efficiency of the network.

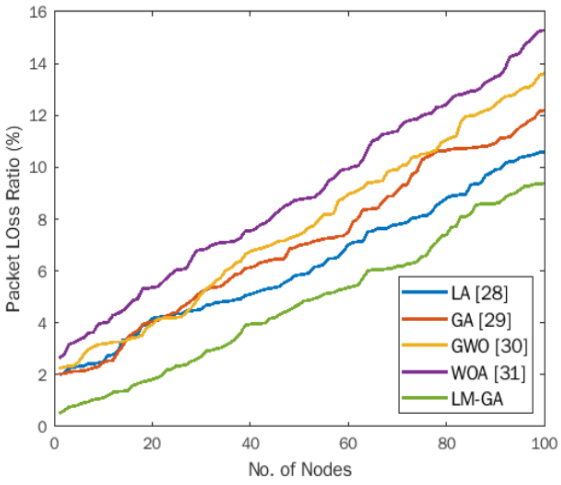


Figure 7: PLR validation

Consequently, we evaluated the packet loss ratio (PLR), which was depicted graphically in Figure 7. The packet loss ratio indicates the proportion of the number of lost packets during transmission to the total number of packets sent. It measures the reliability of the system to deliver data in WSN. The PLR is assessed by increasing the number of nodes in the network to determine the scalability of the proposed algorithm. The existing models including GA, LA, GWO, and WOA achieved PLR of 15%, 13%, 12%, and 10%, while the developed methodology achieved PLR of 9% for 100 nodes. The reduction in proposed model's PLR suggests that it is reliable in transmitting the data packets to the destination. Also, the lower PLR indicates that only few packets are lost during transmission, indicating the better network performance.

The energy consumption measures the amount of energy consumed by the sensor nodes to perform tasks like communication, data processing, sensing, etc., within the WSN. The reduction of energy consumption increases the network lifetime and improves the overall network efficiency. Figure 8 depicts the comparison of energy consumption. Here, the energy consumption metric was evaluated by increasing the number of nodes from 0 to 100. When node count is 100, the energy consumption obtained by the existing techniques like GA, LA, GWO, and WOA is 0.9mJ, 0.85mJ, 0.75mJ, and 0.8mJ, respectively, while the proposed framework consumed minimum energy of 0.55mJ. This minimum energy consumption of the developed framework highlights its applicability in resource-constrained WSN than the existing models. Also, its reduction of energy consumption aids in minimizing the cost function.

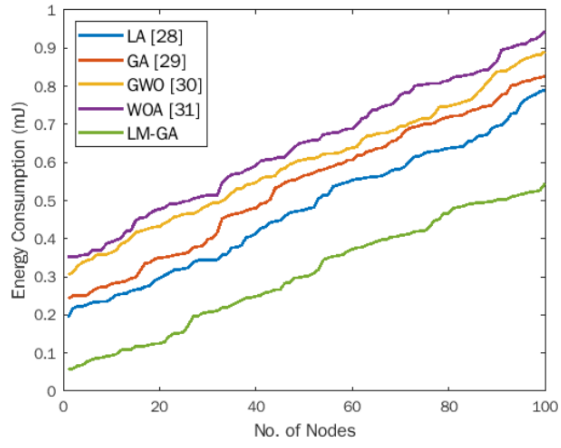


Figure 8: Energy consumption comparison

From this intensive evaluation of proposed model's performances with the conventional models like GA, LA, GWO, and WOA, it is evident that the developed model achieved improved outcomes in terms of PDR, throughput,

achieved improved outcomes in terms of PDR, throughput, and number of alive nodes. On the other hand, the metrics such as energy consumption, packet loss ratio, end-to-end delay, and cost function attained by the proposed technique are less compared to other models. This comprehensive performance evaluation suggests that the proposed method's effectiveness and reliability in providing improved Quality of Service in the WS network. Thus, the proposed model achieved considerable performance over the conventional approach and proved its efficiency.

VI. CONCLUSION

This paper has established a novel ML algorithm that hybridizes optimization concepts with an EA model for energy-efficient CH selection for WSNs. Initially, CHs for every lattice were selected based on the candidate CH selection strategy, which transfers the data. Moreover, unbalanced energy utilization and data redundancy were eliminated via multi-hop communication. For attaining the non-uniform clustering model, the routing among the data packets was done by the efficiency of the hybridized model considering energy, cost, time, network lifetime, and data accuracy. Finally, the performance of the proposed model was verified and validated through a comparative study with the existing models such as WOA, GWO, LA, and GA. The throughput achievement of the proposed model obtained 7.6%, 8.69%, 8.71%, and 13.04% better than LA, GA, GWO, and WOA, respectively. Also, the packet loss ratio attained 23.33%, 29.14%, 37.455, and 39.74% improved than LA, GA, GWO, and WOA, respectively. Thereby, the proposed model performed well and outperformed the conventional approaches. Moreover, the developed framework involves making strategic routing decisions, it ensures tradeoff between the high network lifespan and minimum delay. Also, this framework achieved an optimal balance between the data transmission accuracy and energy efficiency, making it an effective protocol for real-world CH selection scenarios.

Although the proposed work achieved better results, it faces certain limitations. Firstly, the developed algorithm is sensitive to network conditions like node mobility, charging environmental factors, topology variations, etc. These fluctuations decrease the stability and reliability of the proposed framework. Secondly, although the developed approach offered higher scalability than the existing works, it still faces problems in enhancing the scalability to meet the real-time demand in WSN. To overcome these issues, the future study should concentrate on developing adaptive models with continuous learning to resolve these issues.

REFERENCES

[1] Praveen Kumar R, Jennifer S. Raj, and Smys S, "Analysis of dynamic topology wireless sensor networks for the Internet of Things (IoT), " International Journal of Innovations in Engineering and Technology (IJIET), Sept 14 2021.

[2] Praveen Kumar R, Jennifer S. Raj, and Smys S, "Performance Analysis of Hybrid Optimization algorithm for Cluster Head Selection in Wireless Sensor Networks, " in press.

[3] Kale Navnath Dattatraya and K. Raghava Rao, "Hybrid-based cluster head selection for maximizing network lifetime and energy efficiency in WSN, " Journal of King Saud University - Computer and Information Sciences, Available online Apr 4 2019.

[4] Sim Sze Yin, and Yoni Danieli, "A Hybrid Optimization Algorithm on Cluster Head Selection to Extend Network Lifetime in WSN," Journal of Computing in Engineering, pp. 7–18, 2020.

[5] A. Rajagopal, S. Somasundaram, B. Sowmya, Performance Analysis for Efficient Cluster Head Selection in Wireless Sensor Network Using RBFO and Hybrid BFO-BSO, International Journal of Wireless Communications and Mobile Computing. Vol. 6, No. 1, 2018, pp. 1–9.

[6] Umashankar ML, Anitha TN, and Mallikarjunaswamy S. (2021) An efficient hybrid model for cluster head selection to optimize wireless sensor network using simulated annealing algorithm. Indian Journal of Science and Technology. Vol. 14, No. 3, pp. 270–288.

[7] Rajagopal, A., S. Somasundaram and B. Sowmya. "Performance Analysis for Efficient Cluster Head Selection in Wireless Sensor Network Using RBFO and Hybrid BFO-BSO." International Journal of Wireless Communications and Mobile Computing, Vol. 6, No.1, March 2018, pp. 1–9.

[8] N. Lavanya, and T. Shankar, "Hybrid based Energy Efficient Cluster Head Selection using Camel Series Elephant Herding Optimization Algorithm in WSN," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 5, 2020.

[9] J Sathya Priya, Dr. Wael Mohammad Alenazy, and A. R Sathyabama, "Energy Efficient Cluster based Routing Protocol for WSN based on Hybrid BSO-TLBO Optimization Model," Research Square, June

[10] Er. Pankaj Bajaj, 2017, Efficient Cluster Head Selection in Wireless Sensor Networks using Bacteria Foraging Optimization, International Journal Of Engineering Research & Technology (IJERT) NCIETM, Vol. 5, No. 11, 2017.

[11] Rakesh Kumar Yadav, and Rajendra Prasad Mahapatra, "Energy-Aware Optimal Cluster Head Selection Using Hybrid Algorithm for Clustering Routing in Wireless Sensor Networks," International Journal of Intelligent Engineering and Systems, Available Online, Jan 11, 2019.

[12] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand and A. H. Gandomi, "Residual Energy-Based Cluster-Head Selection in WSNs for IoT Application," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5132–5139, June 2019.

[13] A. Ali Y. Ming T. Si S. Iram and S. Chakraborty "Enhancement of RWSN lifetime via firework clustering algorithm validated by ANN" Information vol. 9 no. 3 p. 60 2018.

[14] J. Duan D. Gao D. Yang C. H. Foh and H.-H. Chen "An energy- aware trust derivation scheme with game theoretic approach in wireless sensor networks for IoT applications" IEEE Internet Things J. vol. 1, no. 1, pp. 58–69, Feb. 2014.

[15] P. C. S. Rao P. K. Jana and H. Banka "A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks" Wireless Netw. vol. 23, no. 7, pp. 2005–2020, 2017.

[16] H. Farman et al. "Multi-criteria based zone head selection in Internet of Things based wireless sensor networks" Future Gener. Comput. Syst. vol. 87, pp. 364–371, Oct. 2018.

[17] S. Kallam R. B. Madda C.-Y. Chen R. Patan and D. Cheelu "Low energy-aware communication process in IoT using the green computing approach" IET Netw. vol. 7, no. 4, pp. 258–264, Aug. 2017.

[18] H. Farman H. Javed J. Ahmad B. Jan and M. Zeeshan "Grid-based hybrid network deployment approach for energy-efficient wireless sensor networks" J. Sensors vol. 2016, 2016.

[19] S. Umbreen, D. Shehzad, N. Shafi, B. Khan and U. Habib, "An Energy-Efficient Mobility-Based Cluster Head Selection for Lifetime Enhancement of Wireless Sensor Networks," IEEE Access, vol. 8, pp. 207 779–207 793, 2020.

[20] K. A. Darabkh and L. Al-Jdayeh "A new fixed clustering based algorithm for wireless sensor networks" Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. pp. 71–76, Jun. 2018.

[21] K. Guleria and A. K. Verma "Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks" Wireless Netw. vol. 25 no. 3 pp. 1159–1183 Apr. 2019.

[22] H. El Alami and A. Najid "ECH: An enhanced clustering hierarchy approach to maximize the lifetime of wireless sensor networks" IEEE Access vol. 7, pp. 107 142–107 153, 2019.

[23] Y. H. Robinson E. G. Julie R. Kumar and L. H. Son "Probability-based cluster head selection and fuzzy multipath routing for prolonging lifetime of wireless sensor networks" Peer Netw. Appl. vol. 12 no. 5, pp. 1061–1075, May 2019.

[24] H. Mostafaei "Energy-efficient algorithm for reliable routing of wireless sensor networks" IEEE Trans. Ind. Electron. vol. 66 no. 7, pp. 5567–5575, Jul. 2019.

[25] M. Elhoseny and A. E. Hassanien "Hierarchical and clustering WSN models: Their requirements for complex applications" in Dynamic Wireless Sensor Networks Cham Switzerland: Springer vol. 165 2019.

[26] S. Yahiaoui M. Omar A. Bouabdallah E. Natalizio and Y. Challal" An energy-efficient and QoS aware routing protocol for wireless sensor and actuator networks" AEU—Int. J. Electron. Commun. vol. 83, pp. 193–203, Jan. 2018.

[27] Z. Zhao K. Xu G. Hui and L. Hu "An energy-efficient clustering routing protocol for wireless sensor networks based on AGNES with balanced energy consumption optimization" Sensors vol. 18, no. 11., p. 3938, Nov. 2018.

[28] Rajakumar Boothalingam, "Optimization using lion algorithm: a biological inspiration from lion's social behavior," Evolutionary Intelligence, Vol. 11, pp. 31–52, 2018.

[29] Yi Ding, and Xian Fu, "Kernel-Based Fuzzy C-Means Clustering Algorithm Based on Genetic Algorithm," Neurocomputing, Vol. 188, May 5 2016, pp. 233–238.

[30] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewisa, "Grey Wolf Optimizer," Advances in Engineering Software, Vol. 69, pp. 46–61, March 2014.

[31] Seyedali Mirjalili, and Andrew Lewis, "The Whale Optimization Algorithm", Advances in Engineering Software, Vol. 95, pp. 51–67, May 2016.

[32] Galán-Jiménez, J. (2017, December). Minimization of energy consumption in IP/SDN hybrid networks using genetic algorithms. In 2017 Sustainable Internet and ICT for Sustainability (SustainIT) (pp. 1–5). IEEE.

[33] Phillips, C., Gazo-Cervero, A., & Chen, X. (2011, October). Pro-active energy management for wide area networks. In IET International Conference on Communication Technology and Application (ICCTA 2011) (pp. 317–322). IET.

R. Praveen Kumar, obtained his B.E., from Seethai Ammal Engineering College and M.E., from Sona College of Technology, Anna University, He received his PhD degree from Anna University, India, Chennai. for his research work on Internet of Things.
He is currently a full Associate Professor at Easwari Engineering College, India, Chennai. He is the author and coauthor of more than 25 papers published in prestigious journals and conference proceedings.
R. Praveen Kumar is a member of the IEEE, Fellow member in IETE, Life member in IE. His research interests include Wireless Sensor Networks, Internet of Things, Wireless Networks. To his credit, he has one granted Australian patent and One Indian Patent (Published).

M. P. Prabakaran is currently working as an Associate Professor in the Department of Electronics and Communication Engineering at A.K.T Memorial College of Engineering in Kallakurichi, Tamil Nadu, and India. He received his undergraduate degree in Electronics and Communication Engineering in 2006, as well as his Master of Engineering degree in 2008 from Anna University, Tamil Nadu. He was awarded a gold medal for achieving first rank in his master's program. In 2019, he obtained his PhD in Optical Wireless Communication from Anna University, Tamil Nadu. Dr. M.P. Prabakaran has published several papers in reputable journals and contributed chapters to books. Additionally, he has presented various academic and research papers at national and international conferences. His current research activities focus on optical wireless communication, signal processing, and wireless sensor networks.

Durai Arumugam S S L received the B.Tech. Degree in Information Technology and the M.E. degree in Software Engineering from Anna University, Chennai, India. Currently pursuing Ph.D. in the area of Deep Learning in Anna University, Chennai. He has more than 9 years of teaching and research experience. He is currently working as an Assistant Professor at Easwari Engineering College, Chennai in the Department of Information Technology. Had 10 publications in reputed journals and International Conferences. His current research interests include machine learning and data analytics. He has organized a few conferences, including one virtual conference.

J. Selvakumar completed his Ph.D from Anna University, Chennai in March 2013, Post Graduation in M.E (CSE) from Kongu Engineering College in 2002 and Under graduation in B.E (CSE) from Madras University in 2001. He has more than 21 Years of Teaching and Research Experience. He is presently working as a Professor in Computer Science & Engineering Department in Sri Ramakrishna Engineering College, Coimbatore. He has around 82 National/International Conference and Journal Publications. His research area are Software Engineering, Agile, Dev Ops, Full Stack Development, Green Computing, Orange Computing, Carbon Footprint and Ground Water Recharge Technology. He is IEEE Senior Member (Membership No : 95408388). He has received Research grants from AICTE, Anna University and IEEE. He has also published a Patent.

Direction and Distance Estimation of Sound Sources using Microphone Arrays

Bence Csóka, Péter Fiala, and Péter Rucz

***Abstract*—This paper is concerned with the estimation of the direction and distance of sound sources using the MUSIC (Multiple Signal Classification) beamforming algorithm, and their tracking with the help of Kalman filter. Direction-of-arrival (DOA) estimations are performed by combining acoustical focusing and beamforming. Distance estimation is usually not part of the process, but it is possible through an extension of the beamforming method. MUSIC is a relatively fast and simple method for locating sound sources. It is based on the separation of the cross-spectral matrix of the received signals into signal and noise subspaces. We use the Kalman filter and its extended non-linear version for tracking moving sound sources. We evaluate the performance of these methods by simulations in the MATLAB environment and measurements with unmanned aerial vehicles (UAV). DOA estimations and tracking are possible in both scenarios, but distance estimation is shown to be significantly more problematic in the latter case. We examine the possible causes of the observed errors and discuss possibilities for developing a more robust distance estimation method in the future.**

***Index Terms*—beamforming, Kalman filter, microphone arrays, MUSIC.**

I. Introduction

THE position and trajectory of moving objects can be estimated using numerous remote sensing technologies, including optical and heat cameras or radar. If the object emits sound, it can also be localized by means of an acoustical sensor, i.e., an array of microphones. The latter approach also has its own advantages and drawbacks, and favorable environmental conditions for working reliably. Our objective is to estimate the positions of moving sound sources and track them accurately utilizing an acoustical camera. We implement an algorithm in the MATLAB environment for this task and test it by means of both simulations and measurements.

Submission date: 2023.07.21. Acknowledgement: This work is supported by the Hungarian National Research, Development and Innovation Office under contract No. K–143436.

B. Csóka, P. Fiala and P. Rucz are with the Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics (BME), Budapest, Hungary (e-mail: csokab@hit.bme.hu, fiala@hit.bme.hu, rucz@hit.bme.hu)

The MUSIC algorithm is discussed extensively in the literature. The paper by Xenaki et al. in 2014 [1] details MUSIC alongside conventional and more modern methods and compares them in two-dimensional simulations. Gupta and Kar created a version of the algorithm that is suitable for DOA estimation of coherent sources [2]. Yaning et al. improved the method by decreasing the computational complexity [3]. In general, beamforming algorithms are utilized in various fields, e.g. mining [4], detecting weak signals underwater [5], room acoustics and teleconference systems [6], navigation systems etc.

Estimating the position of sound sources usually only covers the approximation of the direction. For full 3D localization, the distance also needs to be estimated, which is a relatively novel concept in the field of acoustical beamforming. Cai et al. proposed a three-dimensional sound field reconstruction method which combines the use of beamforming and a binocular camera [7]. Valin et al. devised a 3D localization method for a video conference application which worked for up to 3 meters of distance [8]. In 2022, Merino-Martínez et al. presented a distance estimation that is based on asynchronous measurements with the same microphone array at multiple locations, which works for a quasi-stationary sound field [9]. Sarradj used a gridless version of orthogonal beamforming for 3D source-mapping to improve the resolution and reduce the computational cost [10]. Liaquat et al. developed a three-dimensional localization method for a low number of microphones [11]. In contrast, our goal is to devise a purely acoustical method that exploits beamforming with a grid for locating moving sources in a wider range of distances. We use the approach together with a 48-channel microphone array.

First, we discuss the basics of beamforming, its most important principles and concepts. After that, we move on to the MUSIC algorithm, briefly introducing the formulation used in our implementation, addressing its benefits and shortcomings compared to other beamforming methods. An overview of the extension of some beamforming concepts that make distance estimation possible is presented next, before moving on to the Kalman filter and its use in tracking moving sound sources. We conclude the article with presenting our findings from simulations and measurements and evaluating the performance of the proposed algorithm, while looking for potential future improvements.

DOI: 10.36244/ICJ.2024.2.6

II. THE BASICS OF BEAMFORMING

The two main tasks that need to be solved using microphone arrays and beamforming algorithms are acoustical focusing and source localization. Acoustical focusing is based on the Delay-and-Sum method (Figure 1). The received signals of the microphones in the array (which are assumed to be omnidirectional) are amplified and delayed separately, resulting in the amplification of sound waves arriving from the focal direction. Due to the phase relations of the original signals, the sum of these "steered" signals has a higher amplitude in case of waves arriving from the focal direction [12]. Thus, even though the individual microphones have spherical characteristics, the array can have a highly selective directivity that can be designed to suit our needs. The directional characteristics can be further improved (e.g., by the suppression of sidelobes) by individually modifying the amplifications and the delays of the received signals. It is possible to focus on different directions virtually, by applying different delays on the signals, without physically rotating the array.

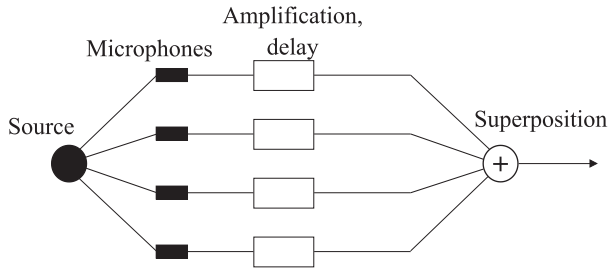


Fig. 1. The Delay-and-Sum method.

The other main task, source localization is solved by means of beamforming algorithms. A group of virtual source points is selected in three-dimensional space, making up the acoustical canvas (also called the scanning grid). By simulating sound propagation from all these virtual source points to the sensors of the array, the location where the similarity of the real (measured) and theoretical (simulated) sound fields is maximal, gives an estimated position. Focusing and source localization can be performed separately, but using them together is advantageous for the direction of arrival estimation of sound sources. By focusing on the virtual source positions one-by-one, an amplitude-like information representing the likelihood of the presence of a sound source at the given coordinate can be attained for each point of the grid using a beamforming algorithm. Thus, an amplitude map (or sound map) is attained, and we can estimate the direction of the source as the point on the canvas having the highest amplitude (likelihood). This way, source localization boils down to finding local maxima on sound maps (Figure 2).

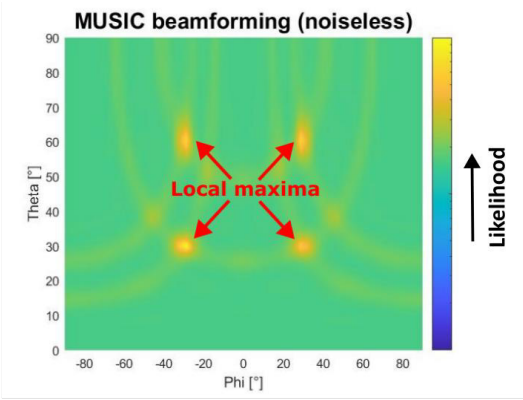


Fig. 2. Sound map created using a beamforming algorithm in a spherical coordinate system, with the azimuth denoted with Phi and the elevation denoted with Theta, having a phase domain of [-90°,90°] and [0°,90°], respectively. Source localization boils down to looking for local maxima on sound maps.

The points of the acoustical canvas are usually placed along an imaginary flat or spherical surface. The microphones of the array are most often in a cross, rectangular grid, circle, or spiral/multi-spiral formation. The distance between the array and the canvas is the focal distance, which can be taken as either finite or infinite (Figure 3). With an infinite focal distance, only the directions of the virtual source positions matter, and the wavefronts are assumed to be flat. The latter assumption holds only if the size of the array is negligible compared to the distance of the source, and hence the angles of incidence are roughly the same for each microphone. The received signals of the microphones also have the same amplitude and only differ in their phase. With a finite focal distance, the exact positions of the virtual sources must be defined, and spherical wavefronts are assumed. The received signals of the microphones differ in amplitude, phase, and angle of incidence.

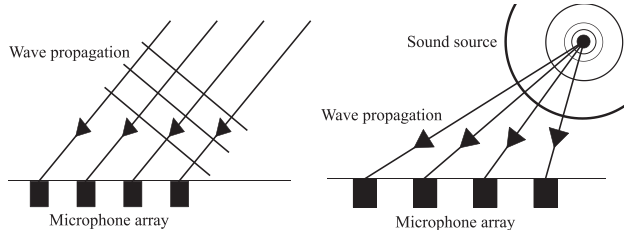


Fig. 3. Infinite and finite focal distance.

Beamforming algorithms in general can be used both in time domain and in frequency domain, and we chose the latter approach for our research. A narrow band is selected from the frequency spectra of short time windows of the received microphone signals, and the energy contained in this narrow band is the amplitude-like information calculated for each point of the acoustical canvas. Choosing the correct center frequency of the band is essential, because too low frequencies result in blurred amplitude maps, while at too high frequencies the principle of spatial sampling is violated, resulting in phantom sources at incorrect positions. This frequency f must satisfy

$$f < \frac{c}{2d} \quad (1)$$

where c is the speed of the sound and d is the distance between adjacent microphones (which, in our case, are placed evenly).

III. MUSIC

Beamforming is used for estimating the source distribution vector ($\mathbf{q}$), which is done making use of the vector of the received signals ($\mathbf{p}$) and a sensing matrix ($\mathbf{S}$). The vectors $\mathbf{q}$ and $\mathbf{p}$ both contain information in the frequency domain, i.e., every element of the vector is the complex amplitude of a component of the emitted or received signal at a given frequency, during a short time window. The number of elements in $\mathbf{q}$ equals the number of virtual source positions on the acoustical canvas, while $\mathbf{p}$ has as many elements as the number of microphones in the array. The sensing matrix connects the two vectors as follows:

$$\mathbf{p} = \mathbf{S}\boldsymbol{q}. \tag{2}$$

The elements of $\mathbf{S}$ can be calculated in 3D space as

$$\mathbf{S}(i,j) = e^{-\mathrm{j}kd_{i,j}}\frac{1}{d_{i,j}}, \tag{3}$$

where j is the imaginary unit, k is the wavenumber (ratio of the angular frequency and the speed of sound), and $d_{i,j}$ is the distance between the i-th microphone and the j-th point of the canvas.

A very important concept for several beamforming algorithms is the cross-spectral matrix (CSM), which is the spectral cross-correlation between the received signals of the microphones. The CSM can be estimated with the received signal in the frequency domain:

$$\mathbf{G} = \frac{\mathbf{P}\mathbf{P}^{\mathrm{H}}}{N}, \tag{4}$$

Here, the $\mathbf{P}$ matrix consists of the $\mathbf{p}$ vectors of the received signals from the last N number of time windows, that is, the estimated CSM is the average cross-correlation between the microphones during the most recent blocks.

MUSIC is a linear algebraic method that is based on the separation of the cross-spectral matrix of the received signals into signal and noise subspaces through eigenvalue decomposition [13]. The eigenvectors corresponding to the largest eigenvalues span the signal subspace, and the rest span the noise subspace. Then, the eigenvectors of the noise subspace $\mathbf{U}_{\mathrm{n}}$, and the sensing matrix $\mathbf{S}$ can be used for estimating the direction of the sound source. The computation is made by means of (4) [1],[2],[14],[15]:

$$\mathbf{P}_{\mathrm{MUSIC}}(j) = \frac{1}{\mathbf{s}(j)^{\mathrm{H}}\mathbf{U}_{\mathrm{n}}\mathbf{U}_{\mathrm{n}}^{\mathrm{H}}\mathbf{s}(j)} \tag{5}$$

where $\mathbf{s}(j)$ is the j-th column of the sensing matrix.

The advantages of the MUSIC algorithm are the higher resolution compared to conventional beamforming methods achieved without considerably larger computational cost and its relatively high noise tolerance. Its disadvantage is that the number of sources must be estimated beforehand. It is also important to mention that the algorithm only works if the sources are uncorrelated.

IV. DISTANCE ESTIMATION

It is important to choose the correct focal distance during acoustical focusing, because the clarity of the sound map greatly depends on it. The greater the difference between the distance of the source and the focal distance, the more blurred the image gets, therefore we can only estimate the position of the observed object with greater variance and uncertainty. This can be used to our advantage, because by extending the acoustical canvas into three dimensions and having virtual source positions at different distances, we are able to estimate the distance of the sound source on top of its direction.

Figure 4 shows the dependence of the amplitude map on the focal distance. In this simulation, the array consists of 48 microphones in a cross formation, the distance between the adjacent sensors is 6 cm, which means that with the speed of sound assumed to be around 340 m/s the upper frequency limit of spatial overlap is slightly over 2.8 kHz as per equation (1). The canvas consists of 91×181 points placed evenly on a quarter of an imaginary spherical surface, whose radius varies. Narrow band beamforming is performed with the center frequency being 2500 Hz. There is one source 5 m from the centre of the microphone array that emits Gaussian white noise. We assume the ambient background noise to be negligible, so that the signal-to-noise ratio is high. As expected, the quality of the image becomes higher when the focal distance is closer to the real distance of the source. As the focal distance becomes less accurate, the peak on the image is more spread out, the sidelobes due to the cross arrangement of the microphones are more prominent, and the level of the background is higher relative to the peak.

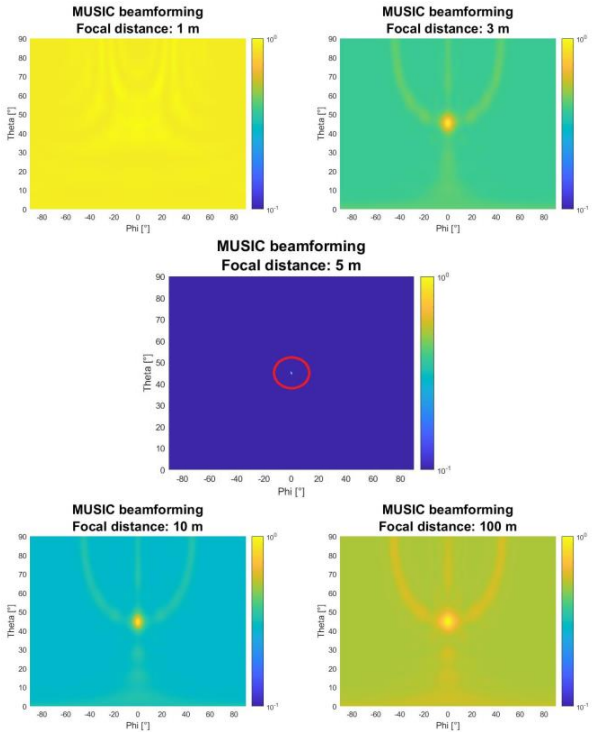


Fig. 4. Amplitude map of the same source distribution with different focal distances. The results of MUSIC beamforming are divided by their maxima in each case, and their logarithms are plotted as a two-dimensional function of the direction.

Extending the acoustical canvas into three dimensions can be done in different ways. One possible approach is to make an initial direction estimation on a primary canvas in the usual way discussed above. After the direction is determined, a secondary canvas is created, in which the virtual source positions are along a straight line in the estimated direction, but at different

distances, thus creating a discretized line as the canvas (Figure 5). We apply beamforming on this secondary canvas, and the point corresponding to the local maximum of the secondary sound map gives the estimated distance of the source. The advantage of this method is that it allows for placing the points on the second canvas very densely without increasing the computational cost significantly.

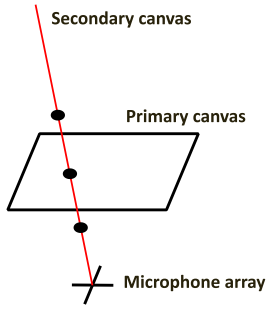


Fig. 5. Extending the initial canvas into three dimensions by multiplying the coordinates of the point that corresponds to the estimated direction.

Figure 6 shows a typical result of a distance estimation (after a successful direction estimation), where one source is located at 5 meters distance, the other is at 50 m. The secondary canvases are made up of 4500 points each, which are placed densely between 0.01 and 1000 m in a partially logarithmic manner (uniform between 0.01 m and 0.1 m, then uniform with a different resolution between 0.1 m and 1 m, and so on). The maxima are at 4.98 m and 49.4 m, respectively, which are the estimated distances. The peak corresponding to the farther source is less sharp than the one to the closer source, because the longer the focal distance, the more similar the situation becomes to infinite focal distance in the sense that the wavefronts arriving at the microphones are closer to planar. This means that the farther a source is from the array, the more difficult and inaccurate its distance estimation becomes.

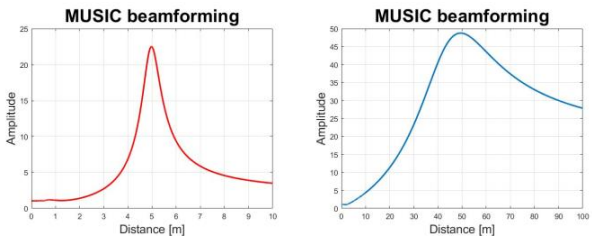


Fig. 6. Distance estimation with one source being at 5 meters (left) and the other being at 50 meters (right).

V. KALMAN FILTER

The Kalman filter estimates the state of temporally dynamic systems [16] (possible alternatives include Particle Filters, Dynamic Data Reconciliation and Double Exponential Smoothing predictors). In this case, the dynamic system is a moving sound source, and its state is its position and velocity, which can be predictively tracked. Not only the current measurements, but the earlier states of the system are also taken into consideration by the algorithm. This characteristic gives the filter higher accuracy compared to methods that only take the present state into account and opens the possibility of predicting the movement of the source. The traditional Kalman filter is an optimal estimator of linear systems, and it can be extended for nonlinear systems as well.

The Kalman filter starts from a state and an output equation, which are used in discrete time in our case:

$$\mathbf{x}(n+1) = F(x(n), u(n), n), \tag{6}$$

$$\mathbf{y}(n) = F(x(n), u(n), n). \tag{7}$$

Assuming that the system is linear and time-invariant, the state equation takes on the following form:

$$\mathbf{x}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{u}(n) + \mathbf{w}(n). \tag{8}$$

Here $\mathbf{x}(n)$ is the state vector in the n-th sampling moment, a vector which has six elements in three dimensions, those corresponding to the position (x, y, and z coordinates) and velocity (v_x, v_y, and v_z) of the sound source. The position can be measured by means of beamforming algorithms. $\mathbf{u}(n)$ is the input excitation vector, $\mathbf{A}$ and $\mathbf{B}$ are system matrices, and $\mathbf{w}(n)$ is the process noise on the input that represents the inaccuracies of the model.

The output vector $\mathbf{y}(n)$is defined by equation (9) in the linear and time-invariant case:

$$\mathbf{y}(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{u}(n) + \mathbf{v}(n), \tag{9}$$

where $\mathbf{C}$ and $\mathbf{D}$ are system matrices ($\mathbf{D}$ is negligible here because the input does not affect the output directly in our case), and $\mathbf{v}(n)$ is the noise vector on the measurements. The two noise vectors $\mathbf{w}(n)$ and $\mathbf{v}(n)$ are uncorrelated and normally distributed, with their expected value being zero and their covariance matrices being $\mathbf{Q}_n$ and $\mathbf{R}_n$, respectively.

The first step is an a-priori estimation of the state and output vectors for the $n+1$-th sampling moment (where the "-" upper index denotes that the estimation is a-priori):

$$\mathbf{x}^- = \mathbf{A}\tilde{\mathbf{x}}(n) + \mathbf{B}\mathbf{u}(n), \tag{10}$$

$$\tilde{\mathbf{y}}(n) = \mathbf{C}\mathbf{x}^-. \tag{11}$$

The difference between the measurement $\mathbf{y}(n)$ and the estimation $\tilde{\mathbf{y}}(n)$:

$$\mathbf{d}(n) = \mathbf{y}(n) - \tilde{\mathbf{y}}(n). \tag{12}$$

This difference can be used for an a-posteriori estimation (where the "+" upper index denotes that the estimation is a-posteriori):

$$\tilde{\mathbf{x}}(n+1) = \mathbf{x}^+ = \mathbf{x}^- + \mathbf{K_n}\mathbf{d}(n), \tag{13}$$

where $\mathbf{K}_n$ is the Kalman gain matrix. The optimal gain $\mathbf{K}_n$ can be found as:

$$\mathbf{P}_n^- = \mathbf{A}\mathbf{P}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}_n, \tag{14}$$

$$\mathbf{P}_n^+ = (\mathbf{I} - \mathbf{K}_n\mathbf{C})\mathbf{P}_n^{-1}(\mathbf{I} - \mathbf{K}_n\mathbf{C})^{\mathrm{T}} + \mathbf{K}_n\mathbf{R}_n\mathbf{K}_n^{\mathrm{T}} =$$
$$= (\mathbf{P}_n^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}_n^{-1}\mathbf{C})^{-1} =$$
$$= (\mathbf{I} - \mathbf{K}_n\mathbf{C})\mathbf{P}_n^-, \tag{15}$$

$$\mathbf{K}_n = \mathbf{P}_n^-\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}_n^-\mathbf{C}^{\mathrm{T}} + \mathbf{R}_n)^{-1} = \mathbf{P}_n^+\mathbf{C}^{\mathrm{T}}\mathbf{R}_n^{-1}. \tag{16}$$

$\mathbf{P}_n^-$ and $\mathbf{P}_n^+$ are the covariance matrices of the a-priori and a-posteriori state vectors, respectively.

While the above method is optimal for the estimation of the state of linear systems, observed systems are often nonlinear in real life. In our implementation, the model for the state vector is linear, but the output measurement $\mathbf{y}(n)$ is given in spherical coordinates, therefore, it is necessary to extend the Kalman filter algorithm for tackling such problems. One such extension is called Unscented Kalman Filter (UKF) [17].

With UKF, the first step is the calculation of $\mathbf{x}^-$ and $\mathbf{P}_n^-$ using equations (10) and (14), respectively. Next, we create $2N$ sigma points around $\mathbf{x}^-$, where N is the number of dimensions in the state space:

$$\mathbf{x}_i^{\sigma*}, \mathbf{x}_{N+i}^{\sigma*} = \mathbf{x}_\mathrm{n} \pm \boldsymbol{\sigma}_\mathrm{i},\ i = 1 \dots N, \tag{17}$$

where $\boldsymbol{\sigma}_i$ is the i-th row of the matrix $\sqrt{N\mathbf{P}_n^-}$. This way, the statistical average and variance of the sigma points have the same values as the state vector and its covariance matrix (and this is why these sigma points are named after the Greek letter that denotes the deviation). The nonlinear output equation (7) is applied on these sigma points, and the obtained points ($\mathbf{y}_i^\sigma$) have an average of $\tilde{\mathbf{y}}$. The covariance and cross-correlation matrices are calculated using equations (18) and (19):

$$\mathbf{P}_{\mathrm{yy}} = \frac{1}{2N}\sum_{i=1}^{2N}(\mathbf{y}_i^\sigma - \tilde{\mathbf{y}})(\mathbf{y}_i^\sigma - \tilde{\mathbf{y}})^\mathrm{T}, \tag{18}$$

$$\mathbf{P}_{\mathrm{xy}} = \frac{1}{2N}\sum_{i=1}^{2N}(\mathbf{x}_i^{\sigma*} - \mathbf{x}^-)(\mathbf{y}_i^\sigma - \tilde{\mathbf{y}})^\mathrm{T}. \tag{19}$$

The Kalman gain matrix is found as

$$\mathbf{K}_\mathrm{n} = \mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{yy}}^{-1}. \tag{20}$$

This correction matrix can now be used for estimating the state vector and its covariance while taking current measurements and previous states into account:

$$\mathbf{x}_{n+1} = \tilde{\mathbf{x}}^+ = \mathbf{x}^- + \mathbf{K}_n(\mathbf{y}_n - \tilde{\mathbf{y}}), \tag{21}$$

$$\mathbf{P}_{n+1} = \mathbf{P}^+ = \mathbf{P}^- + \mathbf{K}_n\big(\mathbf{P}_{\mathrm{yy}} + \mathbf{R}\big)\mathbf{K}_n^\mathrm{T}. \tag{22}$$

VI. SIMULATION EXAMPLE

In this section we test the proposed methods of moving sound source localization with distance estimation by a simulation example.

During the simulations, the array consists of 48 microphones placed in a cross arrangement, and the distance between neighboring ones is 6 cm. This means that the upper frequency limit for the spatial overlap is the same as in section IV. The primary canvas consists of 20000 virtual source positions distributed evenly on a rectangular area, 15 m from the array (Figure 7). The secondary canvas lies along the initially estimated direction and consists of 4500 points distributed in a partially logarithmic manner between 0.01 and 1000 m. There is one simulated point source in the space that emits filtered white noise and moves along a straight line with constant velocity (1, 5, or 10 m/s), parallel to the plane of the sensor array, constantly at either 5, 25 or 50 m (where this distance is from the plane of the array). The SNR is 10 dB, and the time windows are 50 ms long (the SNR here means the ratio of the variances of the "useful" white noise emitted by the source, and the background white noise).

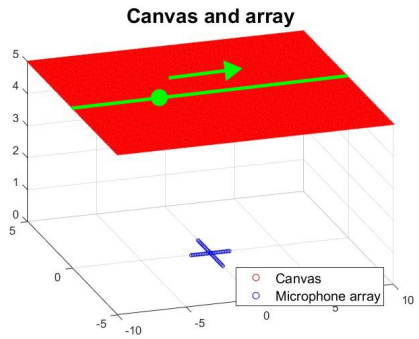


Fig. 7. The canvas (red), the microphone array (blue) and the sound source (green) in the simulation example. The source moves along a straight line with constant velocity.

The presented simulation covers both direction and distance estimation. Direction estimation is successful using the MUSIC algorithm, and the accuracy is further increased by the Kalman filter. Distance estimation is more difficult (because the main lobe of the beamforming is relatively wider along the secondary canvas, thus resulting in an estimation that varies more around the actual distance), but still successful in this simulation example (Figure 8). For farther sources, the method is less accurate, because when the size of the microphone array is much smaller than the distance of the source, the wavefronts are closer to planar and slight changes in distance result in only negligible changes in the angles of incidence. This can be evaluated by investigating the variance of the estimated distance, and this variance is proportionately greater for farther sources. In all three cases, the full 3D position estimation of MUSIC serves as the measurement data for the Kalman filter, and when its parameters are tuned properly, the estimation fluctuates considerably less around the correct distance (the MSE of beamforming is greater than the MSE of the Kalman filter, in most cases by roughly 120-150%, 50-100% and 20-60% for 5, 25 and 50 m, respectively). Thus, this example serves as a promising starting point when proving the adequacy of the distance estimation algorithm.

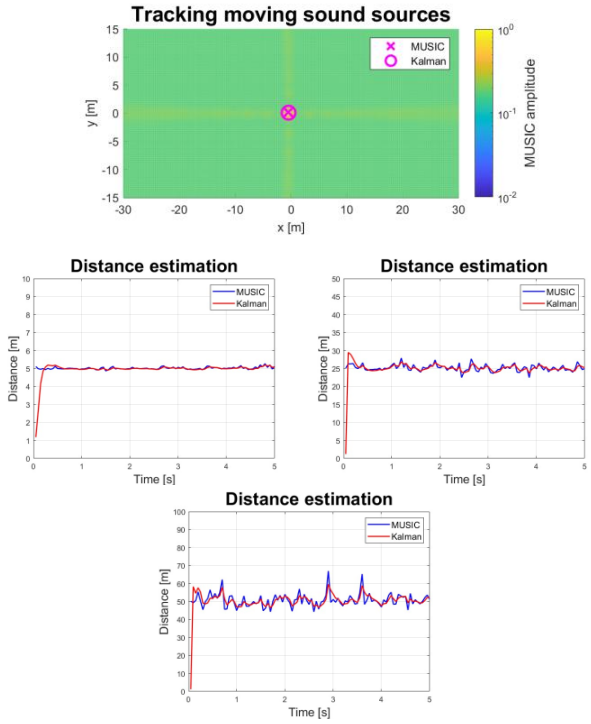


Fig. 8. Direction and distance estimation of a moving sound source using MUSIC and the Kalman filter; the distances are 5, 25 and 50 meters.

VII. OUTDOOR MEASUREMENTS

The validity of the algorithms was proven by simulations, and we demonstrate their results by processing measurements that are closer to real life situations in this section. We performed outdoor measurements as part of a larger project, where the position of unmanned aerial vehicles (rotary wing drones) was estimated by means of different principles. Because the drones were emitting sound, the acoustical method was viable. The two drones measured were of types Secop X8 and Tarot 680.

The microphone array used for the measurements consisted of 48 condenser microphones, placed in the same cross formation as in the previous simulations, stuck firmly in appropriately sized holes on a wooden plank. The array was connected to a PC through an analog/digital converter that sampled all channels simultaneously at rate of 48 kHz. A web camera was placed on the top of the wooden plank. Video capturing was made time-synchronized with the microphone array recordings. Hence, the video recordings of the movement of the drones could be fitted onto the sound maps, so that adequacy of DOA estimations could be confirmed visually.

Direction estimation using MUSIC and Kalman filter is successful in most of the time windows (one snapshot of each of the two drones is seen in Figure 9). The closer the drone is to the microphone array, and the louder its emitted sound is compared to the background noise, the more accurate the estimation becomes. The algorithm, however, does not give a correct estimation in every single time window, as there are a few moments when a strong background disturbance or ground reflection falsifies the result. Distance estimation, on the other hand, is unsuccessful. For Secopx8 we can give a rough estimation around 10 m (which is not nearly accurate enough), but for Tarot680, the estimated distance changes too erratically between time windows to accurately represent the actual distance of the drone.

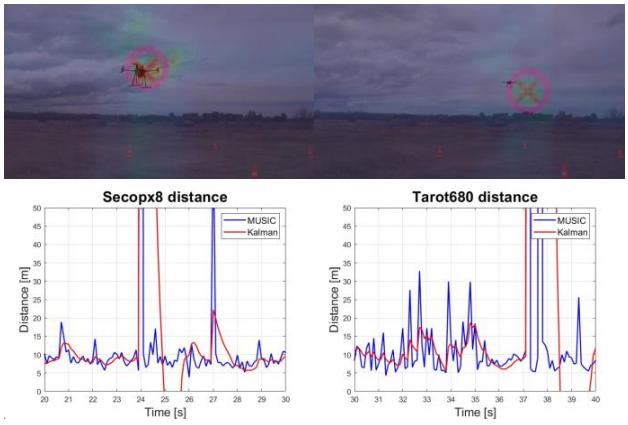


Fig. 9. Direction and distance estimation of Secopx8 (left) and Tarot680 (right).

VIII. COMPARISON AND DIFFERENCES

The beamforming algorithm was capable of estimating direction both in simulations and measurements, but distance estimation was only successful in case of the simulation example. To ensure that the approach works for outdoor measurements, it is important to identify the critical differences, parameters, and conditions we neglected during the simulation that need to be accounted for in a more robust algorithm. Potential critical differences worth investigating include:

- The nature of the sound source. The simulated source is assumed to be a point source, it emits filtered white noise, with a spherical directivity. None of these three assumptions are true for an unmanned aerial vehicle.
- The nature of the background noise. In the simulation, we assumed additive white noise; however, in real measurements, the noise has a fluctuating amplitude and a time-variant spectrum.
- The presence of ground reflections was neglected in the simulation.
- The trajectory and velocity of the source. Drones did not move along a straight line with constant velocity.

So far, we have investigated the emitted sound and the effect of the center frequency (introduced in section II.), and ground reflections.

In the previous simulation, the sound source emitted filtered white noise. Because it is a wideband signal, many choices are possible under the upper frequency limit of spatial overlap for the narrow band detection. However, the noise emitted by a drone has strong tonal components, i.e., the energy in small time windows is concentrated around harmonically related spectral peaks. As the angular velocities of the rotors vary in time, the frequencies of the dominant spectral peaks also change. Therefore, in each time window, the analysis frequency should be fit onto the blade passing frequency or one of its overtones. Therefore, the performance of the distance estimation algorithm is expected to depend heavily on the frequency band.

In the next simulation, the sound source emits a signal that is a sum of sine waves, with a fundamental frequency of 300 Hz and overtones at its multiples up until 2100 Hz, with steadily decreasing amplitudes. The source moves along a straight line with constant velocity, 5 meters from the microphone array (that is placed in the same cross formation as before). The signal-to-noise ratio is set to 10 dB. Figure 10 shows the result of distance estimation depending on the analysis frequency (the analyzed narrow band is extracted with a 3rd order 20 Hz wide bandpass filter). As expected, when the analysis frequency is exactly the frequency of the overtone, the estimation is successful, and as we move farther, the variance of the result increases. At 1500 Hz, there are only minimal deviations from 5 m (at most 0.3 m), but at 1520 Hz, the algorithm is less accurate (the maximal deviation reaches 1 m). At 1540 Hz, 40 Hz from the overtone, the estimation is too unstable to be useful, even though direction estimation still works for most of the simulation. This means that distance estimation is more sensitive to the choice of the narrow frequency band analyzed in comparison to direction estimation.

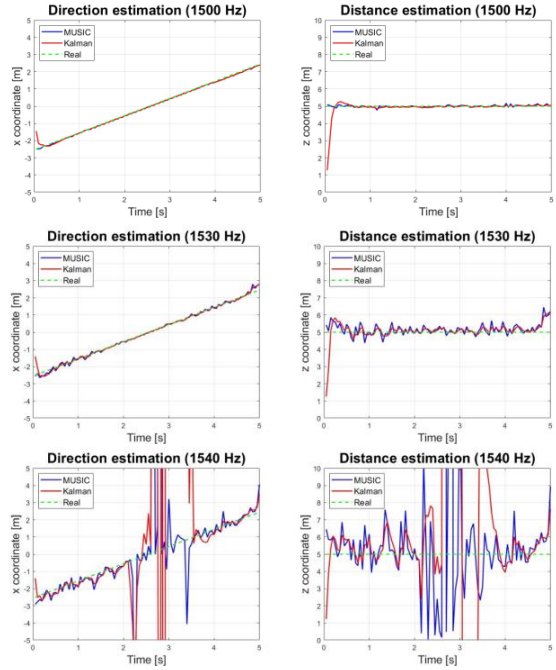


Fig. 10. Direction and distance estimation at different analysis frequencies of a simulated harmonic sound.

A more realistic sound source is incorporated into the simulation by using the sound extracted from a real recording of a moving Secop X8 drone, i.e., the sound signal received by one of the microphones in the array. In this particular recording, an overtone is present fluctuating around 640 – 650 Hz, so the analysis frequency is chosen as 640 Hz. In Figure 11 the comparison between the simulated drone sound and the real outdoor measurements can be seen. The simulated drone sound yields a worse result than the harmonic signal (probably due to the fluctuating frequency of the overtone), but the distance estimation is still much more stable than that in the outdoor measurement. It is worth mentioning that the background noise in the measurements becomes part of the emitted sound of the source in the simulation, which makes the estimation easier. From these results we can conclude that accounting for the waveform of the emitted sound by correctly choosing the analysis frequency does improve distance estimation, but in itself it is not sufficient for arriving at accurate distance estimation. Further critical differences need to be investigated.

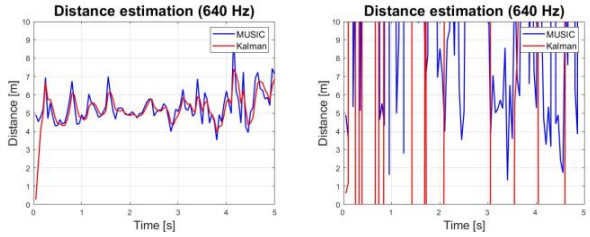


Fig. 11. Comparison of distance estimation during a simulation and a measurement on the same analysis frequency.

During outdoor measurements, the environmental conditions are not ideal, and this includes the sound of the drones being reflected from different surfaces, for example from the ground. If the reflected sound is strong enough, there is a false local maximum in its direction, which can impede correct localization. This issue was neglected in previous simulations, so it is worth evaluating how much negative effect ground reflections have on the performance of the distance estimation method.

The next measurement was carried out in a semi-anechoic chamber, with the same microphone array configuration as before. The sound source was the speaker of a stationary mobile the Secop X8 drone as in the previous simulation. The mobile phone was placed approximately 5 m from the microphone array, first on the reflective ground, to eliminate ground reflections, then on the top of a small table, around 46 cm from the ground, so that both direct and reflected sound waves can reach the microphones. In the former case, by placing the sound source as close as possible to the reflective floor, the difference between the distances of the actual and mirror sources to the same microphone is reduced close to zero, and thus the two signal paths have negligible phase differences.

Figure 12 shows the result of distance estimation, both without and with ground reflections. When reflections are not present, the estimation can be considered successful, even though it fluctuates a bit around the real distance. The measured distance is closer than 5 m, because the wooden plank holding the microphones has a small angle of inclination, and the source was somewhat closer to the plane of the microphone array than 5 m. When reflections are introduced, the peak of MUSIC beamforming on the secondary canvas becomes much flatter, and its magnitude also decreases, therefore the estimated distance has a larger variance. In this case, distance estimation produces a completely unusable result, similar to outdoor measurements. This means that ground reflections are indeed a critical condition that need to be accounted for in the algorithm.

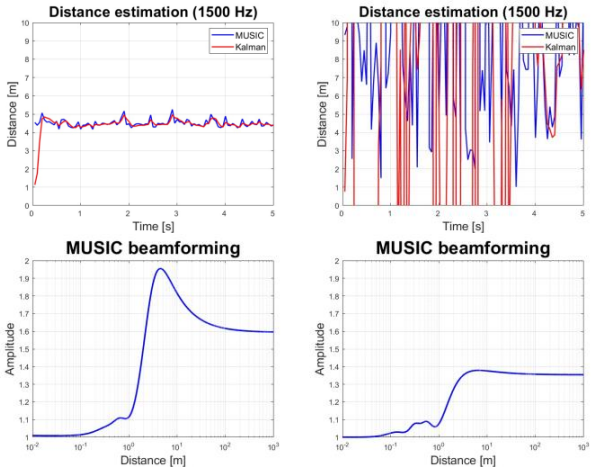


Fig. 12. Comparison of distance estimation during a measurement in a semi-anechoic chamber, without (left) and with (right) ground reflections. The lower figures show the direct, not normalized output of MUSIC on the secondary canvas, as a function of the distance.

To conclude, even though full three-dimensional position estimation is proven possible during measurements, the method still needs further refinement. Even in almost ideal conditions, in a semi-anechoic chamber, the estimation is slightly inaccurate, and the algorithm is not robust enough to handle unfavorable environmental conditions. So far, the only measurement where distance estimation was successful took place in a controlled environment with little to no disturbances.

IX. CONCLUSION AND FUTURE PLANS

In this paper, we discussed the direction and distance estimation of sound sources using microphone arrays and beamforming algorithms. We chose the MUSIC algorithm for beamforming, which was extended by the Kalman filter method for tracking moving sound sources. During simulations and measurements, MUSIC was sufficient for direction estimation, and the Kalman filter improved the results further by smoothing out rapidly oscillating measurement data. However, distance estimation only worked initially during simulations.

Future goals are to investigate the critical differences between simulations and outdoor measurements, and to find the reason for the failure of the distance estimation algorithm. So far, we have investigated the waveform of the emitted sound and concluded that correctly choosing the analysis frequency improves the estimation. We also performed a measurement in a semi-anechoic chamber, where the absence of ground reflections made distance estimation possible, but their presence yielded similarly unstable estimations as those attained in case of outdoor measurements. These two conditions can be accounted for in the future by implementing an adaptive

frequency tracking algorithm and using a method that is robust against ground reflections and correlated signals (for example the SAMV method [18]).

REFERENCES

[1] A. Xenaki, P. Gerstoft, K. Mosegaard: "Compressive beamforming". *The Journal of the Acoustical Society of America*, Vol. 136 (1), 2014, pp. 260–271, DOI: 10.1121/1.4883360.

[2] P. Gupta, S. P. Kar: "MUSIC and improved MUSIC algorithm to estimate direction of arrival". *2015 International Conference on Communications and Signal Processing (ICCSP)*, Melmaruvathur, 2015, pp. 0757–0761, DOI: 10.1109/ICCSP.2015.7322593.

[3] L. Yaning, F. Juntao, R. Xinghao, M. Le: "An improved MUSIC algorithm for DOA estimation of non-coherent signals with planar array". *Journal of Physics: Conference Series*. 1060 012026, 2018, DOI: 10.1088/1742-6596/1060/1/012026.

[4] H. E. Camargo, R. A. Burdisso, P. A. Ravetta, A. K. Smith: "A comparison of beamforming processing techniques for low frequency noise source identification in mining equipment". *American Society of Mechanical Engineers*, 2009, pp. 1–7, https://www.cdc.gov/niosh/mining%5C/works/coversheet1134.html

[5] T. C. Yang: „Deconvolved Conventional Beamforming for a Horizontal Line Array". *IEEE Journal of Oceanic Engineering*, Vol. 43 (1), Jan. 2018, pp. 160–172, DOI: 10.1109/JOE.2017.2680818.

[6] M. Imran, A. Hussain, N. M. Qazi, M. Sadiq: „A methodology for sound source localization and tracking: Development of 3D microphone array for near-field and far-field applications". *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, 2016, pp. 586–591, DOI: 10.1109/IBCAST.2016.7429936.

[7] Y. Cai, X. Liu, Y. Xiong, X. Wu: "Three-Dimensional Sound Field Reconstruction and Sound Power Estimation by Stereo Vision and Beamforming Technology". *Applied Sciences*, 2021, 11(1), 92, DOI: 10.3390/app11010092.

[8] J.-M. Valin, F. Michaud, J. Rouat: "Robust 3D Localization and Tracking of Sound Sources Using Beamforming and Particle Filtering". *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, France, 2006, pp. IV(841)–IV(844), DOI: 10.1109/ICASSP.2006.1661100.

[9] R. Merino-Martínez, B. von de Hoff, D. Morata, M. Snellen: "Three-dimensional acoustic imaging using asynchronous microphone-array measurements". *9th Berlin Beamforming Conference 2022*, https://www.bebec.eu/fileadmin/bebec/downloads/bebec-2022/papers/BeBeC-2022-S08.pdf.

[10] E. Sarradj: "Three-dimensional gridless source mapping using a signal subspace approach". *9th Berlin Beamforming Conference 2022*, https://www.bebec.eu/fileadmin/bebec/downloads/bebec-2022/papers/BeBeC-2022-S06.pdf.

[11] M. U. Liaquat, H. S. Munawar, A. Rahman, Z. Qadir, A. Z. Kouzani, M. A. P. Mahmud: "Sound localization for ad-hoc microphone arrays". *Energies 2021*, 14(12), 3446, DOI: 10.3390/en14123446.

[12] J. Novoa, R. Mahu, A. Díaz, J. Wuth, R. Stern, N. B. Yoma: "Weighted delay-and-sum beamforming guided by visual tracking for human-robot interaction". 2019, DOI: 10.48550/arXiv.1906.07298.

[13] R. Schmidt: "Multiple emitter location and signal parameter estimation". *IEEE Transactions on Antennas and Propagation* Vol. 34, 1986, pp. 276–280, DOI: 10.1109/TAP.1986.1143830.

[14] M. Mohanna, M. L. Rabeh, E. M. Zieur, S. Hekala: "Optimization of MUSIC algorithm for angle of arrival estimation in wireless communications". *NRIAG Journal of Astronomy and Geophysics*, Vol. 2 (1), June 2013, pp. 116–124, DOI: 10.1016/j.nrjag.2013.06.014.

[15] Q. Zhao, W. Liang: "A Modified MUSIC Algorithm Based on Eigen Space". In: Jin D., Lin S. (eds) Advances in *Computer Science, Intelligent System and Environment*. *Advances in Intelligent and Soft Computing*, Vol 104. Springer, Berlin, Heidelberg, 2011, DOI: 10.1007/978-3-642-23777-5_45.

[16] D. Simon: "Optimal State Estimation – Kalman, H∞, and Nonlinear Approaches". John Wiley & Sons, Inc., Hoboken, New Jersey (2006).

[17] Z. Belső, B. Gáti, I. Koller, P. Rucz, A. Turóczi: "Design of a nonlinear state estimator for navigation of autonomous aerial vehicles". *Repüléstudományi közlemények (Aviation scientific publications)* XXVII/3 pp. 255–276 (2015).

[18] H. Abeida, Q. Zhang, J. Li, N. Merabtine: "Iterative Sparse asymptotic minimum variance based approaches for array processing". *IEE Transactions on Signal Processing* 61(4), pp. 933–944, February 2013, DOI: 10.1109/TSP.2012.2231676.

Bence Csóka received his MSc degree in 2021 in Electrical Engineering from Budapest University of Technology and Economics. Currently he is a PhD student at the Laboratory of Acoustics and Studio Technologies at BME and his main areas of research are microphone arrays and beamforming algorithms.

Péter Fiala has received his MSc degree (2002) and PhD (2009) at the Laboratory of Acoustics, Dept of Networked Systems and Services, Budapest University of Technology and Economics. He is active in the fields of computational acoustics, acoustic signal processing and control of noise and vibration.

Péter Rucz received his PhD degree in electrical engineering from Budapest University of Technology (BME) in 2016. His professional interests are related to acoustics, signal processing, and numerical techniques. Currently, he is an associate professor at the Laboratory of Acoustics and Studio Technologies at BME.

Comparison of Auditory and Visual Short-Term Memory Capabilities using a Serious Game Application

György Wersényi*, and Ádám Csapó†

***Abstract*—A comprehensive serious game application has been designed and implemented to examine the capacity and effective- ness of short-term auditory and visual memory, otherwise known as working memory in human subjects. Participants engaged in an adaptation of the well-known paired association game that entails turning over cards and recalling their placement within a 2D matrix structure of various resolutions. Each trial introduced either visual icons (vision-only condition) or auditory objects (audio-only condition). User performance was evaluated through a detailed statistical analysis focusing only on the highest 6x8 resolution condition in the application. Findings suggest that visual memory did not conclusively outperform auditory memory in the context of this game. However, within the scope of auditory stimuli, familiar iconic sounds, such as excerpts of speech and commonplace sounds, were recalled more effectively than unfamiliar, synthetic sounds like parametric waveforms. Furthermore, performance appeared to be influenced by demographic factors, with male and younger subjects yielding superior results.**

***Index Terms*—Auditory memory; visual memory, virtual simulation; gamification; memory game**

I. INTRODUCTION

In this paper, we present a serious game application that we have developed to assess users' capabilities towards retaining different kinds of auditory and visual stimuli in short-term memory. Based on results from the application, we draw conclusions that are relevant to the design of audiovisual user interfaces in a wider technological context.

A. The role of working memory in human cognition and perception

In psychology, working memory is defined as the part of short-term memory that is concerned with immediate conscious perceptual and linguistic processing. It is the cognitive system involved in the temporary storage and processing of a limited amount of information as a given task is being carried out [1]. The most important modalities in this context are the visual and the auditory modalities, with each being characterized by different capabilities [2]–[5].

Many researchers view working memory and short-term memory as significantly overlapping concepts. The difference between the two in the context of this study is that working memory is regarded as an active mechanism (process) for the manipulation and application of memory objects over a short period of time, while short-term memory simply refers to the temporary storage (capacity) of the brain, making information readily available for a short period of time [6]. Apart from this, the two terms may be used interchangeably.

Most of the information kept in short-term memory will be stored for approximately 20 to 30 seconds, or even less. Some information, however, can last in short-term memory for up to minutes, but most information spontaneously decays quite quickly [7], [8].

In contrast with working memory and short-term memory, long-term memory refers to a vast store of knowledge pertaining to prior events. It differs from short-term memory both in terms of duration and capacity. The question of decay / forgetting is still an actively researched area, although some works have argued that there may be a chunk capacity limit even in the case of long-term memory [2].

B. Investigating visual and auditory performance in a modern technological context

A well-known model of working memory is the Baddeley-Hitch model, which differentiates between two components of working memory: a place where visual and spatial information is stored, and another for recording auditory information. According to the model, a central executive part controls and mediates these components [9], [10].

Although a wide range of experiments have already been conducted to test human capabilities and to compare performance in different modalities, recent technical developments have paved the way towards new applications in the digital world with an increasingly extensive use of audiovisual information.

On the one hand, the concepts of Digital Reality and Internet of Digital Reality encompass various kinds of developments that benefit from the broader use of such kinds of enhanced user experience and immersive 3D scenarios in a functionally driven, networked artificial intelligence context [11], [12].

* Széchenyi István University, Győr, Hungary (e-mail: wersenyi@sze.hu)
† Corvinus Institute for Advanced Studies & Institute of Data Analytics and Information Systems Corvinus University of Budapest, Hungary (e-mail: adambalazs.csapo@uni-corvinus.hu)

DOI: 10.36244/ICJ.2024.2.7

Healthcare applications, combat and military simulators, gaming, and various other AR and VR application domains rely on visual and/or auditory information during feedback. Virtual audio displays (VADs) in a wider sense can be applied in simulators, virtual and embedded environments, making extensive use of auditory stimuli.

On the other hand, assistive technologies targeting elderly or disabled individuals using wearables, e.g. for reading, navigation (Electronic Travel Aids) or other use cases, also strongly rely on well-designed auditory stimuli. More generally, cognitive rehabilitation procedures can involve testing and training of the memory capabilities of patients. In all of these applications, the auditory modality and auditory memory play a significant role. Therefore, understanding what type of simulated sound sources are best suited for a given environment – in terms of, e.g. spectral content, duration, amplitude profile – can be crucial.

Serious gaming offers possibilities for testing, training, and scientific data collection using entertaining gaming environments where user involvement and motivation are maintained [1], [13]–[18]. In the era of mobile devices and virtual reality headsets, serious games can reach a large group of subjects both in online and offline scenarios. Using gamification, experiments can be designed in order to collect scientifically relevant data in an entertaining and motivating process. It is an especially useful method if the gamified scenario resembles the original environments and use cases. Gamified applications are useful for both children and elderly users, where maintaining focus and motivation is a key factor. 3D immersive virtual reality scenarios also open new areas for serious gaming where (immersive) audiovisual experience may be extended by haptic/tactile feedback. State-of-the-art gaming applications can be close to reality and virtual training scenarios, and can be suited for testing human cognition and perception performance.

There are different definitions of serious gaming. While some definitions focus almost exclusively on the "learning element" of these games (e.g., [19]), a broader and also very common conception is reflected in multiple definitions, such as "[A serious game is] any form of interactive computer-based game software for one or multiple players to be used on any platform and that has been developed with the intention to be more than entertainment" [20], or "[Serious games are] video games aimed toward problem-solving rather than entertainment" [21]. It is also reported in the literature as being widely accepted that serious games are built using novel technologies, includes at least some game-like features (such as competition, desire to win, or strategy, among others), and are created with a "serious" intention to achieve a concrete objective [22].

On the one hand, all of these areas highlight the importance of the optimization of the presentation of audiovisual information. On the other hand, a gamified environment built for the purposes of evaluating human audio-visual capabilities can in itself be regarded as a serious game. One key aspect in this context is users' ability to recall and remember various properties of audiovisual items, i.e., presence or absence, meaning, semantic connections, temporal or spectral variations, spatial locations, etc. Although the currently presented experiment did not include an evaluation of any possible learning effects, the game design allows for further experiments targeting the learning effect as well, with dedicated experimental design (repeated controlled measurements, inclusion of lower resolutions, etc.). The following subsections provide a brief overview of human visual and auditory working memory.

C. Visual memory

Most research on visual memory in the past has focused on humans' ability to remember visual stimuli over either shorter or longer periods of time, i.e. time was taken to be the key parameter [23]. Thus, it was shown that humans can remember objects seen even for brief exposures or after very long time. The visual working memory is considered as a system that retains and manipulates information over the short term, whilst visual long-term memory is defined as a passive storage of information for longer time periods [24], [25].

The most important property of working memory is its limited capacity, however, it is a core cognitive process supporting human behaviour that relies on temporarily stored visual information [25]. To this end, easily interpretable iconic representations may be most useful. In fact, although object identification and recognition is generally associated with long-term memory, it nevertheless plays a significant role in short-term memory processes as well.

Memory limitation experiments usually apply the so-called change-detection task, where objects are displayed in an array, and after a short break, another array is displayed with changes that have to be identified by the test subjects. Alternatively, single-item presentation instead of an array of items can be used, but in this case the relationship between visual objects in terms of their joint effects on memory cannot be assessed.

Luck and Vogel have suggested that working memory can store only a limited, discrete number of objects [26]. In their experiment, subjects were asked to recall an array of items characterized by a single or multiple features (i.e., color, orientation). After a short delay, another array was shown that was identical or different, and the task was to identify what changes had been applied, if any. Based on the study, Luck and Vogel found that subjects could store only 3-4 objects in working memory, and they could detect changes in both single and across multiple features. Other studies reported that capacity is reduced as feature load increases [27]. Thus, visual memory experiments have to consider both feature load (number of features) and object load (number of objects). This model assumes that subjects can remember all features of the objects within the 3-4 item limit, or will fail completely.

Another model of the memory sees the capacity as information based and limited by a finite resource that can vary unevenly across different items in a display [28]. Here, subjects reported on a continuous scale. For example, a color have to be recalled in form of selecting from a continuous color palette. It was observed that with increasing set size, the

precision of representations decreased. They concluded that subjects could store a continuous amount of information with varying precision: even with set size greater than four items, subjects are able to store more than four items in memory.

All models of the working memory suggest a capacity of 3–4 representations, but this capacity may also be limited by the amount of information load in the display (stimulus complexity) [29]–[32]. Nevertheless, it is possible that items with higher complexity also have higher similarity leading to greater errors.

Former results also showed capacity estimates to be individually different from 1 to 5 objects, furthermore, there is also a large variability within subject in repeated trials [33]. Interestingly, better memory performance was reported when the display consisted of more meaningful stimuli than meaningless images [34]. Moreover, it was demonstrated that training in action video games and even in cognitive training games can contribute to better memory capabilities [35]–[37].

D. Auditory memory

The number of objects or "chunks" humans can remember is limited and depends on the used modality (audio only, visual only, or mixed), presentation method, former training, etc. Working memory refers to the ability to retain stimuli in mind that are no longer physically present and to perform mental operations on them. It allows the temporary storage of relevant information and its task-dependent manipulation.

Regarding the subject of auditory memory, Kaiser summarized the results of several relevant experiments from a neuroscience point of view. Most of these studies have focused on the short-term retention of acoustic information [38]. Auditory memory has to do with the ability to remember words and sounds and to recall information that was received verbally [39], [40]. Zimmermann et al give a good overview of short-term and long-term auditory memory capabilities in different scenarios [41].

Memory capacity limits have been suggested to be "around seven plus or minus two" under various circumstances not limited to auditory tasks indicating a relatively low number of items humans can recall using their memory [42]. Studies show that recognition memory for sounds is usually inferior to memory for visuals [43]–[48]. In [43] four experiments were conducted to examine the nature of auditory and visual memory, including an evaluation of the role of experience in auditory and visual memory, supporting this finding. On the other hand, Lehnert and Zimmer tested the short-term memory of object locations in the auditory and visual modalities pure and mixed, and found the same memory performance in mixed and pure conditions with a very similar decline in performance to the memory load manipulation [49]. They concluded that locations of auditory and visual input are stored in common memory.

Setti et al. compared blind subjects' and sighted subjects' performance using semantic and non-semantic sounds to verify if semantic rather than non-semantic sounds could be better recalled, as well as to see whether exposure to an auditory scene could lead to enhanced memorization skills [50]. In the study, semantic sounds were spatialized in order to reproduce an audio scene. Results showed on the one hand that sighted subjects performed better than blind participants following the exploration of the semantic scene. More generally, although both blind and sighted individuals showed similar audio spatial memory skills, blind participants were found to focus more on the perceived sound positions and less on the information that they gathered on the location of individual items during their initial exploration on of the scene. These findings suggest that whereas visual experience allows the simultaneous processing of multiple stimuli, auditory processing is much more sequential.

In a 2007 experiment, 100 students participated in a learning task, where visual icons had to be associated and learned together with their auditory counterparts [51]. The visual stimuli appeared in two sets of 15 icons arranged in 3 columns and 5 rows, sound stimuli were selected from a set of auditory icons (having a semantic relationship with the visual icon) and earcons. Results showed that participants made faster and more correct matches between visual icons and auditory icons than between visual icons and earcons. This suggests the superiority of auditory icons over (non-familiar) earcons. It was also suggested that localization may be a useful cue for learning the associations between them, however, it was not conclusive.

Current studies reported experiments conducted with musicians [52], [53]. Human communicative sounds could be detected better than other sounds, especially in the case of speech and human-generated vocal sounds. Similarly, song-like vocal phrases can be remembered better, and musical training plays a significant role. If rapid pip-tones were presented to subjects, the auditory memory was found to be sensitive to repeating audio structures [54].

The number of sound events and sonification methods is a central problem of the user interface and the audio modeling as well [55]–[61]. VR environments can influence auditory memory performance based on the context-dependent representation [62].

Connected to the "Sound of Vision" research project, a serious game-based application has been developed for testing memory capabilities [63], [64]. The game was the auditory-only version of the memory game, where players have to pair cards in a matrix arrangement (e.g. Matching Pairs, Find the Pair, etc.). Preliminary results showed that users made fewer pairing errors with familiar than with unfamiliar sounds. However, the number of pairs can have a significant impact on the results.

Generally, memory games test and train visual memory in an entertaining way, using only visual modality. Figure 1 shows a real-world audio version of the game in a museum for children. Different fillings result in different noises by shaking the boxes. The same idea is behind the (non-action-based) serious game application being developed.

Fig. 1. A ”noisy game” in an exploratory installation for children.

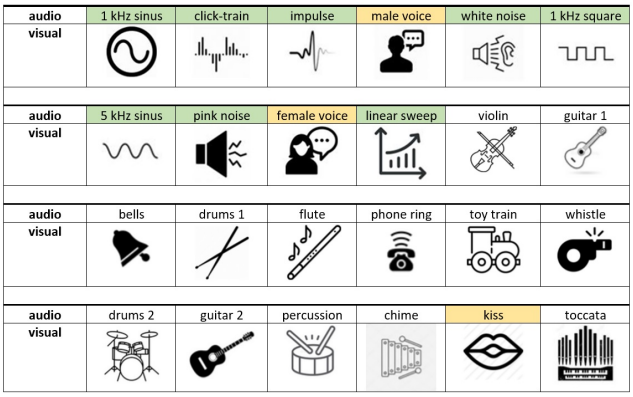


Fig. 2. Summary of the visual and auditory representations on the 6x8 resolution. Yellow marked items are labeled as ”human sounds”, green marked items are ”measurement signals”, and all others are everyday sounds called ”auditory icons”.

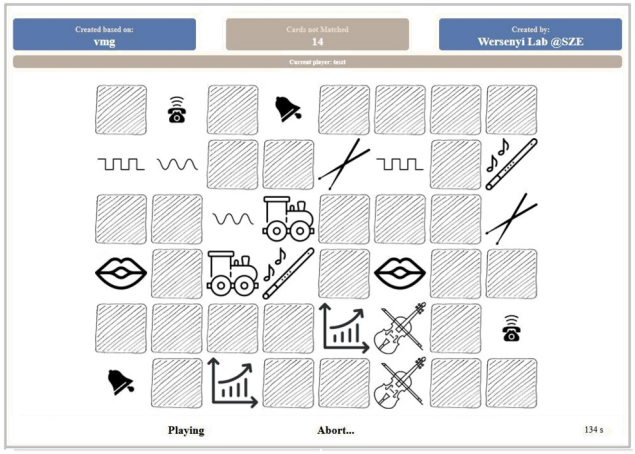


Fig. 3. Screen shot of the user interface during visual gameplay in 6x8 resolution (24 pairs).

The aim of the investigation presented in the next sections targets the evaluation and comparison of

- the memory capabilities of the visual and auditory modality;
- sound stimuli having different frequency content;
- different groups of users (male-female, young-old).

Section 2 introduces the methodology and the measurement setup. In section 3, results will be presented, and section 4 discusses the findings. After the concluding remarks, the main directions for future work will be highlighted.

II. Measurement Setup

A. Game design

The memory game we have developed for the purposes of this experiment relies on a scenario of matrix-arranged cards, with the back of each card initially facing the user. In visual mode, simple black-and-white icons are temporarily shown to users as they flip two cards of their choosing in each round, before the cards return to their original face-down position. In audio-only mode, instead of visual images, short auditory events are played back as the cards are flipped. These iconic sound samples are about 2-4 seconds long each.

The goal of the game is to find matching pairs in each round, by remembering the positions of previously overturned cards, and to do this with the least number of errors (flips). Given that the audio-only mode did not include visual stimuli, a red dot was used to indicate the back of the card that was overturned in each round until both auditory stimuli finished playing. Note that although digital adaptations of this game for single and multiplayer modes have been developed for different platforms, all of them are based on visual representations only. Our application is extended by a modality selector, a set of auditory representations, and the automated logging of results.

Figure 2 shows all visual icons with the corresponding audio sample names as well as their allocation into subgroups. Figure 3 shows an example screenshot of the game in visual mode. In the case of a full HD resolution screen (1080x1920), all cards have a resolution of 80x80 pixels. This size is fixed for all gameplays, independent of the number of cards. The game offers different resolutions from 5x2 up to 6x8, in order to theoretically allow for the comparison of performance as the number of pairs and scope of stimuli types increases. In this experiment, however, only the highest resolution of 6x8 (24 pairs) was used so as to remove the effect of potential confounding factors, like learning effects and/or under- or overrepresentation of stimuli types appearing in multiple resolutions.

After each completed game, a log file was saved on the device with the following records:

- User ID data (nickname, gender, and age);
- Date and time (for controlling later training sessions);
- Time in seconds needed for completing the game;
- Number of total flips (error rate over all pairs);
- Number of flips needed for each sound/icon to be paired.

```
{
    "subjectId": 002,
    "subjectName": "wgy",
    "subjectAge": 47,
    "subjectSex": "Male",
    "timeCost": 134,
    "totalFlips": 52,
    "modality": "audio",
    "rows": 6,
    "cols": 8,
    "timestamp": "Fri Apr 07 2023 12:40:08 CET",
    "numberOfFlipsPerCard": {
     "onekhzsquare": 3,
     "female": 3,
     "impulse1": 7,
     "clicktrain": 5,
     "whitenoise": 4,
     "pinknoise": 3,
     … },
}
```


Fig. 4. Example of the .json file for logging results. This user needed 52 flips and 134 seconds to complete the game in audio only mode. The cards covering the impulse sounds were paired after 7 flips. Data can be easily imported into Excel for statistical analysis.

Figure 4 shows an example log file with recorded data.

B. Participants

All participants were untrained in listening tests, had no formal musical training or any significant gaming experience, but were familiar with the basic idea of the game. No audiometric screening was applied. 40 subjects were included, 20 males and 20 females (mean age 28.85). Furthermore, age groups were created and subjects were allocated to subgroups ”young” (20 subjects below 25) and ”old” (20 subjects above 25). Upon finishing the games, participants were asked informally to give feedback about the experiment (motivation, difficulty, possible changes in the procedure, etc.).

C. Visual and auditory representations

Both visual icons and sound events are included in the levels hierarchically. Every new level (resolution) contains icons/signals from the previous level as well. Because the highest level (6x8) contains all the icons/sounds, it was selected as the basic experiment. Sound events were recorded/generated by the authors, or downloaded from public databases, followed by post-processing. They were selected to represent a variety of sound types, including human sounds, meaningless sounds and everyday sounds. After compiling the sound data base, visual icons were designed with semantic correlation where possible.

D. Methodology

Listening tests were carried out in a silent but non-anechoic room with supervision using the same mobile device (12.4 inch tablet). Users first were informed about the goal of the experiment, but neither the icons nor the sound samples were presented prior the game. All subjects played two games under the exact same circumstances.The first round was always a game in vision-only mode, followed by the audio-only mode. Randomizing the order of modalities was considered, but having the same conditions was considered to be more important given the sample size of the experiment. Moreover, we believe that since the auditory case included no visual stimuli (other than the red dots mentioned previously), there could have been no learning effects or other cross-effects from the visual only mode which was presented to users first. Subjects were instructed to do their best (minimizing the error rate, thus, avoiding wrong flipping of the cards), but otherwise, they were free to choose their gaming strategy and speed. In case of 10 seconds of inactivity, the game would be aborted automatically without logging the results. For maintaining motivation, subjects achieving optimal performance (completing the game without errors) would get a ”perfect game” feedback.

E. Implementation details

In this subsection, we provide a brief summary of some of the technical details of the implementation of the software, both from a technology and algorithmic perspective.

1) Software environment: From a software perspective, the memory game was implemented as a Progressive Web Application (PWA) and shared on Amazon AWS S3. The software was developed based on an application called “vue-memory-game”, shared on Github through the MIT license (https://github.com/leftstick/vue-memory-game). We extended this application with the following features:

- support for pre-defined wave files besides images;
- logging as detailed earlier;
- a “luck management” mechanism detailed in the next subsection;
- a refined user interface.

2) Minimizing the impact of “lucky” initial flips: The game consists of different levels (resolutions). The smallest and easiest is a 5x2 grid with ten cards and 5 pairs to be found. The smaller the resolution, the higher the likelihood that pairs would be found based on pure luck. Finding pairs just by clicking them by chance does not help in evaluating memory capabilities. Although this probability significantly decreases as the number of cards increases (and indeed, in this experiment we used only the highest resolution), the developed software included a correction of this effect of luck.

The algorithm works as follows. Let 2N be the number of cards, thus N is the number of pairs. The main idea is that upon initialization, the cards were not directly connected to the icons/sounds yet – instead, in the first N/2 flips, an icon/sound became associated with the card that was overturned upon demand. In each of these first N/2 flips, the method ensured that no two icons/sounds would be the same.

This method was continued until N/2 if N was even, or (N/2)+1 if N was odd, ensuring that no pairing would be possible until at least half of the icons/sounds were revealed. After this point, the rest of the stimuli were randomly allocated to the remaining cards.

It should be noted that this mechanism operated in a way that was invisible to the users (who could still flip any card

Group	Count	Sum	Mean	Variance	SS
1 kHz sinus	40	280	7,00	8,72	340,00
click-train	40	302	7,55	8,97	349,90
impulse	40	307	7,68	7,56	294,78
male voice	40	276	6,90	5,84	227,60
white noise	40	323	8,08	9,35	364,78
1 kHz square	40	298	7,45	11,43	445,90
5 kHz sinus	40	304	7,60	8,66	337,60
pink noise	40	303	7,58	7,69	299,78
female voice	40	282	7,05	8,92	347,90
linear sweep	40	312	7,80	9,24	360,40
violin	40	291	7,28	6,77	263,98
guitar 1	40	291	7,28	6,82	265,98
bells	40	297	7,43	9,89	385,78
drums 1	40	287	7,18	6,76	263,78
flute	40	293	7,33	11,30	440,78
phone ring	40	303	7,58	9,79	381,78
toy train	40	267	6,68	6,79	264,78
whistle	40	279	6,98	6,85	266,98
drums 2	40	299	7,48	10,41	405,98
guitar 2	40	296	7,40	10,91	425,60
percussion	40	312	7,80	7,86	306,40
chime	40	302	7,55	9,64	375,90
kiss	40	281	7,03	8,59	334,98
toccata	40	300	7,50	9,54	372,00

Fig. 5. Summarized results showing the mean flip values and variances for every visual icon based on 40 participants' gameplays.

Group	Count	Sum	Mean	Variance	SS
1 kHz sinus	40	309	7,72	7,48	291,97
click-train	40	301	7,52	6,66	259,97
impulse	40	311	7,77	8,17	318,97
male voice	40	222	5,55	2,87	111,90
white noise	40	303	7,57	7,17	279,77
1 kHz square	40	305	7,62	9,93	387,37
5 kHz sinus	40	292	7,30	5,75	224,40
pink noise	40	300	7,50	7,28	284,00
female voice	40	218	5,45	3,84	149,90
linear sweep	40	307	7,67	7,60	296,77
violin	40	294	7,35	6,18	241,10
guitar 1	40	319	7,97	5,10	198,97
bells	40	291	7,27	4,97	193,97
drums 1	40	318	7,95	7,99	311,90
flute	40	273	6,82	7,17	279,77
phone ring	40	255	6,37	6,75	263,37
toy train	40	252	6,30	3,44	134,40
whistle	40	277	6,92	9,04	352,77
drums 2	40	303	7,57	6,81	265,77
guitar 2	40	313	7,82	7,27	283,77
percussion	40	282	7,05	6,04	235,90
chime	40	286	7,15	7,26	283,10
kiss	40	226	5,65	2,38	93,10
toccata	40	288	7,20	6,16	240,40

Fig. 6. Summarized results showing the mean flip values and variances for every sound based on 40 participants' gameplays. Red numbers indicate statistically significant difference within the group. The male, female voice samples and the kiss sound could be recalled better than other sounds.

in any location), but it guaranteed that for the first N/2 flips, subjects could not accidentally stumble upon a pair without actually relying on their short-term memory capabilities. Given that the mechanism was also applied to each game in the same way, this introduced no comparative bias into the game at any point.

III. RESULTS

Evaluation of results was based on the number of flips for the total game and for each individual icon/sound using ANOVA and post-hoc analysis using Tukey HSD test. Tukey's HSDF tests all pairwise differences while controlling the probability of making one or more Type I errors. Significance level of 0.05 was set in all paired t-tests and corrected for during the Tukey test.

A. Visual memory

Every participant played the visual-only game first. In order to avoid feature load representations, object load solution was selected. The visual icons (see Figure 2) were similar in size, color and iconic representation. Furthermore, this approach is a modified version of the usual change-detection tasks, where an array of images were shown, but consecutively instead of simultaneously. The statistical analysis showed no difference among the 24 icons (F=0.49; p=0.978). Figure 5 shows descriptive statistics for each icon based on 40 measurements.

Among men, there was a large variability in individual results. 8 out of 10 young and 1 out of 10 old subjects were significantly better than the others. Among females, only one young subject was better. For both genders, younger subjects performed better: males have a mean of 137 (young) and a mean of 193 (old) (F=25.37; p=8.58E-05); females have a mean of 174 (young) and a mean of 193 (old) (F=6.62; p=0.019). Comparing all males and all females, the mean of males (169) is significantly better than the mean of females (189) (F=4.99; p=0.031).

Although mean values of completion time seemed to be quite different, the ANOVA did not support significant difference (246 sec. for males and 275 sec. for females).

B. Auditory memory

Every participant played the audio-only game after completing the visual modality. The audio samples were played back in their entirety after subjects clicked on a given card. Comparing the 24 audio samples, we found three sound samples with the lowest mean values: male voice, female voice, and kiss sound with mean flip numbers of 5.55; 5.45; and 5.65, respectively. All other sounds have means in the range of 6.30 to 7.98 (marked red in Figure 6).

There is, however, no significant difference among these three audio samples. Based on the Tukey-test there was no significant difference among the everyday auditory icons either. It is more important that five of the measurement signals, namely the 1 and 5 kHz sine, click-train, sweep and the 1 kHz square samples were outperformed by the other signals in many (but not all) of the pairwise t-tests. There were significantly higher error rates (number of flips) compared to auditory icons in 6-8 cases for each of these signals.

Among men, only 2 out of 20, and among females, only 1 out of 20 were significantly better than others. In case of males, younger subjects made less errors than older subjects (F= 16.15; p=0.0008), but there was no difference among females. Comparing all males and all females, the mean of males (161) was significantly better than the mean of females (181) (F=7.61; p=0.0087).

Based on the mean completion times, males were significantly faster (mean 440 sec.) than females (mean 491 sec.) based on ANOVA (F=7.30; p=0,010).

Comparing results of all 40 participants between the audio and visual modality, the mean number of flips was equal to 171 for audio-only, and 177 for vision-only. The statistical analysis supported that there was no significant difference between the two modalities (p=0.4).

IV. DISCUSSION

Testing the visual modality supported a-priori assumptions. As expected, there was no difference among the different visual icons, due to the similarity in size and color. The average number of flips needed to find any of the pairs is around 7. The main findings here are that males outperformed females, and younger subjects outperformed older subjects.

There is a vast literature about comparing memory capabilities of different modalities. When viewed together, most of the prior studies focusing on the comparison of visual and auditory memory under various circumstances offer no conclusive results as to one modality being superior to the other. Thus, while a majority of previous experiments showed visual memory to outperform memory in the auditory modality, e.g. [46], [65]–[68]; several other papers reported no difference between the two [69], [70]. In special cases, it has been shown that memory scores could be even higher when processed through the auditory modality in special cases, e.g., for children [71], [72].

In the case of the study reported in this paper, results did not support former findings that visual memory was more reliable than the auditory modality. Comparing the means in Figure 5 and Figure 6, the mean number of flips were around 7 in the case of audio samples as well. Significant improvements were found only in the case of human sounds.

Although not supported by statistical evidence during the paired t-tests in every case, measurement signals also tended to be worse than other audio samples. We can speculate that natural occurring sounds can be recalled and applied better, which supports former findings reported in [51], [55]. It was also expected that similarity would play a significant role. Thus, there were two guitar samples (one distorted, the other not) and three kinds of drum sounds. Furthermore, white and pink noise, sinusoidal and square signals with the same base frequency may have sounded similar, thus, they might have been confused more often according to our expectation. However, this was not supported by the results.

Comparing genders and age groups, males outperformed females. Younger participants delivered better results, but only for men. If we look at the individuals among men and women, there are subjects who are significantly better than others. Interestingly, the variability is greater in case of visual icons: almost half of the males are better than the rest of the group. For females, only 1-2 subjects performed better. This difference between the genders is almost gone in the auditory modality.

If we analyse the factor of age, note that the limit was set to 25 years. Although half of the participants were above this limit, the mean of the age is still low (most of them were below 35). We can assume that a better selection of participants, including more elderly and partitioning them into more age groups (i.e., below 25, between 25-45, above 45) would result in a different outcome. A dedicated experiment is needed for more conclusive results designed for testing the effect of age.

Another factor that was measured is completion time. The two modalities could not be compared, as the mean time is higher for the audio modality due to the playback times of the audio samples. For a conclusive comparison, the visual game should have been delayed after clicking and all sound samples should have been exactly of the same length. Nevertheless, if checking the differences between males and females in the mean completion time in the audio modality, men were significantly faster (F= 7.30; p=0.010).

As mentioned earlier, players are motivated to achieve a ”perfect game”. This occurs if the number of total flips is minimal, all flipped cards could be recalled after the first appearance. According to the informal feedback from the subjects, it is a too difficult task in both modalities in the case of 24 icons/samples. As a matter of fact, even the best players were unable to complete perfect games if the number of pairs exceeds 10 (4x5 resolution). Subjects reported that the game is relatively easy up to 5-6 pairs, but becomes difficult if it has more than 8 pairs.

A follow-up investigation was conducted using all available resolutions. The investigation highlighted the importance of the participant selection process. Specifically, choosing participants from a broader range of age groups and those with diverse experience and training backgrounds can help gather individuals with varying cognitive abilities. In addition, motivation and performance may vary depending on the serious gaming scenario and the user interface used. Comparing gamification methods with traditional memory assessment techniques could reveal their respective advantages and disadvantages in terms of efficiency. A detailed analysis of the follow-up study is planned as future future work.

V. CONCLUSIONS

This paper presented the results of an experiment on the auditory and visual short-term memory in a serious game application. 40 subjects played the memory game of ”finding pairs” in a 6x8 resolution with 24 pairs of visual icons, followed by 24 pairs of auditory representations. Based on the number of total flips, results showed no significant difference between the visual and auditory modalities. Nevertheless, in audio-only mode, human sounds could be recalled the best, followed by familiar everyday auditory icons and unfamiliar measurement signals. Completion time could not be associated with the results. Male and younger subjects delivered better results, however, the age limit of 25-years must be increased, and/or a more detailed set of participants is needed for a conclusive outcome.

Future works include the involvement of different (lower) resolutions, the use of mixed modality (audio and vision together), dedicated sessions to test the effect of training and the age of participants, as well as developing a crowdsourcing module for unsupervised data collection.

REFERENCES

[1] S. Deterding, S. L. Björk, L. E. Nacke, D. Dixon, and E. Lawley, “Designing gamification: creating gameful and playful experiences,” in CHI’13 Extended Abstracts on Human Factors in *Computing Systems*, 2013, pp. 3263–3266.

[2] N. Cowan, “What are the differences between long-term, short-term, and working memory?” *Progress in brain research*, vol. 169, pp. 323–338, 2008. DOI: 10.1016/S0079-6123(07)00020-9

[3] D. Norris, “Short-term memory and long-term memory are still different,” *Psychological bulletin*, vol. 143, no. 9, p. 992, 2017. DOI: 10.1037/bul0000108

[4] D. Burr and D. Alais, “Combining visual and auditory information,” *Progress in brain research*, vol. 155, pp. 243–258, 2006. DOI: 10.1016/S0079-6123(06)55014-9

[5] M. Kubovy and D. Van Valkenburg, “Auditory and visual objects,” *Cognition*, vol. 80, no. 1-2, pp. 97–126, 2001. DOI: 10.1016/S0010-0277(00)00155-4

[6] W. J. Chai, A. I. Abd Hamid, and J. M. Abdullah, “Working mem- ory from the psychological and neurosciences perspectives: a review,” *Frontiers in psychology*, vol. 9, p. 401, 2018. DOI: 10.3389/fp-syg.2018.00401

[7] R. C. Atkinson and R. M. Shiffrin, “Human memory: A proposed system and its control processes,” in *Psychology of learning and motivation*. Elsevier, 1968, vol. 2, pp. 89–195.

[8] P. Kelley, M. D. Evans, and J. Kelley, “Making memories: why time matters,” *Frontiers in human neuroscience*, vol. 12, p. 400, 2018. DOI: 10.3389/fnhum.2018.00400

[9] A. Baddeley, “Working memory,” *Science*, vol. 255, no. 5044, pp. 556–559, 1992. DOI: 10.1126/science.17363

[10] T. Shallice and C. Papagno, “Impairments of auditory-verbal short-term memory: Do selective deficits of the input phonological buffer exist?” *Cortex*, vol. 112, pp. 107–121, 2019. DOI: 10.1016/j.cortex.2018.10.004

[11] P. Baranyi, A. Csapo, T. Budai, and G. Wersényi, “Introducing the concept of internet of digital reality-part I,” *Acta Polytechnica Hungarica*, vol. 18, no. 7, pp. 225–240, 2021. DOI: 10.12700/APH.18.7.2021.7.12

[12] G. Wersényi, A. Csapó, T. Budai, and P. Baranyi, “Internet of digital reality: Infrastructural background–part II,” *Acta Polytechnica Hungarica*, vol. 18, no. 8, pp. 91–104, 2021. DOI: 10.12700/APH.18.8.2021.8.5

[13] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta, “Assessment in and of serious games: an overview,” *Advances in human-computer interaction*, vol. 2013, pp. 1–1, 2013. DOI: 10.1155/2013/136864

[14] J. B. Hauge, I. A. Stanescu, A. Stefan, M. B. Carvalho, T. Lim, S. Louchart, and S. Arnab, “Serious game mechanics and opportunities for reuse,” *eLearning & Software for Education*, vol. 2, pp. 19–27, 2015. DOI: 10.12753/2066-026X-15-000

[15] A. Dimitriadou, N. Djafarova, O. Turetken, M. Verkuyl, and A. Fer- worn, “Challenges in serious game design and development: Educators’ experiences,” *Simulation & Gaming*, vol. 52, no. 2, pp. 132–152, 2021. DOI: 10.1177/1046878120944197

[16] A. C. T. Klock, I. Gasparini, M. S. Pimenta, and J. Hamari, “Tailored gamification: A review of literature,” *International Journal of Human-Computer Studies*, vol. 144, p. 102495, 2020. DOI: 10.1016/j.ijhcs.2020.102495

[17] A. Rapp, F. Hopfgartner, J. Hamari, C. Linehan, and F. Cena, “Strengthening gamification studies: Current trends and future opportunities of gamification research,” *International Journal of Human-Computer Studies*, vol. 127, pp. 1–6, 2019. DOI: 10.1016/j.ijhcs.2018.11.007

[18] D. Djaouti, J. Alvarez, and J.-P. Jessel, “Classifying serious games: the G/P/S model,” in *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*. IGI global, 2011, pp. 118–136.

[19] D. Ralf, S. Göbel, W. Effelsberg, and J. Wiemers, “Serious games: Foundations, concepts and practice,” 2016.

[20] D. Johnson, E. Horton, R. Mulcahy, and M. Foth, “Gamification and serious games within the domain of domestic energy consumption: A systematic review,” *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 249–264, 2017. DOI: 10.1016/j.rser.2017.01.134

[21] B. Pilote and G. Chiniara, “Chapter 2 - the many faces of simulation,” in *Clinical Simulation (Second Edition)*, G. Chiniara, Ed. Academic Press, 2019, pp. 17–32. ISBN 978-0-12-815657-5

[22] P. M. Staccini and J.-P. Fournier, “Chapter 4 - virtual patients and serious games,” in *Clinical Simulation (Second Edition)*, G. Chiniara, Ed. Academic Press, 2019, pp. 41–51. ISBN 978-0-12-815657-5

[23] M. W. Schurgin, “Visual memory, the long and the short of it: A review of visual working memory and long-term memory,” *Attention, Perception, & Psychophysics*, vol. 80, pp. 1035–1056, 2018. DOI: 10.3758/s13414-018-1522-y

[24] T. F. Brady, T. Konkle, J. Gill, A. Oliva, and G. A. Alvarez, “Visual long-term memory has the same limit on fidelity as visual working memory,” *Psychological science*, vol. 24, no. 6, pp. 981–990, 2013. DOI: 10.1177/0956797612465439

[25] W. J. Ma, M. Husain, and P. M. Bays, “Changing concepts of working memory,” *Nature neuroscience*, vol. 17, no. 3, pp. 347–356, 2014. DOI: 10.1038/nn.3655

[26] S. J. Luck and E. K. Vogel, “The capacity of visual working memory for features and conjunctions,” *Nature*, vol. 390, no. 6657, pp. 279–281, 1997. DOI: 10.1038/36846

[27] D. Fougnie, J. W. Suchow, and G. A. Alvarez, “Variability in the quality of visual working memory,” *Nature communications*, vol. 3, no. 1, p. 1229, 2012. DOI: 10.1038/ncomms2237

[28] P. Wilken and W. J. Ma, “A detection theory account of change detection,” *Journal of vision*, vol. 4, no. 12, pp. 11–11, 2004. DOI: 10.1167/4.12.11

[29] G. A. Alvarez and P. Cavanagh, “The capacity of visual short-term memory is set both by visual information load and by number of objects,” *Psychological science*, vol. 15, no. 2, pp. 106–111, 2004. DOI: 10.1111/j.0963-7214.2004.01502006.x

[30] T. F. Brady and G. A. Alvarez, “No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 41, no. 3, p. 921, 2015. DOI: 10.1037/xlm0000075

[31] K. O. Hardman and N. Cowan, "Remembering complex objects in visual working memory: Do capacity limits restrict objects or features?" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 41, no. 2, p. 325, 2015. DOI: 10.1037/xlm0000031

[32] K. Fukuda, E. Awh, and E. K. Vogel, "Discrete capacity limits in visual working memory," *Current opinion in neurobiology*, vol. 20, no. 2, pp. 177–182, 2010. DOI: 10.1016/j.conb.2010.03.005

[33] T. F. Brady, T. Konkle, and G. A. Alvarez, "A review of visual memory capacity: Beyond individual items and toward structured representations," *Journal of vision,* vol. 11, no. 5, pp. 4–4, 2011. DOI: 10.1167/11.5.4

[34] I. E. Asp, V. S. Störmer, and T. F. Brady, "Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful," *Journal of cognitive neuroscience*, vol. 33, no. 5, pp. 902–918, 2021. DOI: 10.1162/jocn a 01693

[35] K. J. Blacker, K. M. Curby, E. Klobusicky, and J. M. Chein, "Effects of action video game training on visual working memory," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 5, p. 1992, 2014. DOI: 10.1037/a0037556

[36] Y. Yao, R. Cui, Y. Li, L. Zeng, J. Jiang, N. Qiu, L. Dong, D. Gong, G. Yan, W. Ma et al., "Action real-time strategy gaming experience related to enhanced capacity of visual working memory," *Frontiers in Human Neuroscience*, vol. 14, p. 333, 2020. DOI: 10.3389/fn-hum.2020.00333

[37] J. Mishra, D. Bavelier, and A. Gazzaley, "How to assess gaming-induced benefits on attention and working memory," *Games for Health: Research, Development, and Clinical Applications*, vol. 1, no. 3, pp. 192–198, 2012. DOI: 10.1089/g4h.2011.0033

[38] J. Kaiser, "Dynamics of auditory working memory," *Frontiers in psychology*, vol. 6, p. 613, 2015. DOI: 10.3389/fpsyg.2015.00613

[39] S. E. McAdams and E. E. Bigand, "Thinking in sound: The cognitive psychology of human audition," in *Based on the fourth workshop in the Tutorial Workshop series organized by the Hearing Group of the French Acoustical Society*. Clarendon Press/Oxford University Press, 1993. DOI: 10.1093/acprof:oso/9780198522577.001.0001

[40] W. Ritter, D. Deacon, H. Gomes, D. C. Javitt, and H. G. Vaughan Jr, "The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review," *Ear and hearing*, vol. 16, no. 1, pp. 52–67, 1995. DOI: 10.1097/00003446-199502000-00005

[41] J. F. Zimmermann, M. Moscovitch, and C. Alain, "Attending to auditory memory," *Brain research*, vol. 1640, pp. 208–221, 2016. DOI: 10.1016/j.brainres.2015.11.032

[42] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and brain sciences*, vol. 24, no. 1, pp. 87–114, 2001. DOI: 10.1017/S0140525X01003922

[43] M. E. Gloede, E. E. Paulauskas, and M. K. Gregg, "Experience and information loss in auditory and visual memory," *Quarterly Journal of Experimental Psychology*, vol. 70, no. 7, pp. 1344–1352, 2017. DOI: 10.1080/17470218.2016.1183686

[44] K. C. Backer and C. Alain, "Attention to memory: orienting attention to sound object representations," Psychological research, vol. 78, no. 3, pp. 439–452, 2014. https://doi.org/10.1007/s00426-013-0531-7

[45] J. L. Burt, D. S. Bartolome, D. W. Burdette, and J. R. Comstock Jr, "A psychophysiological evaluation of the perceived urgency of auditory warning signals," *Ergonomics*, vol. 38, no. 11, pp. 2327–2340, 1995. DOI: 10.1080/00140139508925271

[46] M. A. Cohen, T. S. Horowitz, and J. M. Wolfe, "Auditory recognition memory is inferior to visual recognition memory," *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 6008–6010, 2009. DOI: 10.1073/pnas.0811884106

[47] N. Cowan, "Visual and auditory working memory capacity," *Trends in cognitive sciences*, vol. 2, no. 3, p. 77, 1998. DOI: 10.1016/S1364-6613(98)01144-9

[48] B. G. Shinn-Cunningham, "Object-based auditory and visual attention," *Trends in cognitive sciences*, vol. 12, no. 5, pp. 182–186, 2008. DOI: 10.1016/j.tics.2008.02.003

[49] G. Lehnert and H. D. Zimmer, "Auditory and visual spatial working memory," *Memory & cognition*, vol. 34, pp. 1080–1090, 2006. DOI: 10.3758/BF03193254

[50] W. Setti, L. F. Cuturi, E. Cocchi, and M. Gori, "A novel paradigm to study spatial memory skills in blind individuals through the auditory modality," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018. DOI: 10.1038/s41598-018-31588-y

[51] T. L. Bonebright and M. A. Nees, "Memory for auditory icons and earcons with localization cues," in *13th International Conference on Auditory Display (ICAD13)*, Montréal, Canada, June 26-29, 2007. Georgia Institute of Technology, 2007, pp. 419–422.

[52] C. M. Vanden Bosch der Nederlanden, C. Zaragoza, A. Rubio-Garcia, E. Clarkson, and J. S. Snyder, "Change detection in complex auditory scenes is predicted by auditory memory, pitch perception, and years of musical training," *Psychological Research*, vol. 84, pp. 585–601, 2020. DOI: 10.1007/s00426-018-1072-x

[53] F. Haiduk, C. Quigley, and W. T. Fitch, "Song is more memorable than speech prosody: Discrete pitches aid auditory working memory," *Frontiers in psychology*, vol. 11, p. 586723, 2020. DOI: 10.3389/fp-syg.2020.586723

[54] R. Bianco, P. M. Harrison, M. Hu, C. Bolger, S. Picken, M. T. Pearce, and M. Chait, "Long-term implicit memory for sequential auditory patterns in humans," *ELife*, vol. 9, p. e56073, 2020. DOI: 10.7554/eLife.56073

[55] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *ACM SIGCHI Bulletin*, vol. 19, no. 1, p. 74, 1987. DOI: 10.1145/28189.1044809

[56] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human–Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989. DOI: 10.1207/s15327051hci0401_1

[57] Á. Csapó, G. Wersényi, H. Nagy, and T. Stockman, "A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research," *Journal on Multimodal User Interfaces*, vol. 9, pp. 275–286, 2015. DOI: 10.1007/s12193-015-0182-7

[58] J. Sterkenburg, M. Jeon, and C. Plummer," Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques: 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 16*. Springer, 2014. pp. 633–640. DOI: 10.1007/978-3-319-07230-2_60

[59] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Logos Verlag Berlin, 2011, vol. 1.

[60] J. P. Cabral and G. B. Remijn, "Auditory icons: Design and physical characteristics," *Applied ergonomics*, vol. 78, pp. 224–239, 2019. DOI: 10.1016/j.apergo.2019.02.008

[61] Á. Csapó and G. Wersényi, "Overview of auditory representations in human-machine interfaces," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–23, 2013. DOI: 10.1145/2543581.2543586

[62] A. Shamei and B. Gick, "The effect of virtual reality environments on auditory memory," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2498–2498, 2020. DOI: 10.1121/1.5146928

[63] "Sound of vision. https://soundofvision.net/ (accessed on 20 july 2023)," 2018. [Online]. Available: https://soundofvision.net/

[64] H. Nagy and G. Wersényi, "Evaluation of training to improve auditory memory capabilities on a mobile device based on a serious game application," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[65] A. R. Jensen, "Individual differences in visual and auditory memory," *Journal of Educational Psychology*, vol. 62, no. 2, p. 123, 1971. DOI: 10.1037/h0030655

[66] M. E. Gloede and M. K. Gregg, "The fidelity of visual and auditory memory," *Psychonomic Bulletin & Review*, vol. 26, pp. 1325–1332, 2019. DOI: 10.3758/s13423-019-01597-7

[67] J. Rodríguez, T. Gutiérrez, O. Portillo, and E. J. Sánchez, "Learning force patterns with a multimodal system using contextual cues," *International Journal of Human-Computer Studies*, vol. 110, pp. 86–94, 2018. DOI: 10.1016/j.ijhcs.2017.10.007

[68] K. Lindner, G. Blosser, and K. Cunigan, "Visual versus auditory learning and memory recall performance on short-term versus long-term tests," *Modern Psychological Studies*, vol. 15, no. 1, p. 6, 2009.

[69] K. M. Visscher, E. Kaplan, M. J. Kahana, and R. Sekuler, "Auditory short-term memory behaves like visual short-term memory," *PLoS biology*, vol. 5, no. 3, p. e56, 2007. DOI: 10.1371/journal.pbio.0050056

[70] G. Ward, S. Avons, and L. Melling, "Serial position curves in short-term memory: Functional equivalence across modalities," *Memory*, vol. 13, no. 3-4, pp. 308–317, 2005. DOI: 10.1080/09658210344000279

[71] M. J. Watkins and Z. F. Peynircioğlu, "Interaction between presentation modality and recall order in memory span," *The American journal of psychology*, pp. 315–322, 1983. DOI: 10.2307/1422314

[72] R. Pillai and A. Yathiraj, "Auditory, visual and auditory-visual memory and sequencing performance in typically developing children," *International Journal of Pediatric Otorhinolaryngology*, vol. 100, pp. 23–34, 2017. DOI: 10.1016/j.ijporl.2017.06.010

György Wersényi was born in 1975 in Győr, Hungary. He received his MSc degree in electrical engineering from the Technical University of Budapest in 1998 and PhD degree from the Brandenburg Technical University in Cottbus, Germany. Since 2002 he has been member of the Department of Telecommunications at the Széchenyi István University in Győr. From 2020 to 2022 he was the dean of Faculty of Mechanical Engineering, Informatics and Electrical Engineering, as well as the scientific president of the Digital Development Center at the university. Currently, he is a full professor, member of the European Acoustics Association (EAA) and the Audio Engineering Society (AES). His research focus is on acoustic measurements, virtual and augmented reality solutions, sonification, cognitive infocommunications, and assistive technologies.

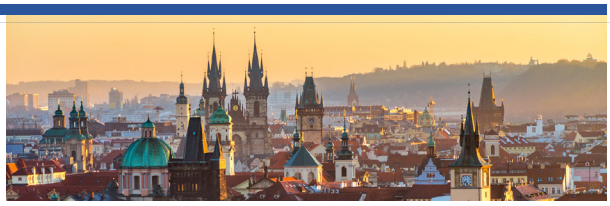
Ádám Csapó obtained his PhD degree at the Budapest University of Technology and Economics in 2014. Between 2016 and 2022, he has worked as an Associate Professor at the Széchenyi István University, and between 2022 and 2023, at Óbuda University in Budapest, Hungary. Currently he is an Associate Professor at Corvinus University of Budapest. Dr. Csapo's research focuses on soft computing tools for developing cognitive infocommunication channels in assistive technologies and virtual collaboration environments, with the goal of enabling users to communicate with each other and with their spatial surroundings in novel and effective ways. He has also participated in the development of a commercial desktop VR platform serving both educational and industrial use cases. Dr. Csapo has over 100 publications, including 1 co-authored book and over 20 journal papers.

20th International Conference on Network and Service Management
Prague, Czech Republic // 28 - 31 October 2024
AI-Powered Network and Service Management for Tomorrow's Digital World

AnServApp 2024

4th International Workshop on Analytics for Service and Application Management (AnServApp)

Monday 28 October 2024



Scope

With enterprise organizations generating petabytes of data each day, their use of, and reliance on data analytics to provide contextual insight into their operations is imperative for improving the implementation, management and delivery of services and applications. Approaches such as predictive data analytics, data mining, machine learning and deep learning are promising mechanisms to harness this immense stream of service and application data to meet the needs of an organization. The main goal of AnServApp is to present research and work-in-progress results in the area of data analytics, machine learning and cognitive science for service and application management. Its interdisciplinary approach offers practical insights and real-world applications that attendees can apply to their work. In addition to regular papers, short papers describing late-breaking advances and work-in-progress reports from ongoing research are also welcomed. The workshop will feature distinguished speakers, and numerous networking opportunities, fostering collaboration and knowledge exchange among researchers, practitioners, and industry experts.

Topics of Interest

Topics of Interest to the workshop include but are not limited to:

- Application Management Services (AMS)
- AI/ML/DL powered solutions
- Network and service security
- Ticket resolution and management
- Event log analysis
- Knowledge management
- Predictive maintenance / Industry 4.0
- Asset tracking and provisioning
- Workload optimization
- Sentiment analysis
- Social media
- Lower carbon foot print applications / services / systems
- Smart cities and smart transportation services / systems
- Social media apps / services / systems
- Smart education services / systems
- Edge, fog, cloud services / systems
- Sustainability and resilience of applications / services / systems
- Digital Twins for networks and services
- Resource management & orchestration
- Multi-Agent Reinforcement Learning (RL)
- Distributed RL over communication networks

Important Dates

Submission Deadline: August 23, 2024
Acceptance Notification: September 9, 2024
Camera Ready: September 16, 2024
Workshop Day: October 28, 2024

All times in Anywhere on Earth (AoE) timezone.

Workshop Chairs

Pal Varga, Budapest University of Technology and Economics, Hungary
José Pedro Pereira dos Santos, Ghent University - imec, Belgium
Guillaume Fraysse, Orange, France
Nur Zincir-Heywood, Dalhousie University, Canada

Paper Submission Guideline

Authors are invited to submit original unpublished papers not under review elsewhere. Papers should be submitted in IEEE 2-column format. Maximum paper lengths, including title, abstract, all figures, tables, and references, are 7 pages for regular papers and 4 pages for short papers. Regular paper length could include up to 7 pages including references. Short papers are accepted as well and must not exceed 4 pages including references.

Papers have to be submitted electronically in PDF format through the EDAS conference management system, accessible via this link: http://www.cnsm-conf.org/2024/AnServApp.html

All submitted papers will be peer-reviewed and selected based on their originality, significance, technical soundness, and relevance to the workshop's theme. For accepted papers, at least one author is expected to register and present the paper in person at the workshop. Accepted and presented papers will be published in the conference proceedings and submitted to IEEE Xplore.

For any questions, please feel free to contact Pal Varga (pvarga@tmit.bme.hu).

2025 IEEE Wireless Communications and Networking Conference

6G Horizons: Bridging Beyond Wireless

24-27 March 2025 // Milan, Italy

Call for Papers

The IEEE Wireless Communications and Networking Conference (WCNC) is a top-ranked, flagship conference of the IEEE Communications Society, bringing together researchers from academia, industry, and government. IEEE WCNC 2025 will be hosted in the vibrant city of Milan, Italy and will be conducted in person, allowing attendees to fully benefit from the conference atmosphere and experience.

Prospective authors are invited to submit their works in the form of research papers describing significant and innovative contributions to the field of wireless communications and networking, in accordance with the four technical tracks listed below. Accepted and presented papers will be published in the IEEE WCNC 2025 Conference Proceedings and submitted to IEEE Xplore.

Proposals for half- or full-day tutorials and workshops are also invited in all communication and networking topics.

Visit Our Website

To learn more about WCNC 2025 in Milan, and how to submit your paper, please visit:

https://wcnc2025.ieee-wcnc.org/

Important Dates:

Paper Submissions Deadline: 2 September 2024

Notification of Acceptance: 20 December 2024

Camera-Ready Papers: 24 January 2025

IEEE ComSoc

IEEE WCNC 2025

ORGANIZING COMMITTEE

General Chair
Maria Luisa Merani, University of Modena and Reggio Emilia, IT

Executive Chair
Hikmet Sari, NJUPT, CN

Vice General Co-Chairs
Antonella Molinaro, University Mediterranea of Reggio Calabria, IT
Ilenia Tinnirello, University of Palermo, IT

TPC Co-Chairs
Marco Chiani, University of Bologna, IT
Giuseppe Bianchi, University of Rome Tor Vergata, IT
Wei Zhang, University of New South Wales, AU

Industry Program Co-Chairs
Slim Peiying Zhu, Huawei, CA
Sinem Coleri, Koc University, TR
Angeliki Alexiou, University of Piraeus, GR

Workshop Co-Chairs
Violet Syrotiuk, Arizona State University, US
Carlo Fischione, KTH Royal Institute of Technology, SE

Tutorial Co-Chairs
Antoine Berthet, Paris-Saclay University, FR
Xavier Costa Perez, NEC Labs Europe, DE

Panel Chair
Muriel Medard, MIT, US
Antonio Capone, Politecnico di Milano, IT

Demo Chairs
Fabrizio Granelli, University of Trento, IT
Ivan Seskar, Rutgers University, US

Operations Chairs
Stefano Bregni, Politecnico di Milano, IT

Finance Chair
Carles Anton-Haro, CTTC, ES

Travel Grants Chair
Rao Venkatesha Prasad, Techn. Univ. Delft, NL

Publication Chair
Mutlu Koca, Bogazici University, TR

Web & Social Media Chair
Ejder Bastug, Nokia Bell Labs, FR

IEEE/IFIP Network Operations and Management Symposium
12 - 16 May 2025 // Honolulu, Hawaii, USA
Managing the Wave to Global Connectivity

CALL FOR PAPERS

The 2025 IEEE/IFIP Network Operations and Management Symposium (NOMS 2025) will be held 12-16 May, 2025, in Honolulu, Hawaii, USA. First organized in 1988, NOMS 2025 follows the 37-year tradition of NOMS and IM as the IEEE Communications Society's primary forum for technical exchange on network and service management, focusing on research, development, integration, standards, service provisioning, and user communities. NOMS 2025 seeks contributions presenting recent developments and technical solutions for dealing with the management of various networks and services. NOMS 2025 program will be under the theme of ***Managing the Wave to Global Connectivity*** and will feature a variety of sessions, including keynotes, tutorials as well as technical, experience, demo, poster, panel, and dissertation sessions.
Authors are invited to submit papers that fall into or are related to the following topics of interests:

Management of Networks
. IP Networks
. Wireless and Cellular Networks
. 5G network and Beyond (6G)
. Quantum Networks
. Optical Networks
. Virtual Networks and SDN
. Home Networks
. Access Networks
. Fog, Edge and Cloud Networks
. Enterprise and Campus Networks
. Data Center Networks
. Industrial Networks
. Vehicular Networks
. IoT, Sensor and M2M Networks
. Information-Centric Networks

Management of Services
. Multimedia Services
. Over-The-Top Services
. Content Delivery Services
. Cloud Computing Services
. Internet Connectivity and Internet Access Services
. Internet of Things Services
. Security Services
. Context-Aware Services
. Information Technology Services
. Service Assurance

Management of Businesses
. Economic Aspects
. Multi-Stakeholder Aspects
. Service Level Agreements
. Lifecycle Aspects
. Process and Workflow Aspects
. Legal Perspective
. Regulatory Perspective
. Privacy Aspects
. Organizational Aspects

Functional Areas
. Fault Management
. Configuration Management
. Accounting Management
. Performance Management
. Security Management

Management Paradigms
. Centralized Management
. Hierarchical Management
. Distributed Management
. Integrated Management
. Federated Management
. Autonomic and Cognitive Management
. Policy- and Intent-Based Management
. Model-Driven Management
. Proactive Management
. Energy-aware Management
. QoE-Centric Management

Management Technologies
. Communication Protocols
. Middleware
. Overlay Networks
. Peer-to-Peer Networks
. Cloud Computing and Cloud Storage
. Data, Information, and Semantic Models
. Information Visualization
. Software-Defined Networking
. Network Function Virtualization
. Orchestration
. Operations and Business Support Systems
. Control and Data Plane Programmability
. Distributed Ledger Technology
. Digital Twins

Methods
. Mathematical Logic and Automated Reasoning
. Optimization Theories
. Control Theory
. Probability Theory, Stochastic Processes, Queuing Theory
. Artificial Intelligence and Machine Learning
. Evolutionary Algorithms
. Economic Theory, Game Theory and Business Models
. Risk Management Methods
. Monitoring and Measurements
. Data Mining and (Big) Data Analysis
. Computer Simulation Experiments
. Testbed Experimentation and Field Trials
. Software Engineering Methodologies

Paper submission guidelines
Authors are invited to submit original contributions, written in English, that have not been published or submitted for publication elsewhere. Technical papers must be formatted using the IEEE 2-column format and not exceed 8 pages (excluding references) for full paper submissions or not exceed 4 pages (excluding references) for short paper submissions. All papers should be submitted through JEMS3 at https://jems3.sbc.org.br/noms2025.
All submitted papers will be peer-reviewed. Accepted and presented papers will be published in the conference proceedings and submitted to IEEE Xplore. Authors of the best accepted papers will be invited to submit extended versions of their papers to IEEE Transactions on Network and Service Management (TNSM).

IMPORTANT DATES

Paper Submission: September 13, 2024
Notification of Acceptance: December 20, 2024
Camera Ready Paper: January 31, 2025

https://noms2025.ieee-noms.org/

General Co-Chairs:
Doug Zuckerman, Telcordia (Retired), USA
Mehmet Ulema, Manhattan College, USA

TPC Co-Chairs:
Noura Limam, University of Waterloo, Canada
Young-Tak Kim, Yeungnam University, Korea

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on

https://journals.ieeeauthorcenter.ieee.org/
Then click: "IEEE Author Tools for Journals"
- "Article Templates"
- "Templates for Transactions".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *Index Terms (Keywords)*. For the final version of accepted papers, please send the short cvs and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue
g) Document Object Identifier (DOI)

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984. DOI: 10.1109/TEI.1984.298778

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to submit their papers electronically via the following portal address:

https://www.ojs.hte.hu/infocommunications_journal/about/submissions

If you have any question about the journal or the submission process, please do not hesitate to contact us via e-mail:

Editor-in-Chief: Pál Varga – pvarga@tmit.bme.hu

Associate Editor-in-Chief:

József Bíró – biro@tmit.bme.hu

László Bacsárdi – bacsardi@hit.bme.hu

IEEE International Conference on Multimedia and Expo

NANTES - FRANCE

ICME

JUNE 30 - JULY 4

2025

JOURNEY TO THE CENTER OF MACHINE IMAGINATION

The IEEE International Conference on Multimedia & Expo (ICME) is the flagship multimedia conference sponsored by four IEEE societies since 2000, promoting the exchange of the latest advances in multimedia technologies, systems, and applications from both research and development perspectives. ICME 2025 will be held in Nantes, France, from June 30 to July 4, 2025. It will bring together leading researchers and practitioners to share the latest developments and advances in the discipline. Featuring high-quality oral and poster sessions, world-class keynotes, exhibitions, demonstrations, and tutorials, the conference will attract leading researchers and global industry figures, providing excellent networking opportunities. In addition, exceptional papers and contributors will be selected and recognized with prestigious awards.

TOPICS OF INTEREST INCLUDE (but are not limited to) :

- 3D multimedia, AR/VR and immersive media
- Emerging multimedia applications and technologies
- Artificial intelligence/machine learning based multimedia
- Image and video processing
- Multimedia analysis and generation
- Multimedia communications, networking and mobility
- Multimedia databases and data mining
- Multimedia quality assessment and metrics
- Multimedia security, privacy and forensics
- Multimedia standards, trends and related research
- Multi-modal media computing, interaction, and human-machine interaction
- Social media analysis and applications
- Speech/audio processing

LOCATION

Nantes, the home city of Jules Verne and the ancient seat of the Dukes of Brittany, is the sixth largest city in France and is renowned for its livability and eco-friendliness. It received the European Green Capital award in 2013. Strategically positioned on the Loire River, Nantes has a rich history as a port and industrial center. Over the past few decades, it has evolved into a technological hub with a high-tech and creative industry. Located only two hours from Paris by high-speed (TGV) train, Nantes is well-connected with direct flights to over 50 international destinations. It has established itself as an emerging metropolis in Europe and a perfect venue for hosting international congresses.

IMPORTANT DATES

22 Nov 2024
- Special session proposals
- Workshop proposals

29 Nov 2024
- Tutorial proposals

15 Nov 2024
- Grand Challenge proposals

13 Dec 2024
- Regular Paper Submissions (including Special Sessions)

14 Mar 2025
- Regular Paper Acceptance notification

25 Mar 2025
- Industry Expo Proposals
- Panels proposals
- Demo proposals submission
- Workshop paper submission
- Industry Technology Workshop proposals

hosted by

IEEE Advancing Technology for Humanity · IEEE COMPUTER SOCIETY · CAS · IEEE ComSoc IEEE Communications Society · IEEE Signal Processing Society CELEBRATING 75 YEARS

For further information, visit the conference website: https://2025.ieeeicme.org/

GENERAL CHAIRS
Patrick Le Callet ▸ University of Nantes, France
Giuseppe Valenzise ▸ CNRS, University Paris-Saclay, France
Weisi Lin ▸ Nanyang Technological University, Singapore

TECHNICAL PROGRAM CHAIR
Ivan Bajic ▸ Simon Fraser University, Canada

TREASURER/FINANCE CHAIR
Toinon Vigier ▸ University of Nantes, France

WORKSHOP CHAIRS
Chaker Larabi ▸ University of Potiers, France
Jiaying Liu ▸ Peking University, China
Shervin Shirmohammadi ▸ University of Ottawa, Canada

INNOVATION PROGRAM (INDUSTRY) CHAIRS
Ioannis Katsavounidis ▸ Meta, USA
Dong Tian ▸ Interdigital, USA
Nick Zacharov ▸ Meta, France

INDUSTRY LIAISON CHAIRS
Balu Adsumilli ▸ Google, USA
Tao Mei ▸ HiDream.ai, China
Sebastian Schwarz ▸ Nokia, Germany

PANEL CHAIRS
Shu-Ching Chen ▸ University of Missouri-Kansas City, USA
Christian Timmerer ▸ University of Klagenfurt, Austria
Maggie Zhu ▸ Purdue University, USA

PLENARY CHAIRS
Moncef Gabbouj ▸ Tampere University, Finland
Jay Kuo ▸ University of Southern California, USA

TUTORIAL CHAIRS
Gene Cheung ▸ York University, Canada
Mylène Farias ▸ Texas State University, USA
Ce Zhu ▸ UESTC, China

GRAND CHALLENGE CHAIRS
Federica Battisti ▸ University of Padova, Italy
Aladine Chetouani ▸ University of Orléans, France
Tanaya Guha ▸ University of Glasgow, UK

SPECIAL SESSION CHAIRS
Wei Hu ▸ Peking University, China
Ichiro Ide ▸ Nagoya University, Japan

EXHIBITS AND DEMO CHAIRS
Jing Li ▸ Alibaba, China
Matthieu Pereira Da Silva ▸ University of Nantes, France
Guo-Jun Qi ▸ Futurwei Technologies, USA

AWARD COMMITTEE CHAIRS
Chia-Wen Lin ▸ National Tsing Hua University, Taiwan
Fernando Pereira ▸ Instituto Superior Técnico, Universidade de Lisboa, Portugal

PUBLICITY CHAIRS
Irene Viola ▸ CWI, Netherlands
Yuming Fang ▸ Jiangxi University of Finance And Economics, Nanchang, China
Ambarish Natu ▸ Australian Government, Australia

PUBLICATION AND WEBSITE CHAIRS
Emin Zerman ▸ Mid Sweden University, Sweden
Lu Zhang ▸ Insa Rennes, France
Hadi Amirpour ▸ University of Klagenfurt, Austria

STUDENT ACTIVITIES CHAIRS
Marco Carli ▸ University Roma Tre, Italy
William Puech ▸ University of Montpellier, France
Jin Zeng ▸ Tongji University, Shanghai, China

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: HÍRKÖZLÉSI ÉS INFORMATIKAI TUDOMÁNYOS EGYESÜLET, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **FERENC VÁGUJHELYI** • *elnok@hte.hu*
Secretary-General: **GÁBOR KOLLÁTH** • *kollath.gabor@hte.hu*
Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502
Phone: +36 1 353 1027
E-mail: *info@hte.hu*, Web: *www.hte.hu*