

Editor: Dana Simian

**Proceedings of
The International Conference on
Applied Informatics**

ICDD

IMAGINATION, CREATIVITY, DESIGN, DEVELOPMENT

November 4 - 6, 2021, Sibiu, Romania

Editor DANA SIMIAN

**IMAGINATION, CREATIVITY,
DESIGN, DEVELOPMENT**

Proceedings of the International Conference
on Applied Informatics
ICDD

**November 04th – 06th, 2021
Sibiu, Romania**

Lucian Blaga University of Sibiu

Lucian Blaga University of Sibiu, 2021

Editor Dana Simian

All papers in this volume were peer review by two independent reviewers

ISSN-L 2069 – 864X

Associated Editor Laura Florentina Stoica

Proceedings of the International Conference on Applied
Informatics, ICDD
November 04th – 06th, 2021, Sibiu, Romania

Copyright @ 2021 All rights reserved to editors and authors

Preface

This volume contains refereed papers presented within the International Conference on Applied Informatics, “Imagination, Creativity, Design, Development“ - ICDD 2021, which was held between November 04th – 06th, at the Faculty of Sciences, Lucian Blaga University of Sibiu, Romania.

The conference is mainly addressed to young researchers from all over the world. The conference gives the participants the opportunity to discuss and present their research on informatics and related fields (like computational algebra, numerical calculus, bioinformatics, etc.). The conference welcomes submissions of original papers on all aspects of informatics and related fields ranging from new concepts and theoretical developments to advanced technologies and innovative applications. Specific topics of the conference included but are not restricted to: Algorithms and Data Structures, Graph Theory and Applications, Formal Languages and Compilers, Cryptography, Modelling and Simulation, Computer Programming, Computer Vision, Computer Graphics, Game Design, Data Mining, Distributed Computing, Artificial Intelligence, Service Oriented Applications, Networking, Grid Computing, Mobile Operating Systems, Scientific Computing, Software Engineering, Bioinformatics, Robotics, Computer Architecture, Evolutionary Computing, Multimedia Systems, Internet Communication and Technologies, Web Applications.

The conference has brought together participants from 6 countries (Bulgaria, Germany, Romania, Serbia, Ukraine and USA).

We thank all the participants for their interesting talks and discussions. We also thank the members of the scientific committee for their help in reviewing the submitted papers and for their contributions to the scientific success of the conference.

December 2021

Dana Simian
Conference Chair

Motto:

“There are no limits, only your imagination”

Scientific committee

Kiril Alexiev - Bulgarian Academy of Sciences, Bulgaria
Vsevolod Arnaut - Moldova State University, Republic of Moldova
Galina Atanasova - Angel Kanchev University of Rousse
Alina Barbulescu - Ovidius University of Constanta, Romania
Arndt Balzer - University of Applied Sciences, Würzburg-Schweinfurt, Germany
Lasse Berntzen - Buskerud and Vestfold University College, Norway
Peter Braun - University of Applied Sciences, Würzburg-Schweinfurt, Germany
Amelia Bucur - Lucian Blaga University of Sibiu, Romania
Stelian Ciurea - Lucian Blaga University of Sibiu, Romania
Nicolae Constantinescu - Lucian Blaga University of Sibiu, Romania
Daniela Danciulescu - University of Craiova, Romania
Lyubomyr Demkiv - Lviv National Polytechnic University and Robotics Lead at SoftServe, Ukraine
Oleksandr Dorokhov - Kharkiv National University of Economics, Ukraine
Dmytro Dosyn - Institute of Computer Science and Information Technologies, Lviv Polytechnic National University
George Eleftherakis - The University of Sheffield International Faculty, City College Thessaloniki, Greece
Michael Emmerich - Leiden Institute of Advanced Computer Science, Leiden University
Calin Enachescu - University of Medicine, Pharmacy, Science and Technology of Targu Mures, Romania
Ralf Fabian - Lucian Blaga University of Sibiu, Romania
Tobias Fertig - University of Applied Sciences Würzburg-Schweinfurt
Stefka Fidanova - Bulgarian Academy of Sciences, Bulgaria
Ulrich Fiedler - Bern University of Applied Science, Switzerland
Adrian Florea - Lucian Blaga University of Sibiu, Romania
Teresa Gonçalves - University of Evora, Portugal
Andrina Granić - University of Split, Croatia
Katalina Grigorova - University of Ruse, Bulgaria
Piroska Haller - University of Medicine, Pharmacy, Science and Technology of Targu Mures, Romania
Daniel Hunyadi - Lucian Blaga University of Sibiu, Romania
Saleema JS - Chris University, Bangalore, India
Milena Lazarova - Technical University of Sofia, Bulgaria
Lixin Liang - Tsinghua University, Beijing, China
Suzana Loskovska - "Ss. Cyril and Methodius" University in Skopje, Republic of Macedonia
Rossitza S. Marinova - Concordia University of Edmonton, Canada
Gabriela Moise - Petroleum-Gas University of Ploiesti, Romania
G. Jose Moses - Raghu Engineering College Visakhapatnam, Andhra Pradesh, India

Mircea Musan - Lucian Blaga University of Sibiu, Romania
Mircea Iosif Neamtu - Lucian Blaga University of Sibiu, Romania
Elena Simona Nicoară - Petroleum-Gas University of Ploiesti
Grażyna Paliwoda-Pękosz - Cracow University of Economics, Poland
Camelia Pinteă - Technical University Cluj-Napoca, Romania
Antoniu Pitic - Lucian Blaga University of Sibiu, Romania
Alina Pitic - Lucian Blaga University of Sibiu, Romania
Cristina Popirlan - University of Craiova, Romania
Anca Ralescu - University of Cincinnati, United States of America
Mohammad Rezai - Sheffield Hallam University, United Kingdom
Cosmin Sabo - Technical University of Cluj-Napoca - North University Center Baia Mare
José Saias - University of Evora, Portugal
Abdel-Badeeh M. Salem - Ain Shams University, Cairo, Egypt
Livia Sangeorzan - Transilvania University of Brasov, Romania
Soraya Sedkaoui - Khemis Miliana University, Algeria
Andreas Siebert - University of Applied Sciences Landshut, Germany
Ioan Silea - Politehnica University of Timișoara
Dana Simian - Lucian Blaga University of Sibiu, Romania
Petrica C. Pop Sitar - Technical University Cluj-Napoca, Romania
Lior Solomovich - Kaye Academic College of Education, Israel
Ansgar Steland - RWTH Aachen University, Germany
Florin Stoica - Lucian Blaga University of Sibiu, Romania
Laura Florentina Stoica - Lucian Blaga University of Sibiu, Romania
Detlef Streitferdt - Ilmenau University of Technology, Software Architectures and Product Lines Group, Germany
Grażyna Suchacka - University of Opole, Poland
Jolanta Tańcula - University of Opole, Poland
Milan Tuba - Singidunum University of Belgrade, Serbia
Eva Tuba - Singidunum University of Belgrade
Anca Vasilescu - Transilvania University of Brasov, Romania
Dana Vasiloaica - Institute of Technology Sligo, Ireland
Sofia Visa - The College of Wooster, United States

Contents

Fuzzy Clustering of Monolingual Embedding Spaces.....	9
Kowshik Bhowmik, Anca Ralescu	
Decentralized open-source cryptocurrency portfolio tracking hybrid application....	17
Adoris-Elian Doran	
Kiosk web application - integrated multimedia information system.....	31
Martin S. Dzhurov	
Pure Pursuit Controller.....	43
Răzvan Gheorghe Filea	
Empirical Study on the Sustainability of Knowledge Transfer on Information Security Awareness Via Virtual Reality Videos Compared to Conventional Videos..	52
Jana Gawrilow, Jule Schumann, Lui Ruck, Patrick Haase	
Digitization in medical teaching.....	66
Susanna Götz	
A comparative study of growing microgreens. Human supervision VS Agritech/Automation	76
Felix Husac	
Hybrid Classifier based Fall Detection in the Elderly using Still Images	86
Srianuradha Kandavel, Anca L. Ralescu	
Design and Implementation of a Node.js Framework for the Development of RESTful APIs	93
Matthias Keckl	
System of modeling the spreading of infectious diseases	106
Yulian Kuryliak	
Keep it Safe	115
Matei Marin	
Improving the overall user satisfaction for applications generating traffic mix over 4G networks	129
Vasile Horia Muntean	
4G LTE and 5G NR - An overview; Technical aspects of the Air interface	137
Vasile Horia Muntean	

Apparent personality analysis based on robust estimation	145
Milić Vukojičić, Mladen Veinović	
List of authors	155
Sponsors	158

Fuzzy Clustering of Monolingual Embedding Spaces

Kowshik Bhowmik, Anca Ralescu

Abstract

Mapping-based methods of learning cross-lingual word embeddings are based on the assumption that independently trained word embeddings of different languages are isomorphic to one another. The isomorphic assumption implies that learning a linear transformation from one language space to another is possible using known translation word pairs. This does not always hold, especially for typologically distant language pairs. While cross-lingual embedding spaces containing resource-rich and related languages have produced impressive results in downstream tasks, the same is not the case for typologically distant or low-resource languages. This discrepancy of performance is an indication that independently trained word embeddings retain linguistic properties. The presented research aims to divide a set of monolingual word embedding spaces into clusters. The degree of isomorphism between the language spaces is utilized as the pairwise inter-lingual distance. Fuzzy C-Means clustering is then applied on the pairwise distances between all the languages in the dataset. The objective of this paper is to investigate the extent to which real-world relationships between languages can be captured from monolingual embedding spaces. The use of a fuzzy clustering algorithm is crucial in quantifying the degree of membership of a language to the language families formed as a cluster of languages.

1 Introduction

Word embeddings are vector representations of words. Cross-lingual or multilingual embedding spaces are a way to represent words belonging to different languages in a shared space. A good cross-lingual embedding space will have semantically equivalent words from other languages in close proximity in the joint embedding space, facilitating tasks such as bilingual lexicon induction, machine translation, etc. Mikolov et al. [1] proposed the linear transformation method of learning cross-lingual word embeddings. This method was inspired by the observation that semantically equivalent words have similar geometric arrangements in independently trained monolingual word embeddings. Hence, utilizing a few known translation pairs or seed lexicon makes learning a linear transformation from one space to another possible. Since then, this line of research has been enriched by many researchers and has yielded impressive results for languages such as English and other resource-rich languages [2, 3, 4, 5]. Recently, it has been noted that the performance of cross-lingual word embeddings in downstream tasks degrades if the space contains low-resource languages [6]. Such languages are also often typologically distant from English, the most commonly used hub space for cross-lingual embedding space. Early research in the area was based on the assumption that, given a set of known translation pairs between any two languages, it is possible to learn a linear mapping from one of those language spaces to the other. This is referred to in the literature as the isomorphic assumption. Since the performance of cross-lingual word embeddings often correlates with typological distance between languages, researchers have agreed that the isomorphic assumption does not always hold, resulting in it being termed as the weak isomorphic assumption [7, 8]. Researchers have also reported improved results for low-resource languages in downstream tasks where they could utilize the relatively richer resource of a known related language [9]. Efforts have been made to incorporate this weak-isomorphic assumption into learning the mapping to improve the cross-lingual word embeddings for

typologically distant languages. Several distance measures have been proposed to quantify the degree of isomorphism between vector spaces of two languages.

The proposed research builds on the efforts at quantifying the degree of isomorphism and use the measures to cluster a given language set in different ways. Three different isomorphism measures have been used to calculate pairwise distance among 12 languages, diverse both in terms of etymology and resource level. These pairwise distances are then passed to a clustering algorithm to group the languages into clusters. The motivation for this is to see which of the three distance measures is better at capturing the cross-lingual distance or similarity encoded in the monolingual word embeddings. The language clusters are compared against our existing knowledge of inter-lingual relationships. A fuzzy clustering algorithm is chosen to perform the clustering since a language can have roots in more than one family of languages. Fuzzy clustering allows a data point to be part of more than one cluster. It assigns the every data point, in our case- a language, a degree of membership to all the clusters which will allow us to take a closer look at the interlingual relationships.

In Section 2, the language set for our experiments is briefly discussed along with the motivation behind the choices. Section 3 outlines the different measures for isomorphism found in the literature and applied in the experiments described in this paper. Fuzzy Clustering algorithm in general, and specifically Fuzzy c-means is discussed in Section 4, as well as our method for choosing the optimal number of clusters for each of the three pairwise language distance representations. Finally, in Section 5, the results of the Fuzzy C-Means clustering algorithm are analyzed. The paper concludes with a brief discussion of the future direction for our proposed work.

2 Language Set

The language set for our experiment includes Azerbaijani(Az), Belarusian(Be), Czech(Cs), English(En), Galician(Gl), Portuguese(Pt), Romanian(Ro), Russian(Ru), Slovak(Sk), Spanish(Es) and Turkish(Tk). The choice of languages is partly motivated by Anastopoulos et al. [10] There are three low-resource languages in the mix(Az, Be, Gl). These low-resource languages have higher-resource related languages (Tk, Ru, Pt/Es). English by default has been chosen by many researchers to be the hub language for their cross-lingual embedding space, meaning all the languages will be mapped to the English space. In more recent researches, it has been found that typologically distant languages are not good candidates for a hub language [11]. Romanian is one of the few Romance languages in Eastern Europe, and so, also has words derived from the Slavic languages.

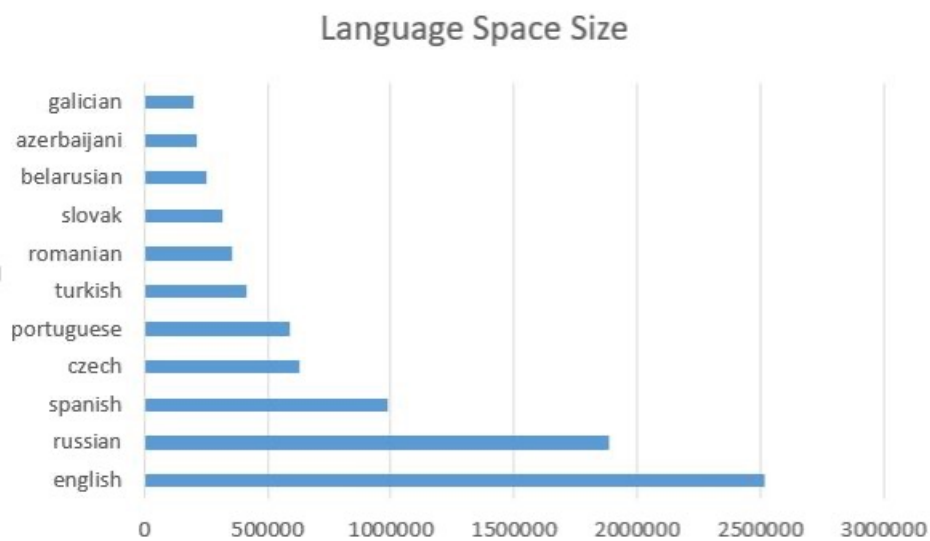


Figure 1: Size of Monolingual Embedding Spaces

The fastText Wiki word vectors have been used for our experiments. These are word vectors pre-trained on Wikipedia of the respective languages. They were trained using the skip-gram model [12], and the default parameters were used in training. All these word vectors have dimension 300. The relative resources on which the embedding spaces in our language set were trained can be seen from Figure 1. The Galician, Azerbaijani and Belarusian spaces can be called low-resource since these spaces have less than 250,000 words. Slovak, Romanian, Turkish, Portuguese, and Czech are moderate-resource languages with less than 600,000 words. Spanish, Russian, English are the resource-rich languages in our language set.

3 Calculating Degree of Isomorphism

While evaluating monolingual embedding spaces is done using intrinsic tasks such as correlations and subspace, the degree of isomorphism is the most critical factor that indicates performance in downstream tasks for cross-lingual embeddings. The degree of isomorphism can be considered a measure of the typological similarity of two vector spaces. The importance of this measure is related to the isomorphic assumption mentioned in Section 1. Recently proposed unsupervised methods for cross-lingual embedding spaces lean more heavily on that assumption. They suggest aligning two monolingual embedding spaces by minimizing the distance between the transformed source language space and the target language space. Various ways have been proposed to quantify isomorphism. For this paper, three of the previously proposed approaches for quantifying isomorphism have been implemented to produce pairwise language distance for the 12 languages in our dataset. The first one is Eigenvector Similarity, a graph-based method proposed by Soggard et al. [6] The second one is Gromov-Hausdorff distance proposed by Patra et al. [7], a metric space based measure. Finally, Relational Similarity is a more straightforward method based on distribution similarity proposed by Vulic et al. [13] The sets of pairwise distances will later be passed as input to a fuzzy clustering algorithm to cluster the languages based on the interlingual distance data. The three approaches to compute the degree of isomorphism are briefly discussed below.

3.1 Eigenvector Similarity

This estimation is based on Laplacian Eigenvalues of the nearest neighbor graphs computed for both the source and target language space from a subset of top N most frequent words. First, the adjacency matrices are built from the nearest neighbor graphs. The number of most informative eigenvalues is determined for both the Laplacian matrices. These numbers, k_1 and k_2 , store at least 90% of the sum of all the eigenvalues of their respective Laplacian matrices. The smaller k is chosen to compute the squared difference of the top k eigenvalues, which is dubbed the eigenvector similarity of the two language spaces. Lower values of this similarity measure indicate similar graphs or more isomorphic vector spaces.

3.2 Gromov-Hausdorff Distance

Hausdorff distance measures the diametric distance between two metric spaces, which intuitively means the distance between neighbors that are the furthest apart. The Gromov-Hausdorff distance, by minimizing the Hausdorff distance over all the possible isometric transformations from one metric space to another, provides an estimation of isometry between the two spaces. This measure has been used to quantify isomorphism between two monolingual embedding spaces by Patra et al. [7]

3.3 Relational Similarity

This is a comparatively straightforward measure proposed by Vulic et al. [13] It is based on the idea that semantically equivalent words in the source and target language will have comparable similarity distribution in their respective embedding spaces. A list is constructed with the identically spelled words shared between the source and target language space. Cosine distance is calculated for every word pair in that list on the source side. The same is repeated for the target side. Pearson correlation coefficient

is computed between the sorted list of similarity scores. Fully isomorphic embeddings will result in a coefficient value of 1, decreasing with lower degrees of isomorphism.

4 Clustering Languages

4.1 Fuzzy C-Means Clustering

The task of assigning a set of objects into groups is known as clustering, where these groups are called clusters. The aim is to partition a collection of objects into clusters. Objects belonging to the same cluster are similar to each other and dissimilar from those belonging to different clusters [14]. Clustering algorithms can be broadly classified into hard and soft (fuzzy) clustering. In hard clustering, each data element can belong to only one cluster. On the other hand, fuzzy clustering allows a data element to belong to multiple clusters with a certain degree of membership. Fuzzy clustering is one of the most widely used clustering algorithms. The fuzzy set theory proposed by Zadeh used the concept of membership function to express the uncertainty of belonging [15]. Fuzzy clustering allows non-unique partitioning of a set of objects into several clusters. Instead of assigning each data point to one cluster as done in hard clustering, fuzzy clustering assigns membership values for each cluster. Their partial membership to all classes is expressed using a degree of membership between 0 and 1. This is particularly useful for objects that fall on the boundaries of two or more classes. Fuzzy c-means is one of the most widely used fuzzy clustering algorithms. It was first reported by Joe Dunn for a special case ($m = 2$) [16]. The general case was developed and later improved by Jim Bezdek [17]. The algorithm assigns cluster membership to data points depending on their distances from the respective cluster centers. The closer a data point is to a cluster center, the greater its degree is to that cluster. The degrees of membership to different clusters add up to one for each data point. The cluster centers and the membership are updated after each iteration.

4.2 Finding Optimal Number of Cluster Centers

Fuzzy c-means clustering algorithm is applied on the pairwise language distance data to find the language clusters. A Python module that implements the algorithm was utilized for the experiments [18]. Clustering is an unsupervised technique. It is the purpose of this research to investigate whether real-life language families will group together to form clusters.

Table 1: Finding Optimal Number of Clusters

Number of Clusters	Validation Metric	Eigenvector Similarity	Gromov-Hausdorff Distance	Relational Similarity
N = 2	PC	0.76	0.704	0.532
	PEC	0.555	0.671	0.953
N = 3	PC	0.806	0.627	0.497
	PEC	0.539	0.916	1.249
N = 4	PC	0.734	0.537	0.473
	PEC	0.763	1.289	1.437
N = 5	PC	0.749	0.549	0.532
	PEC	0.771	1.343	1.403
N = 6	PC	0.781	0.653	0.61
	PEC	0.684	1.073	1.249
N = 7	PC	0.759	0.701	0.685
	PEC	0.784	0.972	1.061

Fuzzy c-means algorithm requires the user to provide the number of clusters to be created. The number of optimal clusters was decided based on the results of cluster validation. Partition Coefficient(PC) and Partition Entropy Coefficient(PEC) were used as validation metrics. Both these metrics were introduced

by Bezdek [19]. PC is defined as

$$PC = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij}^2) \quad (1)$$

where N is the number of samples, c is the number clusters formed and μ_{ij} is the degree of membership of the j^{th} sample to the i^{th} cluster. The range of PC values is $[1/c, 1]$. A value close to $1/c$ indicates lack of grouping tendency in the dataset or the failure of the clustering algorithm in revealing them [20]. PC values closer to unity indicates crisper clusters. The reverse is true for Partition Entropy Coefficient(PEC). Lower PEC values indicate better clusters.

As discussed in Section 3, three different measures were used as interlingual-distance measures. These measures resulted in three different sets of pairwise language distances for our 12 languages of interest. These pairwise distances were each clustered six times with the number of clusters, n changing from 2 to 7. Depending on the results' partition coefficient and partition entropy coefficient, the optimal number of clusters was decided upon. Table 1 shows the values of Partition Coefficient(PC) and Partition Entropy Coefficient(PEC) for different numbers of clusters for all three distance measures. It can be seen that when using Eigenvector similarity, PC is highest and PEC is lowest for $n = 3$. Hence, going forward, we choose the number of clusters to be 3 when pairwise distance is determined using eigenvector similarity. Similarly, we choose $n = 2$ for the Gromov-Hausdorff measure and $n = 7$ for relational similarity.

5 Results and Analysis

5.1 Eigenvector Similarity

Table 2: Clustering Results for Pairwise Eigenvector Similarity Values

Language	Cluster Label	Cluster 0	Cluster 1	Cluster 2
Azerbaijani	0	0.9553749548	0.03289303062	0.01173201461
Galician	0	0.9653123598	0.02378525453	0.0109023857
Belarusian	1	0.07644192888	0.8409271329	0.08263093821
Czech	1	0.05285856821	0.6159181772	0.3312232546
Romanian	1	0.01561677932	0.964432164	0.01995105663
Slovak	1	0.02264262907	0.9525235509	0.02483382006
Turkish	1	0.09718167975	0.7225380166	0.1802803036
English	2	0.002174238263	0.00823760348	0.9895881583
Portuguese	2	0.0264887177	0.123870656	0.8496406263
Russian	2	0.04558537082	0.1091271527	0.8452874765
Spanish	2	0.003913274723	0.01643892104	0.9796478042

Table 2 shows the results of Fuzzy c-means algorithm on the pairwise Eigenvalue similarity measure for the 12 languages in our dataset. As discussed in Section 4.2 the optimal number of clusters for this distance measure was set to 3. Accordingly, 3 clusters were formed from the 12 languages. The languages Azerbaijani and Galician form one of the three clusters. Both the languages have a high degree of membership to Cluster 0(0.95, 0.97). Belarusian, Czech, Romanian, Slovak, and Turkish form the second cluster. While Romanian, Belarusian, and Slovak have a high degree of membership to Cluster 1, Czech and Turkish have significant membership to Cluster 2 as well(0.33, 0.18). Finally, Cluster 2 includes English, Portuguese, Russian and Spanish. English and Spanish belong almost exclusively to Cluster 2 with membership values of 0.99 and 0.98, respectively. Russian and Portuguese show a small degree of membership to cluster 1 as well (0.11, 0.12). The size of the language spaces may be one of the factors impacting the clusters. According to Vulic et al. [13] the corpus size on which embedding spaces are trained on can impact the degree of isomorphism. It can be seen from the results that Azerbaijani and Galician, two low-resource languages have the same cluster labels. While the resource rich languages English, Portuguese, Russian and Spanish have the same cluster labels. In these cases, the similarity in

resource availability may have resulted in higher degree of isomorphism among languages that do not necessarily have typological similarities.

5.2 Gromov-Hausdorff Distance

Table 3: Clustering Results for Pairwise Gromov-Hausdorff Values

Language	Cluster Label	0	1
Azerbaijani	0	0.6415238438	0.3584761562
Belarusian	0	0.9314493988	0.06855060124
Czech	0	0.9243313151	0.07566868486
Romanian	0	0.7548381654	0.2451618346
Russian	0	0.7612092276	0.2387907724
Slovak	0	0.8924948002	0.1075051998
Spanish	0	0.7066968061	0.2933031939
Turkish	0	0.8792561145	0.1207438855
English	1	0.1574221578	0.8425778422
Galician	1	0.2479783351	0.7520216649
Portuguese	1	0.2187419051	0.7812580949

Table 3 shows the result of the Fuzzy c-means clustering algorithm on pairwise Gromov-Hausdorff distances. The optimal number of clusters for the Gromov-Hausdorff measure is found to be 2. The first cluster is made up of 9 of the 12 languages in the dataset. Of these, Belarusian, Czech, and Slovak have the highest degree of membership to Cluster 0. Azerbaijani, although labeled as cluster 0, have a significant degree of membership to Cluster 1 (0.36), followed by Spanish (0.29), Romanian (0.25), and Russian (0.24). On the other hand, Cluster 1 includes English, Galician and Portuguese, and all these languages have some degree of membership to Cluster 0 (0.16, 0.25, 0.21). The size of the monolingual embedding spaces does not visibly affect the clusters formed, since Galician- one of the smallest embedding space in our dataset- is clustered with the much larger English and Portuguese embedding spaces.

5.3 Relational Similarity

Table 4: Clustering Results for Pairwise Relational Similarity Values

Language	Cluster Label	0	1	2	3	4	5	6
Galician	0	0.987	0.002	0.002	0.001	0.002	0.001	0.002
Azerbaijani	1	0.160	0.255	0.148	0.100	0.113	0.080	0.140
Turkish	1	0.004	0.977	0.005	0.001	0.003	0.001	0.005
Romanian	2	0.002	0.003	0.985	0.001	0.002	0.001	0.003
English	3	0.000	0.000	0.000	0.997	0.000	0.000	0.0003
Portuguese	4	0.040	0.036	0.048	0.030	0.791	0.013	0.038
Spanish	4	0.044	0.038	0.050	0.040	0.769	0.014	0.040
Belarusian	5	0.001	0.001	0.001	0.000	0.000	0.992	0.001
Russian	5	0.164	0.122	0.145	0.081	0.118	0.209	0.158
Czech	6	0.013	0.015	0.016	0.005	0.010	0.005	0.932
Slovak	6	0.122	0.137	0.150	0.047	0.081	0.068	0.391

The number of optimal clusters for the Relational Similarity measure was found to be 7. Table 4 shows the result of the Fuzzy c-means clustering algorithm on pairwise relational similarity measures. Galician, Romanian and English form their own single-language cluster with a high degree of membership to the cluster they're labeled as (0.99). Interestingly, Azerbaijani is grouped with its related higher-resource language, Turkish. The former, however, has a membership degree of 0.26 to its assigned cluster, while

the latter has a much higher degree of membership(0.98). A similar trend can be seen for Russian and Belarusian, Slovak and Czech. They are known to be related languages and have the same cluster labels. But they are not members of their assigned cluster to the same degree. Spanish and Portuguese, another pair of languages, also seem to be reflecting their real-world similarities in their cluster label. Unlike the previous cases, they both have a similar degree of membership (0.79, 0.77) to their assigned cluster label. Azerbaijani and Turkish are disparate in terms of the size of their respective embedding spaces, as are Belarusian and Russian. Their grouping together should be an indication that the corpus size on which these spaces were trained on does not impact the clusters when Relational Similarity is used as a distance measure.

6 Conclusion

This research applies fuzzy c-means algorithm on independently trained monolingual-word embedding spaces and analyses the results. Three different measures of isomorphism between vector spaces was utilized as pairwise distance between the languages. The principal goal was to observe whether real-world relationship between languages are encoded in their word embedding spaces, and which of the three measures can capture them best. From the results, it can be said that, the relatively simpler approach to measuring degree of isomorphism- Relational Similarity, showed the most promising results. The clusters formed using Relational Similarity as a distance measure best reflect our real-world knowledge about the languages in the language set. Employing a fuzzy clustering algorithm such as fuzzy c-means allowed us to study how different languages can belong to more than one cluster with varying degrees of membership. This information would have been lost had we used a hard clustering algorithm to cluster the language spaces.

In future, these language cluster knowledge can be used to build better cross-lingual word embedding spaces. Hub language is an important factor in deriving cross-lingual embedding spaces. Existing literature empirically show that a typologically distant hub language can impact the performance in downstream tasks. It can be hypothesized that the resource-rich languages in a particular cluster reported in this research can serve as a good choice for a hub space for the other languages that have the same cluster label while the different hub languages can be mapped to a shared embedding space.

References

- [1] Tomas Mikolov, Quoc V Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. arXiv preprint arXiv:1309.4168, 2013.
- [2] Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. Normalized word embedding and orthogonal transform for bilingual word translation. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1006-1011, 2015.
- [3] Artetxe, Mikel, Gorka Labaka, and Eneko Agirre. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 2289-2294. 2016.
- [4] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. Learning bilingual word embeddings with (almost) no bilingual data. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 451-462. 2017.
- [5] Samuel L. Smith, David HP Turban, Steven Hamblin, and Nils Y. Hammerla. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. arXiv preprint arXiv:1702.03859 (2017).
- [6] Anders Søgaard, Sebastian Ruder, and Ivan Vulić. On the limitations of unsupervised bilingual dictionary induction. arXiv preprint arXiv:1805.03620, 2018.

- [7] Barun Patra, Joel Ruben Antony Moniz, Sarthak Garg, Matthew R. Gormley, and Graham Neubig. Bilingual lexicon induction with semi-supervision in non-isometric embedding spaces. arXiv preprint arXiv:1908.06625, 2019.
- [8] Kowshik Bhowmik and Anca Ralescu. Leveraging Vector Space Similarity for Learning Cross-Lingual Word Embeddings: A Systematic Review. *Digital* 1, no. 3 (2021): 145-161.
- [9] Ndapandula Nakashole and Raphael Flauger. Knowledge distillation for bilingual dictionary induction. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 2497-2506, 2017.
- [10] Antonios Anastasopoulos and Graham Neubig. Should All Cross-Lingual Embeddings Speak English? arXiv preprint arXiv:1911.03058, 2019.
- [11] Geert Heyman, Bregt Verreet, Ivan Vulić, and Marie-Francine Moens. Learning unsupervised multilingual word embeddings with incremental multilingual hubs. 2019.
- [12] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017): 135-146.
- [13] Ivan Vulić, Sebastian Ruder, and Anders Søgaard. Are All Good Word Vector Spaces Isomorphic?. arXiv preprint arXiv:2004.04070, 2020.
- [14] Suganya, R., and R. Shanthi. Fuzzy c-means algorithm-a review. *International Journal of Scientific and Research Publications* 2, no. 11 (2012): 1.
- [15] Lotfi A. Zadeh. Fuzzy sets. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*. 1996. 394-432.
- [16] Joseph C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. 1973. 32-57.
- [17] James C. Bezdek, Robert Ehrlich, and William Full. FCM: The fuzzy c-means clustering algorithm. *Computers geosciences* 10.2-3 (1984): 191-203.
- [18] Madson Dias. *fuzzy-c-means: Python module implementing the Fuzzy C-means clustering algorithm*, 2019.
- [19] James C. Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science Business Media, 2013.
- [20] Vania C. Mota, Flavio A. Damasceno, and Daniel F. Leite. Fuzzy clustering and fuzzy validity measures for knowledge discovery and decision making in agricultural engineering. *Computers and electronics in agriculture* 150 (2018): 118-124.

Kowshik Bhowmik
College of Wooster
Mathematical and Computational
Sciences
1189 Beall Ave, Wooster, OH-
44691
USA
E-mail: kbhowmik@wooster.edu

Anca Ralescu
University of Cincinnati
EECS
2600 Clifton Ave, Cincinnati,
OH-45221
USA
E-mail: ralescal@ucmail.uc.edu

Decentralized open-source cryptocurrency portfolio tracking hybrid application

Adoris-Elia Doran

Abstract

Considering the recent rapid expansion of the cryptocurrency world, it has become increasingly more difficult for the average investor to keep track of the performance of their investment, considering the large number of assets and different accounts. This article explores a possibility to track the performance of a cryptocurrency portfolio by automatically retrieving portfolio data from various cryptocurrency exchanges and wallets, as well as tracking the price evolution of different cryptocurrency assets.

1 Introduction

According to a study done by *BDC Consulting* in 2019 carried out on 800 respondents, it was determined that only 18.91% of cryptocurrency owners have a single account. The vast majority, of 40.47% had at least three accounts. [1]

In a similar fashion, according to *Statista*, back in February 2021 there were 4,501 different cryptocurrencies. In November 2021, that number has grown quite significantly, to 7,557 cryptocurrencies. [2]

Due to this vast diversification of the cryptocurrency world, it has become increasingly more difficult for the common investor to track the balances of their cryptocurrency assets across multiple accounts and with a large number of possible assets. With this scenario in mind, this paper proposes an application which would assist investors in understanding the performance of their portfolio by automatically obtaining balance information from various accounts and displaying current price information for their owned assets.

The core functionality of the application is to allow the user to view the balances and the actual value of each of the cryptocurrencies owned and stored into multiple cryptocurrency accounts. As mentioned in the introductory section, the accounts an user can have categorized separately in the application. These account types reflect in the application by the means of a navigation bar at the bottom of the screen.

Since the scope of the application is to automatically obtain the user's portfolio information, the paper proposes a mechanism for obtaining this information, as described in section 7. For the initial implementation, multiple cryptocurrency exchanges (*Binance*, *Bitfinex*), individual cryptocurrency wallets (*Elrond*, *Ravencoin*), as well as decentralized exchanges (*PancakeSwap*). More information about the types of accounts and how they relate to the application are described in section 3.

During the research of how to assist an investor in tracking their portfolio, there could be identified two different approaches:

1. **The analysis of individual transactions**, to determine profit and loss. This approach allows the user to view the performance of each owned asset, however it is more prone to errors due to information that is sometimes no longer available (for example, *Binance* only keep three months worth of transaction logs). One such application is *Delta*, a mobile application owned by *eToro* (which is also an investment and cryptocurrency broker).
2. **The analysis of the resulting balance**. This calculation is much simpler and thus easier to implement. For this to work the balance of each cryptocurrency owned is required, as well as the current price information these. The current value of all the cryptocurrencies is then compared against the initial investment (which can be manually entered and easily determined by the user via bank statements). The downside is that the user is not able to see performance of individual positions within the portfolio.

For the purpose of this paper and the initial implementation, the second approach was used. The rationale was as follows:

- Although a good technique in the investment world, analyzing transactions in the cryptocurrency space is significantly difficult due to the large amounts of edge cases (exchanging, withdrawing, staking), which can result in erroneous values.
- For an average investor, knowing the direction of the overall portfolio should be sufficient. Information about individual positions tends to be more towards professional or commercial traders, a demographic that this paper does not target.

Since the application is implemented as a hybrid application, it has support for the following platforms:

- Mobile applications, which can be installed on both iOS and Android devices. Packaged as a hybrid application using *Apache Cordova*¹.
- Desktop applications, supporting all three major operating systems (*Windows*, *Linux* and *macOS*), packaged as a hybrid application using *Electron*².
- Directly accessed from a web browser with a responsive interface, which can be accessed from any modern smartphone, tablet or desktop device.

¹*Apache Cordova* allows the creation of mobile applications for both *Android* and *iOS* using the fundamental web technologies (HTML, CSS and JavaScript). More details about *Apache Cordova* can be found on its official website: <https://cordova.apache.org/>

²*Electron* allows using web technologies to create desktop applications for *Windows*, *Linux* and *macOS*, as well as have access to APIs that are not normally accessible to web applications. For more information: <https://www.electronjs.org/>

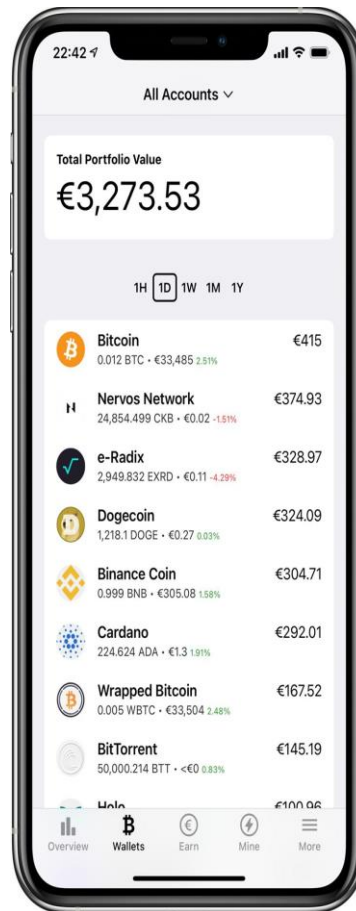


Figure 1: Screenshot of the "Wallets" section of the application, where the user can see information about exchange accounts, as well as individual wallets

2 Functional Overview

2.1 Spot Wallets

One of the more common types of accounts are “spot” wallets. These are split into two subtypes:

1. Exchange accounts – these are created by cryptocurrency exchange platforms such as Binance in order to store the converted cryptocurrencies. These are usually not for long-term storage, as individual wallets might be safer.
2. Individual wallets – these are cold or hot storage wallets created by the user for a particular cryptocurrency network (blockchain). Some wallets might only support one cryptocurrency.

From the application point of view, there is no difference between these two types of accounts, as they all can contain one or more cryptocurrency assets.

For all the sections of the application where cryptocurrencies are displayed, the list of owned assets will display the name of the cryptocurrency, the quantity owned, the current price of the asset, the price change and the current total value of the asset.

There is also a selection bar to choose the time frame for the percentage change, from the last 24 hours to the last year.

In order for the user to have a general view of the assets owned, an aggregate account (called “All Accounts”) which sums up all the account balances. The user can choose to view either all the owned assets via the aggregate account or select an individual account by pressing on the title section.

At the end of the asset list there is also a chart which displays the percentage allocation of all the cryptocurrencies owned by the user, as long as they take up more than 5%. The assets with under 5% are grouped under a dedicated entry called “Others”. The purpose of this chart is to allow the investor to calculate their exposure to risk by not diversifying their portfolio enough.

When switching between accounts, the number of assets as well as the total value in the user’s preferred currency and the name of the account set by the user is displayed in order to allow for easy identification of the accounts.

2.2 Investment accounts

An alternative to holding a particular cryptocurrency is to invest it by the means of deposits at an exchange or by participating in pools via decentralized exchanges.

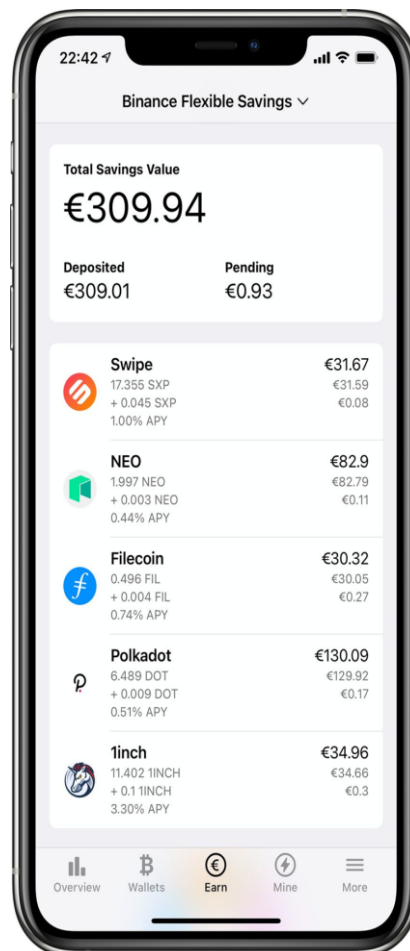


Figure 2: Screenshot of the "Earn" section, which shows investment account balances, as well as the pending rewards.

The main difference when compared to wallet accounts being that these accounts also generate income. In general, the following account types are supported:

1. **Individual cryptocurrency deposits.** Here the user deposits one or more cryptocurrencies or tokens for a fixed or variable duration, similar to a bank deposit. The interest (or “reward”) is paid back in the same cryptocurrency as the initial investment. For this kind of investment, the following additional information is displayed: the initial deposit (as deposited amount, as well as the current value in the user’s preferred currency), the total interest earned on that particular deposit, and the estimated yearly gain.
2. **Individual cryptocurrency deposits, with a different reward token.** This is a very similar investment account to individual cryptocurrency deposits, however the reward is received in a different asset. In order to accommodate this difference, the application displays the name and icons of both the initial and the rewarded asset.
3. **Paired cryptocurrency investments.** These can often be found in decentralized exchanges, as liquidity pools, in which the initial investment consists of two different cryptocurrencies which have an approximately value and are used for providing liquidity to the markets. The investment of the two cryptocurrencies is represented as a single token, and that token is also used as a reward. For this kind of account, the application displays the name and icon of both cryptocurrencies belonging to the pair, the initial investment of each asset (including the current value of the initial investment), as well as the reward.
4. **Paired cryptocurrency investments, with a different reward token.** Similarly to paired cryptocurrency investments, there are two assets which are invested initially by the user, with a third cryptocurrency representing the reward, usually a token specific to the exchange. The user is presented with the same information as previously, with just the reward being displayed in a separate cryptocurrency.

2.3 Mining accounts

The mining section has a different approach when compared to the Spot and Earn accounts. The main difference is that only one cryptocurrency mining activity can be viewed at one time. For each activity (see Figure 3) the following information is presented to the user:

1. **Pending rewards,** which represent the amount the user has mined but are not yet paid out by the mining pool. The application will display the name and illustration of the asset, as well as the actual pending amount and its current value in the user’s preferred currency.
2. **General information about the workers.** This section displays information about the devices which the user is mining with, such as its name, the date and time it was last seen (it has last contributed to the mining pool), as well as the mining rate (also known as the hash-rate).
3. **The list of the last transactions (payouts).** The application will list the last 14 transactions that were created by the mining pool in order to reward the user for the mining activity. Here the actual amounts, the value in the preferred currency, as well as the date of the transaction are displayed.

Due to the nature of these types of accounts, there are no aggregate account which combines the information into a general summary. However, the value of all the mining accounts is presented in the account switcher, as well as the general overview section.

Similarly to the “Wallets” and “Earn” section, the pending rewards are counted towards the total portfolio value. However, the “Payouts” are not counted since they might have already been spent or exchanged by the user. Instead, the mined amounts will show up in the “Wallets” section if the correct account was added by the user.

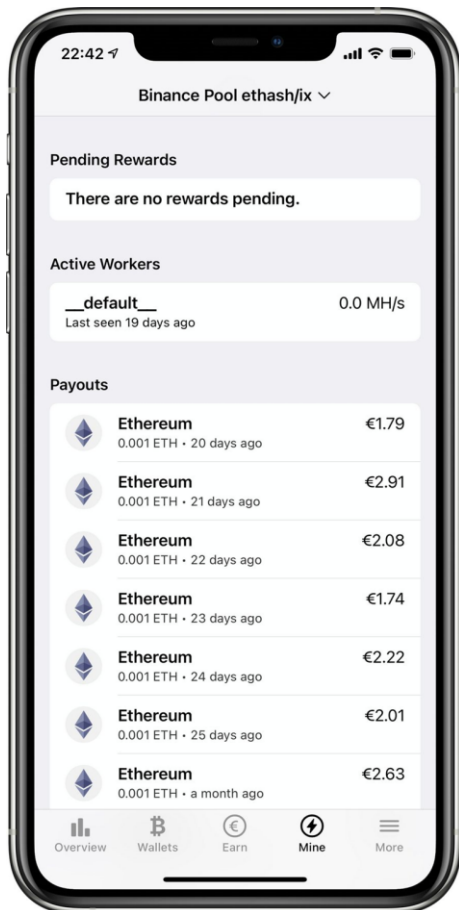


Figure 3: Screenshot of the "Mine" section where information about active mining workers and pending transactions are shown.

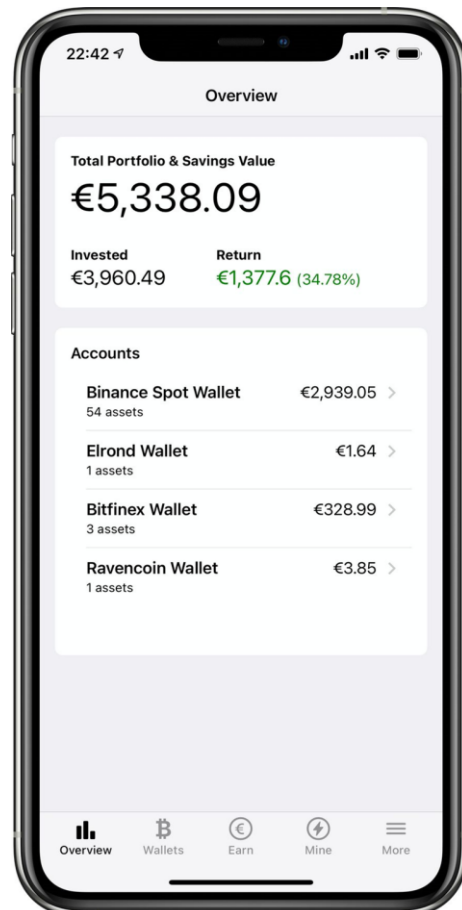


Figure 4: Screenshot of the "Overview" section, summarizing the total portfolio value as well as offering an account switcher.

2.4 Overview section

The purpose of the overview section is to allow the user to better understand the distribution of their investment, as well as to be able to quickly understand the direction the portfolio is heading.

Currently, the following information (see Figure 4) is displayed in this section:

1. **The total value of the accounts.** This sums up all the cryptocurrencies owned by the user in all the three main sections of the application and displays it in the user's preferred currency.

2. **The value of the initial investment.** At this time, since the application is unable to track transactions, it has to be defined manually by the user and it is shown for reference purposes.
3. **The unrealized returns of the investment.** This represents the difference between the total value of the accounts (1) and the initial investment (2). The return is unrealized because it refers to the opened positions only. For quick reference, the color of the text reflects whether the user is currently at a gain, or at a loss.
4. **The account switcher.** This allows the user to quickly navigate to any of the accounts, as well as view the individual values for them.

As a feature improvement, the overview section is planned for an overhaul which adds multiple functions such as a chart to track the unrealized returns across multiple time frames.

2.5 The “Additional options” section

This section allows the user to configure the application, as well as add integrations to new cryptocurrency accounts. The following features are present:

1. **Deposits and withdrawals.** This section allows the user to enter manually the information about each deposit of cash into the account, as well as withdrawals of cash. The purpose of this is to be able to calculate the return on the investment.
2. **Appearance options.** This allows the user to toggle the display of particular information (such as hiding unwanted cards), the number of decimals to be displayed, etc.
3. **Preferred currency.** This will change the currency in which the total value of a particular cryptocurrency, as well as the portfolio overview totals are displayed. The currencies from which the user can choose are limited to the currencies provided by the data retriever, as there has to be a known correspondence between the fiat currency and the cryptocurrency.
4. **Themes.** Allows the user to change between a light and a dark theme. There is also an automatic option which toggles between the two themes based on the device settings. This preference is saved locally, so it can be set on a per-device basis.
5. **Account management.** Here the user can view a list of the account integrations, as well as add or edit existing integrations. When adding a new account integration, the user is asked about the account type and a list of supported platforms is displayed. When selecting a particular account/supported platform, information about the connection details will be requested from the user. Depending on the type of the account, the user will be generally asked for an API key, and possibly a private key. The user can also assign a friendly name which will be used to describe the account when navigating through the application.

3 Retrieving cryptocurrency information

In order for the application to be able to display the user’s portfolio, it needs to obtain the relevant data from multiple services and combine them in a meaningful way. In order to do so, the application is split into multiple modules with different responsibilities:

1. **Account providers** which connect to different cryptocurrency exchanges such as *Binance*, or individual wallets. They provide the application with a list of cryptocurrencies, as well as the owned balances.
2. **Data providers** offer the application information about a particular cryptocurrency, such as the current price in a particular currency, but also static information such as the full name of the asset, its pictogram, etc.

3. **The asset manager** which is in charge of creating the correspondence between the cryptocurrency symbols identified by the account providers and the actual cryptocurrency information obtained from the data provider.

3.1 Obtaining balance information from accounts

In order to obtain information about the user's balance for any given account, the application must communicate with the service where the user is storing their cryptocurrency (be it centralized cryptocurrency exchanges, decentralized exchanges or individual wallets) in order to obtain the balance information.

Because of the nature of the accounts, there is usually a different API for each platform, as well as for each type of account (wallet accounts, investment/earn accounts or mining accounts). In the context of the application, these implementations are called account providers.

When the application is connecting to a particular service, it will have to authenticate. The authentication methods also differ from platform to platform, with the most common methods being entering an API key, with additionally a private key. These credentials are entered by the user during the configuration process.

When connecting to an individual wallet, the address of the wallet should also be provided. This uniquely identifies the wallet and is formatted as a hexadecimal string, with a slightly different format for each cryptocurrency network.

In the case of decentralized exchanges, an intermediate service such as *Yieldwatch* is used which analyzes the performance of the portfolio. For these kind of services, there is also an option to improve the speed of the data retrieval by selecting which platforms are used by the user (such as *PancakeSwap*).

In order to protect the user from various man-in-the-middle attacks, the application will refuse any API key created by the user that has more than "read" rights (to obtain balance information) such as the ability to open or close transactions. This means that even if an attacker gains access to the API key, they will only be able to obtain balance information and publicly available information.

While the account provider can provide accurate balance information for accounts owned by the user, the platforms usually do not provide price information. In addition, a cryptocurrency can be identified by different symbols across different platforms, meaning that account provider has to work with the asset manager to correctly identify the cryptocurrency for a particular balance.

At the moment, the following cryptocurrency platforms are supported by the application:

1. **Binance** is, at the time of writing, is the exchange with the highest market capitalization in the cryptocurrency space. [3] Due to this, it was the first candidate for which to implement an account provider. This platform offers multiple tools for investors such as exchanging cryptocurrency pairs, investment accounts, mining accounts, [4] all of which were implemented in the application.
2. **Binance Smart Chain** is an alternative to the *Ethereum blockchain*, developed by *Binance*, with the reason of reducing fees and improving the speed of the transactions. The native cryptocurrency for this chain is the *Binance Coin* (BNB). This network is also compatible with *Ethereum* wallets, however it runs on a distinct network. [5] The application is able to track all the cryptocurrencies/tokens owned by the user.
3. **Bitfinex** is another cryptocurrency exchange similar to *Binance* offering cryptocurrency pair exchanges, as well as margin trading accounts. [6] It is also one of the top exchanges by market capitalization. [7] Support for spot accounts (wallets) was added to the application.

4. **PancakeSwap** is a decentralized exchange which allows exchange via cryptocurrency pairs, as well as investment opportunities such as participating in liquidity pools (see 4). This platform runs as a decentralized application on the the *Binance Smart Chain*. [8]
5. **Beefy.Finance** is another decentralized application that acts as an investment optimizer. Through smart contracts, it will try to find the most profitable investment platforms given certain cryptocurrencies. [9] Similarly to *PancakeSwap*, the initial investment and rewards are tracked by the application.
6. **Yield Watch** is an investment tracking application which analyses decentralized applications such as *PancakeSwap* and *Beefy.Finance*. [10] This platform is used by the application in order to obtain detailed information about the initial investment and rewards of the two platforms.
7. **Elrond** is a relatively recent addition to the cryptocurrency space, that can be used for currency purposes, as well as smart contracts. There was a recent introduction of the *Elrond* mobile wallet application called *Maiar*. [11] The integration with *Elrond* allows viewing the balance of the user's *Elrond* wallet.
8. **Ravencoin** is a cryptocurrency based on Bitcoin which adds support for fungible and non-fungible tokens, similarly to *Ethereum*. [12]
9. **Flypool** is a set of mining pools for different cryptocurrencies such as *Ethereum*, *Zcash* and *Ravencoin*. [13] By participating to the mining pools with computational resources, users are able to receive mining rewards. These rewards can be tracked in the proposed application (see 5).

3.2 Obtaining cryptocurrency information from data providers

After the application has identified the user's balance, the application must also obtain information about each of the cryptocurrencies owned by the user, especially regarding the current value of a particular asset. This information is gathered via external services using a module of the application called the data provider.

The information needs regarding a particular asset are as follows:

1. Static information such as the full name of the cryptocurrency, a pictogram identifying it visually and the symbol. This information can also be cached by the application in order to improve the speed of data updates.
2. Dynamic information such as the current value of the cryptocurrency, as well as the percentage change for multiple time frames (e.g. last 24 hours, last week, last month).

On the market there are multiple services suitable for a data provider. The ones considered for the purpose of this application were *CoinMarketCap* and *CoinGecko*. These provide both a user-friendly interface to cryptocurrency information by means of a dedicated website, as well as an API to access this information programmatically.

The *CoinMarketCap* was the first service to be considered for the application as it is highly popular and backed by a large institution in the cryptocurrency space (*Binance Capital Mngmt.*). However, the API service is provided on a per-subscription basis; there is a free plan with the following limitations, at the time of writing:

- A limit of 10.000 calls per month. For a general user, this should be enough as it amounts to around 1.240 synchronizations per month (considering 4 calls per synchronization). However, this would imply that the application can only be used by one user, so users would have to maintain their own API keys (thus, making the installation process much more complicated).

- Only 9 of the 22 services are available. This means that no historical price information can be obtained (such as obtaining the price information for the time frame proposed by the application: last 24 hours, last week, last month, etc.), only the current price information. This is a significant downside for a data provider, as it would disable an important feature of the application.

Considering the limitations presented above, a different platform had to be chosen. A good alternative that is also popular is *CoinGecko*. According to the official statement, the goal of *CoinGecko* is to democratize cryptocurrency data, which fits well with the scope of this application.

Unlike *CoinMarketCap*, *CoinGecko* does not impose any rate limits and it also offers historical price information. One unique aspect of *CoinGecko* is that it does not identify cryptocurrencies by their symbol, but by a unique ID that is specific to *CoinGecko*. This means that the application must do an additional mapping between symbols and the IDs of *CoinGecko*.

Even if *CoinGecko* was selected as the primary data provider, the application is structured in such a way that it can use any data provider as long as there is an implementation for it. This makes the application more durable in case *CoinGecko* changes their service policy, or it can also use a fallback in case the main platform is experiencing down-time.

3.3 Uniquely identifying assets across different platforms

When an account provider offers information about the user's portfolio, the cryptocurrencies are not yet identified, as each platform has a different identification method and there is no universal way to identify cryptocurrencies.

In order to create a correspondence between the information obtained by the account providers and the data providers, a dedicated module was created. This module is called the asset manager, and unlike the account and data providers, it has a single implementation.

In order for the asset manager to correctly identify a cryptocurrency from a user's balance information, the following information is needed:

1. The symbol of the cryptocurrency, as it is identified by a particular platform. This can be used to immediately identify common cryptocurrencies with symbols such as BTC (*Bitcoin*), ETH (*Ethereum*), RVN (*Ravencoin*), etc.
2. Optionally, a full or partial name of the cryptocurrency, as it is reported by a particular platform. This can help resolve conflicts where there are multiple cryptocurrencies sharing the same symbol.

The asset manager is able to use the information detailed above and determine the corresponding cryptocurrency by means of the following algorithm:

1. Download a full list of cryptocurrencies are reported by the data provider. This information can later be cached for improving further data retrievals.
2. When a request is made to identify a cryptocurrency, the asset looks into the list of all cryptocurrencies and tries to find all the assets having the same symbol as the requested one. Because the same symbol can be by multiple assets, the algorithm will create a list of all the assets matching that particular symbol.
3. Based on the number of assets identified as having the same symbol as the desired one, the following cases can occur:
 - In case no assets could be identified, the asset manager will create a blank asset which will contain the name and symbols that were reported by the account provider. This will

mean that the user will not be able to view price information (since it cannot be matched with an existing cryptocurrency), but will be able to view the balance information.

- In case exactly one asset could be identified, then that particular asset is going to be used. This is the best case scenario, since there is a very high likelihood that the correct cryptocurrency has been identified.
- If multiple assets would be identified, then the asset manager will have to select one of them. Since the symbols are identical, only the name of the cryptocurrency can be used for identification purposes. However, names of cryptocurrencies can widely differ between platforms. In order to cope with this, the application will apply the Levenshtein distance algorithm in order to find the name which is most similar to the corresponding cryptocurrency name. The asset whose name has the least edit distance is most likely to be the correct asset.

4 Architecture of the application

During the design of the application, a client-server architecture was selected. In this scenario, the user would most likely have to host their own server either on a personal device or on a third-party cloud service. There is also the possibility of the application having its own cloud service in the future, but that would be outside the scope of this paper.

In the future, in order to aid the decentralization of information, the mobile and desktop applications would be able to integrate the server component as well, being able to run independently of a server.

One of the main reasons to select a client-server architecture was related to the synchronization of data. Since user configuration such as the credentials, as well as the cached data from the data and account providers is stored on the server, it is much easier for the user to access the platform from different devices. In addition, the server would allow accessing the platform also by means of a web browser, not just by a mobile or desktop application.

4.1 Server architecture

The role of the server in the context of this application can be defined as follows:

1. Storage of the credentials for the user's accounts, as well as storing the preferences.
2. Connect to the various data providers to obtain cryptocurrency information, as well as account providers to obtain information about the user's balance.
3. Offer an API which allows a programmatic access of the user's portfolio performance information, as well as the balances and cryptocurrency price information. This can be used either by the mobile and desktop applications or third-party software such as a widget to be displayed on a desktop environment.

The server is responsible of managing the following configuration:

- The user's cryptocurrency accounts, which are for a given cryptocurrency platform (e.g. *Binance*, *Bitfinex*) and can have custom name selected by the user. The name does not have to be unique as it is not used for identification purposes. In order to connect to the account,

an API key has to be generated by the user and provided here. For some types of accounts a private key or a wallet address is needed.

- Information about the deposits and withdrawals of the user. These are needed in order to determine the user's portfolio performance, when compared against the current value of their investment.

At the moment the user preferences and credentials are stored in a *JSON* format, however there is an on-going effort to use a database instead to be more easily scalable and configurable via the client application. The client application already has an implementation for the configuration section, which has to be linked to the server.

The main attribute of the server is to use the known configuration to connect to different services and obtain the information about the user's cryptocurrency account to be displayed in the client application. As such, the server hosts the account providers (see 8), data providers (see 9) and asset manager (see 10) which are described in their own particular subsection.

From a technical point of view, the server was implemented in *Node.js*, an execution environment which allows back-end development using the JavaScript programming language. For managing dependencies, the de facto *Node.js* package manager was used, *NPM (Node Package Manager)*.

The server uses the *Express.js* framework which allows the implementation of a *REST-ful* service. The server exposes two resources via the REST API:

1. A way to obtain the cached portfolio information. This will allow the application to quickly display the portfolio information to the user without waiting for the synchronization to complete (which can take seconds or minutes based on the number of the accounts).
2. A way to synchronize the portfolio, in order to obtain the latest balances and the latest price information for the cryptocurrencies.

4.2 Client architecture

The client application retrieves the information about the user's cryptocurrency performance and displays them in a meaningful way to the end user.

The web application is implemented as a single-page application in order to provide better user experience and to be closer to mobile applications. The web application was developed using *Framework7*, a library which provides a near-native-looking user interface for web applications.

In order to aid the development of the single-page application, the *Svelte* web framework was used. It allows reusing components of both *Framework7* and custom reusable components specific to the application.

More details about how the actual user interface of the client application is presented in section 3.

5 Conclusions

This article presented the functional and graphical specifications of implementing an application which can assist investors in tracking the performance of their portfolio automatically.

Although the core functionality of the application has already been implemented, in order for the application to become a viable product, it would require some of the short and long term improvements listed below.

Since the mission of this project is to be decentralized and open-source, the entire source code of this application is publicly available on GitHub.³

Future directions:

Short-term improvements:

- Improve the user experience by adding more ways to display the existing information, as well as to display additional information about cryptocurrencies.
- Add more integrations based on the most popular trends in the cryptocurrency space.

Medium-term improvements:

- In subsection 11, it is mentioned that the server architecture currently uses *JSON* as a storage mechanism for both user preferences and credentials.

Long-term improvements:

- As mentioned in section 11, an improvement towards the decentralized mission of the application would be to allow the mobile and desktop applications to run without a server. This would reduce the need to connect to an additional service and would make it easier for users not having to set up their own server.

Acknowledgment: This work was supervised by Professor Ph.D. Dana Simian, from Dept. of Mathematics and Informatics, Faculty of Sciences of the *Lucian Blaga University of Sibiu*.

References

- [1] BDC Consulting, *Cryptotraders survey*. Oct. 2019. [Online]. Available: <https://bdc.consulting/insights/cryptocurrency/survey-crypto-traders>
- [2] R. de Best, *Number of crypto coins 2013-2021*, Statista. <https://www.statista.com/statistics/863917/number-crypto-coins-tokens/>
- [3] CoinMarketCap, *Top Cryptocurrency Exchanges Ranked By Volume*, CoinMarketCap. <https://coinmarketcap.com/rankings/exchanges/>
- [4] Binance, *Bitcoin Exchange | Cryptocurrency Exchange | Binance*. <https://www.binance.com/en>
- [5] Binance, *Binance Smart Chain: A Parallel Binance Chain to Enable Smart Contracts*. Apr. 17, 2020. [Online]. Available: https://dex-bin.bnbstatic.com/static/Whitepaper_%20Binance%20Smart%20Chain.pdf
- [6] Bitfinex, *Bitfinex | Cryptocurrency Exchange | Bitcoin Trading | Futures Trading | Margin Trading*. <https://www.bitfinex.com/>

³The source code of the portfolio application can be publicly accessed on GitHub: <https://github.com/eliandoran/portfolio/>

- [7] CoinMarketCap, *Cryptocurrency Prices, Charts And Market Capitalizations*, CoinMarketCap.
<https://coinmarketcap.com/ro/>
- [8] Binance Academy, *A Guide to PancakeSwap*, Binance Academy.
<https://academy.binance.com/en/articles/a-guide-to-pancakeswap>
- [9] Beefy.Finance, *Beefy.Finance Overview*. <https://docs.beefy.finance/beefyfinance/>
- [10] yieldwatch.net, *yieldwatch FAQs*, Medium, Mar. 03, 2021. <https://yieldwatch.medium.com/yieldwatch-faqs-93c2cde244bf>
- [11] Yahoo! Finance, *Elrond Announces Wallet and Global Payments App Launch on 31st January 2021*.
<https://finance.yahoo.com/news/elrond-announces-wallet-global-payments-201400661.html>
- [12] Raven, *Home, Ravencoin*. <http://ravencoin.org/>
- [13] Bitfly, *General Terms of Operation*. May 18, 2021. [Online]. Available: https://bitfly.at/GTO_v1.5.pdf

Adoris-Eliau Doran

Lucian Blaga University of Sibiu
Faculty of Sciences, Dept. of Informatics
Str. Dr. I. Rațiu, No. 5-7, Sibiu 550012
ROMANIA
E-mail: elian.doran@ulbsibiu.ro

Kiosk web application - integrated multimedia information system

Martin S. Dzhurov

Abstract

This paper presents kiosk web application developed for the needs of the multimedia information system, located in the campus of University of Rousse "Angel Kanchev", Bulgaria. The paper aims to show the flaws of the previously used software for this purpose and to show the benefits of the newly developed one. These days more and more institutions are aiming to automate everyday tasks, thus making the workflow easier for the personnel and the client. One way of automating the process of giving and spreading information in any modern institution is integrating kiosk-terminals with suitable kiosk-software [1]. This idea was implemented successfully by my university over 15 years ago, and since then the kiosk-terminals have been a huge information-helper for students, staff and university visitors. For example: students can search fast and easily their weekly schedule, they can orientate themselves in the university campus thanks to the interactive campus map, they can read news from the university newsfeed and many others. The idea and the implementation were perfect, but in the coming years many of the old programmers who created the old kiosk-application left the university, leaving the software with little to none options for updates and support. As a young and enthusiastic person, I saw a huge problem in the now dying information system, so I got to work in order to create a new modern and fully supported kiosk web application from ground zero. In this paper I will discuss the stages of development of a whole new system, starting with the lowest level - the hardware, building to OS level, and finalizing with a whole new web application with full support and no paid licenses. I will present to you the application structure, views, layers of access and the content management system, providing a dynamic content editing environment.

1 Introduction

More than 20 kiosk-terminal devices are located on the territory of the university. There are two varieties – two "large" and a number of "small" kiosks. They are located mainly in the Main building and in the newly built Second Building, with one terminal in each of the other buildings. The information system was installed at the university 15 years ago and has performed well in the learning process of its time. Due to certain circumstances, as well as due to the already outdated hardware and software, for a long time this system was unnecessary rather than helping the learning process.

The idea for the application came to my mind not long after I had started working for the university in ICSC (Information and Computer Service Center), Multimedia Department. I was gradually beginning to get acquainted with the existing system implemented in the kiosks. The more I became interested the more I realized that this system had long been ineffective and did

not help students and faculty [2], and that kiosks had become mere advertising screens for clubs held at the university, as they are often overlooked. I felt it was time for a change, and then I was inspired to build this new system from scratch.

The complete system, which was integrated on the kiosk terminals, was created 14 years ago, which is a long time, especially when it comes to software. One of the reasons for the "decline" of the system is the lack of a team to maintain it. The kiosks had Windows XP SP3 (later upgraded to Windows7, 8.1, 10), and each machine had its own hard disk drive (HDD). The software used for the kiosk application was a combination of the long-outdated Microsoft Internet Explorer 9 and a web application written on already unsupported web technology, also owned by Microsoft.

The old kiosk system has been creating a lot of problems and currently has a number of shortcomings. Here are some of them:

- Common cases of misuse of the kiosk device – opening web pages and applications to which the end user should not have access;
- Reaching terminal cases where the user cannot return to the kiosk web application;
- Frequent cases of unexpected interruption of the application itself and/or the browser on which it was running;
- Frequent servicing of the device due to instabilities and malfunctions related to the operating system itself;
- Outdated and unfriendly design, which leads to a brief retention of the user's attention;
- Lack of responsive web design – designing the user interface so that it changes according to the screen size of the respective device;
- Lack of application inactivity mechanism – It has often happened that a kiosk terminal stays in the same view/page for days until the next user arrives and he returns it to its original view.

In summary, it can be said that the system was not up-to-date, modern and most importantly - does not fulfill its purpose, namely to serve students, faculty and staff of the university. Problems with legacy systems like this often cannot be fixed by updating, upgrading, etc. In this case, the lack of access to the source code of the old application required the construction of a new system from scratch.

2 Stages of development

2.1 Overview

The construction process started from scratch, so I decided to start from the lowest layer of the system - the hardware. I looked at the available hardware base of the devices and moved on to exploring the possibility of implementing an operating system that takes up minimum hardware resources of the devices. Then I moved on to the selection of a methodology for delivering the kiosk application to users. It was best to use a web browser running in "kiosk" mode and visualize a web application on it. The longest stage of the development of the whole system was the stage of designing and writing the web application. The last stage, which was the most dynamic of all, was the commissioning of the system on the territory of the university. Let's discuss them in more details:

2.2 Hardware inspection and analysis

As I said earlier, I've started developing the new system from ground zero, so I needed to know with what hardware base I was working on. During this stage, a detailed analysis of the hardware resources of each device located on campus was made. This included a site visit and use of the available operating system or system BIOS to collect information about hardware components such as the CPU, Motherboard, RAM, Network Card, make and model of the touchscreen panel on the screen and others. After analyzing the information gathered, I found that most devices have

outdated hardware. At the moment there were no available recourses in order to update the hardware, so this situation guided the idea for the next step in the development of the system - finding a stable, light and secure operating system that was going to support the new web application.

2.3 Operating system choice

When choosing an operating system, I had three key requirements to cover: it has to be very lightweight (minimum hardware resources consumption); it has to be stable – in order to avoid unexpected OS crashes; it has to be secure for obvious reasons – kiosks are PCs for public use and they have to be well prepared for unexpected “attacks”. Having this in mind, left me with several Linux distributions to choose from - Lubuntu, Xubuntu, Ubuntu Mate, Linux Mint, Zorin OS Lite. Two of the distributions met the requirements - Lubuntu and Ubuntu Mate. Although Lubuntu has a few percent lower resource consumption (Fig.1.2), I chose to use Ubuntu Mate [3] because of the operating system's graphical shell (Fig.1.1). The user interface to support touchscreen technology was crucial. Ubuntu Mate uses the Gnome Shell – a common GUI shell in the Ubuntu ecosystem. This user interface is built with integrated touchscreen functionality, making it ideal for this purpose. As a Linux distribution, Ubuntu Mate was stable and secure, and in combination of its GUI shell, this made a perfect choice for suitable operating system.



Fig. 1.1: Ubuntu Mate – Gnome GUI shell

CPU: 2 core / 2 threads – 1GHz at least
(1c/2t works fine as well);

RAM: 1GB (2 GB would be better) – works fine even on DDR2 RAM;

Storage: 8 GB – for minimal install
(2 GB after stripping it down to the bare minimum);

Fig. 1.2: Ubuntu Mate – System requirements

2.4 In search of suitable drivers

After analyzing the hardware and choosing an operating system, I came to choosing the appropriate drivers. Fortunately, the minimal installation of Ubuntu Mate, like any other Linux distribution, offers a drivers package suitable for the available hardware devices. The problem at this stage of development was finding suitable drivers for the additionally applied touchscreen matrix on some of the kiosk screens (Fig.2). Because the technology was not built into the screen, but was later upgraded, Ubuntu Mate did not detect the screen as a touchscreen and therefore did not apply a driver to manage the touch-sensitive matrix. This posed the difficult task of finding compatible drivers for each of the matrices, as well as applying and configuring them. Unfortunately, due to the fact that the matrices are of Asian and old origin, finding the right drivers was a real pain, and for some screens no suitable drivers have been found to this day. Lately there has been some progress but the battle still wages on.

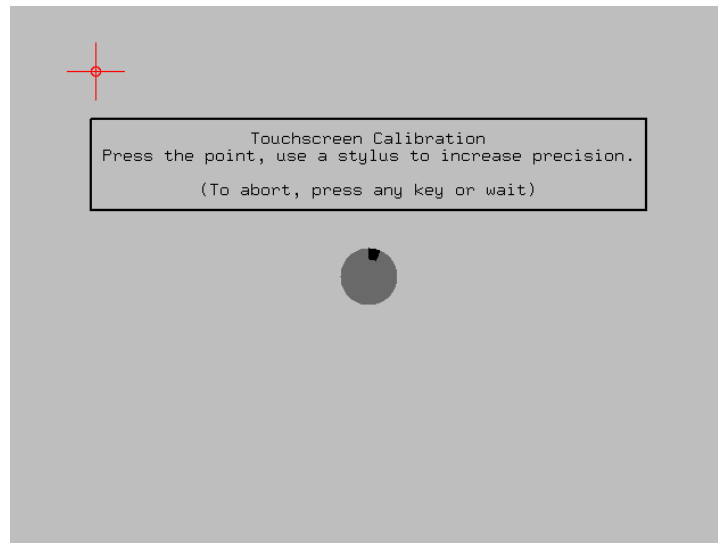


Fig. 2: Ubuntu Mate – built in Linux touchscreen calibrator

2.5 Network OS image – PXE Boot

One of the problems I wanted to solve with the new system is the labor-intensive maintenance of the software (and parts of the hardware) of the kiosk terminals. In the event of a software problem or a desire to upgrade the system, a site visit had to be made, which was not always possible to happen quickly and unnoticeably. I started researching options to eliminate the problem, finding a solution through the PXE Boot network technology. Pre eXecution Environment [4] is a method for booting a client computer(s) using only its network card, therefore the presence of a storage device is not required. This technology allows the device(s) to be started via a signal sent over a local network and loading OS Image, in order to use the centralized distribution of one operating system and software configuration to all other machines on the network (Fig.3). This solves the problem of system maintenance and upgrades by eliminating one hardware component - the available HDD or SSD on the machine, and the use of a centralized OS Image solves the problem of easily distributing the update to the machines. This method works similar to client-server web application.

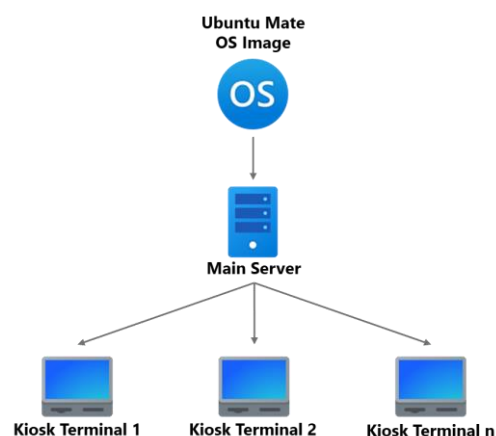


Fig. 3: Summary diagram of OS image loaded from network on each kiosk terminal via PXE Boot

2.6 Web browser choice

Choosing a browser was not difficult task at all. I knew from the beginning that I wanted to use a Chromium-based web browser, because I had worked with its V8 Engine [5] for a long time before and I knew its capabilities well. Google Chrome was chosen because of its stable and trouble-free operation, as well as because of its wide range of browser extensions. The web browser is used to visualize the web application running at the kiosk (Fig.4). Due to the built-in kiosk mode, the integration of the browser for the purposes of the kiosk happened easily and fast. It is running in portable state on the OS image, so the updates of the browser are handled smoothly and unnoticeably.



Fig. 4: Hierarchical diagram of the deployment of the web application

3 Application structure/architecture

The application that I have developed applies the “client-server” architecture. The **Client side** is represented by the kiosk-terminal and its instance of the web application. It can communicate with the **Server side** via a request-response method (handled by the Apache Web Server), or it can send messages in real time, which are handled by the PHP Web Socket Server (Fig.5). This gives the application the opportunity to use full duplex bidirectional communication, which means that the flow of information is in real time. This also means that it is not only the web application that can send information to the server, but also the other way around. The application also utilizes SQL based Database (MariaDB), that holds important information needed by the app to operate correctly. Those three combined, form the **Server side** of the application.

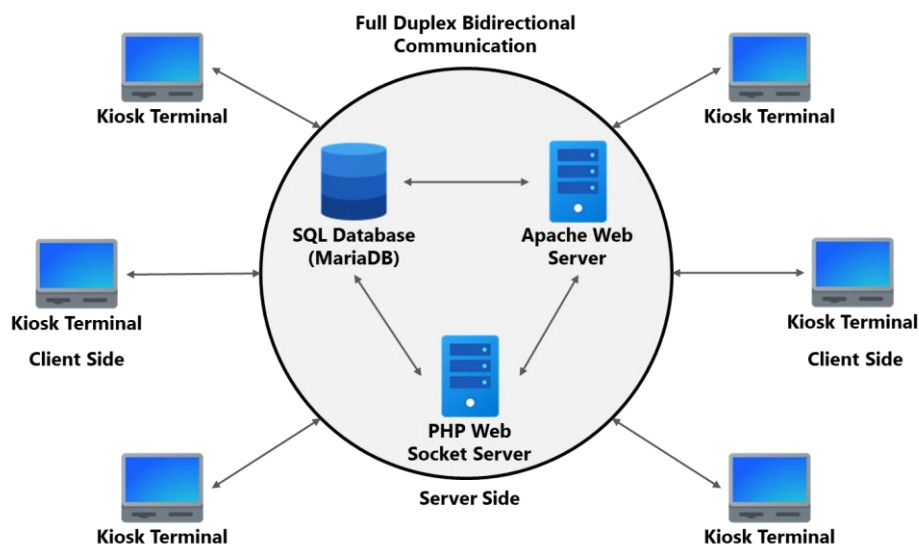


Fig. 5: Diagram showing the full duplex bidirectional communication between server-side and client-side.

The application is developed with the following web technologies:

- **HTML** – Hyper Text Markup language – used for describing the files structure in readable form for the web browser [6];
- **CSS [7]** – Cascading Style Sheets – used for styling the HTML elements. I have used the SASS [8] pre-compiler, so the styling sheet is written in SCSS, but the pre-compiler converts it back to the well-known CSS language, readable by the browser.
- **JavaScript [9]** – used for interaction and implementing functionality for the html elements (for example button clicks). I have used the JQuery [10] library, which makes JS code writing nice and easy.
- **PHP** – PHP: Hypertext Preprocessor is the backend language of choice. It handles the Web Socket Server (Ratchet), and manages backed information processing and security. It handles tasks like user authentication, data manipulation, data visualization etc. The whole CMS backend is also supported by PHP [11].

4 Application views and access layers

4.1 Device configuration

The “Device configuration” view (Fig.6) represents something like installation for the web application. This is one time setup process. When the system administrator starts the kiosk application for the first time on each kiosk terminal, he will be presented with this view, where he can enter a unique name for the kiosk terminal, a password for access and also a description, which is partially generated by the system (IP & MAC address). On completing the information input, the admin can press the button “Configure”, the data will be saved in the database, the window will refresh and present the admin/user with the home screen – the application’s main navigation.

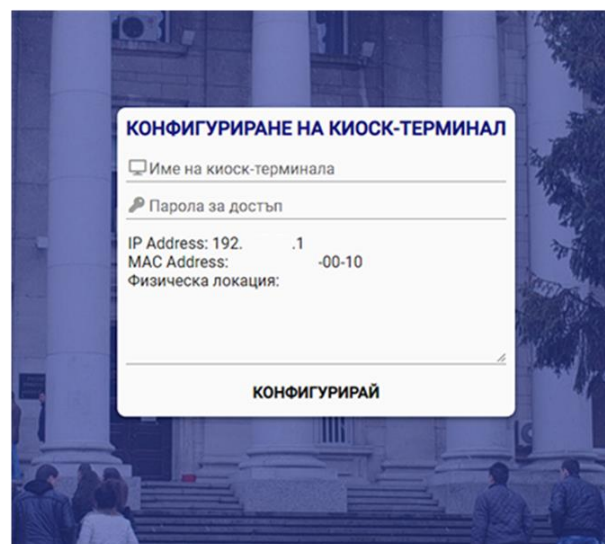


Fig. 6: Kiosk-terminal configurator-view

4.2 Home screen / Application navigation

The home screen (Fig.7) is divided in 4 imaginary parts (top to bottom) – News feed panel, Header panel, Navigation grid, Translate button.

The "News feed panel" is a panel, filled with sliding (right to left) news titles, imported from the university website into this system. Each title is a clickable link, which leads the user to a page, where he can read the full contents of the news article.

The "Header panel" contains the logo of the university, Title and Subtitle, exact time and date, real time weather widget. All of those can be changed and edited via the Content Management System.

The "Navigation grid" contains all important links to local, public or custom pages. Those buttons/links can also be edited via the Content Management System.

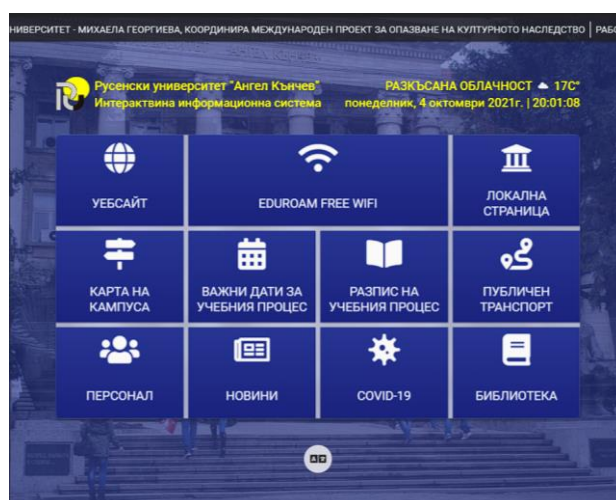


Fig. 7: Kiosk-terminal Home-view / Main navigation

The "Google Translate" button helps the user to translate the application user interface in his/her native language. The google translate widget support more than 80 languages.

4.3 Page navigator

The "Navigator" is a page/view that serves as a container for other web pages, local pages or custom pages, that serve information delivering for the kiosk user (Fig. 8-11). It is divided in two main sections – navigation header panel and page container. The navigation panel has almost the same elements as the header in the home view – university logo, weather widget and exact time and date. It also has three navigation buttons – home, forward and backward. They work the same way as the buttons in the web browser. The home button leads to the home view and the other two move back and forward based on the user history. The other section is used for a container-panel to hold the content of the corresponding page.



Fig. 8: Eduroam Network – connection instructions

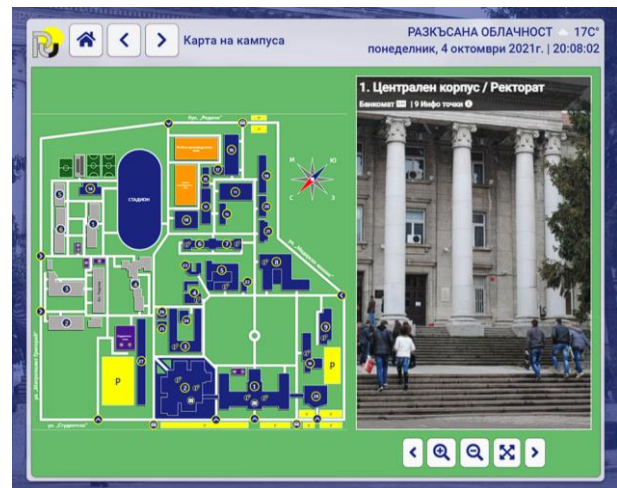


Fig. 9: Interactive campus map

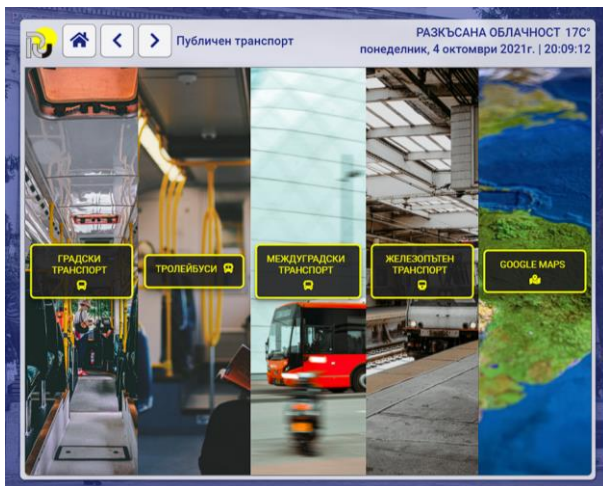


Fig. 10: All kinds of public transport schedules

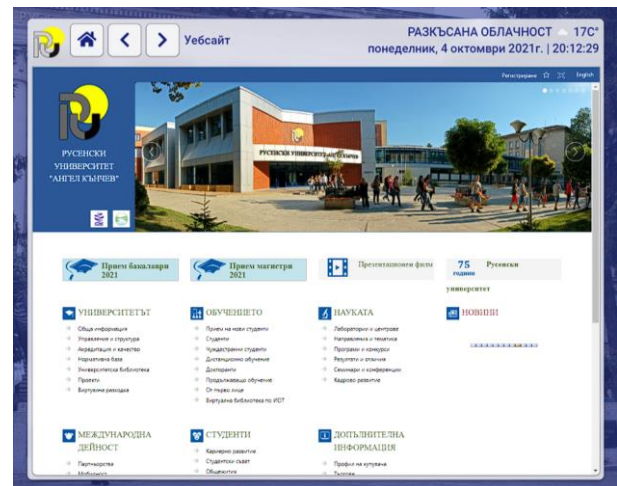


Fig. 11: University website

4.4 Content Management System(CMS) / Administration panel

As you can see so far, this web application has many flexible aspects, so this dynamic information needs some type of management and editing system. In order to achieve this goal, I have created CMS that covers all needed functionality for easy system management and support.

In order to access this CMS, I have created a simple hidden log-in panel (Fig.12), which can be accessed only by users who know about it. Only a limited number of people have access to the system itself, as they have pre-created profiles with given credentials – those are the profiles of my colleagues from ICSC – Multimedia.

The system has 3 layers of access (3 types of users). They are – Administrator, Editor and Ticket center manager.

The administrator has the highest level of access and he can operate all of the menus shown below. The editor has access to menus, only related to content creation and editing. The ticket center manager has only one menu, which is used by only one person, for pptx file upload of an interactive upcoming theatricals preview.

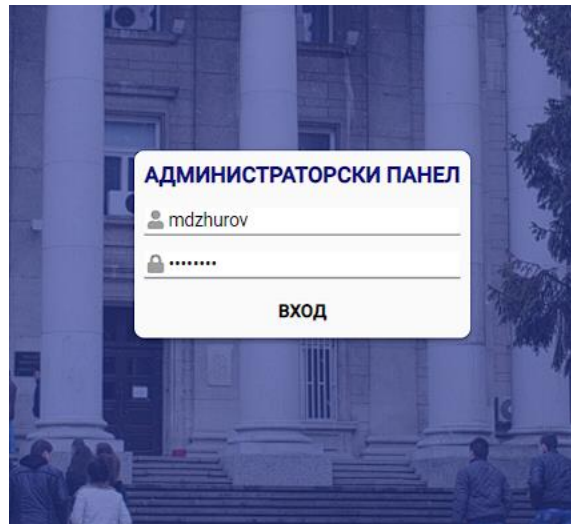


Fig. 12: Admin-panel / CMS log-in form

The administrator profile is created when the application is installed on a certain pc or a server, and it is embedded into the database with default credentials. The administrator is the only type of user that can create, delete or promote other users to higher or lower positions.

The CMS has all necessary menus and tools to manage successfully all of the application aspects. I will describe the system and its menus from the perspective of an administrator, because he has access to all of the menus available:

- **System preview** – here the user can see the final look of the system before publishing it for final use. He can also test all the links and the navigator – it is like having a virtual environment for testing.
- **Active devices** – This is where the administrator can track the active devices in the following network, where the application is deployed. Thanks to the web socket communication between client and server, if there is a problem and any of the terminals goes offline, the administrator will be immediately notified and actions will be taken. The full duplex communication also allows the admin to send messages to all or certain clients and perform tasks like refreshing, restarting or shutting down the device. These are some of the possibilities that such architecture allows, and there will be further implementation of this feature.
- **Home page** – It has three submenus:
 - **News feed** – The control panel of the sliding news titles, which appear on the home view. Here the user can add, edit or delete news articles.
 - **Header** – Here the user can upload/change/delete logos, edit the title and subtitle and also hide/show the weather and exact time and date widgets.

- **Navigation grid** – Here the user can control the navigation grid – rows, columns and grid gap. He can also add/edit/delete/change position of the navigation buttons (Fig.13).
- **Background images** – The background images are programmed to crossfade every n seconds, so here the user can upload or delete the images that appear in the app background (Fig.14).
- **Schedule of the learning process** – There are two panels in this menu – for full-time and part-time degree programmes. Here the user can upload files for the two types, and these files can be reached by the end user through the menus in the home view.
- **Weekly schedule of every degree programme** – Here the user is presented with a table, with all courses and degree programmes. Here he can add/edit/delete or upload a new file to existing record (Fig.15).
- **Users** – This is a menu reachable only by the administrator. Here he can add a new user, edit an existing one or delete.
- **Activity logs** – This is a menu, where the administrator can see a list of activities performed by the users of the system (Fig.16). In case of a problem or system misuse it will be easier for the admin to find the origin of the problem and the responsible user.
- **Console** – Terminal, simulating a console like environment, where the administrator can track messages send by the web socket server. Future feature implementation will also include a text field from which commands will be sent through SSH for a full real time communication between the admin and the server. The generated data is stored on the server in the form of simple .txt files.
- **Server logs** – This is a list, where the administrator can read previously saved server logs. In order to avoid server memory overflow, logs older than 7 days are deleted as they are no longer up-to-date and needed.
- **Ticket centre** – This is the simplest menu of all. It is a menu that has only one button that is used for pptx file upload by user type “Ticket centre manager”.

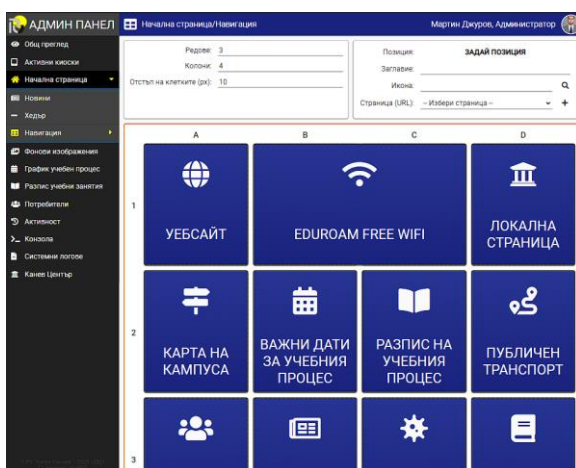


Fig. 13: Navigation buttons editor

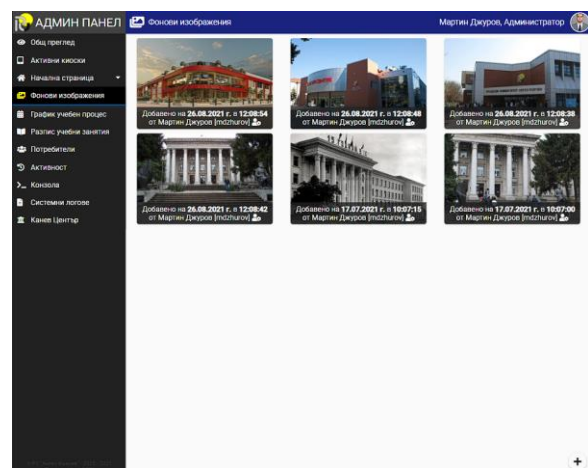


Fig. 14: Background images management

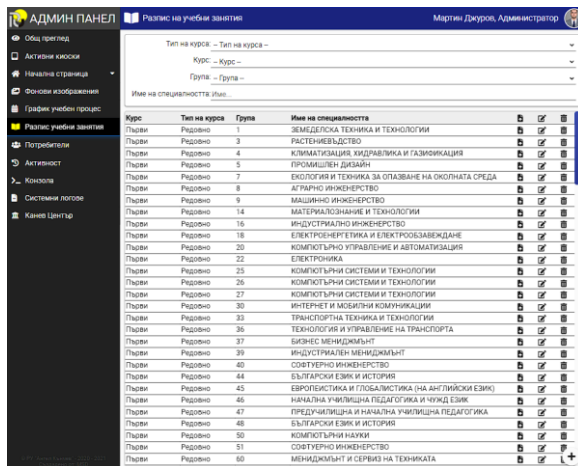


Fig. 15: Learning schedules management

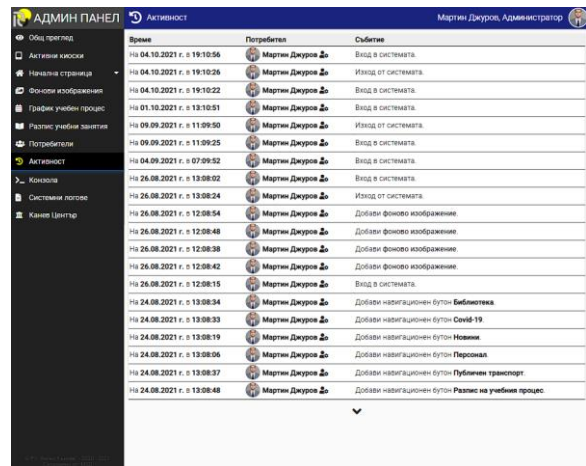


Fig. 16: Activity log

5 Conclusion & Future development plans

This system was developed from ground zero for a little over 6 months by a single student (me) with the goal of creating stable and usable software in order to revive the long-lost glory of the kiosk terminals. Nowadays, it has been serving the university for almost a whole year without any serious issues. Since the system was implemented successfully the uptime of every kiosk terminal has exceeded the uptimes of the old system by months.

The kiosk terminals themselves are “relieved” as their hardware is not stressed and used at its full potential all the time, since the software is lightweight and stable. The cost of hardware support has also dropped severely because of the lowered hardware consumption and the elimination of local storage devices (HDD or SSD) as the OS image is loaded via PXE boot through the network. In the near future, there is a plan to drop the support of conventional PC components and replace them with Raspberry PI 4B devices (Single Board Computer). This SBC will reduce the value per kiosk terminal almost 3 times and will also support the kiosks for at least 10 years in the future. This step is considered, but still under the stage of discussion and further development.

The kiosk web application is a complete system that covers full support – from the OS image, through the kiosk terminal setup and exploitation, to the CMS – a complete system for information editing and management, with the addition of an administrative aspect. The system also supports different type of users, and each type has different properties and access points.

The research I have conducted showed that there was no suitable existing software to cover all of the aspects that were to be covered in the current situation. At least – there was no packaged solution that includes OS and application with suitable specifications.

The system I have devised is not limited to only one purpose and this is deliberate. The reason is that in the future I plan to make it more flexible, with more predefined modules and functionalities, in order to be used easily by other type of institutions – not only schools or universities. This system has a great potential for significant development, and I want to bring it to its limits.

Acknowledgement: The development of this paper was supervised by Assoc. Prof. Galina Atanasova, PhD, Department of Technical and Natural Sciences, University of Rouse “Angel Kanchev”, Bulgaria.

References

- [1] Ivanov, S., & Webster, C. (2017). *Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies – a cost-benefit analysis*. International Scientific Conference “Contemporary tourism – traditions and innovations”, Sofia
- [2] Kallweit, K., Spreer, P., & Toporowski, W. (2014). *Why do customers use self-service information technologies in retail? The mediating effect of perceived service quality*. Journal of Retailing and Consumer Services, 21 (3), 268-276. <https://doi.org/10.1016/j.jretconser.2014.02.002>
- [3] Ubuntu MATE for a retrospective future. (2021). URL: <https://ubuntu-mate.org/>.
- [4] PXE Boot - Preboot Execution Environment. URL: https://en.wikipedia.org/wiki/Preboot_Execution_Environment
- [5] V8 documentation. (2021). URL: <https://v8.dev/docs>
- [6] HTML documentation. URL: <https://devdocs.io/html/>
- [7] CSS documentation. URL: <https://devdocs.io/css/>
- [8] SASS & SCSS documentation. URL: <https://devdocs.io/sass/>
- [9] JavaScript documentation. URL: <https://devdocs.io/javascript/>
- [10] JQuery documentation. URL: <https://devdocs.io/jquery/>
- [11] PHP documentation. URL: <https://www.php.net/manual/en/>

Martin S. Dzhurov
University of Rouse “Angel Kanchev”
Department of Natural Sciences
8 Studentska str., POB 7017, Ruse
BULGARIA
E-mail: martindzhurov2000@gmail.com

Pure Pursuit Controller

Răzvan Gheorghe Filea

Abstract

This work describes the implementation of a Pure Pursuit Controller and the tools used to design it. Pure Pursuit is a tracking algorithm, most often used in robotics, that computes the target velocities based on the error of the robot's current position and some target position. It allows the robot to follow a path quickly, smoothly and accurately using only the vehicle's position. On top of that, the algorithm's flexibility and dynamic nature allow it to efficiently deal with any path deviations.

1 Introduction

The general Pure-Pursuit path tracking algorithm, which first appeared in 1985 in the paper "First Results in Robot Road-Following" [1], reaches its current form only 7 years later when the algorithm was further developed in the paper "Implementation of the pure pursuit path tracking algorithm" [2]. Its goal is to move a robot along a predefined path.

The program is implemented as an Android app, developed in *Android Studio*, in the *Kotlin* programming language. It was developed with the initial intention of being used in the FTC -*FIRST TECH CHALLENGE* - high school robotics competition, thus using originally the FTC competition SDK (Software Development Kit). The goal of the algorithm is to control the movement of an autonomous robot efficiently and robustly, with the robot having the ability to continue after a collision or deviation from the predefined path took place.

This paper is structured as follows. Section 2 presents the idea behind the algorithm. Section 3 describes the computation of the intersection of a circle and segment, while Section 4 describes the theory of moving a robot, with a Mecanum chassis, to a point in the plane. Section 5 describes some optimizations applied in the implementation of the algorithm. Section 6 presents the most important parts of the algorithm in pseudo-code form and Section 7 mentions the added functionalities of the algorithm. Sections 8 and 9 describe the tools developed and used to implement and debug the algorithm. Finally, section 10 highlights the main conclusions and proposes some directions for further development of the application.

2 Algorithm concept

In terms of input and output data, the algorithm receives both the path - represented as a list of points with x and y coordinates - and the current position of the robot. Initially the program validates the input data and then generates a list of segments forming the sections of the path to be traversed Fig. 1.

Given the current position of the robot, the intersection of a circle with the center in the middle of the robot - with a predetermined radius - and the previously generated segments is calculated. Since several intersections may result from this process, the intersection closest to the current destination is always chosen. Using proportional controllers for x, y and rotation, the speeds required for movement are obtained, and the inverse kinematics for the Mecanum wheels convert the planar motion into speeds for the robot wheels.

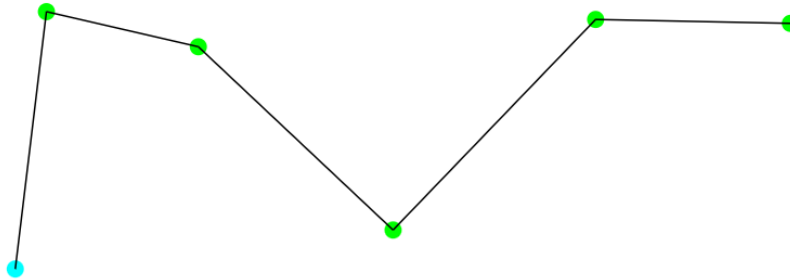


Figure 1: Path divided into segments

2.1 Circle-segment intersections

There have been several attempts to solve this problem effectively. It has been concluded, that it is simpler to find the intersection of the circle with a straight line [3], Fig. 2, and then check if the intersections found fall within the limits of the said segment.

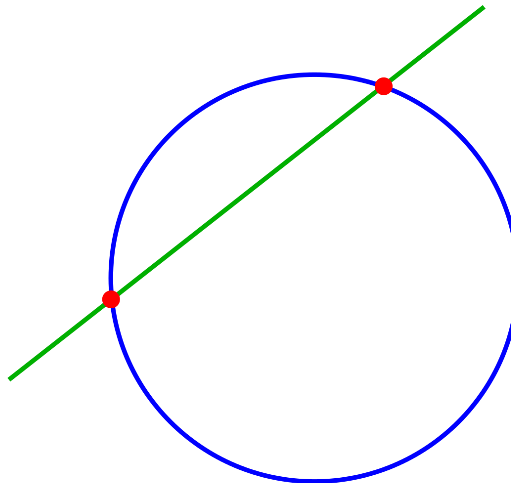


Figure 2: Circle-line intersections

It was also realized that to reduce the complexity of the algorithm, it can be assumed that the center of the circle is at the origin; if not, it is translated there. Thus, we have a circle centered at coordinates $(0, 0)$ and radius r .

$$ax + by + c = 0 \tag{1}$$

Knowing the equation of the line (1) and the coordinates of the two points forming the segment, we calculate a, b and c as in the sequence below:

```
// Line equation using a segment  
a = p2.y - p1.y  
b = p1.x - p2.x  
c = - (a * p1.x + b * p1.y)
```

After this step, a discriminant (2) is calculated, which allows us to distinguish between three different situations from the outset:

- The right does not intersect the circle at all
- The straight line is tangent to the circle, which means that there is only one intersection between the straight line and the circle
- The right cuts the circle into two different sections, so the closest point to the target position must be determined.

$$\Delta = r^2 * (a^2 + b^2) \quad (2)$$

As mentioned at the beginning, the next step of this part is to check if the points are actually inside the segment, because up to this point, a straight-circle intersection has been calculated. To check this, we test whether the sum of the distances between the intersection points and the segment is equal to the length of the segment, while also accounting for floating-point arithmetic imprecision.

If there are two intersection points, it must be determined which is closest to the next target point. To do this we calculate the distance between the intersections and the current target point and choose the intersection that is at the smaller distance. In Fig. 3 this is exemplified, points A and B being the intersection points with the circle and C being the point we want to move to.

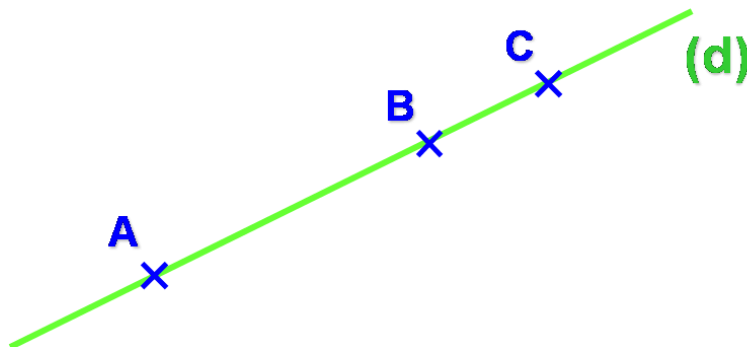


Figure 3: Line with intersections

2.2 Holonomic movement towards a point

Once the point to follow has been found, the robot is moved to this new location in the most efficient way possible. In the first step, the absolute angle of the intersection and the distance between the robot and this point are found. Next, this information is used to calculate the relative x and y plane velocities for the robot using formula (3), where α is the relative angle to the target and d is the distance between the intersection and the robot.

$$\begin{cases} x = \sin(\alpha) * d \\ y = \cos(\alpha) * d \end{cases} \quad (3)$$

The algorithm being developed on a Mecanum wheel robot, the relative x, y and angular velocities are fed into a function that calculates the velocity of each wheel using the inverse kinematics for Mecanum wheels. Of course, this function can later be adapted for other chassis types.

The conversion from the desired plane and angular velocity we obtain to the actual wheel speed is done by finding the inverse kinematics [4] of a 4-wheel Mecanum drive. In Fig. 5 we have the following notations: ω_{1-4} wheel speed, r radius of a Mecanum wheel, v_x horizontal plane speed, v_y vertical plane speed and ω angular velocity of the robot.

$$\begin{cases} \omega_1 = \frac{1}{r}(v_x - v_y - (l_x + l_y)\omega), \\ \omega_2 = \frac{1}{r}(v_x + v_y + (l_x + l_y)\omega), \\ \omega_3 = \frac{1}{r}(v_x + v_y - (l_x + l_y)\omega), \\ \omega_4 = \frac{1}{r}(v_x - v_y + (l_x + l_y)\omega). \end{cases}$$

Figure 4: Path divided into segments

Below is the implementation of the inverse kinematics, as a function that takes as parameters the position of the robot and the point towards which it moves. The *angleWrap()* function normalizes an angle in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$. And the function *holonomicInverseKinematics()* implements the equations in Fig. 4.

```

segment = Segment(robotPos, intersectionToPursue)
absoluteAngleToTarget = segment.slope()
distanceToIntersection = segment.length()

relativeAngleToTarget = angleWrap(robotPos.angle - absoluteAngleToTarget)
relativeX = sin(relativeAngleToTarget) * distanceToIntersection
relativeY = cos(relativeAngleToTarget) * distanceToIntersection

totalPower = abs(relativeAngleToTarget) + abs(relativeX) + abs(relativeY)

if (totalPower < 0.01) return [0, 0, 0, 0]

return holonomicInverseKinematics(
    relativeX,
    relativeY,
    angleWrap(relativeAngleToTarget - angleOffset) / 2
)

```

3 Algorithm optimization

One first obvious way to improve the performance of an algorithm is to reduce its complexity. The algorithm described in this paper has linear complexity, the order of complexity being $O(n)$, where n is the number of points on the predefined path. It was proposed to try to reduce this complexity to constant time.

To reduce the complexity, the index of the current segment on which the robot is located is remembered. In this way, only intersections with the current and next segment can theoretically be considered, although this idea created some unforeseen problems when it was implemented.

The proposed solution was to look not only at the current and next segment but also at the previous one. This way there are no problems when moving to the next waypoint of the route and it fixes another existing problem: if two waypoints were too close to each other, but one was further along the route, the algorithm would simply move to the latter one skipping part of the route.

4 Algorithm implementation

As mentioned in section 2 when initializing the algorithm, it builds a list of segments from the list of points received as input data.

```
// Construct list of segments
segmentsList = []
segmentsList += Segment(initialRobotLocation, waypoints.first().point)

for (i in 0 until waypoints.size - 1)
  segmentsList += Segment(waypoints[i].point, waypoints[i + 1].point)

// Extend the last segment to stop the robot from
// to oscillate in the case of the last segment
expandedRadius = waypoints.last().followRadius * 1.5
segmentsList += segmentsList.last().expandBy(expandedRadius)
```

The algorithm being a closed loop has a main function that is called each time the robot position changes, returning the speed required to reach the destination corresponding to each wheel. The *computeSegmentCircleIntersection()* function used below is the implementation of section 3.

```
if (isFinished()) return [0, 0, 0, 0]

currentWaypoint = waypoints[currentWaypointIdx]
distanceToWaypoint = robotPos.distanceTo(currentWaypoint.point)

// Proceed to the next point after arrival at the current crossing point
if (distanceToWaypoint <= currentWaypoint.triggerRadius)
  currentWaypointIdx++

// Calculate all intersections with all route segments
intersections = pathSegments.flatMap { segment ->
  computeSegmentCircleIntersection(
    segment,
    robotPos,
    currentWaypoint.followRadius
  )
}

// Find the nearest intersection to the crossing point
intersectionToPursue = intersections.minBy { intersection ->
  intersection.distanceTo(currentWaypoint.point)
}

return holonomicMoveTowardsPoint(
  intersectionToPursue,
  robotPos
)
```


5 Further developments

After completing the initial stage of the algorithm, it was decided to implement some features specifically designed for use in the robotics competition.

5.1 Recovery techniques

To solve a potential problem where the robot may lose its way, several recovery techniques have been implemented. While in a perfect system it is impossible for the robot to lose its path, in a robotics competition many unexpected things can happen.

A first method, called *NextWaypoint*: when the robot loses its route, it moves directly to the next waypoint at full speed. This had some limitations and was later replaced by *PursuitNextWaypoint*: which generates a segment between the current position and the waypoint.

Another such recovery technique called *LastKnownIntersection*: always saves the last intersection, found by the algorithm, between the segment and the circle. If the path is lost, the algorithm moves the robot to this saved location.

5.2 Types of waypoints

The algorithm is extremely flexible, allowing many improvements to be added. One of these is to have more than one type of waypoint. Instead of such a point being a simple two-dimensional point in the plane, a lot of other properties can be assigned.

Most of the properties of waypoints allow for advanced movements, for example for the robot to move at a certain angle to the target point. In addition, each waypoint also supports an action. This action, represented in memory as a lambda function, will be executed automatically and asynchronously (on a separate execution thread) upon reaching the waypoint.

The following types of crossing points have been implemented:

- Target-Angle-Waypoint: the robot moves towards the target point at a preset angle
- Reverse-Waypoint: moves the robot backwards towards the waypoint
- Strafe-Waypoint: the robot moves sideways left or right in the direction of the waypoint
- Pivot-Waypoint: moves the robot towards the target while pointing its face in the direction of a pivot point.
- Interrupt-Waypoint: when the waypoint is reached, an action is executed on another execution thread, while the robot stops until the action is completed.

6 Integration tests

The modular architecture, used in the project, allowed for testing of the robot movement and algorithm directly from the computer. These are called integration tests because their purpose is to test as many components as possible and how they behave together. All the tests share a common basis, which is a physical simulation that converts wheel speed into global planar motion. Using this technique, it is easy to write such tests, as we only need to provide a list of the desired path. The integration tests also allowed us to export data and visualize it in specialized programs such as *Microsoft Excel*. This fundamental achievement in testing and debugging the algorithm clarified the next step in the development of the project: building a stand-alone simulation program.

The first attempt to visualize the coordinates transmitted by the integration tests was made in *Microsoft Excel*, the result can be seen in Fig. 5. This was done by exporting all the positions through which the robot moved into a CSV format file, which is then imported into Excel and used to generate a graph.

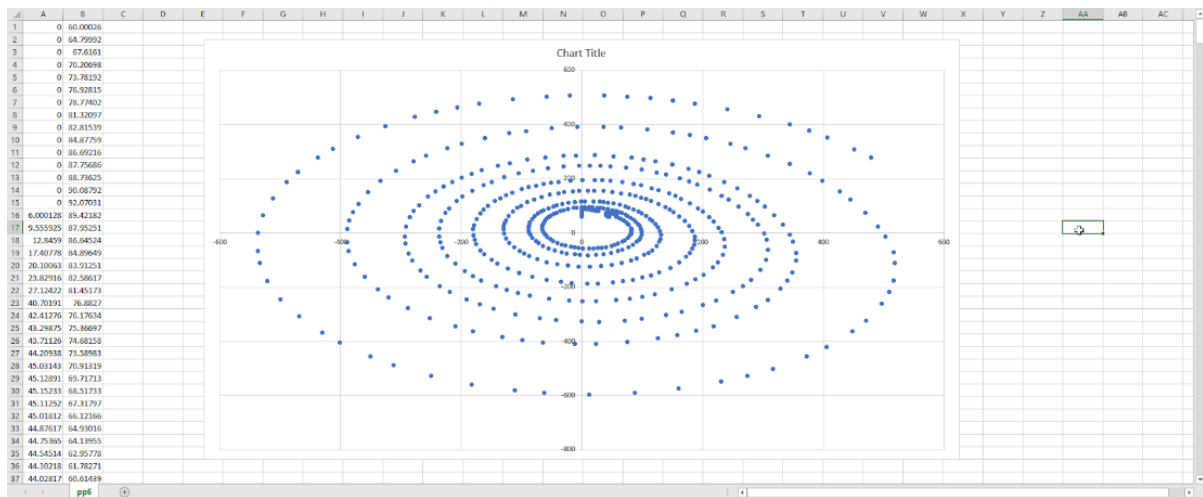


Figure 5: Visualizing a test in Microsoft Excel

7 The Simulator

Realizing the limitations of using integration tests and Microsoft Excel for visualization, it was decided to develop a simulation program alongside the algorithm. The *IntelliJ IDEA Editor* was used for its development and the *Jetpack Compose Desktop Framework* created by *JetBrains* was used to create the interface.

This simulator aims not only to visualize and debug the algorithm but also to help improve the algorithm in the future. Initially the simulator was created only to display the route points and the path taken by the algorithm in real time, visible in Fig. 6. A variation of the code written for the integration tests was used to run the simulation. Later the possibility to create and modify paths from within the program was added (Fig. 7).

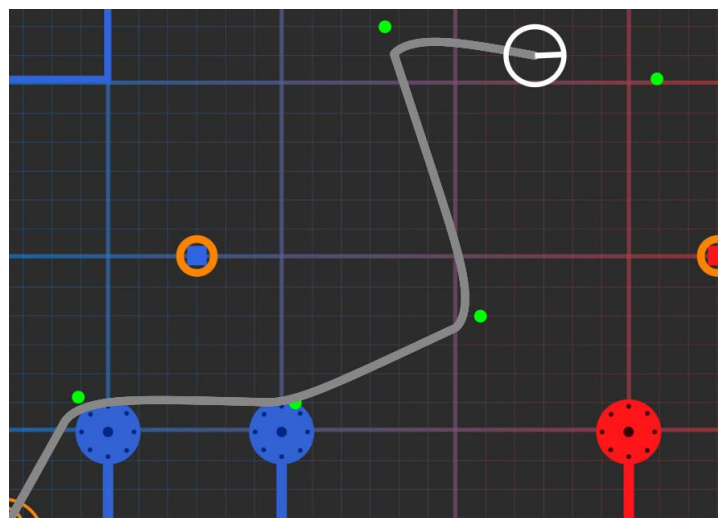


Figure 6: Visualizing a path in the simulator program

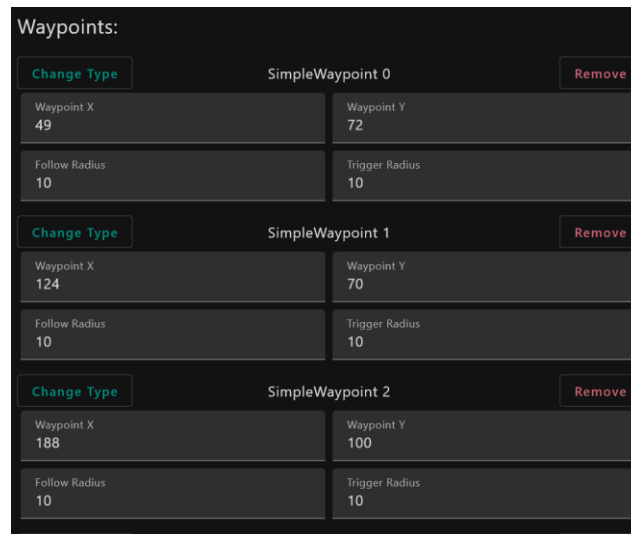


Figure 7: Editing a path from inside the simulator

Together with the algorithm, the simulator has also been progressively improved. Over time, many functionalities have been added such as editing all types of waypoints, specifying the righting technique and many other routes display settings.

In addition, an option has been added that allows to control the speed of the simulation and the possibility to save and edit later a specific trace, which is exported in JSON format using the *KotlinX Serialization* code library.

Another improvement to the simulator that was added later is the consideration of robot inertia. The simulation saves among the state variables not only the current position but also the current velocity. The output of the algorithm is taken by the simulator and treated as acceleration. The accelerations are multiplied by a constant representing the mass of the robot according to the second principle of mechanics. From the resulting value the friction force is subtracted. The consequence of this process is that the robot moves more realistically and can skid in tight curves.

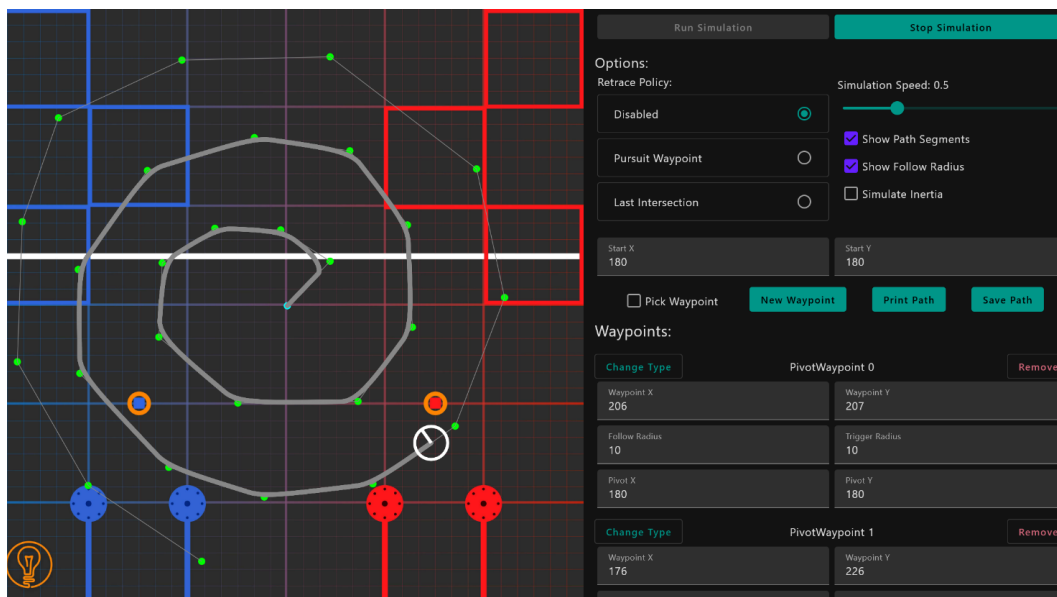


Figure 8: The final version of the simulator program

8 Conclusion

There are undoubtedly a multitude of motion algorithms in the field of robotics, but all of them contain both advantages and disadvantages. Although it does not seem as efficient as generating spline curves and following them, Pure Pursuit is a good choice for almost any kind of robot navigation because the algorithm is by its nature much more dynamic and robust, being able to adapt to any unforeseen challenges encountered by the robot in its path. The algorithm also allows for rapid modification of the algorithm as well as the addition of new functionality to meet specific needs.

This paper implements such an algorithm, which allows a robot to move knowing only its position in the plane. It also describes the addition of specific functionalities such as retracing techniques and multiple waypoint types.

In the future it is intended to implement a sub-algorithm that calculates the radius of the circle used to calculate intersections in real time, depending on the curvature of the tracked route. In contrast to the current implementation which requires the circle radius to be specified for each point of the path from the input data.

Acknowledgement: This work was supervised by Professor *Delilah Florea*, from *National College "Samuel von Brukenthal"*, Sibiu, Romania.

References

- [1] R. Wallace, A. Stentz, C. E. Thorpe, H. Maravec, W. Whittaker, and T. Kanade, "First Results in Robot Road-Following," in IJCAI, 1985, pp. 1089-1095.
- [2] R. C. Coulter, "Implementation of the pure pursuit path tracking algorithm," DTIC Document 1992.
- [3] Circle-Line Intersection, algorithm explanation - <https://cp-algorithms.com/geometry/circle-line-intersection.html>
- [4] Taheri, Hamid, Bing Qiao, and Nurallah Ghaeminezhad. "Kinematic model of a four mecanum wheeled mobile robot." *International journal of computer applications* 113.3 (2015): 6-9.

Filea Răzvan Gheorghe
Babes-Bolyai University
Informatics
Str. Mihail Kogălniceanu, nr. 1, Cluj-Napoca,
ROMANIA
E-mail: razvan.filea@gmail.com

Empirical Study on the Sustainability of Knowledge Transfer on Information Security Awareness Via Virtual Reality Videos Compared to Conventional Videos

Jana Gawrilow, Jule Schumann, Lui Ruck, Patrick Haase

Abstract

In today's information society, security plays a significant role. Increasing importance is being attached to information security and data protection. It is therefore in the interest of companies, but also of private individuals, to demonstrate the best possible Information Security Awareness (ISA). To expand and promote awareness of ISA among individuals, appropriate knowledge transfer is required. Since Virtual Reality (VR) is becoming increasingly important in education, we investigated the learning effect on ISA using informative VR video training. At the same time, this work represents a continuation of "360 Degrees of Security" by Fertig et al. [8]. We determined how a didactic-oriented VR learning video compares to a simple VR video in terms of learning success. Likewise, a comparison of the knowledge transfer with a conventional two-dimensional video is drawn. The evaluation of the learning success was realized by questionnaires, which the test persons filled out in time intervals. The results show that the learning video based on didactic guidelines generally performs better. Based on these findings, we developed guidelines for the optimal creation of a VR learning video.

1 Introduction

Numerous surveys and studies have repeatedly shown that the majority of all information security incidents occur due to mistakes made by employees [1]. Usually, these mistakes happen by accident and are without malicious intent; the employees are simply not sufficiently sensitized to security issues. For this reason, employees with low Information Security Awareness (ISA) levels are currently one of the biggest threats to their own companies [29]. A study conducted in 2019 on industrial espionage and data theft in German companies confirms this assumption [1]. For example, one in four companies reported that they have been victim to social engineering at least once between 2017 and 2019 [1]. This makes social engineering the second most common attack method used by hackers [1]. The trend proves that even the best technical measures to ensure corporate information security are useless if employees remain vulnerable to social engineering. To reduce this vulnerability, employees must also be integrated into the information security process [29].

Sufficiently trained employees can adapt to a changing risk landscape faster than technical and procedural countermeasures [29]. However, to do so, they must be integrated into the organization's security concepts [29]. To achieve this, security awareness programs and ISA trainings must be implemented. However, security threats that exploit human behavior are becoming more sophisticated, which is why ISA trainings must not be exclusively oriented towards new attack

methods but must be fundamentally designed to be of a higher quality to better reach the groups to be targeted [4,24].

VR represents one possibility here. VR offers a fundamentally new way of imparting knowledge. This new way of learning is examined in more detail by Fertig et al. [8]. The authors compare two groups: Subjects in the first group are trained on ISA-compliant behavior traditionally via a two-dimensional (2D) video, while the second group is shown a VR video with the same content. Both groups must then apply what they have learned in various tests at regular intervals. Subsequently, the performance of both groups is compared, and it is shown that subjects who watched the traditional learning video perform significantly better. However, it should be noted here that the 2D video was created according to basic learning didactic principles, while this aspect was neglected in the VR video. This raises the question: are 2D videos per se better for learning ISA content than VR videos? To answer this question, the findings of Fertig et al. [8] were further investigated.

In the beginning, the research questions and hypotheses are defined. Subsequently, basics about ISA, social engineering, VR, and learning theory based on established literature are elaborated. Thereupon, best practices for creating a VR learning video that follows didactic guidelines are presented in the Section "Guidelines". The Section "Methodology" describes the general approach to this work. The approach to the creation of the video is described in the following section "Making the VR Learning Video". The penultimate section, "Conducting the Study," describes both the process of the study and the analysis of the results. Finally, the results of the conducted study are critically discussed and put in relation to the previous work of Fertig et al. [8].

2 Research Questions

The aim of the project work was to find out whether VR videos can be used to convey ISA content in a sustainable manner. To investigate this, learning didactic means were developed and based on this a VR video was shot and edited. The content of this video was designed to convey information on the topic of social engineering in the best possible way. The content of the film is based on the two videos by Fertig et al. [8]. This allows comparisons to be made between the different videos. Finally, we created a study to answer the following research questions:

RQ1. Does a VR video based on learning-didactic guidelines result in a higher knowledge increase about social engineering than a VR video that is not based on those guidelines?

RQ2. Does a VR video based on learning-didactic guidelines achieve a more sustainable learning outcome than a VR video that is not based on those guidelines?

RQ3. Does a VR video based on learning-didactic guidelines result in a higher knowledge increase about social engineering than a traditional 2D video?

RQ4. Does a VR video based on learning-didactic guidelines achieve a more sustainable learning outcome than a traditional 2D video?

3 Related Work

3.1 Information Security Awareness

The term "Information Security Awareness" focuses not only on digital information but also on analogue information. As a result, it is indispensable that the knowledge of the relevant individuals is also included [32]. For this reason, the following issue arises: The issue is not the computer system itself, but how people handle it. If employees are not adequately sensitized, data leaks or hacker attacks can occur.

According to Helisch and Pokoyski [12], ISA is an interplay of cognition, organization, and intention to act. This interaction is a process, which has no clear end. Therefore, the company's goal is to ensure that their employees are constantly educated and that their awareness of data security is increased. However, it must be mentioned that it does not only concern companies, but also private individuals. To achieve this goal, it is important to view people as impressionable beings. Thus, ISA can be seen as a cognitive construct, as it requires thinking, knowing, feeling, perceiving, and acting [32]. This is understood to mean, for example, that individuals need to know that they must keep login data secure. Both their thinking and perception, as well as their feelings and actions, should be guided by the understanding of security. Interpersonal relationships can also lead to confidential information being disclosed to third parties. For this reason, it is important to study human behavior [32]. From this, it can be deduced that negative environmental influences must be averted. In addition, the necessity to recognize certain weak points must be made clear. Here, whether it is a person, or a system is not relevant.

3.2 Social Engineering

To ensure the ISA, we focused on the social engineering forms of phishing and pretexting. Characteristics and methods of how victims can protect themselves from attacks and how perpetrators proceed with their attacks were also shown. That in social engineering it is not the system that is attacked, but the people who operate these computer systems must be emphasized here [3]. Due to certain negative environmental influences, people can be manipulated. Therefore, that human error can cause a system to be hacked can be concluded. Not only the hacking itself is relevant here, but furthermore, the acquisition of internal company and/or sensitive information.

However, different forms of social engineering attacks are distinguished here. One of the best-known and most common methods is the phishing attack. The procedure here is that a perpetrator sends a malicious e-mail to his victim. The goal is to influence and deceive the victim's consciousness. Through this manipulation, the attacker hopes that sensitive or confidential data will be forwarded to him. However, a second purpose can also be a direct attack on the system. This can be done by downloading an attachment or pressing a hyperlink within an e-mail that unknowingly infects the computer system [21]. We mainly focused on pretexting, which is also a social engineering approach. Here, the attacker invents a fictitious scenario to obtain sensitive information [18]. Pretexting consists of four phases: Information gathering, trust-building, manipulation and exploitation, and execution of the damaging procedure [20].

The first phase describes that an attacker gets familiar with the victim. This usually happens via social networks. Through these, the perpetrator manages to gather information about his attack target [22]. In the next step, trust-building takes place. The attacker uses targeted investigative techniques to engage in conversation with his victim and gain his trust [20]. After this step is completed, the attacker will move on to the third phase. Here, he manipulates his target to make him feel the need to help him. He convinces the person to perform a certain action. In this step, the goal of obtaining confidential information is implemented [20]. In the last stage, the social engineer uses the gathered information for his benefit [20]. As mentioned in the beginning, both mentioned social engineering forms and measures were dealt with in this work, which is used for protection. By preventing social engineering, ISA can be provided.

3.3 Virtual Reality as a Learning Environment

In today's world, much more so in the digitalized age, there are a few ways in which learning content can be delivered. There are many innovative methods. One possibility is VR. This transports the learner into a three-dimensional space. Here, the user gets the impression of being in a different place, as various images, sensory impressions, and sounds have an impact on them [27].

Based on this, the teaching of educational content has been taking place since the 1970s [6]. VR can be presented by using VR glasses. The technical content of a topic is conveyed through the didactic means of VR glasses. In this case, interactive knowledge elements are made available to the learner. This learning method is used in various fields because an actual real environment can be simulated. Such an environment is not dangerous for the user and there are no other costs involved [13]. Likewise, it offers the possibility to promote different competencies, such as spatial awareness or critical thinking [2]. The terms "Immersion" and "Presence" often appear in connection with VR. Immersion is understood as the extent of a realistic representation that a technology can depict. This can be measured directly since it depends on objective aspects. Objective aspects can be the display or the software [25]. Not only the scope but also the quality of the sensory impressions that a technology can convey are recorded and considered. Also, the interaction possibilities play a major role [30]. The term "Presence", on the other hand, refers to the subjective feeling of being part of a virtual world. Although the user physically is not [25]. The subjective characteristics of the user, such as emotions, motivations but also cognitive abilities, influence the Presence. In this context, it must be mentioned that immersion also has a strong impact on the user's experienced presence [7]. In summary, these are the inevitable factors that must be successfully implemented in a VR illustration. Through these factors, a foundation for a promising and beneficial learning environment is created.

3.4 Theory of Learning

Learning is a continuous process that accompanies people throughout their lives. In the digitalized age, it is easier to gain knowledge and build up a pool of knowledge. Along with this, education must be constantly reformed. This is especially noticeable in school education, as new research and thus new learning content often emerges within a few years [10]. Thus, learning has a high importance, especially today. Learning describes a process of change, with the result of the change of the behavioral potential. Only with changes in behavior learning becomes measurable. Through this conclusion, the learning process becomes visible [15]. Memory plays a major role in this context, as it is the basis for the learning of content. The memory process includes different aspects, such as the attention and concentration of the learner. Through a chain of processes in the brain, processing, and storage of information finally takes place. This information is stored in either short-term or long-term memory [16].

Since the term learning offers a lot of room for interpretation, several learning theories have been developed. Three of the most popular theories are Behaviorism, Cognitivism, and Constructivism. Behaviorist theory views humans as stimulus-response beings that can be conditioned [10]. In this learning theory, the goal is to change observable human behavior [26]. Constructivism, on the other hand, views learning as, for each person, an individual construct of knowledge [10]. Here, the humans are perceived as an active and self-directed beings [26]. Since cognitive learning theory is often referred to in the context of VR, it was also applied in this research project. Cognitivism focuses on information processing. Human reasoning, decision-making, problem-solving ability, and memory play a significant role in this theory [14]. The learning process in Cognitivism consists of information intake, information processing, and information storage in memory [19]. In this learning theory, cognition and memory research are the focus. Analogies are drawn here to the information processing of a computer [26]. For example, Cognitivism compares short-term memory to the random-access memory and long-term memory to the hard drive of a computer [26]. Within Cognitivism, research was conducted on the extent to which linguistic or pictorial knowledge is stored in memory. In addition, it has been examined how many units of information can be present in working memory at the same time and how the learning of new content, affects this [26]. The learner is considered in Cognitivism as an active part of his learning process [19]. The goal of this theory is to describe knowledge structures, considering cognitivist assumptions, and to optimize learning processes [26]. Most of the current learning research is based on theoretical approaches, but VR can be used to realize the cognitivist learning theory.

4 Guidelines

The aim of the project was to impart as much knowledge as possible to the test persons in the long-term using a VR video that was shot and edited based on didactic guidelines. Thus, we informed ourselves intensively about the use of didactic means in learning videos in VR environment. Our findings and applied methods are explained in the following in the form of guidelines.

Since the viewing direction in VR videos is freely selectable in contrast to classic films, this also increases the risk that the viewer overlooks some details. According to Sylvia Rothe, the extent to which the viewer's attention and viewing direction should be directed when watching VR videos depends on various factors [27]. These include both the type of film and the goal that is to be achieved by the film. Since our video is an educational video, appropriate guidance of the subjects is mandatory. This ensures important information is not overlooked by the viewer. The methods used to guide the viewer are called guiding methods. Some guiding methods are specifically designed for VR movies, while others can be used in classic movies as well. However, these approaches should only be used in VR movies to guide the viewer in the current field of view [27]. The hints used to draw attention to specific regions are called cues. Section 4.1 discusses the guiding methods we used in more detail. Finally, Section 4.2. deals with further didactic means whose use is recommended in learning videos, besides the different guiding methods.

4.1 Guiding Methods

As in classic films, a simple but very effective method is to guide the viewer through the film by letting him follow a story and the associated sounds and movements [27]. This can prevent the viewer from missing important information because it is built into the story. However, in VR videos, it is important to note that the region of interest to the viewer may also be outside of the viewer's field of view. In such cases, it is advisable working with cues that make the viewer change his viewing direction. Auditory cues are particularly useful here, such as a sound or a spoken sentence coming from a different direction. Another useful method, which is similarly used in traditional film production, is to use blurring [27]. This can be used to effectively direct the viewer's line of sight. If only certain areas of the film are in focus and thus clearly visible, the viewer automatically tends to turn his attention to these regions. According to Hata et al. [11], at certain thresholds it is even possible that the viewer does not notice the blur at all, but still turns his attention to the focused areas. If this method is used to direct the focus to superimposed images and to convey information about them, it makes sense to support the whole thing auditorily. This is based on Paivio's Dual Coding Theory, which is explained in more detail in Section 4.2.

Another method is the use of overlays [27]. These include overlays such as arrows or other characters. With these cues, it is important to note that those that are not directly on the target, such as arrows, require a longer processing time. This is because they must be interpreted by the viewer first. These are so-called endogenous cues. Cues that are part of the target, such as a colored border, on the other hand, are found much more quickly and are called exogenous cues. In contrast to exogenous cues, endogenous cues must be interpreted first, but they improve the processing of the event and enable longer-term storage of the information.

The viewer's gaze can also be guided by so-called stylistic rendering [27]. This means using stylistic elements such as depth of field, colors, and brightness to alter certain areas. These variations can be used to direct the viewer's gaze to these regions.

As mentioned earlier, auditory cues play a significant role also, especially in VR environments [27]. In contrast to visual cues, they can also draw the viewer's attention to events outside the field of

view. Various experiments have also shown that viewers find visual cues faster by combining them with an auditory cue.

4.2 Other Didactic Tools for Use in Learning Videos

To support images, both static and moving such as animations or films, auditorily often makes sense. This is due to the Dual Coding Theory by Paivio [23]), which states that people have better memory performance when the information is received both visually and verbally. The underlying assumption here is that there are two different systems for processing information. First, there is the verbal system, which processes linguistic information. The units in which this information is stored are called logogens. On the other hand, there is the non-verbal system, which is responsible for storing pictorial information in the so-called *imagenes*. According to Paivio [23], if both systems are activated simultaneously, this increases the probability that information will be retained. However, it is important to make sure that the audio is also related to the images shown, as thereby their meaning can be clarified.

To include text in videos also often makes sense. According to Langer et al. [17], if text is inserted in the video to convey learning content, some points should be considered. On the one hand, the wording should be as short and simple as possible, and the facts should be described concisely. To bring a clear structure into the text so that the reader keeps track is also important. For the same reason, you should limit yourself to what is necessary, avoid unnecessary repetitions and leave out inessentials. By embellishing the text again and again with various examples or by incorporating content and information into a story, arousing the reader's interest and possibly even involve him or her personally is possible. This is important to motivate the reader and keep his attention during the entire reading process.

5 Methodology

To answer the research questions of the study defined at the beginning, it was necessary to work out a clearly structured procedure first. While planning this procedure, the individual steps to be carried out were evaluated and assigned to four different phases.

5.1 Preparation

The first phase focused on the preparation of essential prior knowledge. Since the research questions refer to an improvement of the learning success of the test persons, it was necessary to research literature and other scientific publications, which deal with the learning theory in the environment of VR learning videos. This resulted in a particular focus on the use of cognitivism to engage subjects more intensively in virtual reality. To produce the VR learning video, it was very important to place multimedia cues that support the subject, such as the deliberate blurring of insignificant gaze directions or the integration of additional overlying image and video elements. From this, technical and stylistic possibilities for the use of multimedia cues were derived and summarized in guidelines. Another preparatory step was the analysis of the previous VR learning video. The learning content to be conveyed was identified and its potential for optimization was examined. Based on the resulting findings, the overlying topics of "phishing" and "pretexting" were established, and particularly relevant points were highlighted. Based on the previous observations and the previously established guidelines, a script was then created that dealt with the topic areas in an illustrative manner with the help of a fictitious scenario. Multimedia cues were planned directly into the respective VR situation so that the learning process of the subsequent test subject is cognitively supported.

5.2 Making the VR Video

The second phase was about making the VR learning video. For this purpose, the created script was first divided into individual coherent sections to plan the recordings sensibly on different days. Particular attention was paid to assigning roles in a way that made sense, so that a smooth flow could be ensured. Following the multimedia guidelines, care was also taken during the filming process to ensure that sufficient opportunities were created for the placement of clues during post-production. After the recording work was completed, the next step was to sift through and classify the material, as the recordings were all filmed several times for quality assurance reasons. Using the video editing program "Adobe Premiere Pro", the sorted footage was then put into chronological order and cuts were made according to the script. In the post-production phase, the VR learning video was then edited according to the multimedia guidelines so that the subject was guided through the topics and cognitively supported in the best possible way, as had been worked out in the preparation phase. After editing and processing of the VR learning video was completed, the result was checked about the requirements arising from the preparation to be able to make any improvements before a final VR learning video was rendered.

5.3 Conducting the Study

In the third phase of processing, a study was prepared to test the subjects about the knowledge imparted. For this purpose, the contents of previous surveys were checked and adopted to be able to make an objective comparison of the results. The prepared study was the same as the study that Fertig et al. [8] used in their work. A copy can be found here (<https://ruck.limesurvey.net/961421>). Twelve test persons were selected to participate in the study and a total of 20 points could be achieved. Appointments were made for all subjects to use the premises and equipment of our university for the demonstration of the VR learning video. At the respective appointments, the subjects began working on the first survey before viewing the VR learning video to be able to observe any prior knowledge. The VR learning video was then presented using a VR headset, followed by the second survey to note direct progress made by the subjects. The other surveys were sent to the test subjects at different intervals of two, four and eight weeks to check the consolidation of the taught topics. The surveys were conducted via the portal "Limesurvey" (<https://www.limesurvey.org/>), which offers various options for exporting the data later.

5.4 Data Analysis

For the evaluation of the data, it was necessary to export the survey results from the portal "Limesurvey" to be able to evaluate them concerning the research questions. For this purpose, the anonymous data was analyzed using various formulas to be able to make statements about the values.

6 The VR Video

6.1 Structure and Theme of the Video

To be able to make an objective comparison of the study results, it was necessary to design the content to be conveyed analogously to the previous VR learning videos. As already described in Section 5, this content was accordingly implemented with the help of a detailed and didactically prepared script. The goal was to create added value through the way in which knowledge was conveyed, rather than changing the content itself. For this reason, the entire script was structured in such a way that a fictitious company is portrayed by the test person switching into the viewpoint of two different employees. During this representation, various memories of these employees are lived through, which are intended to clarify the events within the company. Every scene tries to explain different aspects of information security

awareness that are needed to complete the questionnaire. A link to the video can be found here (https://youtu.be/zbE_abTSrCA, <https://youtu.be/0VxyCAjOOMk>, <https://youtu.be/tpImCIJQZsc>).

6.2 Post-Production and Used Effects

To involve the subjects more in the VR learning video on a didactic level as well, appropriate multimedia cues were used during post-processing, among other things. Here, particular attention was paid to adherence to the previously elaborated multimedia guidelines. To make both the fade-in and fade-out of individual scenes vivid, immersive transition effects were used. These prevent abrupt changes that could negatively affect the subjects' VR experience.

A VR video converts the entire content of the visible panorama to a spherical layer, making it perceived as a virtual reality [5]. Therefore, to make immersive content more interactive, it is recommended to display it using multiple layers [5] to be able to target subjects' attention. In the presentation of the different memories, multiple layers were used to place multimedia cues. These consisted of externally inserted and spherically modified two-dimensional content. Using this technique, the presentation of "phishing" was supported by directing the subjects' attention to the computer on which various contents were displayed up to the virus warning that occurred. Likewise, multiple layers were applied to show the screens of various other devices enlarged and highlighted. For this purpose, a two-dimensional image was again converted into a spherical form so that it could be viewed as a superordinate layer in virtual space. Both the representation of a "phishing e-mail" and the visualization of the first step of the "pretexting" process, "obtaining information", was illustrated in this way. By using these multiple layers, subjects become more involved in the learning content [5]. Another method used in post-processing was the blurring of irrelevant viewing layers. Here, research revealed that viewers tend to view regions with little or no blurred content when other areas are in a blurred state [28]. When this effectuation occurred, explanations in audio format were added in addition to the visual overlay to educate subjects about the content shown. In addition, to provide subjects with guidance during the scene changes, multimedia cues in the form of text areas were spherically inserted. These contain information about the respective following scene, such as the designation of the current step of the "pretexting", or also temporal information. The texts were also inserted into the sphere of the VR in the form of a multiple layers. To increase the focus on these elements, the other layers were either blacked out or blurred during the display. Additionally, multimedia cues in the form of prominent red circles were included to highlight specific objects within the VR learning video. While diegetic cues are based on elements that belong to the narrative environment [28], such encirclements are non-diegetic cues. In the post-processing of the VR learning video, these were applied to highlight the suspicious USB stick, as non-diegetic cues can be applied to guide subjects to an important detail [28].

7 The Study

The study was conducted with a group consisting of a total of twelve unpaid subjects. In the following, this group will be referred to as VR2. The study results of the VR2 group will be compared with the results of the groups non-VR and VR from the study by Fertig et al. [8]. The conduct and procedure of the study were similar for all three groups; for reasons of clarity, only the procedure of the VR2 study will be described in more detail below. The detailed procedure of the other two studies is explained in detail in "360 Degrees of Security" by Fertig et al. [8].

The VR2 group received the VR video created by the authors of this paper with special attention to didactic learning correctness. To achieve this level of correctness, a lot of work went into researching how humans learn and implementing the results into our video, which is explained in section 6.2. In this paper, a video without that research is called standard or traditional. The participants completed a total of five tests. The first one is before the viewing of the respective video. The second immediately following the viewing of the video. Tests 3, 4, and 5 took place two,

four, and eight weeks, respectively, after viewing the video. These periods were chosen to be able to determine a possible decrease or increase in performance. While physical presence was not required for the last three tests, participants were required to be present for the first two tests and the viewing of the video. In the beginning, the procedure of the study was explained to the group, making them aware that the video would only be viewed once. Then, the structure of the test sheet was explained. Any questions that arose should be asked directly to avoid ambiguity. Participants were told in the introduction that it was okay if they were unable to answer the test questions during the first processing. This was to allow a direct comparison of the knowledge level before and after viewing the learning material so that the exact knowledge acquisition could be determined. After completing the first questionnaire, the subjects took a seat on a chair and were familiarized with the use of the HTC Vive Pro VR goggles. This was followed by viewing the video, and then the second test. Instead of the socio-demographic question section, this now included a section on the virtual reality experience. For the last three tests, participants were simply sent links to the surveys so that the tests could be completed remotely.

7.1 Study Population and Sampling

As stated earlier, the study was conducted with twelve participants ($N = 12$). Nine subjects (75%) were between 18 and 25 years old, two (16.67%) between 26 and 35, and one participant (8.33%) between 51 and 65.

Five participants (41.67%) were in permanent employment at the time of the study, and five others (41.67%) were in a degree program or other full-time continuing education program. Two participants (16.67%) indicated that neither applied to them.

The following results emerged from the results of the socio-demographic questions: the average participant in the VR2 group is between 18 and 25 years old, uses a computer every day and his smartphone or tablet for several hours a day. He has nothing to do with information security issues in either his private or professional life and does not know what status information security has in his company's corporate policy. Nevertheless, he considers the efforts of his employer or his educational institution to actively promote information security to be rather endeavouring. The average participant rates his knowledge of information security as rather less good.

7.2 Study Results and Analysis

Since the performed Kolmogorov-Smirnov tests [9] indicated that the data were normally distributed, parametric tests were applied. To test whether subjects' test scores differed significantly before participation in the session and after, the paired t-test was applied within each group [9]. To compare follow-up tests with their preceding tests to measure retention performance over the eight-week period, the paired t-test was also used [9]. A comparison of the two scores of the no-VR and VR groups was made possible by the unpaired t-test [9]. All tests performed were two-sided tests with significance level of $p \leq 0.05$. Each result provided:

- the significance level p .
- the t value: t can be negative or positive. Negative means mean one is less than mean two. A positive value means the opposite.
- the h -value: h indicates whether the null hypothesis of equal medians can be rejected. Here, $h = 1$ indicates that the null hypothesis of equal medians can be rejected and $h = 0$ indicates that it cannot be rejected.

In the first test, participants in the VR2 group scored an average of 4.125 points ($\sigma = 2.112$), and in the second, 17 ($\sigma = 1.595$). A significant difference in scores in the first and second runs of the tests was demonstrated by paired t-test with $p \leq 0.05$ (cf. Table 1). As also evidenced by the t -value, the scores were significantly lower in the first run.

Table 1: Results of the paired t-test of the first and second test of group VR2

Group	Paired t-test		
VR2	p = 0,0000000001	h = 1	t = -22,945

In the third run, the average score was 16.333 ($\sigma = 1.435$), in the fourth 16.292 ($\sigma = 1.602$), and in the last 14.958 ($\sigma = 1.616$).

On average, subjects' performance did not change in the second and third runs. However, the t-value shows a slight decrease in the median.

Table 2: Results of the paired t-test of the second and third test of group VR2

Group	Paired t-test		
VR2	p = 0,071	h = 0	t = 2

Also, from the third to the fourth run, the performance of the subjects did not change seriously. Again, the t-value shows a slight drop in the median.

Table 3: Results of the paired t-test of the third and fourth test of group VR2

Group	Paired t-test		
VR2	p = 0,820	h = 0	t = 0,233

In the last test, a clear drop in performance was evident. The median drops and the paired t-test confirms the observations. The t-value also shows a lower mean in test 5.

Table 4: Results of the paired t-test of the fourth and fifth test of group VR2

Group	Paired t-test		
VR2	p = 0,0002	h = 1	t = 5,360

Comparing the results of the second test among the three groups, it is clear from the paired t-tests with $p \leq 0.05$ that VR2 performs measurably better than VR. This is also shown by the t-value. No large difference is evident between VR and no-VR, as well as between VR2 and no-VR. Overall, however, VR performs the worst, while VR2 performs the best.

Table 5: Comparison of the results of the second test of all groups with each other

Group	Paired t-test		
VR and VR2	p = 0,027	h = 1	t = 2,376
VR2 and no-VR	p = 0,332	h = 0	t = 0,993
no-VR und VR	p = 0,252	h = 0	t = 1,178

In the third test, the results of the three groups converge somewhat. There is no longer a measurable difference, as can be seen from the t-values and the p-values. VR2 still performs best, while VR ranks last.

Table 6: Comparison of the results of the third test of all groups with each other

Group	Paired t-test		
VR and VR2	p = 0,211	h = 0	t = 1,296
VR2 and no-VR	p = 0,693	h = 0	t = 0,400
no-VR and VR	p = 0,369	h = 0	t = 0,919

In the fourth test, the results continue to converge. The t-values become even smaller, proving that there is no significant difference. The rank distribution remains the same: VR2 performs best, followed by no-VR and VR in last place.

Table 7: Comparison of the results of the fourth test of all groups with each other

Group	Paired t-test		
VR and VR2	p = 0,414	h = 0	t = 0,832
VR2 and no-VR	p = 0,509	h = 0	t = 0,673
no-VR and VR	p = 0,956	h = 0	t = 0,056

In the last test, a slightly different picture emerges. Between no-VR and VR a significant difference is provable using the paired t-test. The t-value also confirms this. The ranking changes in favor of no-VR. This group now occupies first place, followed by VR2 and VR.

Table 8: Comparison of the results of the fifth test of all groups with each other

Group	Paired t-test		
VR and VR2	p = 0,128	h = 0	t = 1,580
VR2 and no-VR	p = 0,444	h = 0	t = -0,780
no-VR and VR	p = 0,036	h = 1	t = 2,239

From the findings obtained, subjects trained by a learning-didactically correct VR video perform better than subjects trained by a simple VR video or by conventional 2D videos.

8 Discussion

In the following section, all mentioned research questions (cf. Section 2) are discussed and finally answered based on the available results of the study. Research question RQ1 can be answered in the affirmative. Since the participants of both studies had no prior knowledge regarding the topic of social engineering, comparable, correspondingly low scores were obtained by both groups in the first survey. Both group VR and group VR2 showed a significant improvement in the second test, after viewing the video. However, participants in the VR2 group reached an even higher number of points overall. Direct comparison to the first survey also reveals that subjects who had viewed the learning didactic video improved more: group VR increased by an average of 11.67 points from the first to the second test, while VR2 was able to show an average increase of 12.88 points. Research question RQ2 can also be answered in the affirmative. First, the further course of the two studies must be considered. They developed very differently. In group VR2, the forgetting curve designed by Dr. Ebbinghaus can be seen clearly [33]. It describes how much information a person has learned is forgotten again over time and what percentage of the knowledge is retained. As early as the third survey of the study conducted, a drop

in performance can be seen, as also depicted by the forgetting curve. This continues during the subsequent surveys and increases again significantly in the last survey. In the VR group, on the other hand, an improvement can even be observed between the second and fourth test. This initially surprising phenomenon was also observed in the work of Ulsamer et al. [31]. This course of the two studies initially suggests that the VR group even achieved better learning results in the long term. However, it is now important to note that despite an increasing decline in performance, the VR group, nevertheless, achieved a higher score on each of the Tests 2 to 5 than the participants in the VR group. Thus, sustained better performance can be observed with the didactically correct video and research question RQ2 can be answered with yes. Question RQ3 can also be answered with yes. This can be explained primarily by the fact that group VR2 achieved higher average scores on the second test after viewing the video than group non-VR. It is important to note that the participants in group VR2 achieved these better results even though, unlike the non-VR group, they had no prior knowledge of social engineering and had nothing to do with information security in their daily lives. Also, the subjects who had seen the VR video increased their scores by an average of 12.88 points between Test 1 and Test 2, while the non-VR group only saw an increase of 8.79. However, what somewhat mitigates the significance of this fact is that the non-VR group already achieved significantly better results in Test 1 due to their prior knowledge. Thus, they had less opportunity to improve further than subjects in the VR study. Research question RQ4 cannot be answered unambiguously. In general, both groups showed a drop in performance over time from Test 2 to Test 5 according to Dr. Ebbinghaus' forgetting curve [33]. In test 3 and test 4, the subjects of group VR2 each performed slightly better on average, but in test 5, the participants of the non-VR group achieved better results. If we consider this again under the circumstance of the lack of prior knowledge of the participants of group VR2, we can assume that the didactically correct VR video nevertheless achieved a more sustainable learning result than the conventional learning video. However, this statement cannot be proven clearly.

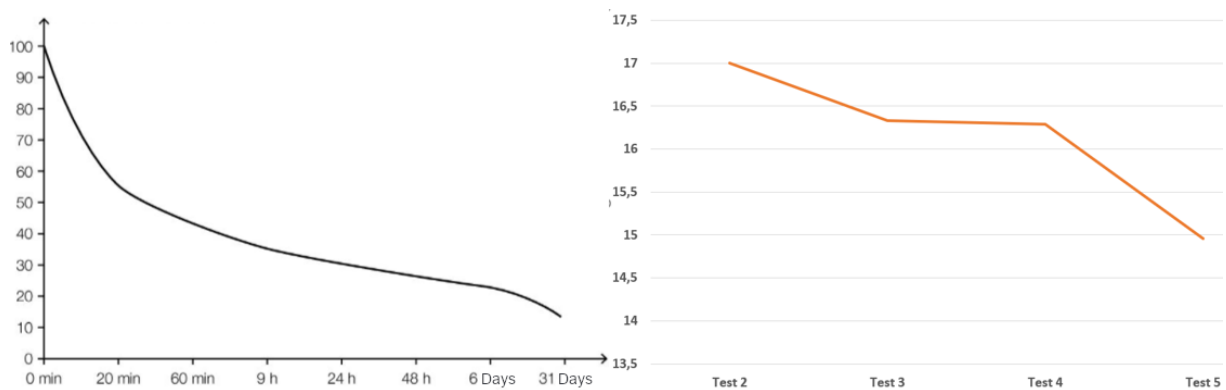


Figure 1: Comparing the Ebbinghaus Curve on the left to the results of group VR2 on the right

9 Conclusion

After conducting the study, a significant improvement in Information Security Awareness (ISA) can be observed among the test subjects when VR is used for the training videos. Both the test subjects who were trained using a conventional VR training video and the test subjects who were trained using a didactically correct VR training video achieved better learning results in the long term than with a training video without VR, thus confirming the results of previous studies on this topic. In the context of this study, the different learning developments between a group with a conventional VR learning video and a didactically prepared VR learning video were compared. Initially, similar positive

developments were found, although the use of the didactically correctly prepared VR learning video was able to achieve even better results on average in the surveys. Even if in both test groups a similarly beginning forgetting curve could be registered, the obtained scores in the surveys speak nevertheless for an improvement of the results with use of a didactically correctly structured VR learning video.

To be able to reproduce the production of such a didactically correctly constructed VR learning video, multimedia guidelines (cf. Section 4) were elaborated in the context of this study, with the help of which qualitative specifications for future VR learning videos of this kind can be adhered to. Here, special attention was paid to the correct elaboration of didactic means, such as the use of cognitivism, which has already been scientifically presented in Section 3.4. To be able to register further progress in this area of the learning theory, future studies should therefore deal with the counteracting of the forgetting curve, which occurred both with the subjects of the conventional VR learning video, as well as with those of the learning didactically correctly developed VR learning video and which negatively influenced the results.

In summary, the results of this study show that VR can offer significant added value for use in a training environment and that a didactically correct presentation contributes positively to the results of the test persons. In the future, this may become more important, especially for companies that want to educate their employees about various complex topics with a didactically developed concept.

Acknowledgement: This work was supervised by Research Assistant *Tobias Fertig*, from the *University of Applied Sciences Würzburg-Schweinfurt*.

References

- [1] Berg, A. & Niemeier, M. (2019). *Wirtschaftsschutz in der digitalen Welt*. Statista, Berlin. URL: <https://de.statista.com/statistik/daten/studie/164300/umfrage/arten-von-computerkriminalitaet-in-unternehmen-in-deutschland/> (visited on September 11, 2021).
- [2] Buchner, J. & Aretz, D. (2020). "Lernen mit immersiver Virtual Reality: Didaktisches Design und Lessons Learned", *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 17 (Jahrbuch Medienpädagogik)*, 195-216.
- [3] Bullée, J.W. & Junger, M. (2020). "Social Engineering", in: Holt, T. & Bossler, A. (eds.) *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Cham: Palgrave Macmillan.
- [4] Bundesamt für Sicherheit in der Informationstechnik (2012). *Leitfaden Informationssicherheit IT-Grundschutz kompakt*.
- [5] Choi, K., Yoon, Y., Song, O. & Choi, S. (2019). "Interactive and Immersive Learning Using 360° Virtual Reality Contents on Mobile Platforms", *Mobile Information Systems, vol. 2018*.
- [6] Daniela, L. (2020). *New Perspectives on Virtual and Augmented Reality: Finding New Ways to Teach in a Transformed Learning Environment*. Abingdon: Routledge.
- [7] Dengel, A. & Maegdefrau, J. (2019). "Presence Is the Key to Understanding Immersive Learning", in: Beck, D., Peña-Rios, A., Ogle, T., Economou, D., Mentzelopoulos, M., Morgado, L., Eckhardt, C., Pirker, J., Koitz-Hristov, R., Richter, J., Gütl, C. & Gardner, M. (eds.) *Immersive Learning Research Network*. Cham: Springer.
- [8] Fertig, T., Henkelmann, D. & Schütz, A. E. (2022). *360 Degrees of Security: Can VR Increase the Sustainability of ISA Trainings?*
- [9] Field, A. P. (2009). *Discovering statistics using SPSS: and sex, drugs and rock "n" roll*, 3rd Edition. Los Angeles: SAGE Publications.
- [10] Franken, S. (2019). *Verhaltensorientierte Führung: Handeln, Lernen und Diversity in Unternehmen*, 4th Edition. Wiesbaden: Springer Gabler.
- [11] Hata, H., Koike, H., Sato, Y. (2016). "Visual Guidance with Unnoticed Blur Effect", in: Tortora, G., Levialdi, S., Tucci, M. (eds.) *Proceedings of the International Working Conference on Advanced Visual Interfaces*. New York: Association for Computing Machinery.
- [12] Helisch, M. & Pokoyski, D. (2009). *Security Awareness: Neue Wege zur erfolgreichen Mitarbeiter-Sensibilisierung*, 1st Edition. Wiesbaden: Vieweg + Teubner.

- [13] Hellriegel, J. & Cubela, D. (2018). “Das Potenzial von Virtual Reality für den schulischen Unterricht - Eine konstruktivische Sicht”, *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 58-80.
- [14] Jadin, T. (2018). “Medienpsychologie Psychologische Aspekte des Lernens mit digitalen Medien”, *Psychologie in Österreich* 5, 368-374.
- [15] Kiesel, A. & Koch, I. (2012). *Lernen: Grundlagen der Lernpsychologie*, 1st Edition. Wiesbaden: VS Verlag für Sozialwissenschaften.
- [16] Kullmann, H. M. & Seidel, E. (2005). “Lernen und Gedächtnis im Erwachsenenalter”, *Perspektive Praxis*. Bielefeld: W. Bertelsmann Verlag.
- [17] Langer, I., Schulz von Thun, F., Tausch, R. (2011) *Sich verständlich ausdrücken*, 9th Edition. München: Ernst Reinhardt Verlag.
- [18] Lohani, S. (2018). “Social Engineering: Hacking into Humans”, *International Journal of Advanced Studies of Scientific Research*.
- [19] Meir, S. (2016). *Didaktischer Hintergrund Lerntheorien*. URL: https://lehrerfortbildung-bw.de/st_digital/elearning/moodle/praxis/einfuehrung/material/2_meir_9-19.pdf (visited on August 15, 2021).
- [20] Mitnick, K. D. & Simon, W. L. (2002). *The Art of Deception: Controlling the Human Element of Security*. New York: Wiley & Sons.
- [21] Nikolov, L. G., Dimov, R. & Zlatinov, B. N. (2020). “Malware in Social Engineering”, *CONFSEC*. Borovetz, Bulgarien.
- [22] Ozkaya, E. (2018). *Learn Social Engineering: Learn the art of human hacking with internationally renowned experts*. Birmingham: Packt Publishing Ltd.
- [23] Paivio, A. (1990) *Mental Representations: a Dual Coding Approach*, *Oxford Psychology series*. New York: Oxford University Press.
- [24] Prantner, K. (2009). “Awareness und Lernen”, in: Helisch, M. & Pokoyski, D. (eds.), *Security Awareness: neue Wege zur erfolgreichen Mitarbeiter-Sensibilisierung*. Wiesbaden: Vieweg + Teubner, 29-55.
- [25] Rau, V., Niemöller, D. & Berkemeier, L. (2019). “Ist weniger mehr? - Designprinzipien für Virtual Reality Training aus kognitionspsychologischer Sicht”, *HMD* 56, 809-822.
- [26] Reinders, H., Ditton, H., Gräsel, C. & Gniewosz, B. (2015). *Empirische Bildungsforschung: Gegenstandsbereiche*, 2nd Edition. Wiesbaden: Springer Fachmedien Wiesbaden.
- [27] Rothe, S. (2020). *Konzepte und Guidelines für Applikationen in Cinematic Virtual Reality*. Dissertation, Ludwig-Maximilians-Universität München.
- [28] Rothe, S., Buschek, D. & Hußmann, H. (2019). “Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges”, *Multimodal Technologies and Interaction 2019*.
- [29] Schütz, A. (2018). “Information Security Awareness: It’s Time to Change Minds!”, *International Conference on Applied Informatics ICDD*. Sibiu, Rumänien.
- [30] Steuer, J. (2006). “Defining Virtual Reality: Dimensions Determining Telepresence”, *Journal of Communication* 42(4), 73-93.
- [31] Ulsamer, P., Schütz, A., Fertig, T. & Keller, L. (2021). “Immersive Storytelling for Information Security Awareness Training in Virtual Reality”, *Hawaii International Conference on System Sciences*.
- [32] Weber, K., Schütz, A. E. & Fertig, T. (2019). *Grundlagen und Anwendung von Information Security Awareness: Mitarbeiter zielgerichtet für Informationssicherheit sensibilisieren*. Wiesbaden: Springer Fachmedien Wiesbaden.
- [33] Wendt, M. & Becker-Carus, C. (2017). “Methoden und Befunde der frühen experimentellen Gedächtnisforschung”, *Allgemeine Psychologie*, 2nd Edition. Heidelberg: Springer.

Jana Gawrilow
University of Applied Sciences
Würzburg-Schweinfurt
Faculty of Computer Science
and Business
Information Systems
Sanderheinrichsleitenweg 20,
97074 Würzburg
GERMANY
E-mail:
jana.gawrilow@student.fhws.de

Jule Schumann
University of Applied Sciences
Würzburg-Schweinfurt
Faculty of Computer Science and
Business Information
Systems
Sanderheinrichsleitenweg 20,
97074 Würzburg
GERMANY
E-mail:
jule.schumann@student.fhws.de

Lui Ruck
University of Applied Sciences
Würzburg-Schweinfurt
Faculty of Computer Science
and Business
Information Systems
Sanderheinrichsleitenweg 20,
97074 Würzburg
GERMANY
E-mail:
lui.ruck@student.fhws.de

Patrick Haase
University of Applied Sciences
Würzburg-Schweinfurt
Faculty of Computer Science and
Business Information Systems
Sanderheinrichsleitenweg 20, 97074
Würzburg
GERMANY
Email: patrick.haase@student.fhws.de

Digitization in medical teaching

Susanna Götz

Abstract

From the research on digitalization in medical education, an application emerges that is intended to make it easier for prospective physicians to get started with infarct localization. The Easy ECG virtual reality app was developed for this purpose; it shows the correlations between ECG signals and the anatomically correct heart. The user can display different infarcts and thus learn about differences and similarities between the infarcts with the Oculus Rift. It is also possible to plot the conduction of impulses to the heart, which is closely related to the localization of the infarct. This helps to understand infarcts and their effects on the heart as a function of time. Because Easy ECG can provide an immersive and accessible learning experience for medical students, the app is an impetus to bring a breath of fresh air and change to conservative medical teaching through digitization. The virtual reality application is written in C# and runs on the Oculus Rift with the Unity game engine version 2018.4.36f1 (LTS).

1 Motivation

Books for medical students are complicated and often difficult to read. For example: "Excitation onset. If the neighboring cell of a cell that is still at rest is electrically excited, this leads to small charge shifts between the surfaces of the two cells and, via the gap junctions, also between the cell interiors, because charge changes have occurred at the excited cell as a result of the action potential [1]." Sentences like this are not uncommon in introductory books for future doctors. Often one has to read them several times to fully understand their content. This is a particular problem in infarct localization, where a large amount of information has to be linked mentally so that abnormalities in cardiac activity can be interpreted correctly. To overcome this problem, we decided to create an application that graphically depicts what could previously only be understood mentally.

In an interview with students [2] and doctors [3] about the visualization of medical data, the ECG signal came up again and again. It was confirmed several times that many doctors have problems with the evaluation of ECG signals. Serious errors occur in 4 to 33 % of the cases. This is partly due to the fact that cardiologists detect abnormalities in the ECG only 53 to 96 % of the time. Medical students still only recognize 17 to 63 % of pathologies in an ECG [4]. Since the electrocardiogram allows for a variety of diagnoses and its correct interpretation requires a lot of practice, research focused on the ECG signal. The ambition was to work out an application that would have a meaningful use in medicine, namely to support the correct evaluation of ECG signals.

To achieve this goal, it is necessary to expand our knowledge of electrocardiography. It is difficult to acquire the necessary expertise, especially in the field of cardiac conduction. If you want to understand how the ECG signal and the stimulation currents in the heart are connected, you have to put a lot of thought into it. Because understanding the connection is difficult not only for non-medical students, but also for medical students who already have a little prior knowledge. It turned out that there are often difficulties in mentally transferring the two-dimensional representation of the books to a three-dimensional, anatomically correct heart [5].

Since it is not possible to depict a heart in such a way that every wall with its vessels is visible, it is difficult to mentally project the signals onto the heart at different angles and directions. Another problem for me was that the models shown in the textbooks were often greatly reduced [6, S. 72]. Again and again the students had to flip back and forth and mentally superimpose models so that they had all the necessary learning-relevant information together. At the time of the research (September 2019), there were no applications that showed the interaction of different ECG signals to localize different infarcts in VR. In June 2021, a VR application was created by MIT, but it deals exclusively with the placement of the electrodes and the associated effects on the signal; it does not address the heart, its angle, or different infarcts [7]. In our Easy ECG application, the anatomically correct representation of the heart plays a major role, because the position of the coronary vessels is necessary to understand what type of infarction is present [8].

In order to analyze the difficulties more deeply, an interview with Christian Koch, a trainee doctor in his practical year was conducted. He is currently working at the Heart Centre Leipzig. He supported the thesis that many students have difficulties in understanding the relationship between ECG signal and conduction. Christian K. reported in the interview: It took him a long time to be able to imagine what actually happens with an ECG signal [9]. Even whole books have been written for this reason, as you can read in this book dedication from *ECG Course for Isabel*: "The story of Isabel [...] When she studied medicine, she - like many medical students - had difficulties with the reporting and interpretation of electrocardiograms. One day, her father and Hans-Joachim Trappe decided to help her and wrote an ECG course for her in 27 lessons [6, S. COVER]."

The app was designed to provide an easier introduction to electrocardiography for aspiring physicians. Since several interviews revealed how difficult it is to understand the content of electrocardiography, the app is intended to help young physicians who still have the ECG course ahead of them.

This task resulted in the following requirements. On the one hand, a learning environment must be generated in which the learning content of infarct localization is presented to the user. The learning content must be brought in a direct, visual context with an anatomically correct, three-dimensional heart. In addition, it must be possible to display all supplementary content (coronary vessels, leads, position of electrodes and position and location within the corpus) to the user for a better understanding of the subject area.

2 Basics

When we acquired the expertise, the authors used the same teaching materials as a medical student has to use when taking the ECG course. Even though it has been proven that in the field of electrocardiography teaching, computer-supported teaching aids achieve better learning results than the usual established teaching aids, most prospective medical students resort to textbooks (e.g. ECG course for Isabel) or teaching platforms (e.g. amboss.de), whereby the most commonly used teaching platforms are designed like the books in terms of structure and content [10, S.

1291]. There are no animations, films or interactive parts built in that could explain what is shown to the user better than a drawing.(e.g. amboss.de and viamedici.thieme.de.)

The usual teaching media in medicine are limited to very reduced two-dimensional representations. These are neither anatomically correct nor do they contain all the necessary information to be able to grasp the complex facts simply and quickly. (Only the teaching of electrocardiography in Germany was examined) However, if one were to display everything that is necessary to localize an infarction (coronary vessels, leads, position of the heart in the body), the representation would also very quickly become very confusing.

The mental leap that a prospective physician has to make from the signal to the conduction of the stimulus and its localization in the heart is great. The two-dimensional mode of representation requires a new way of thinking if the knowledge learned is to be transferred to an anatomical heart. The curves of the individual leads must be assigned to the heart regions of the 2D representation and then mentally transferred to a three-dimensional heart in order to find out which coronary vessels are affected.

So why should we continue to reduce the representation of such a complex organ to two dimensions when there is now also the possibility of representing the heart in 3D and anatomically correct? With the application, the authors want to give students the opportunity to learn directly on the anatomically correct heart. That would be a step in the right direction and a digitization where digitization is really necessary. Merely putting the content of a book online should not be seen as progress or even digitization. It has never been easier to improve, or at least supplement, teaching methods such as electrocardiography with the help of programs. Therefore, some suggestions for improvement are presented in the following chapter.

3 Methods

The goal was to design an application for medical students that would facilitate their introduction to electrocardiography and especially infarct localization. In this application, we wanted to show the direct connection between signal and myocardial excitation in the best possible way. It also should show the loop that the stimulation current makes through the heart directly on the anatomically correct, three-dimensional heart as a function of the curves shown.

The heart had to be scalable and rotatable so that every region of the heart and every coronary vessel could be seen. Since the organ is related to the body, it was important to represent the body in relation to the heart as well. The rib cage also played a role because it determines where the electrodes of the chest wall lead (V1-V6) are located. In order to be able to see at first glance where the areas affected by the infarction are located, it had to be possible to mark the associated leads and the coronary vessels lying within them. The correlation between signal and heart excitation was to be represented by corresponding highlights on the curves as well as on the model of the heart. The representation in VR (Virtual Reality) should give the user a better feeling for the connections, because these do not take place in front of their "inner eye" as in a textbook, but are virtually visible to them.

We also wanted to add sound to the application so that the user gets a short explanation for each infarct they view. This should provide an immersive, audio-visual learning experience that is more memorable than previous teaching methods. Within the time-limited framework of the research, the authors could not, of course, go into detail about every infarct. Therefore, infarcts of the lateral (side), posterior (back) and anterior (front) wall are treated as examples.

Together with medical students, we therefore determined infarcts that best illustrate the principle of infarct localization. A total of seven infarcts were selected: anterior wall infarct, supra-apical infarct, anteroseptal infarct, anterolateral infarct, inferior (diaphragmatic) infarct, posterior infarct and lateral infarct shown in Figure 1. For each of these infarcts, the acute and subacute stages of all 12 leads should be shown. This is important because the sub-acute stage

shows much more inconspicuous abnormalities than the acute stage [11]. Overall, the application is structured in two blocks: the 3D model of the heart and the charts of the individual infarcts.

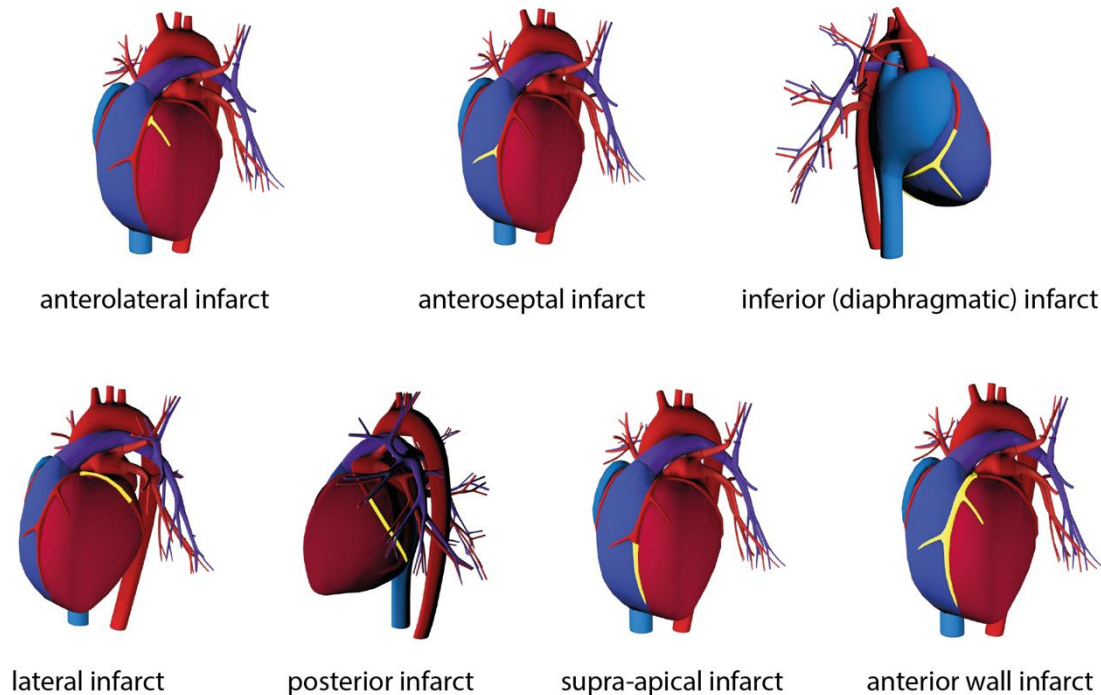


Figure 1: The seven infarcts, shown on the anatomically correct heart. The yellow coloring shows the affected coronary vessels.

4 Implementation

4.1 Design implementation

We decided to develop a Virtual Reality Application because it is much easier to understand an organ as complex as the heart in three-dimensional space than in 2D (using the *Oculus Rift*) [12]. Learning and understanding in VR was a big topic at the 4th Future Congress of the Federal Ministry of Education and Research. In a panel discussion on the topic of "The future in an augmented / virtual world", the technology was touted by Axel Steinkuhle - designer, futurist and CEO of evrbit - as immersive and thus more visually memorable than usual teaching tools.

The authors used the *Oculus Rift* for the application. Unlike most VR goggles, the *Oculus Rift* is a model designed for seated (desk) use. These VR glasses were chosen because a learning application is used while studying, which is mostly done at a desk.

The setup of the *Oculus* is also comparatively simple. To locate it in the room, you only need to place two 20 cm high sensors to the left and right of the screen. The *Oculus Rift* is also much more comfortable to wear than related VR glasses, such as the *HTC Vive*. This is a big advantage because you usually learn for longer than 30 minutes. In addition, cabling is easier because the goggles and sensors do not need an external power supply, only a USB connection.

Especially in VR, the user can quickly lose orientation in the 360-degree space. Therefore, it is all the more important to navigate with good menu navigation and spatial orientation.

As already mentioned, the application consists of two blocks, the block of the 3D heart (right hand) and the block of the infarct charts (left hand). The control of the blocks was assigned to the

respective controller: Accordingly, the left controller controls the infarct charts attached to the left hand and the right controller controls the 3D model. Figure 2 shows these two blocks attached to the controllers. It is also shown how the heart can be displayed within the corpus with the electrodes to be applied.

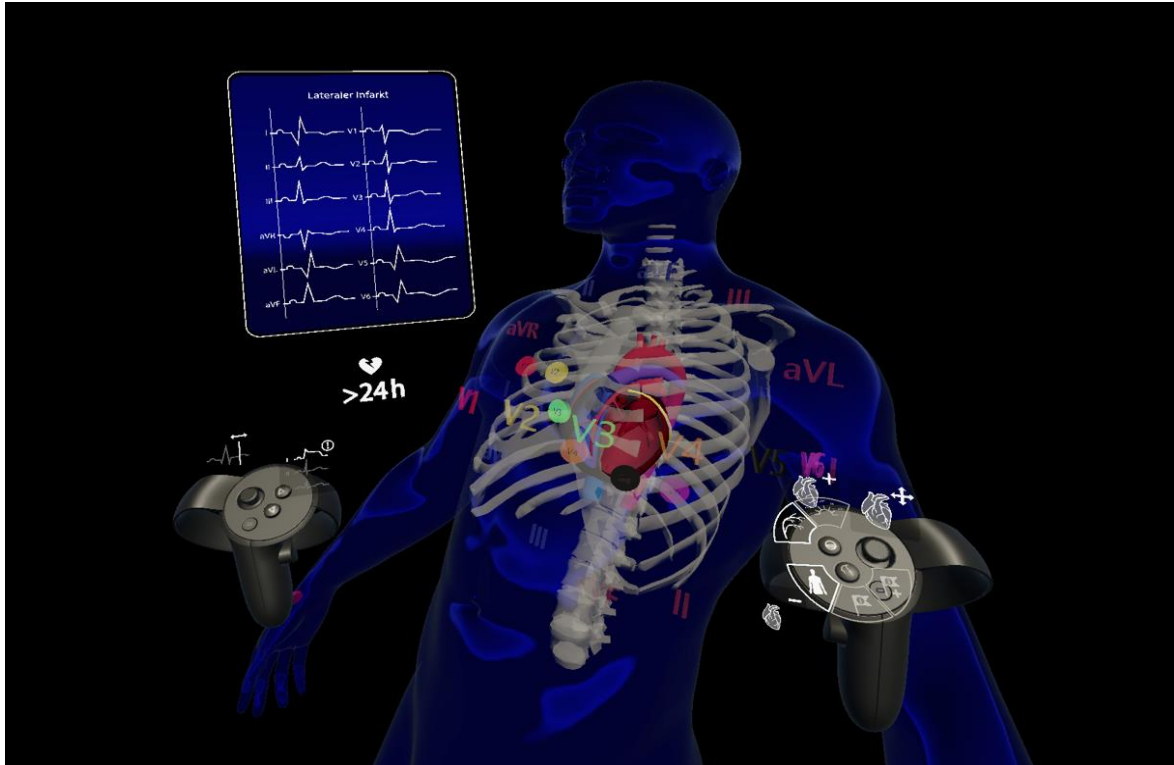


Figure 2: The Setup of the Application with the infarct charts and the body with the ECG electrodes and heart. (Source: Screenshot of the Virtual Reality Application)

The joystick of the left hand is used to select the charts. Each chart has an active and an idle mode. In Active mode, a slightly luminous frame appears around the chart that is currently selected. In Idle mode, this frame is transparent. The selection is confirmed by pressing the OK button. The menu then disappears and the selected chart with the signals is displayed enlarged. To be able to view the heart and the signal at the same time, the chart with the signals is still attached to the left hand.

If you have now selected an infarct, you can move the scanner (time "t") with the joystick and have the stimulus transmission to the selected infarct plotted in real time. As soon as $t > 0$, the 3D model of the heart becomes transparent to better see the course of the loop. The scanner and the resulting plot can be reset to the initial position by clicking on the joystick.

If you hold down the highlight button "X", the leads in which abnormalities can be found are marked in red. We have refrained from marking the affected signals from the beginning because a greater learning effect can develop by reflecting on the infarct. The coloring of the curves is only meant to be an aid.

The three-dimensional model also shows where the infarct is located after selecting the chart. An audio file describing the infarct starts automatically.

In the second interaction zone of the right hand, you can now choose between the display "Mark coronary vessels" and "Mark drainage zones". This is done by pressing the "A" key. In the menu of the right controller, the respective active status is represented by a highlight of the

respective icon. By pressing the "B" button, a body, including the thorax and electrodes, important designations of the anatomy or their abbreviations can be displayed. The 3D model can also be rotated and scaled using the joystick of the right hand. A click on the joystick sets the model to its initial position and size.

For a quick introduction to the usability and because the resolution of the *Oculus Rift* is in process of enhancement, we have included tooltips that can be seen in Figure 3. These provide pictographic information on the respective function of the buttons and joysticks. Pictograms are easy to interpret and help the user to quickly understand the function [13]. In Table 1 the assignment of the application buttons is listed.

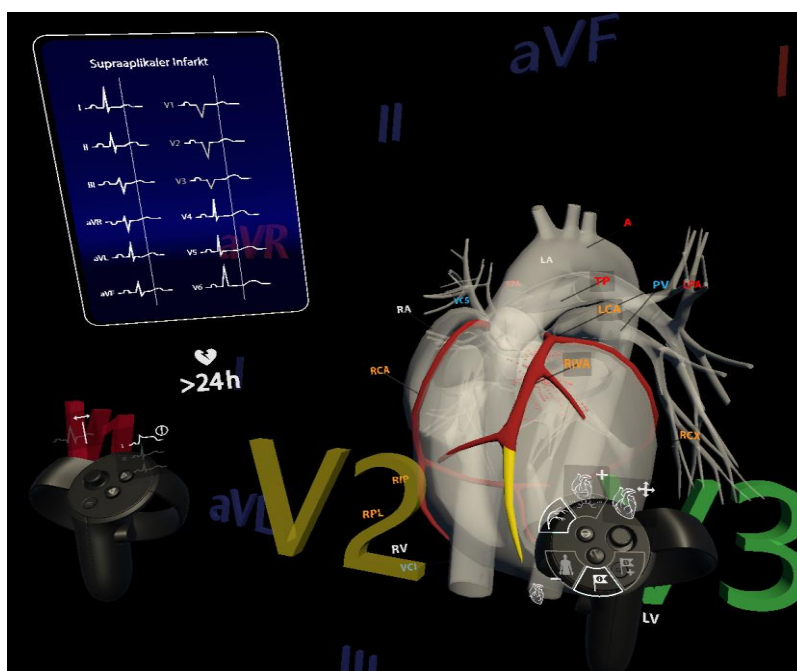


Figure 3: The UI of the Application with the tooltips attached to the controllers. (Source: Screenshot of the Virtual Reality Application)

Table 1: The Application menu description of Easy ECG on Oculus Rift

The left controller (Interaction level one)	
Joystick	Case selection
Button X	Confirm Case selection
Interaction level two (after chart selection)	
Joystick	Timeline control
Button X	Highlight affected derivatives
Button Y	Back to interaction level one
Trigger Button	Fade in of the acute stage
The right controller	

Joystick	Rotate heart in XYZ axis
Trigger Button	Zoom in
Grab Button	Zoom out
Button A – Status 0	Highlight affected coronary vessels
Button A – Status 1	Highlight affected lead sections
Button B – Status 0	No enhanced information
Button B – Status 10	Show body with chest and electrodes
Button B – Status 2	Show technical terms abbreviated
Button B – Status 3	Show technical terms written out

In some tests, it became apparent that especially users with an affinity for game consoles instinctively used the controls differently. As it turned out, some used the controllers of the *Oculus Rift* like an *X-Box* or *PlayStation* controller. Since the application was initially divided into two separate interaction areas, this led to confusion in the beginning. However, this changed when the tooltips were displayed. With the tooltips to help them, most of the test persons needed less than two minutes to use the application correctly.

The colors of the application were mostly kept simple and restrained. Only individual highlights draw the user's attention to the desired point through color highlighting. These highlighted points are, for example, abnormalities in the curves of the individual leads or coronary vessels in the heart affected by infarction.

The colors of the heart correspond to those of the usual teaching aids for medicine. The right atrium and the right ventricle are colored blue, the left atrium and the left ventricle are colored red. For design reasons, the color values were slightly adjusted. In doing so, the colors were edited so that they can be better displayed in VR and fit more harmoniously into their surroundings. The material of the charts attached to the left hand is slightly transparent so as not to obscure the view of the heart.

4.2 Technical implementation

In order for the infarcts to be displayed in the app, data was needed. We still had some "physical ECG strips" (physical ECG strips are recordings of healthy heart signals that do not show any abnormalities) left over from previous research - these are ECGs that are plotted directly from the ECG device onto a graph paper. The authors wanted to vectorize these curves and digitize their values. However, it proved difficult to get hold of ECG strips showing an infarct.

When the experiment was started with a self-made ECG, we first had the problem of not being able to record any data. The remedy was a small script that stored the values in a CSV file. However, as described above, the ECG could only output one derivative. In addition, the signal had too many voltage peaks and noise-effect.

Since the authors wanted to work three-dimensionally, 12 leads were needed. Using an online database called *Physionet*, we came across various data sets that had been recorded with 12 derivatives. A program in *Matlab* was written that could plot individual records. This way the authors decided which of the recordings were best suited for the experiments, i.e. which curves deviated the least from the standard curve. To be able to implement a training app for infarct localization, data on infarcts are of course needed. Unfortunately, no ECG recordings of infarctions could be found via the *Physionet* database.

Because optimal ECG curves were needed for the training application, we decided to use curves from the textbooks [6, S. 72–78]. To do this, the ECG curves of the different infarcts were scanned and traced the paths in *Adobe Illustrator*. We used the data obtained from this process in two ways. Firstly, three-dimensional curves were generated from these paths with the help of the Extrude tool in *Cinema4D*. It was possible to assign these to the individual infarct charts. Secondly, the paths were used to determine the curve values as a function of the time axis using a program written in *Matlab*. The authors were then able to save these values for each infarct case and for each of the derivatives in a CSV file. This was necessary to be able to calculate the vector loops there. Some of the curves had a shifted offset (zero line), which we had to compensate manually.

As described earlier, the derivatives are axes that pass through the heart at different angles and planes. If we now consider each derivative individually as a distance (in Millivolt) that can be mapped in the coordinate system at time t , simple mathematical formulas make it possible to calculate the voltage curve in the heart. By calculating the respective points on the X and Y axis of the vertical plane via sine and cosine and linking these with the values of the chest wall derivations in the horizontal plane, a three-dimensional structure is obtained as a result. This describes the conduction and propagation of impulses in the heart.

The authors tested the calculations with a Python-Script with *Matplot*, which was executed via a Virtual-Machine in the terminal. *Matplot* could be used to plot the resulting three-dimensional curves as can be seen in Figure 4.

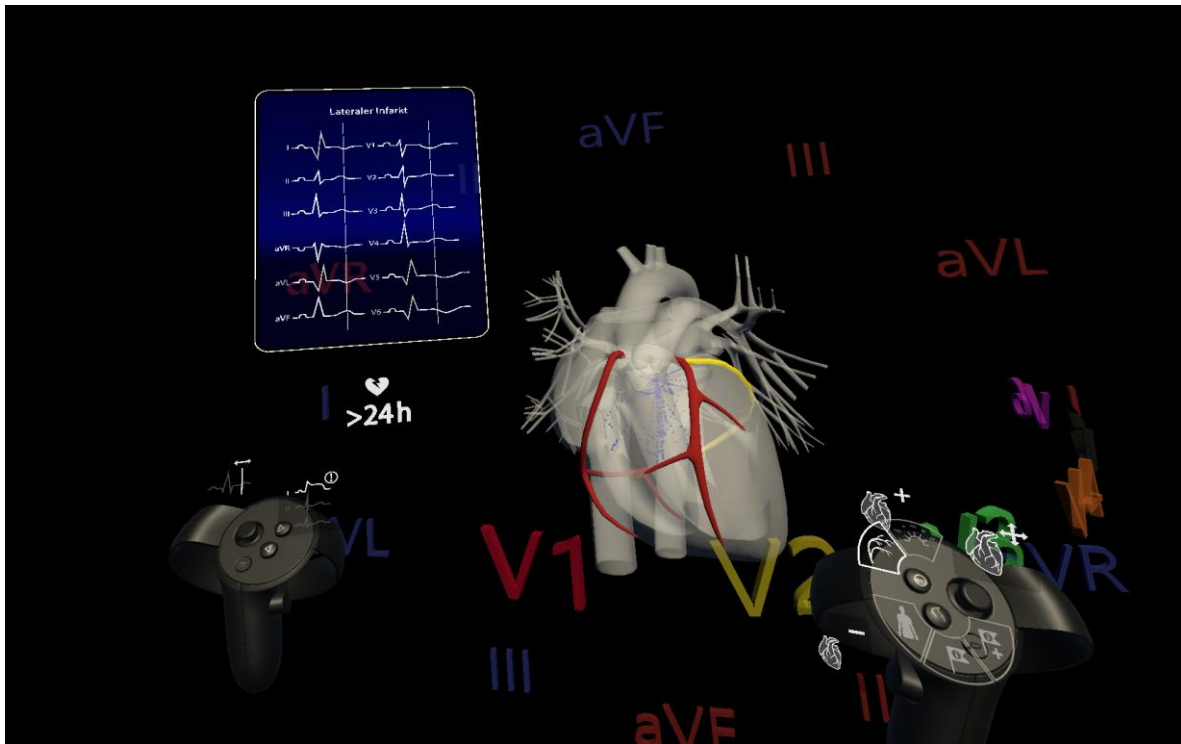


Figure 4: Virtual Reality Application with plotted propagation of impulses in the heart. (Source: Screenshot of the Virtual Reality Application)

5 Conclusion and Outlook

In conclusion, the requirements for the application were met. The application is an enrichment for medical teaching. It visualizes what could only be imagined before. It achieves the goal of demonstrating the connection between signal and heart muscle excitation in a comprehensible way. The next step of this research is an evaluation of the gained knowledge of a test person.

The current state of the application makes it easy for aspiring physicians to get started with infarct localization. The user can choose between seven infarct cases. For these infarcts, he can display the corresponding abnormalities on the anatomically correct heart. The user can view the infarct-specific heart regions/coronary vessels and the corresponding abnormalities in the leads. In addition to these functions, it is also possible to have the conduction of impulses "plotted" in the heart. This is particularly helpful in understanding cardiac excitation and how it relates to the ECG signals.

In the future, we would have especially liked to improve the interaction of the application. To enable instinctive interaction, the authors would have to take more account of the operation of common gaming consoles, i.e. follow more commonly used interaction concepts of a gaming controller. In the case of the X-Box controller, this would usually be the following concept: confirm at the bottom left, back at the top left and the main action button at the bottom right.

Leaving interaction out of the equation, it would also be possible to run the app in VR on any smartphone. This would make the app usable regardless of location, and one could conveniently refresh one's knowledge of infarct localization anytime and anywhere. In the future, complete courses on infarct localization could also be held in VR. It would also be good to include more infarct cases. For this reason, the application was programmed in such a way that it can be expanded by as many cases as desired. In addition, we would find a "professional mode" appropriate, which works with real ECG strips. Since these often contain noise or other signal interference, it is more difficult to find abnormalities in the leads.

From this and other potential for expansion or improvement, this application is certainly capable of further development. That is why the authors are considering pushing this application further. Even if the app already has certain advantages over a textbook, it will still need to be supplemented with some tools before it can complement or even completely replace the textbook. A complete ECG course in VR would be a goal that could be pursued in the future [14, 15].

References

- [1] R. F. Schmidt, F. Lang, und M. Heckmann, Hrsg., *Physiologie des Menschen: Mit Pathophysiologie*, 31. Aufl. Berlin Heidelberg: Springer-Verlag, 2011. doi: 10.1007/978-3-642-01651-6.
- [2] M. B. und D. B., „Visualisierung in der Medizinischen Lehre“, Sep. 15, 2019.
- [3] M. Lehnhardt, „Visualisierung in der Medizinischen Lehre II“, Sep. 18, 2019.
- [4] S. M. Salerno, P. C. Alguire, und H. S. Waxman, „Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence“, *Ann. Intern. Med.*, Bd. 138, Nr. 9, S. 751–760, Mai 2003, doi: 10.7326/0003-4819-138-9-200305060-00013.
- [5] U. Dardon, V. Vieira, S. B. S. Prestes, T. de C. C. Parméra, L. Cotts, und L. P. Bergqvist, „TWO-DIMENSIONAL AND THREE-DIMENSIONAL MODELS USED FOR TEACHING HUMAN EVOLUTION IN SECONDARY SCHOOLS. LEARNING PROFICIENCY ASSESSMENT. A CASE STUDY“, *J. Sediment. Environ.*, Bd. 1, Nr. 2, Art. Nr. 2, 2016, doi: 10.12957/jse.2016.23547.
- [6] H.-J. Trappe und H.-P. Schuster, *EKG-Kurs für Isabel*, 6. Aufl. Stuttgart New York: Thieme, 2013.
- [7] „IMMERSED IN: Medicine—Performing and understanding ECGs in virtual reality—June 30 | MIT.nano“. <https://mitnano.mit.edu/immersed-seminar-series/medicine> (last visited: Dec. 06, 2021).

- [8] „Visualization of Medical Data - Bachelor Project on Behance“. <https://www.behance.net/gallery/86189737/Visualization-of-Medical-Data-Bachelor-Project> (last visited: Dec. 06, 2021).
- [9] C. K., „Visualisierung in der Medizinischen Lehre III“, Dez. 03, 2019.
- [10] R. E. Fincher, A. M. Abdulla, M. R. Sridharan, J. L. Houghton, und J. S. Henke, „Computer-assisted learning compared with weekly seminars for teaching fundamental electrocardiography to junior medical students“, *South. Med. J.*, Bd. 81, Nr. 10, S. 1291–1294, Okt. 1988, doi: 10.1097/00007611-198810000-00020.
- [11] K. Masek, „ECG Diagnosis: Acute Pericarditis“, *Perm. J.*, S. e146–e146, Dez. 2013, doi: 10.7812/TPP/13-044.
- [12] S. W. Greenwald u. a., „Technology and Applications for Collaborative Learning in Virtual Reality“, Philadelphia, PA: International Society of the Learning Sciences., 2017. doi: 10.22318/csc12017.115.
- [13] C. Tijus, J. Barcenilla, B. C. de Lavalette, und J.-G. Meunier, „Chapter 2: The Design, Understanding and Usage of Pictograms“, *Writ. Doc. Workplace*, S. 17–31, Jan. 2007, doi: 10.1163/9789004253254_003.
- [14] G. Székely und R. M. Satava, „Virtual reality in medicine“, *BMJ*, Bd. 319, Nr. 7220, S. 1305, Nov. 1999.
- [15] J. Majerník, M. Maďar, und J. Mojžišová, „Integration of virtual patients in education of veterinary medicine“, in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sep. 2017, S. 185–188. doi: 10.15439/2017F134.

Susanna Götz
University of Applied Sciences Würzburg-Schweinfurt
Faculty of Design
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: susanna.goetz@fhws.de

A comparative study of growing microgreens. Human supervision VS Agritech/Automation

Felix Husac

Abstract

Agriculture has been the main food source for humans and livestock since prehistoric times. Along the ages, many technologies have revolutionized the way humans grow plants and animals for food, but progress has been slow over the last couple centuries. Despite the tremendous progress in technology, agritech solutions are sparse, complicated and oftentimes too expensive for what they offer. The agricultural niche that is gaining traction in today's world through mediums like fine dining, world renowned chefs and health enthusiasts alike is the growing of microgreens. Microgreens are deemed one of the most nutritious foods on the planet. They are easy to grow, especially with the help of automation, and they make for quite a lucrative business. This paper aims to study the different types of growth mediums and conditions possible for microplants, creating a special environment for their development as a cheap alternative to the expensive and sometimes incomplete solutions already on the market. The future of agriculture is heading towards automation, hydroponic cultures and vertical scalability, because of both cost and environmental concerns. The proposed techniques and technologies used are scalable, and a future development is automating a whole greenhouse, from planting the seeds to harvesting the results and to analyzing the data generated for future optimization.

1 Introduction

According to a CNBC report, agritech companies raised around 5 billion USD in 2020 [1], and as a whole, the agriculture industry is worth around 1.109 trillion USD in the US alone. [2] From lifestyle gurus to haute cuisine world renowned chefs, microplants are gaining attention and interest, being regarded as a superfood. Growing microplants is a really lucrative business, as an article from growcycle.com mentions, and the return on investment for a pound of final product is the most expensive legal plant crop. [3]

This paper aims to bring forth an alternative to the current small scale farming practices on the one hand, and on the other a comparison between different methods and possibilities for the growth of microgreens in particular. By describing an innovative approach of building a scalable micro-greenhouse and evaluating the crop yield for different nutritious substrates and technical configurations, the reader will understand the use and the endless possibilities of such a system. In order to achieve this goal, a study of the current market solution for micro-greenhouses and tabletop plant growth solutions was developed.

The main reason for this proposed solution is personal interest. Microgreens are considered a superfood by many self proclaimed health gurus [4], their cultivation is less environmentally impactful because microgreens are in most cases vertically farmed, sustainable and require 95% less land than

traditional agriculture [5]. With a small upfront investment, the return will be significant. Microgreens are the most expensive legal crop [3].

Also, they require less human interaction and care than adult plants due to their short lifespan. Another benefit of cultivating them is that they are more nutrient dense than their adult counterparts and are just as low in energy content. This makes them the perfect addition to a healthy and nutrient packed diet. [6]

Another reason for taking this endeavor is the funding kindly provided by the Științescu Fund. This idea started when I was asked to join in a highschool science project funded by the Științecu Fund through the "Tara Fagarasului" Community Foundation and headed by a group of teachers, specialists and enthusiasts.

The Științescu Fund is a financing program for educational projects in the fields of STEAM (sciences, technology, engineering, arts, mathematics), which aim to cultivate the passion of children aged 6-19 for these fields. [7]

This year (2021), through "Științescu Țara Făgărașului", the Community Foundation Țara Făgărașului grants funding amounting to 78,317 lei (out of the total budget of 78,880 lei) for the 7 educational projects that have passed the judging stage, and the teams selected to receive funding will benefit from support in communication, in motivating and involving students and in project management. [8]

The jury that awarded the funding was composed of representatives of academia, entrepreneurs, representatives of companies and NGOs. Through their impartiality and experience gained in various fields, the members of the Jury ensured a selection process relevant to the direct beneficiaries of the selected projects.

The selected projects are in the fields of: ecology and recycling, IT, robotics and engineering, astronomy, biodiversity, physics, mechanics, biology, experimental and applied chemistry, agriculture and permaculture, and promote both collaboration between educators and students, between rural and urban areas and the connection between various branches of exact sciences.

The above mentioned project is split into two parts. The first part is a mechanics and engine design class for highschool students, where a professional mechanic will show them how an internal combustion engine works by building a mockup with them, and using a working motor to create a waste shredder. The waste shredder will be used to compost common household waste such as fallen leaves, branches and biodegradable materials. The second part of the project was born from the idea of using the compost made by the waste shredder to create something that will further expand the learning opportunities of the students involved. Thus, the title of the project was born: "Waste with a purpose". The compost generated by the shredder will be left to ferment and decompose and can then be used as a nutrient rich substrate for plant growth. [8]

The results stipulated in the project proposal, that must be produced by the project deadline, are as follows: the creation, assembly and use of two micro-greenhouses, the writing and design of a book containing a detailed tutorial for building and programming the micro-greenhouse, culinary recipes with microgreens as base ingredients and a study of different substrates and their effect on plant growth.

The proposed solution is a polycarbonate box, created on a wooden structure. It has two growth stages, one design specifically for seed germination and the other stage for growth and development. An array of sensors, like temperature sensors, water flow sensors and pHsensors monitor the life support conditions in real time, making sure that the plants have a constant supply of fresh water, nutrients, fresh air and the optimal amount of light with the help of fans, water pumps, LED growth lights and so on. The micro-greenhouse will also log the data from the sensors and send it to the computer for long term storage or for viewing and interpreting on a GUI.

The rest of this paper will describe the current state of the art in micro scale and small scale plant raising in the first chapter, with the financial incentives and agriculture grant opportunities for such projects presented in the second chapter. The third chapter will deal with the actual design of the micro-greenhouse - from concept to rough draft to actual build and also a short overview of different growth

mediums in an artificial environment.. The fourth chapter will discuss the challenges posed by the information flow between the modules and sensors used to power, control and track the operation of the micro-greenhouse.

The conclusion and future development possibilities will prove once again the utility and simplicity of such a system and also the real possibility of turning similar agritech systems into a profitable business.

2 Brief overview of small scale Agritech solutions

Smart plant monitoring systems are a popular subject among tech enthusiasts with a passion for botanics. For example, the Flower CARE smart monitor [9] is an inexpensive gadget that is inserted in the soil of your favourite pot plant, and it is capable of sending back real time readings of ambient light, soil moisture, soil nutrients and air temperature via bluetooth.

The problems start arriving when there is more than one plant involved. Systems that accommodate multiple plants and adapt their conditions in real time to suit the needs of the plants are still expensive. Even simple table-top solutions that can grow about 12 plants at the same time are pricey considering the true cost of the materials and the electronics used in their manufacturing. For example, the cheapest option in a niche online store, that incorporates an LED and a water recirculation pump, is a product over 100EUR (480 RON). [10]

Another solution that promises full automation (meaning watering and light) is almost three times as expensive 260EUR (1220 RON). [11]

A simple hydroponic growth bed with an integrated water pump is over 150EUR [12], a price much too high considering a 12V submersible water pump is around 10EUR or less [13].

Most of the above systems, regardless of price range, are equipped with just one single button: one that turns the machine on or off. The integrated LED and the water pump are turned on and off automatically, but the user has no control over it, nor is he provided with feedback and information in order to monitor his crops. The above-mentioned products have minimal functionality, and are mostly designed to just sustain the life of the plant crop seeded in them and not to monitor or help improve it. For large scale industrial farming, systems such as those described above barely scratch the surface for the needs they need to meet. Plants need to be tracked and monitored from seed, to plant, to plate. Such big undertakings are achieved via satellite images for crop growth monitoring, soil level sensors and smart modules that create a mesh network in the field and relay data back and forth in real time, robotic heavy equipment such as harvester robots and the sorts and even hyperspectral cameras that are used to detect plagues and pests before they have time to develop and ruin a crop. [14, 15, 16]

The average small scale consumer user will never be able to buy or use such systems because of their complexity, maintenance costs and the lack of a need for such an over the top solution for a small space and modest possibilities.

The proposed solution offers a hybrid between the two extremes, centered around minimal complexity for maximum efficiency. The micro-greenhouse design proposed in this paper is cheap, easy to build, modular and scalable in regards to the end user's needs and offers a way to log and monitor sensor data and to also automate the process of plant growth.

3 Financial Incentive

The agrifood industry revolves around every aspect of food production: from growth to sale and delivery.[17]

It is the most important and also the oldest industry in the world. Since the Neolithic Age, humans have kept agriculture in high regard. Over the centuries, techniques such as making flour from cereals, cheese making, selective breeding of bacteria for fermenting beverages and many other techniques have their roots in agriculture. [18]

The agrifood industry is a profitable one, with around 8 trillion USD raised by startups in the US alone in 2018, [19] and in the EU, it generated an estimated gross added value of approximately 177billion EUR in 2020 alone [20].

Starting in agriculture could be quite an expensive endeavor. It all depends on the scale of operations. Cost attenuates with scale, and subsidies have the same behaviour as compound interest, but these are only guaranteed for large scale production. Luckily, there are many incentive plans, subsidies and grants offered by private investors and countries alike in order to secure a prosperous future and the uninterrupted production of food. In recent years, there has been a change of priorities in food production, with many countries aiming to reduce the climate impact of farming by using smart technologies and promoting growth in the agritech, biotech and IoT sectors via funding, and also the return to organic farming and land restauration [21]. Several examples of such funding opportunities will follow.

For example, the European Union has two funds that cover agricultural expenditures, the EAGF (European Agricultural Guarantee Fund) - that directly finances farmers and regulates certain market activities and the EAFRD (European Agricultural Fund for Rural Development) that helps develop rural areas of member states. [22]

The EIT Food's main web page heading states that its purpose is to be "Europe's leading food innovation initiative, working to make the food system more sustainable, healthy and trusted". The EIT is holding a funding call on new and innovative agricultural technologies, as well as raising awareness and advocating for a better food system. [23, 24]

A Horizon2020 funded programme called Smart AgriHubs that raised 20M EURO, unites over 164 partners in the AgriFood sector, aiming to "realise the digitization of European agriculture by fostering an agricultural innovation ecosystem dedicated to excellence, sustainability and success". This project aims to deliver "80 new digital solutions to the market, raise 30M EUR additional funding [...] and plans to help digitize over 2 million farms [...] across Europe". [25]

Many other pilot projects such as the LIFE GAIA Sense project - that is currently deployed in over 18 demonstration farms and helped reduce the use of pesticides and water by 25% - and the ICT-AGRI-FOOD fund aim to create a more sustainable and technology advanced style of farming, one that protects the environment and cuts down on costs without sacrificing production volume or quality. [26]

4 The proposed solution

4.1 Solution overview

To create a scalable and vertically stackable micro-greenhouse, the decided shape for the final design is a rectangle. The modularity is provided by the sensor arrays and microcontroller nodes, that can work independently from each other, or can work together if connected. If the end user desires to place the micro-greenhouses on top of each other, there is no need for extra water circuits, all the hoses being connected to the main pumps via T shaped fittings.

The actual micro-greenhouse is a rectangular box, with a wooden structure and polycarbonate walls. It is 50cm tall, 50cm wide and 100cm long. The box is divided into two main zones: the germination stage and the growth stage. The germination stage occupies the lower 25cm of the whole micro-greenhouse. It's an opaque, humid environment, ideal for the first stages of life of the plant being cultivated. After the plant sprouts, it will be moved to the upper stage, called the growth stage. The growth stage represents the upper half of the box, and is constructed of clear polycarbonate walls for good illumination. It also features LEDs for better lighting control. The baby plants will be spending their time in the growth stage until they are ready to harvest. (Fig. 1)

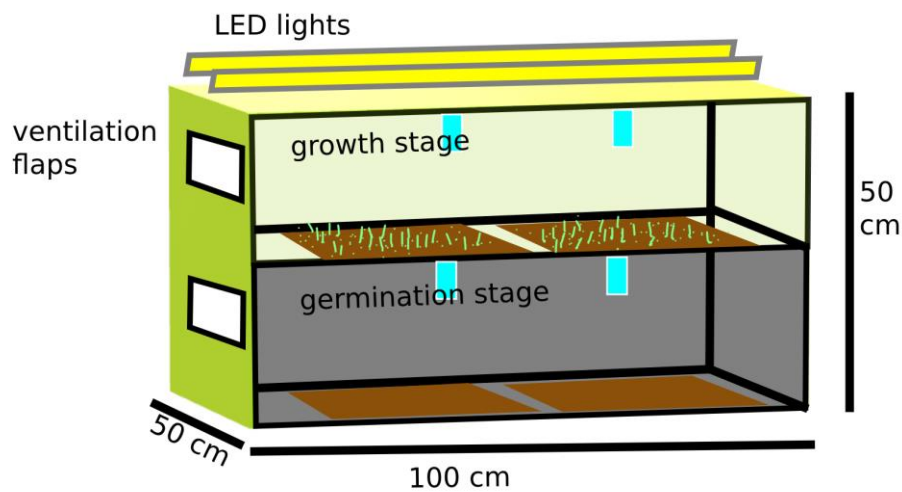


Figure 1: Simplified Schematics

The two compartments occupy the length of the box. There are ventilation flaps on each of the stages that are controlled by servo motors. They can be opened or closed. The two rectangles on the top of each stage represent the atomizers that are used to spray water and fertilizer over the crops. The LEDs that provide an intense light source necessary for photosynthesis are depicted as two rectangles that span the length of the top stage of the box.

The pipes used for water circulation are food grade clear PVC pipes, in multiple diameters, depending on purpose: the main hose that provides clear water to the recirculating tank and the hose that disposes of the used or over-fertilized water are of a larger diameter (19mm), and the pipes connecting the fertilizer tanks, the recirculating tank and feed the atomizers are of a smaller diameter (8mm). These smaller tubes have debit sensors attached to calculate the flow of water.

The growth substrate is positioned in plastic trays, in both levels of the micro-greenhouse. The trays have drainage holes drilled in them so that water never stagnates and is safely flowing around the seeds.

4.2 Mechanical components

The micro-greenhouse is equipped with several servo operated flaps that raise to let fresh air inside, fans that circulate the air inside the micro-greenhouse and around the plants and pumps that control the flow of water. The ultrasonic atomizers are used to maintain the optimal level of humidity and to gently water the seeds without flushing them out in the recirculating tank with the water. The water pumps are 12V pumps and they are used to transfer water from the fresh tank to the recirculating tank, the plant trays and the waste water tank. Peristaltic pumps are used to transfer precise amounts of fertilizer and antifungic solution to the atomizers when the pH level of the water indicates so (plants enjoy a slightly acidic substrate, generally pH 7 and slightly below [27]).

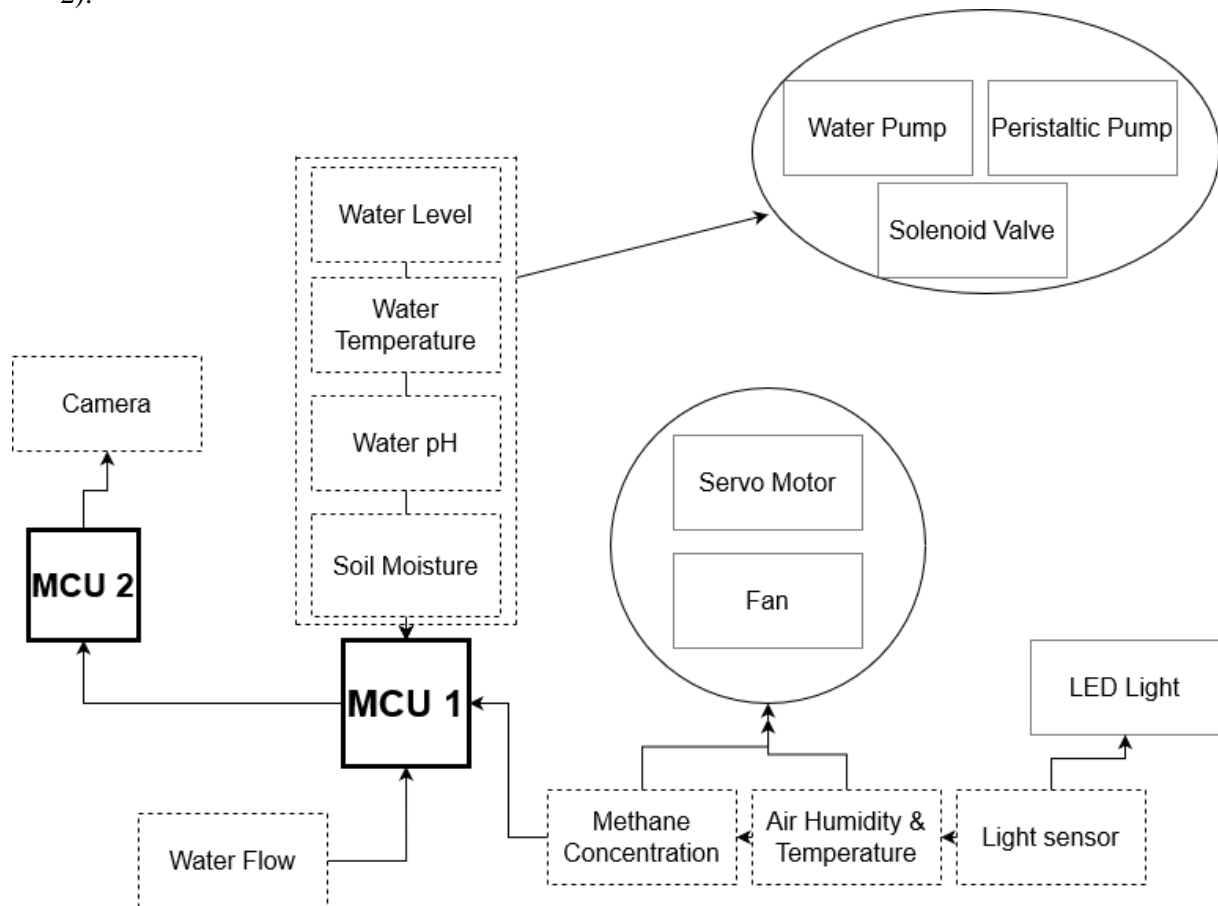
4.3 Sensors

In order to properly control and monitor the growth environment of the microplants, the following sensors were used:

- DHT22 sensor - for air humidity levels and air temperature
- thermometer - for monitoring water temperature
- soil moisture sensor - for checking the water content of solid substrates
- water level sensor - for hydroponic crops

- air methane sensor - for checking methane concentrations inside the micro-greenhouse (plants produce methane [28])
- photodiode - as a light sensor, both to make sure that no light enters the lower germinations stage and that the quantity of light in the upper stage is adequate
- water flow sensor - for gathering water consumption metrics and to detect possible clogs in the system
- pH sensor - for measuring the pH of the water being recirculated in the system

Their interaction with the microcontrollers and actuators are depicted in the following diagram (Fig. 2):



Legend: Dashed Line – sensors; Narrow Straight Line – actuators; Bold Contour Line - microcontrollers

Figure 2: Interaction between sensors and actuators.

As depicted above, multiple sensors can interact with the same actuator (water level, temperature etc.), some sensors are only used for gathering data and have no control input (camera, water flow sensor) and others have only one purpose, such as the light sensor, for turning the LED on or off.

4.4 Microcontrollers

In order to interpret the data from the sensors, and to gather and store the information gathered from the life support environment, two or more microcontrollers are needed, depending on the user's choice of splitting the modules into command modules and slave modules. One or more of the microcontrollers

are used for real-time data gathering and life support system adjustment, while one microcontroller is responsible for relaying the data from all the modules it oversees to the computer for long term storage. For this project, the microcontrollers of choice were the Arduino Mega 2560 and the ESP32. The Arduino was chosen especially for its large number of input and output pins, necessary for reading data from all the sensors and sending commands to ventilators, servos, pumps and lights. The ESP32 was chosen for its WiFi capabilities and camera module. It will provide data coming from the Arduino to the GUI via its web interface, and also live pictures from inside the micro-greenhouse. The data generated will be stored in a database on the computer and used for further analysis. The Arduino will monitor and adjust the functions of the micro-greenhouse in order to maintain optimal conditions for life support and plant growth.

4.5 Growth substrates

The comparative study aims to discover if there are any differences between the following growth substrates:

- paper towel - least expensive material
- pure hydroponic - most convenient alternative
- coco coir - fibrous and permeable, organic and reusable
- compost - dense and filled with inorganic matter

Several trials (without the micro-greenhouse) and many online microgreen specialists recommend coco coir for best results [29]. If unavailable, paper towels are the next best choice [30]. When the seeds germinate, they need to anchor themselves in the substrate in order to grow their sprout properly. Hydroponic culture can only be tolerated by more mature plants, or by seeds that are grown in fine mesh sieves. [31, 32]

A substrate too rich in nutrients and too dense for water to flow through freely is again a less than ideal choice, because young plants are very easily susceptible to diseases and mold. The humid environment of the micro-greenhouse is kept in balance by fans, ventilation flaps and atomizers, and any water that stagnates in the excessively nutrient rich soil can have a damaging effect on the microplant crop. Excess water deprives the plants of oxygen and the young sprouts die. [33, 34]

4.6 Software challenges

Building the structure and connecting the components together is less demanding than sending, receiving, storing and interpreting meaningful telemetry data from the micro-greenhouse. There are of course free and paid services that deal with IoT devices and communications, and provide ready-made graphs and charts and graphs, like <https://thingspeak.com/>, a live IoT analytics platform that helps its users visualize and interpret the data received, all in the cloud, while also being able to set up alerts and trigger actions [16]. The only downside to this service is that the free tier is limited in read and update frequency (one update every 15 seconds) and it has a low message limit.

The free alternatives that are proposed for this project are the MQTT protocol, ArduinoJSON, MySQL and Grafana.

MqTT stands for Mosquitto protocol. It was created to facilitate data exchanges between a server and constrained devices. It is ideal for IoT and machine-to-machine communication due to the fact that it has low bandwidth requirements, it mitigates unreliable network connections and doesn't require complex configuring. It is in essence a "publish/subscribe protocol that allows edge-of-network devices to publish to a server" [35, 36]. In this application, MqTT is responsible for the communication between the ESP32 and the computer.

ArduinoJSON is a powerful serialization library available for Arduino IDE. It helps with storing data locally and also sharing data via serial, WiFi and ethernet (if available). One of its major benefits is that it has a fixed memory allocation, meaning that devices with little memory can still use it and not run into heap fragmentation problems [37]. This library helps with storing and sending data from the

Arduino to the ESP32, and the Json objects that arrive will be sent further along, to the computer, to be stored into a MySQL database.

MySQL is a widely used, widely known RDBMS based on SQL. It is most commonly used for data storage. MySQL will be used to store the Json documents incoming from the Arduino. MySQL version 8 has a special data type named “JSON Data Type” that is able to store Json objects and also perform queries on them. [38]

Grafana is a dashboarding tool that is free and open source. It comes with a built-in data source plugin that makes queries and visualizes data directly from MySQL. It is used to create interactive, beautiful graphs, reports and dashboards. [39]

Another challenge is daisy chaining the microcontrollers. For this, I chose to use serial communication/UART. It is one of the easiest ways to move data across multiple microcontrollers. The Arduino is connected to the ESP32 via the RX and TX pins. The data will flow from the Arduino to the ESP32 via serial, for every reading of the sensors, as a Json object, and the ESP32 will send back the appropriate commands if needed. [40, 41]

6 Conclusion and future developments

This paper presents a simple and cost effective alternative to similar current products on the market. It's aimed not toward industrial farming and agriculture, but toward the normal enthusiast and hobby farmer. It is modular and scalable, and complies with the latest trends in agritech: automation and low environmental impact. Being designed as a multi stage crate, the end user can stack multiple units to create a rack arrangement, thus increasing the cultivation area without needing more soil or a larger footprint device. The best growth medium for microgreens in general is coco coir substrate, but the most cost effective option is paper towels. Cultivating microgreens on denser substrates poses a threat to their wellbeing by retaining more water and encouraging mold growth.

The possibilities of future developments are limited only by the end user's needs. The micro-greenhouse can easily include a robotic harvesting arm, equipped with a gripper, a pair of scissors and a camera for automatically harvesting the microgreens. Being scalable and modular, this project can be expanded to control a larger area, for example a vegetable garden or greenhouse. The sensor modules can be bundled together with a microcontroller and a bluetooth module to create independent subunits that can be placed in any desired area of the garden, in order to relay information to a central command center that actuates the ventilation, watering and nutrient dispersion systems.

From the metadata collected from running the automated vegetable garden, one can extract training data for machine learning models, in order to create autonomous, real time responsive systems that adapt to the type of crops and soil condition, guaranteeing a better yield, more control and less human intervention.

This system can be included in a business model or it can be developed as a standalone business too. For instance, restaurants that focus on healthy eating or haute cuisine can now create their fresh ingredients on their own. Thus they are no longer dependent on deliveries of fresh herbs that become less and less fragrant and flavorful with every hour passing since their picking. This can significantly increase the quality and taste of dishes. For a normal user, such as a health enthusiast, such a system is a simple way to guarantee a daily intake of high quality nutrients.

Acknowledgement: This work was supervised by Professor *Dana Simian*, from “*Lucian Blaga*” *University of Sibiu* and made possible by funds kindly provided by the *Științescu Fund*.

References

- [1] Laura Colodny, *Covid and 'peak cow' created a boom for food and agriculture tech in 2020*, CNBC accessed on 18.10.2021, <https://www.cnbc.com/2021/04/30/food-and-agriculture-start-ups-raised-record-22point3-billion-in-2020.html>
- [2] Export Enterprises, *The economic context of the United States*, Nordea, accessed on 18.10.2021, <https://www.nordeatrade.com/en/explore-new-market/united-states/economical-context>
- [3] Craig Wallin, *Top 10 Questions About Growing Microgreens For Profit*, Profitable Plants Digest, accessed on 18.10.2021, <https://www.profitableplantsdigest.com/top-10-questions-about-growing-microgreens-for-profit/>
- [4] Kaitlynn Fenley, *The Gut Health Benefits of Microgreens*, Cultured Guru, accessed on 18.10.2021, <https://cultured.guru/blog/the-gut-health-benefits-of-microgreens>
- [5] *What Experts Say About Microgreens*, Infinite Harvest, accessed on 18.10.2021, <https://infinite-harvest.com/what-experts-say-about-microgreens/>
- [6] Duane Mellor, Ekavi Georgousopoulo, Nenad Naumovski, *Health Check: are microgreens better for you than regular greens?*, The Conversation, accessed on 18.10.2021, <https://theconversation.com/health-check-are-microgreens-better-for-you-than-regular-greens-73950>
- [7] <https://stiintescu.ro/#>, Fondul Științescu, accessed on 18.10.2021,
- [8] <https://fagaras.stiintescu.ro/proiecte-2021/>, Concursul de proiecte 2021 Făgăraș, accessed on 18.10.2021
- [9] *Flower CARE smart monitor*, accessed on 18.10.2021, https://www.led-grower.eu/flower-care-smart-monitor/?gclid=Cj0KCQjww4OMBhCUARIsAILndv6tEU2obj0cPTkQ9REVZ9zhZer8VTz7XEDbTIAGfKkMk3SGiewjhm0aAnHdEALw_wcB
- [10] *Vegebox Smart Garden*, accessed on 18.10.2021, https://www.luminam.ro/vegebox-smart-garden-9-planting-vacancy-led-21w-230v/?gclid=Cj0KCQjww4OMBhCUARIsAILndv4GAN25HyO1GjYHTKtbrA3O4yOTqIH5pi01H2wKdEUWyh-9oOAI8oaAsUxEALw_wcB
- [11] *GrowIt Farm*, accessed on 18.10.2021, https://www.klarstein.ro/Aparate-de-bucatarie/Ghivece-inteligente/GrowIt-Farm-gradina-inteligenta-interioara-28-de-plante-48-W-LED-8-litri-28-de-plante.html?gclid=Cj0KCQjww4OMBhCUARIsAILndv7a-7ZSks-Rk3Bzf3B8SRGhBBGBjWJFLPT5AzTOeClzLI9U6dZgwAAaAl80EALw_wcB
- [12] *Urban Bamboo*, accessed on 18.10.2021, https://www.electronic-star.ro/Pentru-casa/Casa/Urban-Bamboo-gradina-hidroponica-12-plante-24-W-LED-7-litri-bambus.html?gclid=Cj0KCQjww4OMBhCUARIsAILndv5Qh_o9-cbkBVf3DUIRCoSIGIGpcSnZQrCoG23NFq6riEnp450NZOYkaAox4EALw_wcB
- [13] *Mini DC12V Water Pump*, accessed on 18.10.2021, https://www.banggood.com/Mini-DC12V-3M-Micro-Quiet-Brushless-Motor-Submersible-Water-Pump-p-1182617.html?cur_warehouse=CN&rmmms=buy
- [14] Jade Liu, *Imec and XIMEA launch industrial grade hyperspectral camera solution*, Imec Int Press release, accessed on 18.10.2021, <https://www.imec-int.com/en/press/imec-and-ximea-launch-industrial-grade-hyperspectral-camera-solution-addressing-high-quality>
- [15] *Complete farm management solution*, Cropin, accessed on 18.10.2021, <https://www.cropin.com/smartfarm/>
- [16] *Data collection in the cloud with advanced data analysis using MATLAB*, ThingSpeak, accessed on 18.10.2021, <https://thingspeak.com/>
- [17] *Archives of the Agricultural Experience*, UManitoba, accessed on 18.10.2021, https://www.umanitoba.ca/libraries/units/archives/exhibits/agric_exper/agri-food/industry_index.html
- [18] Jean-Luis Rastoin, *The agri-food industry at the heart of the global food system*, accessed on 18.10.2021, <http://regardssurlaterre.com/en/agri-food-industry-heart-global-food-system>
- [19] Francesco Castellano, *Feeding the Future: An Overview of the Agrifood Industry*, TopTal, accessed on 18.10.2021, <https://www.toptal.com/finance/market-research-analysts/agrifood-industry>
- [20] *Performance of the agricultural sector*, EuroStat, accessed on 18.10.2021, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Performance_of_the_agricultural_sector
- [21] Kate Abnett, *EU strikes deal on huge farm subsidies, ending three years of negotiations*, Reuters, accessed on 18.10.2021, <https://www.reuters.com/world/europe/late-night-breakthrough-brings-eu-closer-deal-farm-subsidies-2021-06-25/>

- [22] *European Agricultural Fund*, EU Funds, accessed on 18.10.2021, <https://eufunds.gov.mt/en/EU%20Funds%20Programmes/Agricultural%20Fisheries%20Fund/Pages/European-Agricultural-Fund.aspx>
- [23] *Healthy and sustainable food for all*, EIT Food, accessed on 18.10.2021, <https://www.eitfood.eu/>
- [24] *EIT Food Open Call 2022*, EIT Food, accessed on 18.10.2021, <https://www.eitfood.eu/projects/eit-food-call-for-proposals-2022>
- [25] *Connecting the dots in the agri-tech sector*, Smart AgriHubs, accessed on 18.10.2021, <https://www.smartagrihubs.eu/about>
- [26] *Innovation to boost farming and protect the environment*, European Commission, European Commission, accessed on 18.10.2021, https://ec.europa.eu/info/strategy/recovery-plan-europe/recovery-coronavirus-success-stories/digital/innovation-boost-farming-and-protect-environment_en#smart-farming
- [27] *Acid or alkaline soil: Modifying pH*, Sunset, accessed on 18.10.2021, <https://www.sunset.com/garden/garden-basics/acid-alkaline-soil-modifying-ph>
- [28] Christopher Joyce, *Plants Found to Produce Methane*, NPR, accessed on 18.10.2021, <https://www.npr.org/templates/story/story.php?storyId=5151093&t=1636677969933>
- [29] Nick Greens Grow Team, *Understanding coco coir as a grow medium*, Nick Greens, accessed on 18.10.2021, <https://www.nickgreens.com/post/2018/01/23/how-to-grow-microgreens-using-coco-coir>
- [30] *Grow Microgreens on Paper Towels in 7 Simple Steps*, GroCycle, accessed on 18.10.2021, <https://grocycle.com/growing-microgreens-on-paper-towels/>
- [31] Sara Bakmo, *Sprouting in a Strainer*, Sara Bakmo, accessed on 18.10.2021, <https://www.sarabackmo.com/growing-sprouts-in-a-strainer/>
- [32] *How to use the Stainless Steel Hydroponic Microgreens Grower!*, accessed on 18.10.2021, <https://www.youtube.com/watch?v=ndl4uO8qdyY>
- [33] *Simptome de carenta si de exces in elementele nutritive*, AgroSemFert, accessed on 18.10.2021, <https://agrosemfert.ro/Simptome-de-carenta-si-de-exces-in-elementele-nutritive-la-plantele-legumicole>
- [34] *Overwatering*, Missouri Botanical Garden, accessed on 18.10.2021, <https://www.missouribotanicalgarden.org/gardens-gardening/your-garden/help-for-the-home-gardener/advice-tips-resources/pests-and-problems/environmental/overwatering.aspx>
- [35] *The Mosquito Protocol*, MQTT, accessed on 18.10.2021, <https://mqtt.org/>
- [36] *15 Frequently Asked MQTT Questions*, HiveMQ, accessed on 18.10.2021, <https://www.hivemq.com/blog/15-frequently-asked-mqtt-questions/>
- [37] *The best JSON library for Arduino*, ArduinoJSON, accessed on 18.10.2021, <https://arduinojson.org/>
- [38] *The JSON Data Type*, MySQL, accessed on 18.10.2021, <https://dev.mysql.com/doc/refman/8.0/en/json.html>
- [39] *Dashboard anything. Observe everything.*, Grafana, accessed on 18.10.2021, <https://grafana.com/>
- [40] *Daisy Chaining Arduinos via serial*, Arduino Forum, accessed on 18.10.2021, <https://forum.arduino.cc/t/daisy-chaining-arduinos-via-serial/529802>
- [41] John Boxall, *Daisy-chain Arduinos for very simple networking*, Freetronics, accessed on 18.10.2021, <https://www.freetronics.com.au/blogs/news/6994498-daisy-chain-arduinos-for-very-simple-networking>

Felix HUSAC
Lucian Blaga University of Sibiu
Sciences Department
Str. Doctor Ion Rațiu, Nr.5-7, Sibiu, 550012,
ROMANIA
E-mail:felix.husac@ulbsibiu.ro

Hybrid Classifier based Fall Detection in the Elderly using Still Images

Srianuradha Kandavel, Anca L. Ralescu

Abstract

According to World Health Organization, elderly individuals aged 60 years, or more are at increased risk of suffering often fatal falls. This triggers the need to monitor the activities of elderly and carefully assess them for immediate detection of fall. Video surveillance when combined with machine learning techniques are proven more efficient than wearable devices, although they might cost additional memory space and computation. In this paper, a hybrid classifier that uses still images to identify the human body without bounding boxes and detect a fall action is proposed. The VGG-19 Network is used as a feature extractor, the extracted features are classified using K-Nearest Neighbors classifier. This hybrid classifier has shown approximately 60% accuracy. Its performance may be improved when optimized with hyper tuning of parameters.

1. Introduction

According to the demographic survey of United Nations[1], among the world population, one in 11 people were aged over 65 in the year 2019. In the United States more than 16.9% of the population are senior citizens [2]. At this rate by 2050, there will be 1 senior citizen in every six people across the world. Fall-related injuries rank second in causing fatalities among this age group. Occurrence of fall may also lead to disability, hospitalization, living in nursing home under surveillance etc. thereby reducing the quality of life. This has increased the necessity to identify and monitor the actions of the elderly.

There are various studies for elderly monitoring using wearable devices. However, visual sensors have gained importance due to their availability and convenience. When combined with the computational power of Convolutional Neural Networks (CNN), visual monitoring of fall events has proven reliable. Most of the visual monitoring studies use video surveillance. Even though the accuracy of such systems can be optimized, the issue of managing cost and storage of huge volume of data and computation prevails. This paper provides a methodology for detecting fall events from still images using a CNN for feature extraction and simple classifier for detection. From this point on the paper is organized as follows: Section 2 discusses the background for the study. The study objectives and methodology are described in Section 3. Section 4 explains results and discussion, followed by the conclusion in Section 5.

2. Background

The devices available for fall detection in the field can be majorly divided into wearable and non-wearable devices.

2.1 Wearable Devices

Most commonly available devices for fall detection are Personal Emergency Response Systems (PERS) that consists of accelerometer or gyroscope sensors. Accelerometers detect through the sudden change in the magnitude of acceleration of human body [3]. Gyroscope sensors reads the rotational velocity of individuals during a fall [4]. Barometric pressure sensors are also available to record the differential pressure change that occurs at time of fall event [5]. The elders are expected to wear the devices containing these sensors all the time. This might not be possible due to the degenerative memory that comes with ageing. These PERS are designed in such a way that individual might have to raise the alarm even in emergency cases, which is not feasible[6, 7].

2.2 Non-Wearable Devices

To overcome the shortcomings of wearable devices, scientists explored the use of environmentally mounted sensors such as Infra-Red (IR) sensors, acoustic sensors, floor vibration sensors and video-based sensors. Video-based sensors gained popularity due to the advancement in computer vision field. Traditional cameras and depth imaging sensors such as Microsoft Kinect came into play for providing video surveillance in home and healthcare environments. Main concern among the elders regarding video-based monitoring was invasion of privacy which can be addressed by using the silhouettes of the images. Obtrusion of vision is a common problem faced while using visual-based sensors. Multiple cameras can be set in the living environment of elders to obtain the 360-degree view of the environment and to obtain unobtrusive visuals of individuals. Microsoft Kinect provides various modes for capturing the images such as Red Green Blue (RGB), RGB-Depth (RGB-D), IR images and skeletal tracking mode. The captured videos are processed and applied to convoluted networks for detecting falls [8].

2.3 Related Works

A few of relatable works done with visual sensors and machine learning algorithms in fall detection are as follows. Action recognition in still images depend majorly on the surrounding of the human subject. Liu et al. have proposed a human-mask loss in the still images to identify the human using heat maps [9]. Their system consists of two branches one for locating the human subject and the other to classify the acquired feature map for action recognition based on the class activation map by Zhou et al [10]. They used Stanford40 dataset and MPII Human Pose Dataset to obtain a mean average precision of 94.06% and 40.65% respectively. They have not addressed the possibility of obtrusion by object and their interaction [10].

Zhang et al. proposed a 3D depth image analysis method depending on the relative distance between human depth and floor plane. If the floor plane is near the recognized human depth, fall event is detected. Human in the frame is represented by an ellipse derived from foreground extraction. Based on the transformation of distance and angles from imaging plane to 3D plane fall detection is performed. They used a video sequence filmed in controlled environment [11].

Adhikari et al. extracted the human silhouette from background subtraction followed by feature extraction and activity recognition by a VGG net based CNN [12]. Five human poses were taken into consideration including lying. In most cases, the final pose after the fall event is lying on the ground. So, it is crucial for the model to be sensitive to lying pose. RGB-D dataset with 21,499 images at 30fps is used here. They achieved 81% accuracy at the 17th epoch with 99% sensitivity to lying pose [12].

Fall detection could be considered as a part of Human Action Recognition to identify the occurrence of fall. Sargano et al. adopted a transfer learning technique from AlexNet on KTH and UCF101Action recognition datasets [13]. Weights from the pretrained AlexNet are used for feature extraction as the architecture itself contains 60 million parameters. A SVM-KNN (Support Vector

Machine- K-Nearest Neighbor) classifier is used for classifying the extracted features. 98.15% accuracy has been achieved by this method [13].

3. Objectives and Methodology

Video surveillance has a main disadvantage of invasion of privacy which can be overcome by using the RGB-Depth input for fall detection instead of the RGB input. Video sequences occupy more space and cost more computation than still images. Hence this paper focuses on using RGB-Depth images obtained from the experimental setup by Kwolek et al. from two Kinect sensors [14]. Fall detection methods follow two major algorithms, threshold-based and machine learning based. In threshold-based algorithms, a threshold is fixed based on calculations. When the value from sensors are above the upper threshold and next value falls below the lower frequency, a fall is detected. Most of the wearable devices and sensors use this algorithm. Before 2012, 19 out of 20 fall detection mechanisms have used threshold-based algorithms [15]. Machine learning algorithms have gained recent popularity in fall detection using classifiers and clustering algorithms. Studies have been made for detecting fall using a combination of threshold based and machine learning based algorithms. Visual sensors such as cameras use classifiers and clustering algorithms for identifying a fall event. In this paper, we are dealing with a hybrid classifier made up of a CNN for feature extractions and a KNN for classification of extracted features from RGB-D still images. Instead of training an entire CNN, the weights of VGG-19 Net are used by transfer learning for extracting the features out of RGB-D images. In order to overcome the limited availability of large fall image datasets, available small datasets are used with KNN for classification.

Transfer learning has gained importance after the success of Convolutional Neural Networks in ImageNet competitions. Transfer learning is a technique of using weight from models trained on a particular dataset that can be used to train on completely different datasets. VGG-19 Net is one of the top performing networks trained on 1000 classes with training set of 1.2 million images. These pretrained weights are available in open source to save the user from painstaking task of optimizing the weights for the recognition. Features extracted from the pre-trained network are then fed as an input to the KNN classifier. *Figure 1* shows the proposed methodology for fall detection.



Figure 1: Proposed methodology for fall detection

3.1 VGG-19 Network Architecture

VGG-19 is a version of VGG-16 which is a successor of AlexNet. VGG-19 consists of 16 convoluted layers and 3 fully connected layers [16]. This network takes 224×224 RGB images as input and filter of size 3×3 . Each convolution block consists of multiple filters that slide through the input to produce feature maps. After every layer, a max pooling layer is used to extract sharp and smooth features from the convolutions. For example, in the first convolution block, when an input of size 224×224 is passed through 64 filters an output of $112 \times 112 \times 64$ is obtained after the max pooling layer. To exploit the benefit of using CNN for feature extraction, the output of the final convoluted layer (Block 5) is used ($28 \times 28 \times 512$). In order to use the VGG19 for feature extraction, the output of the three fully connected layers is not required. Hence, they can be eliminated. The extracted feature vector is flattened to prepare as an input for KNN classifier. Visualization of the features from the first convoluted block can be seen in the feature map plotted in **Error! Reference source not found.**

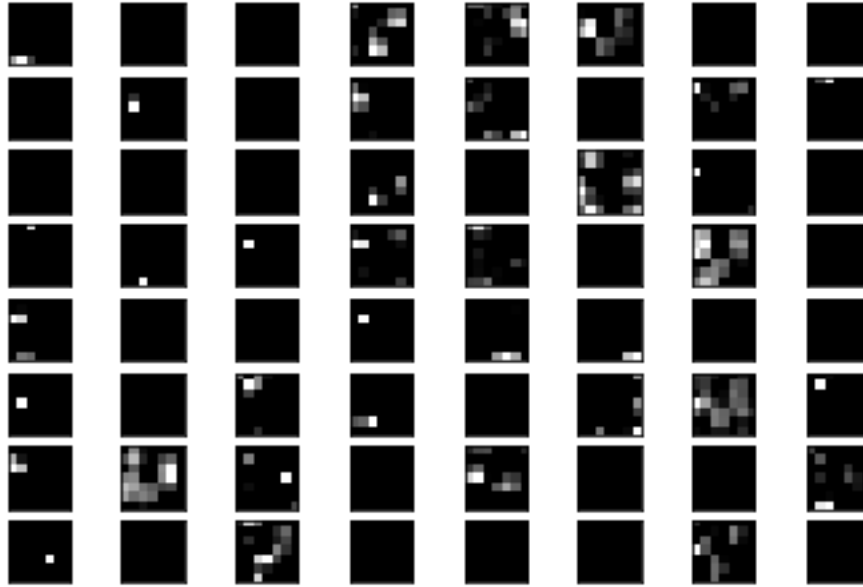


Figure 2: Feature Map extracted from RGB-Depth images

3.2 K- Nearest Neighbors Classifier

K-Nearest Neighbor (KNN) is a simple classifier available in machine learning that has found great use in image classifications. The logic behind KNN is that images of the same category are clustered together and classified based on the distance between them. There are many distance functions such as Euclidean, Manhattan, Minkowski, Chebyshev, Cosine, hamming etc. that can be used with KNN. This study uses Euclidean distance to find the nearest neighbors. The parameter ‘K’ decides the number of nearest neighbors used to vote for the classification of a data point. It is selected by the cross-validation method.

3.3 The Dataset

From the available literatures on RGB-Depth datasets, there is no standard publicly available real-time datasets from Kinect with human actions, let alone still images. Hence researchers usually reproduce fall action in a controlled environment along with non-fall actions. For this proposed method, UR(University of Rzeszow) Fall Detection video dataset has been used [14]. Two Microsoft Kinect sensors were used to record the data and the sensor provides RGB-Depth images at the rate of 30fps. The frames of fall action and daily activities are separated and labeled into two classes labeled as 0 and 1, where 1 indicates fall event. Images are then resized to fit the target size of VGG-19 Net. Two Kinect sensor for capturing fall action provide images in two different perspectives thus increasing the number of available samples. After preprocessing the dataset, 964 fall actions and 2188 non-fall action images are obtained.

Figure 3 shows a sample of RGB-Depth images with daily activities and fall actions from the UR dataset

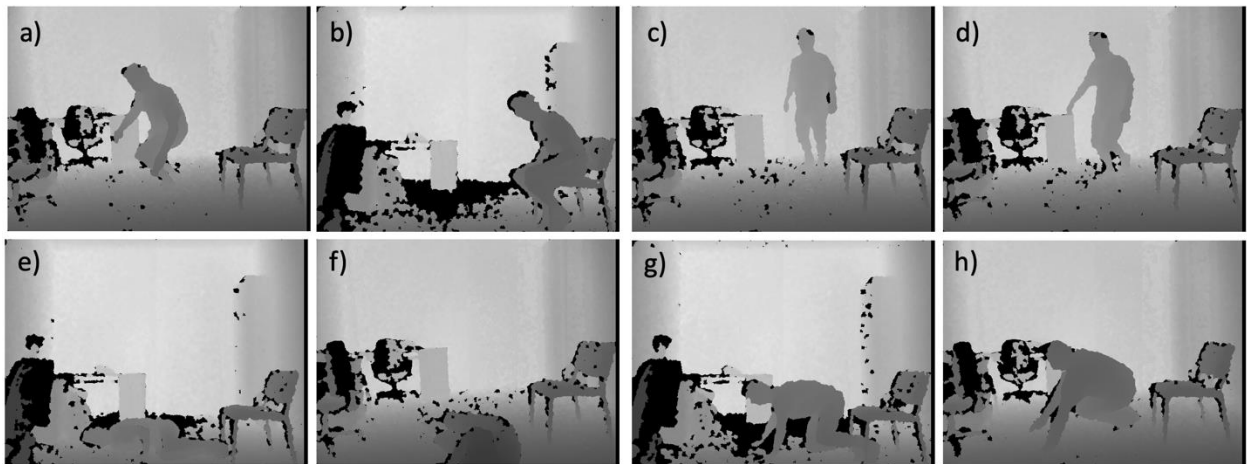


Figure 3: RGB-Depth images from UR Fall Detection Dataset. Images a) to d) represent non-fall actions such as daily activities. Images e) to h) represent fall action

4. Results

The original dataset consists of 30 fall and 40 daily activity sequences [14]. Data is collected in a way to increase the samples from the frames and label them accordingly. Dataset is divided into training, cross validation and testing sets in a proportion of 60-20-20. Among the training data, a small number of images are withheld for cross validation to determine the value of K. Therefore, the training dataset is used to find the nearest neighbors, cross-validation dataset to find best value of K as 3 and the test data to evaluate the model.

Figure 4 below provides the confusion matrix for actual and predicted classes.

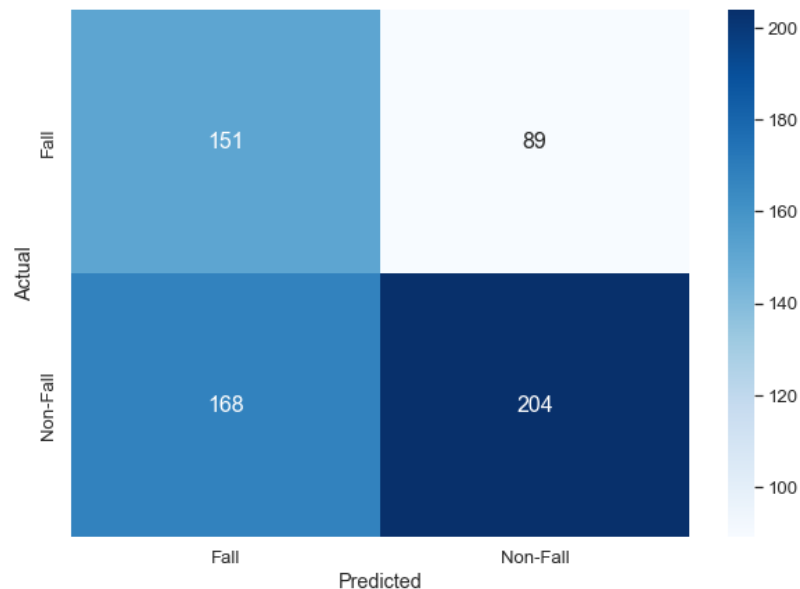


Figure 4: Confusion matrix for fall and non-fall actions. This matrix shows the true positives and true negatives on the diagonal while false positive is represented in 2,1 and false negatives are represented in 1,2

The proposed model was able to achieve 58.83% accuracy with the proposed classifier with depth images from UR fall dataset. The same dataset took a hit when applied to KNN image classifier as it considers the pixel values of the image rather than the extracted features. When applied to the fully connected layers of VGG19 Net, the CNN performed better than KNN classifier, however below the expected performance of the CNN. This might be due to the limited availability of data, as a CNN requires huge data for training. To understand the importance of RGB-Depth dataset, the proposed hybrid model is evaluated with Stanford40 dataset. It can be seen from Table 1, that depth images provide better results when compared to the RGB images.

Table 1: Comparison of proposed model with some of the other classifiers

Model	Accuracy
Proposed Model with UR fall detection data[14]	58.83%
Proposed Model with Stanford40[17]	55%
Simple Classifier (KNN)	47.98%
VGG19 Net[16]	52.7%

5. Conclusions and Future Work

At times when availability of data is limited, transfer learning helps train a model from already established networks by using their weights. The proposed model can reduce the computation complexity and makes it easier to process still images rather than video sequences to detect fall in the elderly. Optimization of parameters can improve the performance of the model. As already mentioned, only 964 fall action frames were available causing an imbalance in the dataset. Data augmentation of the images can increase the number of samples, thus providing a balance in the dataset. Yet, it cannot provide the different types of falling poses. In order to find the parameter K in KNN, other methods of cross validations can be employed. However, this might also increase the computational complexity of the model. Along with these, for future improvements, the hybrid classifier can incorporate other classifiers such as SVM, K-means which may improve the accuracy and specificity of the system.

References

- [1] United Nations, Department of Economic and Social Affairs, and Population Division, *World population ageing, 2019 highlights*. 2020.
- [2] “2019 Profile of Older Americans,” Administration of Aging, May 2020.
- [3] A. Sucerquia, J. López, and J. Vargas-Bonilla, “Real-Life/Real-Time Elderly Fall Detection with a Triaxial Accelerometer,” *Sensors*, vol. 18, no. 4, p. 1101, Apr. 2018, doi: 10.3390/s18041101.
- [4] Q. T. Huynh, U. D. Nguyen, L. B. Irazabal, N. Ghassemian, and B. Q. Tran, “Optimization of an Accelerometer and Gyroscope-Based Fall Detection Algorithm,” *J. Sens.*, vol. 2015, pp. 1–8, 2015, doi: 10.1155/2015/452078.
- [5] W. Lu, C. Wang, M. C. Stevens, S. J. Redmond, and N. H. Lovell, “Low-power operation of a barometric pressure sensor for use in an automatic fall detector,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 2010–2013. doi: 10.1109/EMBC.2016.7591120.
- [6] J. Fleming and C. Brayne, “Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90,” *BMJ*, vol. 337, p. a2227, Nov. 2008, doi: 10.1136/bmj.a2227.

- [7] D. Kunkel, R. M. Pickering, and A. M. Ashburn, “Comparison of retrospective interviews and prospective diaries to facilitate fall reports among people with stroke,” *Age Ageing*, vol. 40, no. 2, pp. 277–280, Mar. 2011, doi: 10.1093/ageing/afq177.
- [8] L. Alhimale, H. Zedan, and A. Al-Bayatti, “The implementation of an intelligent and video-based fall detection system using a neural network,” *Appl. Soft Comput.*, vol. 18, pp. 59–69, May 2014, doi: 10.1016/j.asoc.2014.01.024.
- [9] L. Liu, R. T. Tan, and S. You, “Loss Guided Activation for Action Recognition in Still Images,” *ArXiv181204194 Cs*, Dec. 2018, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1812.04194>
- [10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” *ArXiv151204150 Cs*, Dec. 2015, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1512.04150>
- [11] L. Yang, Y. Ren, and W. Zhang, “3D depth image analysis for indoor fall detection of elderly people,” *Digit. Commun. Netw.*, vol. 2, no. 1, pp. 24–34, Feb. 2016, doi: 10.1016/j.dcan.2015.12.001.
- [12] K. Adhikari, H. Bouchachia, and H. Nait-Charif, “Activity recognition for indoor fall detection using convolutional neural network,” in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, Nagoya, Japan, May 2017, pp. 81–84. doi: 10.23919/MVA.2017.7986795.
- [13] A. B. Sargano, X. Wang, P. Angelov, and Z. Habib, “Human action recognition using transfer learning with deep representations,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 463–469. doi: 10.1109/IJCNN.2017.7965890.
- [14] B. Kwolek and M. Kepski, “Human fall detection on embedded platform using depth maps and wireless accelerometer,” *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 489–501, Dec. 2014, doi: 10.1016/j.cmpb.2014.09.005.
- [15] X. Wang, J. Ellul, and G. Azzopardi, “Elderly Fall Detection Systems: A Literature Survey,” *Front. Robot. AI*, vol. 7, p. 71, Jun. 2020, doi: 10.3389/frobt.2020.00071.
- [16] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv14091556 Cs*, Apr. 2015, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [17] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei, “Human action recognition by learning bases of action attributes and parts,” in *2011 International Conference on Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 1331–1338. doi: 10.1109/ICCV.2011.6126386.

Srianuradha Kandavel
MLCI Laboratory,
EECS Department, University of Cincinnati
OH 45220
USA
E-mail: kandavs@mail.uc.edu

Anca L. Ralescu
MLCI Laboratory,
EECS Department, University of Cincinnati
OH 45220
USA
E-mail: ralesca@ucmail.uc.edu

Design and Implementation of a Node.js Framework for the Development of RESTful APIs

Matthias Keckl

Abstract

Although REST was not widely used to build web services for the first few years after its release, nowadays, most new public-facing APIs are at least trying to follow the REST ideas and principles [2, xvii]. But, unfortunately, although the resulting APIs claim to be REST-compliant, they are often not, as Fielding criticised in his blog post [4]. To ensure REST compliant development, we propose a new Node.js framework that should make it harder to harm the REST constraint than comply with them. We also provide an analysis of currently available Node.js frameworks to demonstrate the lack of tools in the Node.js ecosystem. The conducted empirical study in this work indicates that the proposed framework offers significantly better support for REST API development than existing frameworks.

1 Introduction

Like Meinel and Sack [16, p. 22] describe it, whenever a client, for example with the help of a browser, requests a document from a server, some form of interaction must take place between the two. Thereby, this communication is based on the so-called Hypertext Transport Protocol (HTTP). As Tilkov et al. [3, p. 9-10] points out, Roy Fielding has been involved in its standardization since 1994. For versions 1.0 and 1.1 of HTTP, he used his self-developed HTTP Object Model as a foundation for its design. Then in 2000, he has taken a step further in abstraction and developed the abstract architectural style Representational State Transfer (REST) from the HTTP Object Model in his doctoral thesis [1]. This architectural style is an abstraction level above HTTP since HTTP is a concrete implementation of this style. However, in practice, many HTTP-based APIs are called REST APIs, even though they violate fundamental REST constraints, as Fielding criticizes in his blog post [4]. This is problematic in that one can only use an architecture (REST) optimally if one uses the concrete implementation (HTTP) as intended by the architect (Fielding) [3, p. 10]. Several reasons could be why so-called REST APIs often violate REST constraints, but we will focus on two reasons here. The first reason concerns the style in which Fielding's work is written. The academic style and the high level of abstraction make it hard for the average developer to understand it fully [2, p. 343]. The second reason is the lack of tools to help developers not only interact with the HTTP protocol but also comply with the REST constraints [5]. To address these two issues, we propose a Node.js framework that enables developers to implement REST APIs effectively and at the same time makes it overall harder for developers to violate the REST constraints than comply with them. The strict nature of the framework will still allow developers who do not fully understand the concepts of REST to develop in a REST compliant manner since it forces the developer to obey the constraints. Related work achieving the same object is the model-based approach from Schreibmann and Braun [5].

This paper will first provide a short overview of the Representational State Transfer and its constraints. After that, a few selected currently existing Node.js frameworks are analysed to demonstrate that the lack of tools also applies to the Node.js ecosystem. Then, after presenting the proposed framework, the

implemented framework is evaluated through an empirical study. Thereby, the process and results of this study will be presented. Finally, a conclusion and an outlook on future research topics are given.

2 Representational State Transfer

Before analysing the existing Node.js framework, we will give a brief overview of the architectural style of REST so that we have an idea of the aspects we will consider when examining the framework.

2.1 Architectural Constraints

Richardson et al. [2, p. 350 ff.] give an excellent analysis of the structure and content of Fielding’s work on which the following section is based. In the third chapter of his work, Roy Fielding decomposes different possible network architectures into their architectural properties. These properties can be then combined to describe other common architectures. Afterwards, in chapter five, he defines the Representational State Transfer by composing six architectural properties (constraints), which were previously described in chapter three. Five of these six constraints must be fulfilled for an API to be RESTful since the **Code on Demand** one is optional. The five mandatory ones will be briefly explained in the following.

- **Client-Server** This constraint should be well known since the client-server architecture is the most dominant network architecture.
- **Stateless** The application state is held by the client and is sent with every request the client sends to the server. If some part of the application state should be known by the server (an online shopping cart for example), it should be converted into a resource. Or as Fielding writes in his work [1, p. 78-79], “each request from client to server must contain all of the information necessary to understand the request, and cannot take advantage of any stored context on the server”.
- **Cache** Fielding describes very understandably what this constraint means. “Cache constraints require that the data within a response to a request be implicitly or explicitly labelled as cacheable or non-cacheable. If a response is cacheable, then a client cache is given the right to reuse that response data for later, equivalent requests” [1, p. 79]. The two constraints Stateless and Self-descriptive Messages are responsible for the fact that it can be decided without any further context whether two requests are identical.
- **Layered System** “The layered system constraint is less about proxies and gateways, and more about the fact that adding one between client and server is nearly a transparent operation. A client doesn’t know whether it’s talking directly to a server, or whether it’s talking to a proxy that talks to a proxy that talks to a gateway” [2, p. 353].
- **Uniform Interface** “The central feature that distinguishes the REST architectural style from other network-based styles is its emphasis on a uniform interface between components” [1, p. 81] and “[t]he uniform interface constraint covers most of the things that make the Web unique” [2, p. 350] underlines the great relevance of this constraint. Therefore, in the following we provide a detailed description of the four constraints that make up the Uniform Interface based on Richardson et al. [2, p. 347-350].
 - **Identification of Resources** Every resource must be identified by a URI and even though if the resource state may be changed the URI remains unchanged. Thereby, a resource can be any information that can be named, containing a document, an image or a collection of other resources [1, p. 88].
 - **Manipulation of Resources Through Representations** A resource can model any kind of information used in the business logic of a Web API. For example, one could have a user modelled as a resource. But since the user, maybe currently held in the main memory, cannot be sent over the network. Therefore we need the concept of a representation. “A representation is a sequence of bytes, plus representation metadata to describes those bytes” [1, p. 90]. A

common representation format is the JavaScript Object Notation (JSON). So one could convert the user in the main memory to JSON and send it to the client that way (so this would be the sequence of bytes). The Content-Type header set to application/json would then be the metadata describing the bytes.

- **Self-Descriptive Messages** “An HTTP message contains all the information necessary for the recipient to understand it. There’s no free-floating documentation nearby that clients are also expected to understand” [2, p. 349].
- **Hypermedia as the Engine of Application State (HATEOAS)** This constraint makes it mandatory for the server to include hypermedia controls in each response to the client, thereby hypermedia controls describe possible state transitions the client could make [2, p. 360]. To put it in one sentence, the server proposes possible state transitions and the client decided which one it takes. HTML links or forms are common example that represents hypermedia controls [2, p. 360].

2.2 Key Principles

Since, for example, the **Layered System** constraint is more important when it comes to the deployment of web APIs anyway [2, p. 351], we think Tilkov et al. [3, p. 11 ff.] provide a more practical and short and easy to understand explanation of the key principles of REST.

- **Resources with unique identification** One should use URIs to identify all the instances of all the essential abstractions of an application that are worth it.
- **Links/Hypermedia** Use hypermedia to define relationships between resources wherever possible and control the application flow through hypermedia controls.
- **Standard methods** All resources should implement the application protocol (HTTP) correctly. Thus, a GET request should always be safe and idempotent for example. This way all clients know how to interact with the application’s resources.
- **Different representations** Different representations of the same resource should be provided for different requirements. Each client can then, through the content negotiation mechanism, request the desired representation format.
- **Stateless communication** The application state should be held by the client or transformed to a resource. The server should only be aware of the client for the duration of a single request and forget about it immediately afterwards.

3 Analysis of existing Node.js Frameworks

To demonstrate that there are no appropriate tools for developing RESTful APIs in Node.js either, we provide an analysis of existing frameworks. Therefore, we first create a template with which the selected frameworks are analysed. Afterwards, we choose four frameworks representing the currently available Node.js Frameworks. Finally, the analysis results are presented.

3.1 Template Creation

When creating the template, we only consider the aspects of the frameworks that support the developers in the concrete implementation of REST APIs. For this reason, for example, principles such as **Resources with unique identification** and **Client-Server** are not considered, as these are primarily relevant in the design of REST APIs. So our template consists of 6 aspects, which we consider the most practical relevant for the RESTful development.

- **Standard method (M)** The framework should support the developer in using the HTTP methods RFC 7231 [6] compliant. For example, the GET method should not be used to generate a resource. Likewise, the server's response to a POST request, for example, should include a location header that contains the URI of the resource just created.
- **Hypermedia (HM)** As HATEOAS is a mandatory constraint, a framework should support the developers to define relationships between resources and to provide the possible state transitions for each state.
- **Pagination (P)** Pagination affects only collection resources, so a list of single resources. Such a collection could contain millions of resources, but since the server does not want to serve all these resources in a single response, it should implement pagination strategies [2, p. 101]. And once a server implements a pagination strategy, it must provide the client links with the correct relation types (see RFC 5988 [7]) to navigate through the collection. Accordingly, we examine how the frameworks help to implement different strategies and generate the corresponding links.
- **HTTP Caching (C)** Caching is also a mandatory constraint. Like Tilkov et al. [3, p. 138] suggest, we assume that the caching is delegated to the infrastructure via the Cache-Control header. Thus, we analyse whether the frameworks support the developers to use the HTTP caching mechanisms defined in RFC 7234 [8].
- **Conditional Requests (CR)** Defined in RFC 7232 [9], conditional requests are an effective HTTP mechanism for updating caches and to prevent the lost update problem. Since these are also used in conjunction with the HTTP caching mechanisms, a framework should provide support for this as well.
- **Content Negotiation (CN)** As already mentioned in the previous section, one should provide different representations for different requirements. The HTTP mechanism used to let the client decide which representation it needs is called content negotiation, defined in RFC 7231 [6]. Since a different representation could mean another data format (e.g. JSON and XML) but also a different view on the same resource [2, p. 239], the framework should not only implement content negotiation RFC compliant but also allows the developer to define different routes handler for different representations, because various views may require deviating processing steps like additional database calls.

3.2 Analysed Frameworks

Before we present the results of the analysis, we first provide an overview of the frameworks to be examined.

3.2.1 Express

Express [13] reach version 1.0.0 already in 2010 and thus the oldest analysed framework. It is the first analysed framework because “Express.js is sometimes considered the de facto solution when it comes to building a web application in Node.js, much like Ruby on Rails was for Ruby for a long time” [10, p.31].

3.2.2 NestJS

NestJS [15] is, in contrast to the other three analysed frameworks, a very opinionated and feature-rich framework. Internally, NestJS uses Express or Fastify as an HTTP server framework. As the website states, “Nest provides a level of abstraction above these common Node.js frameworks (Express/Fastify) but also exposes their API directly to the developer. This gives developers the freedom to use the myriad of third-party modules which are available for the underlying platform”. Consequently, we will focus on the features offered by NestJS and not on these provided by Express or Fastify under the hood.

3.2.3 Fastify

Fastify [12] is built with a focus on performance and developer experience. According to their benchmarks Fastify is one of the fastest Node.js frameworks and up to 4.6 times faster than Express. Fastify is built so that it has very few built-in features but can be extended by a variety of plugins developed and maintained by the Fastify team. We consider these less as external dependency and more as built-in functionality since they are developed by the Fastify team.

3.2.4 Restify

Restify [14] was developed with the goal of providing a framework that enables the developers to build “semantically correct RESTful web services ready for production use at scale” as it is stated on its website. Another goal is to be as compliant as possible with the RFCs.

3.3 Results

For the evaluation of the individual aspects, a scaling rate is used, which Fielding also uses in his work [1]. Thereby, the more plus an aspect has, the better the respective framework supports the developers in this area. The same only the other way around applies to the minus. To the best of our knowledge, there is no literature that addresses the evaluation of frameworks and their support for REST APIs. For this reason, the current state of the source code, the documentation and the website, if available, is used as a basis for the analysis

Framework	M	HM	P	C	CR	CN
Express	-	-	+	+	+	+
NestJS	-	-	-	-	--	-
Fastify	-	-	-	+	++	+
Restify	-	++	+	-	-	++

Table 1: Results of the framework analysis

As one can see in Table 1, none of the analysed frameworks supports the developer with all aspects. For example, even Restify, developed for RESTful development, lacks support for conditional requests, caching, and the correct usage of the HTTP methods.

4 State-based Framework

Our proposed State-based framework pursues two objectives. First, it should support the developers in applying the aspects mentioned above to their API and making it harder to harm the REST constraints than to comply with them.

Before we demonstrate the exemplary development with the implemented framework, the fundamental design principles and the state templates are presented.

4.1 Design Fundamentals

As described above, the goal is to design a framework that enables developers to build RESTful APIs efficiently and thereby also enforces strict compliance with the REST constraints. Therefore we will first have a look at the framework design that enforces the REST compliant development.

4.1.1 HTTP Server Framework

One design decision is to use an existing framework to reduce the development overhead. Around this framework, we build a wrapper to ensure REST compliant development. Fastify was chosen as the base HTTP server framework because it is easy to extend and its performance. Thereby it was a conscious decision against Restify, although it scored best in the analysis. The reason for this is the poorer extensibility compared to Fastify. Since neither Fastify nor Restify offers adequate support, the decision fell on the framework with the easiest extensibility.

4.1.2 The Idea of State Templates

While developing a Web API that is REST and HTTP compliant, one will recognize two circumstances. The first one is the repetitive tasks that must be implemented at each endpoint. These include the link generation for the pagination or the implementation of conditional requests, for example. The second one is that the sequence in which the tasks are processed is usually the same. To overcome the repetitive tasks, each endpoint is implemented as a whole class instead of a single method. This design allows common tasks to be refactored into an abstract base class. We apply the template pattern [11, p. 325 ff.] to the abstract class to tackle the second issue. This skeleton, specified in the template method, consists of concrete methods, implementing the repetitive tasks, and abstract methods. The developers must then implement these abstract methods to do the request specific tasks, such as the database access. We obtain five different classes if this approach is applied to the four HTTP CRUD methods since each method requires a different sequence flow. When these five classes are now subsequently analysed, one can see that many functionalities are implemented in all classes. Since there is no need for the duplicated code, further abstract classes are created, which build the base classes for the previous five classes. These contain all common functionalities used by all subclasses, such as authentication and methods that enable hypermedia. These steps result in the class hierarchy seen in Figure 1. Due to this, the developers are supported to develop RESTful APIs efficiently since the most required functionalities are already implemented in the base classes. Further, the development is also less error-prone because the framework and not the developers implement most functionalities. Additionally, they are supported and even enforced to consider the most important aspects, like hypermedia, through the need to implement the abstract methods defined in the base classes. Thereby, the developers can use hook methods to extend and change the behaviour of the states. This is a common approach for the template pattern and is described in [11, p. 326 ff.]. As one can see in figure 1 the classes are named like States. This is because we see one endpoint as one state in the application since a REST API should behave like a virtual state machine [1, p. 109].

4.1.3 Router

The next part that should enforce the correct development is the so-called Router. Router refers to the component that allows developers to register endpoints that listen to specific paths. The Router does not enforce any structure regarding the registered paths, but it does enforce how an HTTP method-specific request gets processed very strictly. To provide an example, an endpoint that is listening for GET requests would have to return an instance of a class that inherits from AbstractGetState. This is because the request processing happens in the state itself. This has the advantage that the developers can be guided more easily within the state by forcing them to implement specific abstract methods.

4.2 State Templates in Detail

The HTTP method-specific states will be described after introducing the two base states AbstractState and AbstractStateWithCaching, with their common functionalities. Thereby the focus will be on the sequence of tasks they define in their template method. To give a graphical overview of this, we define a graphical model to represent it. This model is shown in Figure 2. This model is used to represent the sequence of the individual tasks. A distinction is made between tasks that are already implemented and tasks that must be implemented by the developers in the respective concrete class implementation. It also shows when and with which HTTP status code the processing of the request is completed.

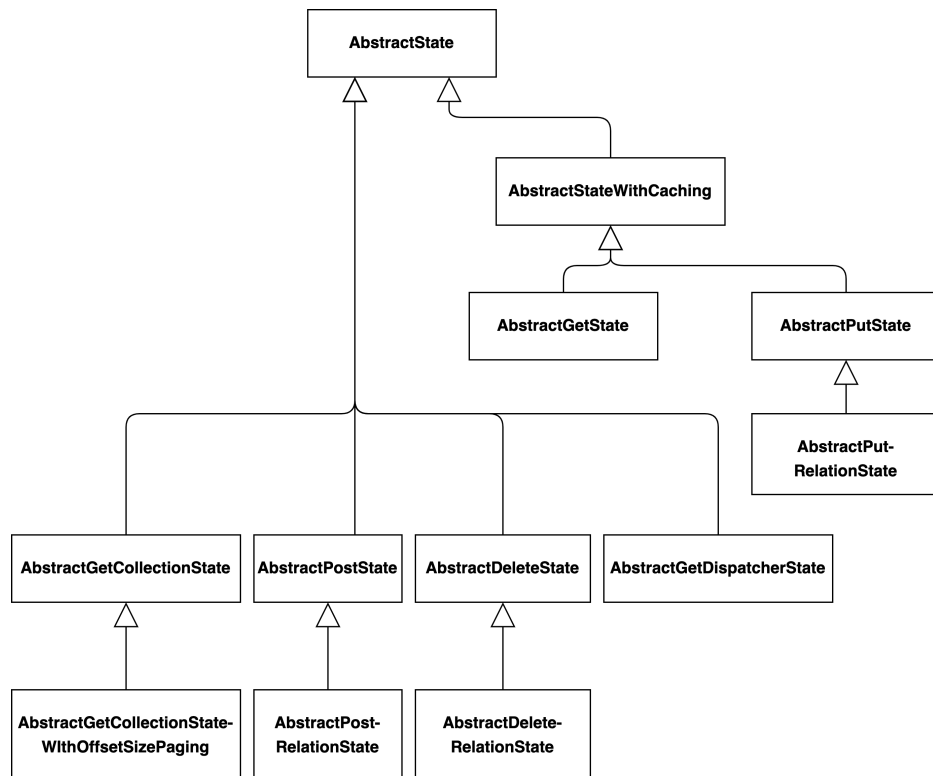


Figure 1: The resulting class hierarchy when refactoring repetitive tasks and common functionalities into abstract base classes for the CRUD HTTP methods.

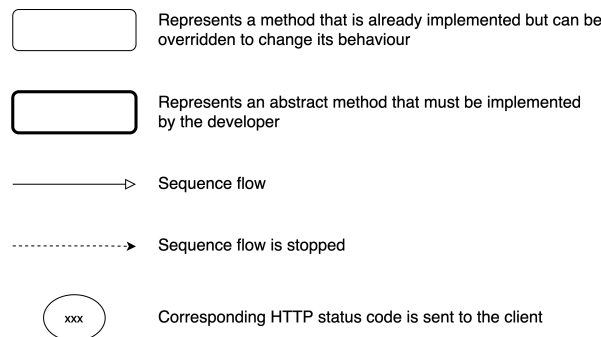


Figure 2: Definition of the graphical model for the visualization of the sequence flows of the state classes.

4.2.1 AbstractState

AbstractState is the base class of any other state class in the framework. The key aspect of this class is its very limited public interface. It only provides two public methods. On the one hand, this is a method to give the state access to the request and response object and, on the other hand, the Template method, which initiates the processing of the request. Through the request and response object, all other methods within the state can access the information and data of the request. This includes the request payload, query parameters, path parameters and headers. Through the response object, among other things, the response payload, the status code and response headers can be set within the state. The reason for this restricted public interface is to give developers as few opportunities as possible to make mistakes. With the current implementation, they only have to call the method, which passes the request and response object to the state, and then the template method to process the request. Everything else is done in the

state classes, in which the abstract methods guide the developers. Besides the two public ones, there are protected methods that support the developer to implement role-based access control, define the possible state transitions, define state entry constraints, convert links, and extract the required data from the request.

4.2.2 AbstractStateWithCaching

The `AbstractStateWithCaching` extends the `AbstractState` by caching utilities. Hence, the class provides therefore four different caching types:

- Deactivate caching by appropriate Cache-Control header
- Define Caching behavior via Cache-Control header
- Define Caching behavior via Cache-Control and ETag header
- Define Caching behavior via Cache-Control and Last-Modified header

4.2.3 AbstractGetState

The `AbstractGetState` is the base class for every state that represents GET endpoints. Since such a request always corresponds to a resource, the developer has to provide the resource class as a generic parameter. The sequence flow is shown in Figure 3. As one can see in the sequence flow, this state does support conditional requests and sends automatically a 304 Not Modified response if the client still knows the current state of the resource. Additionally, the state checks if the resource with the requested id exists, and if not, it sends a 404 Not Found response. Thereby state entry constraints are conditions that must be fulfilled so that the client is granted access to this application state. These conditions are checked in the `verifyAllStateEntryConstraints` method. Due to the scope of this work, we cannot go into more detail here about other methods within the sequence. For a detailed explanation see Keckl [17].

4.2.4 AbstractPostState

The `AbstractPostState` is used to create resources. Since the client does not send the resource directly but a representation of the resource, this class requires two generic parameters: the resource class and the representation/view class. This state also supports the process of creating the resource model from the view model. The logic for this is placed in the `mergeViewModelToDatabaseModel` method. Therefore, the method first receives an instance of the resource class from the abstract method `createDatabaseModel`. In this method, the developers can set values required to create the resource, but which are not allowed to be set by the client, like user roles, for example. The next step is to merge the view class with the resource class by the merge method. The method signature matches precisely this behaviour. An instance of the view class can be taken and merged with an instance of the resource class and retrieve an instance of the resource class as a result. This result, saved in the class attribute `modelToStoreInDatabase`, is then used within the `createModelInDatabase` method to save it in the database layer. Besides this, the state also sets the Location header automatically using the identifier created by the database layer.

4.2.5 AbstractPutState

This state is used when the client updates a resource. For this purpose, it is required first to fetch the current state of the resource from the database. This model is then merged with the representation sent by the client. The merge method used for this takes each view property and sets its value in the resource to be updated. If a property was intentionally not set by the client, the corresponding property is set to undefined in the resource. The result of the merge process should be afterwards saved within the `updateModelInDatabase` method. The response to such a request is either empty or contains the updated resource.

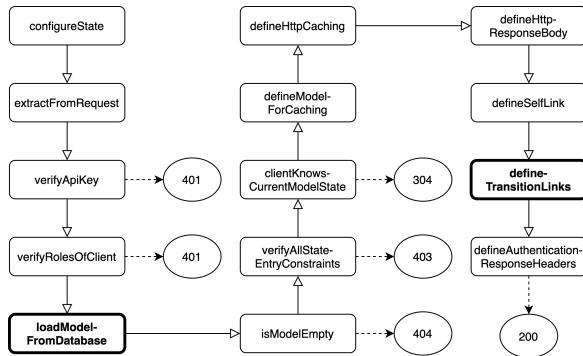


Figure 3: Sequence flow AbstractGetState.

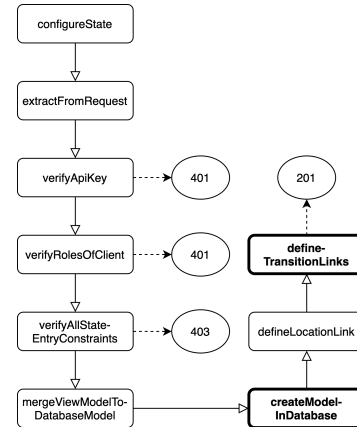


Figure 4: Sequence flow AbstractPostState.

4.2.6 AbstractDeleteState

Just like the AbstractGetState the AbstractDeleteState refers to a resource, and therefore a generic parameter is required. Before the resource with the specified identifier is deleted, the state completes two checks besides the authentication and authorization. First, it verifies whether or not the resource exists. In case the resource is missing, a 404 Not Found response is sent. After that, it checks if the client knows the current state of the resource. If it does not, it will receive a 412 Precondition Failed response. For this purpose, the method *createEntityTagOfResult* is used to generate an ETag of the resource fetched from the database layer. This ETag is then compared with the ETag sent by the client. If it is configured so, the response payload of the respective request will contain the current state of the resource before it was deleted. By default, the response payload is empty.

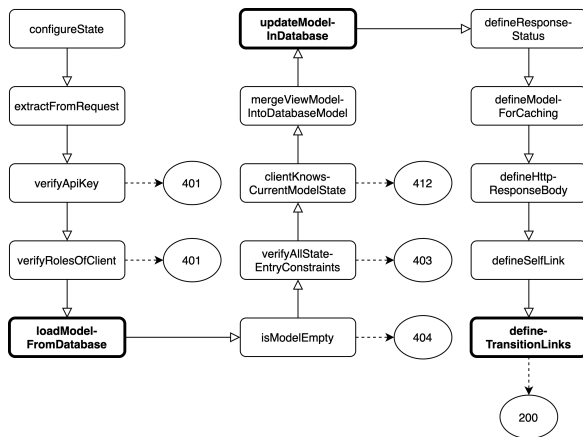


Figure 5: Sequence flow AbstractPutState.

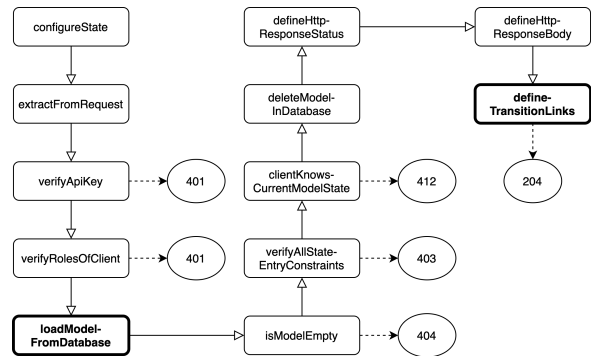


Figure 6: Sequence flow AbstractDeleteState.

4.2.7 AbstractGetCollectionState

The AbstractGetCollectionState represents a request to get a collection of resources. Since such a collection resource could contain millions of entries, the server should implement pagination strategies in a single response, but the server does not want to serve these in a single response [19, p. 101]. Hence, the developer has to specify a pagination strategy via the *definePagingBehaviour* method. With the returned class, which must be a subclass of the *AbstractPagingBehaviour*, the links to navigate through the collection are added within the *definePagingLinks* method as HTTP Link headers. Next to the links,

HTTP headers are also set to provide the client with information about the number of results and the total number of results. In addition, the framework provides the developers with the *PagingBehaviourUsingOffsetSize* as a built-in pagination strategy. Finally, the caching behaviour is set by default to "no-store, no-cache" since the frequency of changes is very high for collection resources.

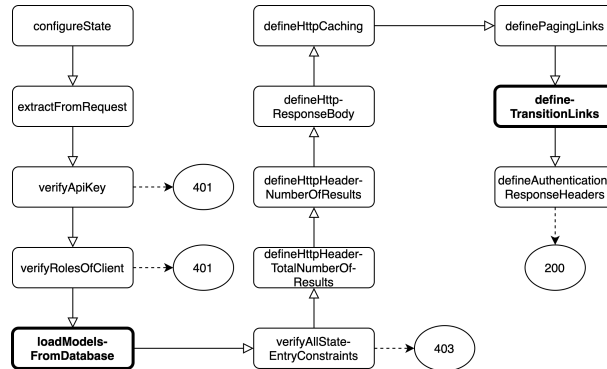


Figure 7: Sequence flow AbstractGetCollectionState.

4.2.8 RelationStates

As one can see in Figure 1, there are several RelationStates. These are used when sub-resources are processed, so when the client wants to interact with a resource related to another resource. An example of this could be a client requesting all comments from a specific user. The HTTP request could then look like below, for example. In this example, the requested list of comments are the sub-resource.

```
GET /users/4392/comments HTTP/1.1
Host: localhost:8080
Accept: application/json
```

4.2.9 AbstractGetDispatcherState

The AbstractGetDispatcherState acts as the entry point to the API. The primary responsibility is to provide the client with all possible state transitions to start interacting with the API. This could be the link to the login page, for example.

4.3 Using the Framework

After explaining the framework's design in the previous sections, the following one presents the exemplary development process with this framework. Therefore, a use case with the User resource is considered. For this resource, we want to implement an endpoint to receive a single user. With this endpoint, one receives a JSON representation of the requested user resource. To do this, the user class must be defined first. Afterwards, the corresponding class GetSingleUser have to be implemented. The database access, as well as the definition of the possible state transitions, takes place in this. The generation of the ETag and the setting of the Cache-Control header happen automatically within the AbstractGetState class. Finally, the state must be registered by the router. Therefore the developer must provide the state class access to the request and response object, defining a media type and give a view converter, which converts the user to JSON.

```
class User {
    id: number
    name: string
    password: string
}
```

```
@State()
export class GetSingleUser extends AbstractGetState<User> {
  constructor(private readonly userRepository?: UserRepository) {
    super()
  }

  protected defineTransitionLinks(): Promise<void> | void {
    this.addLink('/users', 'getAllUsers', 'application/vnd.user+json')
    this.addLink('/users/{id}', 'deleteUser', [this.requestedId])
    this.addLink('/users/{id}', 'updateUser', 'application/vnd.user+json', [
      this.requestedId,
    ])
  }

  protected loadModelFromDatabase(): Promise<SingleModelDatabaseResult<User>> {
    return this.userRepository.readById(this.requestedId)
  }
}

@Controller()
class UserController {
  @Get<GetSingleUser>({
    produces: 'application/vnd.user+json',
    // userViewConverter converts the User class to JSON
    viewConverter: userViewConverter,
  })
  public getSingleUser(
    req: HttpRequest,
    res: HttpResponse
  ): Configured<GetSingleUser> {
    return new GetSingleUser().configure(req, res)
  }
}
```

5 Empirical Evaluation of the implemented Framework

This study has only a small number of subjects, and accordingly, the results are only partially meaningful. Nevertheless, it can be taken as the first indication for further studies. Accordingly, we only provide a brief overview of the study procedure and discuss the most important findings.

5.1 Survey Setup

To present the general structure of the study, we provide an overview of the five steps the study consists of.

1. The 8 study participants filled out an initial questionnaire about their knowledge of the technologies required for the study (Node.js, TypeScript, REST, general software development experience)
2. After this initial questionnaire, all participants attended a one hour workshop. In this workshop, the most important aspects of REST were presented to create a basic understanding of this topic among every participant.
3. As third step, all participants had to work on a coding challenge. In this task, they had to develop a small REST API that allows them to create a book resource, read it, and read a list of books.

Thereby an offset-size pagination strategy should be implemented. Special attention was also paid to the correct implementation of the hypermedia principle.

4. The participants were then divided into two groups. One group (group A) solved the task with Express.js and the other (group B) with the framework developed here.
5. Finally, all participants had to fill out one final questionnaire.

It must be noted that of the initial 8 experimental participants, only 2 completed the coding task. However, 3 participants have at least started the coding task. Therefore, the data of these participants was also taken into account.

5.2 Discussion of Key Results

We identify and discuss three questions whose answers and results are particularly interesting. The numerical results are given as a 2-tuple defined as follows: (\bar{x}, σ)

- **From 1 (very easy) to 7 (very hard), how easy did you find it to develop an API using REST principles?** Participants from group A (4, 1.41) found it slightly harder to develop REST compliant compared to group B (3.33, 0.58). Even though there is no significant difference between these two groups, it is a first indicator that the developed framework supports the implementation of the REST principles. However, regardless of the framework used, study participants still tended to find it challenging to be REST compliant since both values are above 3.
- **How much did the framework you used from 1 (very supported) to 7 (not supported at all) help you develop REST compliantly?** However, if the participants are asked how much the framework used in each case helped with REST compliant development, one gets a much more evident result. While group A (6.5, 0.71) rated the support of Express.js very poorly, the participants from group B (1.67, 0.58) rated the support of the framework developed here as very good.
- **From 1 (very easy) to 7 (very hard) how challenging was it to implement the hypermedia principles?** As Richardson et al. [2, xviii] points out, hypermedia is a topic that many developers still do not fully understand. Therefore, the results on this question are particularly interesting. And also, in this question, it appears that the framework developed here successfully supports RESTful development, also related to the more difficult hypermedia aspect. Results for group A was (4, 0) and for group B (2.67, 2.08).

6 Conclusion and Future Work

In this paper, we proposed a new Node.js framework for the efficient development of RESTful web APIs. With the very strict state-class-based approach of the framework, we try to overcome the two problems: the lack of tools and the hard-to-understand work of Fielding and thus also the misunderstanding principles of REST. To demonstrate that the lack of tools can also be applied to the Node.js ecosystem, we provided an analysis of existing frameworks. The result of this analysis was that none of the tools provided satisfying support. The key difference between these analysed frameworks and the one presented here is the class-based approach. It provides an abstract base class for each HTTP CRUD method. These classes force the developers to work REST compliant. Furthermore, they automate repetitive tasks and so increase development efficiency. However, the framework only supports the work with JSON representations. Thus, future work could include support for other representation formats such as HTML, XML or binary formats like JPEG. The presented empirical study indicated that the realised framework does support the developers while implementing REST APIs. However, because only a few participants took part in the study, future work should conduct studies with a larger number of participants. Furthermore, when designing the coding tasks, care should be given to ensure that they are not too time-consuming for the participants, as this would lead to dropouts, as we experienced in our study.

Acknowledgement: This work was supervised by Prof. Dr. *Peter Braun*, from *University of Applied Sciences Würzburg-Schweinfurt*. This work benefits from funds given by Prof. Dr. Peter Braun.

References

- [1] Roy T. Fielding, *Architectural Styles and the Design of Network-based Software Architectures*, PhD thesis, University of California, 2000
- [2] Leonard Richardson et. al., *RESTful Web APIs. Services for a Changing World*, O'Reilly, 2013
- [3] Stefan Tilkov et al., *REST und HTTP*, dpunkt.verlag, 2015
- [4] Roy T. Fielding, *REST APIs must be hypertext-driven*, 2008, URL: <https://roy.gbiv.com/untangled/2008/rest-apis-must-be-hypertext-driven> (visited on 11/09/2021)
- [5] Vitaly Schreiber and Peter Braun, Model-driven Development of RESTful API, *WEBIS*, pp. 5-17, 2015
- [6] Roy T. Fielding and Julian Reschke, *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*, RFC Editor, 2014, URL: <https://rfc-editor.org/rfc/rfc7231.txt> (visited on 11/09/2021)
- [7] Mark Nottingham, *Web Linking*, RFC Editor, 2010, URL: <https://rfc-editor.org/rfc/rfc5988.txt> (visited on 11/09/2021)
- [8] Roy T. Fielding et al., *Hypertext Transfer Protocol (HTTP/1.1): Caching*, RFC Editor, 2014, URL: <https://rfc-editor.org/rfc/rfc7234.txt> (visited on 11/09/2021)
- [9] Roy T. Fielding and Julian Reschke, *Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests*, RFC Editor, 2014, URL: <https://rfc-editor.org/rfc/rfc7232.txt> (visited on 11/09/2021)
- [10] Doglio Fernando, *REST API Development with Node.js*, Apress, 2018
- [11] Erich Gamma et al., *Design Patterns*, Addison-Wesley professional computing series, 1994
- [12] Fastify, URL: <https://github.com/fastify/fastify> (visited on 11/09/2021)
- [13] Express, URL: <https://github.com/expressjs/express> (visited on 11/09/2021)
- [14] Restify, URL: <https://github.com/restify/node-restify> (visited on 11/09/2021)
- [15] NestJS, URL: <https://github.com/nestjs/nest> (visited on 11/09/2021)
- [16] Christoph Meinel and Harald Sack, *WWW*, Springer, 2004
- [17] Matthias Keckl, *Design and Implementation of a Node.js Framework for the Development of RESTful APIs*, 2021, URL: https://github.com/matthyk/bachelor_thesis/blob/main/bachelor_thesis.pdf (visited on 25/11/2021)

Matthias KECKL
University of Applied Sciences Würzburg-Schweinfurt
Faculty of Computer Science and Business Information Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: matthias.keckl@outlook.de

System of modeling the spreading of infectious diseases

Yulian Kuryliak

Abstract

This paper describes open source software that is designed to predict an epidemic outbreak. The system was developed with Python programming language and allows simulations for such epidemiological models as SI, SIS, SIR, and SEIR, as well as models of the Erdős–Rényi, Watts-Strogatz, and Barabási–Albert networks. The system, which accepts a large number of parameters, has a beneficial impact on the research activities, but at the same time, it has a simple interface in the form of a dashboard that can be used by an average person. These parameters include the number of nodes and edges in the network, the disease infection rate, Time a person spent in a particular state, the demographic distribution, and statistics on the mortality of the disease, and so on. The simulation is performed using the accurate and efficient Gillespie algorithm.

1 Introduction

Given the recent outbreaks of epidemics, predicting the possible consequences of them and developing methods of control is an extremely important issue. Most epidemic outbreak simulation programs and literature are based on differential equations, mean-field theory, and asymptotical analysis [12]. Some of the stochastic methods use methods with discrete-time models [3], which does not allow to realistically predict the time between infections. All these methods are not realistic enough. Also, the a lot of other systems focus on the number of reproductions R , but not on the time to the next infection. In addition, classical epidemiology and the vast majority of research [13, 15, 10] use a complete graph, graphs with the same number of contacts or a random Erdős–Rényi graph [5] as a network in which the disease spreads, and does not often focus on more realistic networks, such as the Watts-Strogatz small-world model [14] or the Barabási–Albert scale-free model [1].

Therefore, the challenge is to develop a software that could realistically predict the time between infections and estimate the number of infected nodes at a given point in time. Also, the software has to be understandable to the average user. The developed software makes it possible to concentrate on the peak number of infected and critically infected. In addition, the software has to pay into account the contact network and provide an opportunity to choose an epidemiological model [7] (SI, SIS, SIR, SEIR, where S is susceptible, I is infected, R is removed, E is in the incubation period), and to set the time spent in states to simulate outbreaks of various diseases. Such software can be used as an example in the medical field to prepare hospitals to deal with the consequences of a disease and prepare places in a time.

In this work, the Gillespie algorithm [6] is used to simulate an epidemic outbreak, which, unlike methods that use Markov chains with discrete-time, makes it possible to realistically predict the time to next event (in our case, the event is an infection of a node). This algorithm allows taking into account the network topology with its possible changes in the simulation process.

2 Description of the simulation algorithm

To simulate the spread of the disease, the stochastic Gillespie algorithm is used, which is based on Markov chains with continuous-time [11] for realistic simulation in time. This algorithm allows determining the

time until the next infection of a node depending on the total number of contacts of infected nodes with susceptible nodes in the network. Therefore, this method allows focusing on a particular node of the network, taking into account its features and position in the network. Also, the advantage of this method is the ability to perform various actions on network nodes in the simulation process (for example, to introduce quarantine restrictions at a certain point in time).

The Markov chain is given by a matrix of transitions Q with dimension $2^N \times 2^N$, where q_{ij} is the characteristic of the probability of transition from state i to state j , q_{ii} is the characteristic of the time after which the transition to the next state will take place (infection of a new node will take place), N is the number of nodes. At one chosen point of time possible to be only one infection and in our implementation recovery of node takes place after a constant period of time. Thus, most of the values of the transition matrix are 0, so, we do not calculate probabilities of transitions to impossible states but only for possible states for the next transition and avoid state space explosion in this way. Therefore, Gillespie's algorithm allows the performance of accurate and efficient simulation.

Let i denote the current state and i_{+j} denote a network state where node $j \in [1, N]$ is a node that potentially gets infected in addition to the previously infected nodes in state i . Then

$$q_{i,i_{+j}} = \begin{cases} \lambda v(j), & \text{if node } j \text{ is susceptible;} \\ \sum_{l, l \neq i} q_{i,i_{+l}}, & \text{if } i == i_{+j}; \\ 0, & \text{if node } j \text{ is not susceptible.} \end{cases}$$

Here λ is the infection rate of the virus, $v(j)$ is the number of infected neighbors of the vertex j in state i . The $v(j)$ for all nodes is calculated by the multiplication of the incidence matrix by the column vector of infectious nodes, in which 1 is contagious, 0 is not contagious. The resulting string vector is multiplied element by element by a string of susceptible nodes in which 1 is susceptible, 0 is not susceptible.

The time to the next infection is calculated by exponential distribution $F(x) = \theta e^{-\theta x}$, for $x \geq 0$, where $\theta = q_{ii}$. The expected value M is given by $1/\theta$, and in the CTMC by $1/q_{ii}$. (ie, the mathematical expectation of time to the next infection will decrease with increasing sum of infected neighbors of each of the susceptible network nodes). An example of time prediction is shown in Fig.1.

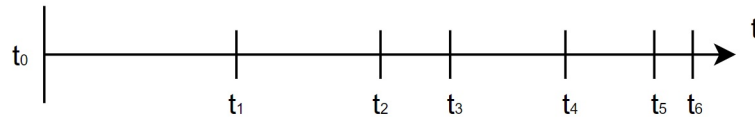


Figure 1: Example of expected time

The probability of infection p_{ij} is determined by the formula: $p_{ij} = q_{ij} / \sum_{l, l \neq i} q_{il} = q_{ij} / q_{ii}$. The state to which the system enters (ie, the node to be infected) is selected by "roulette wheel" [9]. According to this principle, we generate a random number z in $[0, 1)$ on a uniform distribution, and choose such k (the vertex to be infected) that the inequality holds: $\sum_j^{k-1} p_{ij} < z < \sum_j^k p_{ij}$. The state of the nodes is updated before each transition to a new state, as well as the collection of information about the state of the network.

3 System analysis

3.1 Purpose of the system

The main purpose of the system is to provide opportunities for simple research to predict the spread of the disease in various networks. At the same time, the system should be simple enough for ordinary, untrained users to understand, to assess the possible consequences caused by a particular disease. That is, the purpose of the system is both research and promotion of epidemiology. For a better understanding, the system should provide the ability to present the results in both tabular and graphical form.

3.2 Use cases of the system

There are two main options for using the system: make simulations and check the results of simulations. The use case diagram of the system is given in Fig.2. Parameters must be entered both for the simulation and for checking its results. The parameters are divided into epidemiological, network parameters, and simulation parameters.

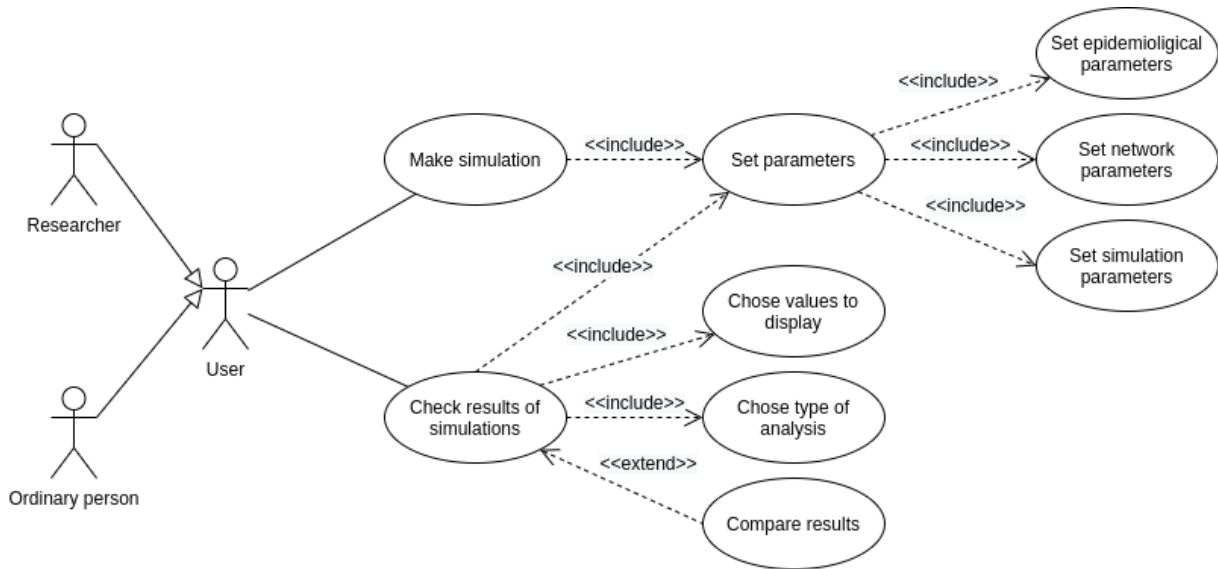


Figure 2: Use case diagram

Epidemiological parameters include: epidemiological model, infection rate, and other parameters that depend on the chosen epidemiological model, such as periods a person spent in infected and critically infected states, incubation period, demographic distribution data, and disease and mortality statistics.

The network parameters include: network model (complete graph and Erdős–Rényi, Watts–Strogatz, and Barabási–Albert models), the number of nodes in the network, and the number of edges in it or the initial number of edges for the node depending on the model.

The simulation parameters include the maximum simulation time (conditional(simulation) days), the period of collecting information about the state of network nodes and the number of simulations to be performed with the obtained parameters.

Checking simulation results also includes selecting values to display and selecting the type of analysis. In addition, it is possible to compare simulations with different parameters, which is done by adding a new curve to the graph.

Under the selection of values to display we understand the parameters for constructing curves. Such curves can show the number of infected, critically infected, susceptible, and dead nodes at each discrete point in time, given by the period of information collection.

The analysis can be averaged from several simulations with the same parameters and focused on the maximum values. Averaging can be done by finding the arithmetic mean or median values.

3.3 Description of the logical structure of the system

The developed software is an analytical panel using which you can set parameters for providing simulation and checking results. Its modules are scripts that run at user request with user-entered parameters from the main module Dashboard. The collaboration diagram is shown in Fig.3.

System modules:

- Run_simulation;
- Analysis;

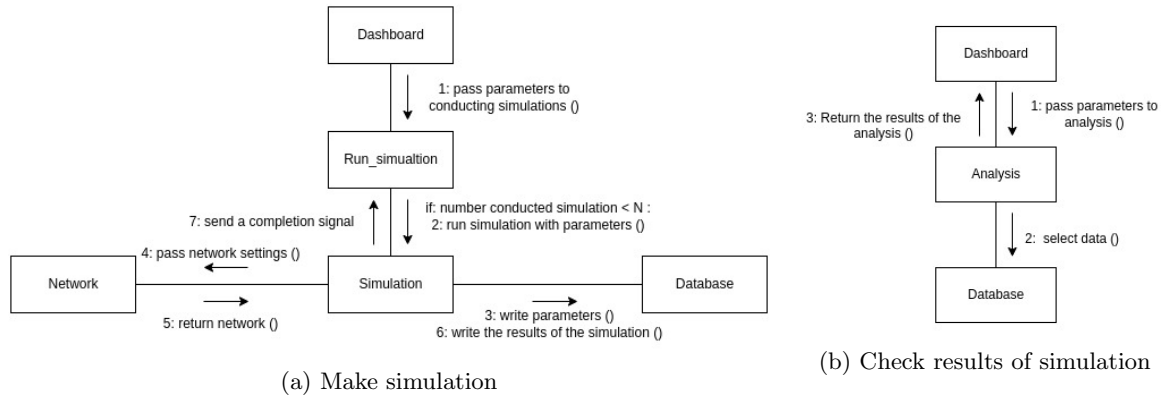


Figure 3: Collaboration diagram

- Simulation;
- Network;
- Database.

Dashboard module is responsible for interacting with the user, that is, receiving parameters from the user for simulation or analysis, and displaying their results. This module is responsible for directly running Run_simulation and Analysis modules, and receives information from the latter for display. Information about the simulations already performed, to display the appropriate fields for user input parameters, this module takes from Database.

Run_simulation module is a script that runs module Simulation with the specified parameters a specified number of times, each time in a new thread (maximum is working 4 threads in the same time). It can be started independently or using the Dashboard module from which also receives parameters for carrying out simulation.

The analysis module performs data selecting, averaging, and finding deviations. Also, this module prepares the processed data to display. It gives the prepared data back to the Dashboard module. It is started using the Dashboard module from which accepts parameters of simulations for the analysis.

Simulation is a script that is run by the Run_simulation script and receives epidemiological parameters and network data. It simulates the spreading of the virus, keeps track of all actions performed on the nodes, and also collects information about the state of the network after each specified period of simulation time, and records the data to the database.

Network - a script for creating a network, which is called by the Simulation script and builds the network following the model selected by the user and the network parameters set by them, and then returns it to the Simulation module.

Database - Non-relational database implemented using a file system.

4 Software description

4.1 Software specification

The largest unit of data over which the action is performed is the adjacency matrix of dimension $n \times n$, where n is the number of nodes in the network, therefore, memory usage is proportional to n^2 . The running time of the program to infect all nodes of the network is proportional to n^3 , where n is the number of nodes because the most difficult part is finding the number of infected neighbors, which is implemented by multiplying the adjacency matrix by the vector column of contagious nodes, and it is done n times to infect all nodes.

Table 1: Resources

Amount of nodes	RAM (MiB)	Time (second)
200	86	0.07
500	92	0.42
1000	129	2.5
2000	277	19.5
5000	1075	242.3

Table 1 shows time and memory usage for different amount of nodes in the network. Data on the operation of the program was obtained using an AMD Ryzen 5 3500U processor. Python version 3.9.7. The arithmetic mean of time was taken from 10 simulations for the complete graph in which all nodes were infected. All simulations were performed in single-threaded mode.

4.2 Used development tools

- Python was used as a programming language for developing the system;
- Igraph was used for building network models and visualization;
- Numpy was used for working with arrays and matrices;
- Pandas was used for working with data;
- Dash was used for designing a dashboard;
- Plotly was used for plotting.

4.3 Description of the program interface

The graphical interface of the program is made in the form of a dashboard(analytical panel) that works using a browser. At the top of each page is a navigation menu, through which the user can go to the pages "Simulation" and "Analysis".

Once on the "Simulation" page, the user will see the elements with which he can enter the parameters for the simulation(Fig.4). Input fields are created dynamically, depending on the selected epidemic model and network model.

Figure 4: Parameters entry page for simulation

When the simulation is complete, the network states history at each time point, with a user-specified step will be displayed. The display is in the form of a graph on which each state of the node has a certain color and a graph of the number of nodes in a certain state according to point of time(Fig.5). States colors: white - susceptible, green - infected, blue - not susceptible, red - critically infected, purple - dead. The edges of the graph show the connections between network nodes.

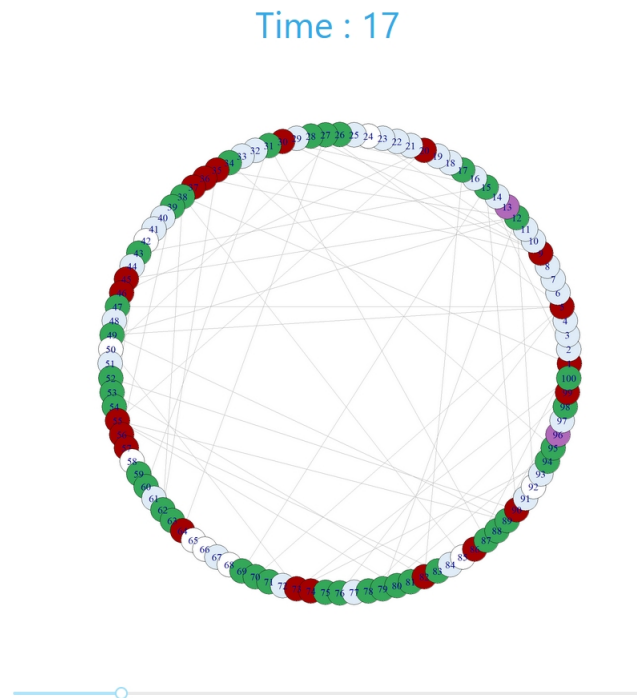


Figure 5: Display the state of the network as a graph

On the analysis page(Fig.6), the user can choose the analysis of the average or peak values or one specific simulation. Fields are created dynamically, depending on the selected models, as in the case of the simulation page.

Epidemic Outbreak PredictorSimulation Analysis

Select kind of analysis

Separated analysis of eac...▼

Select epidemiological model	Select network model	Select amount of nodes	Select critically infection rate
<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">SIR × ▼</div>	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Barabasi-Albert × ▼</div>	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">100 × ▼</div>	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">same × ▼</div>
Select threatment time	Select critically threatment time	Select initial amount of nodes	
<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Select... ▼</div>	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Select... ▼</div>	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Select... ▼</div>	

Submit

Figure 6: The form of entering parameters for analysis

In the case of analysis of average values, the program allows the user to build graphs of the number of infected, critically infected, susceptible, recovered, and dead at all points of time, as well as to compare graphs for different parameters by adding new curves to the graph(Fig.7). In addition, the possibility of displaying the error with a user-specified deviation is implemented. The graph of the analysis of an individual simulation is displayed as well as the graph with averaged values, except for the possibility of displaying the error and the averaging method.

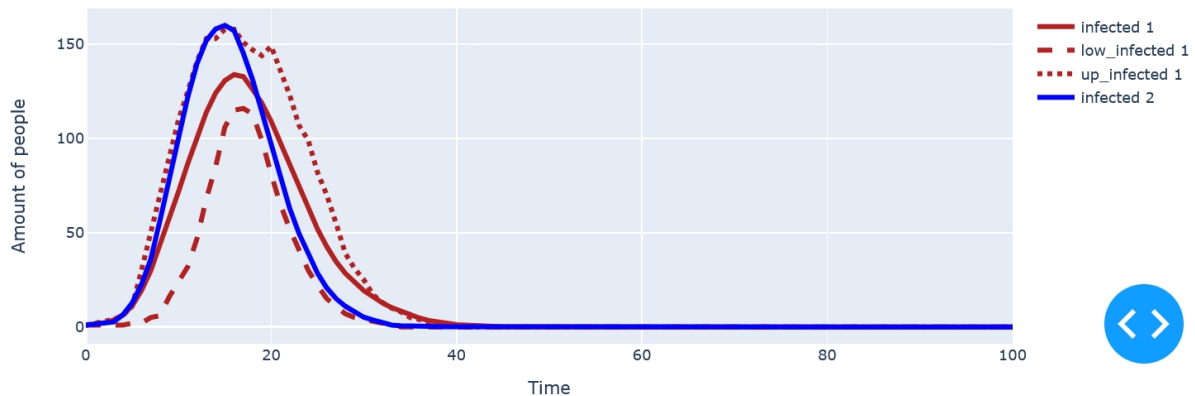


Figure 7: Curves of arithmetic mean values infected with and without error display

In the case of analysis of peak values, it is possible to analyze the number of people in a certain state for the given parameters depending on the coefficient of infectivity of the disease. The user is allowed to choose the same parameters as for the analysis of average values except for the infection rate. For convenience, the possibility of choosing a linear or logarithmic scale for measurements is implemented (Fig.8).

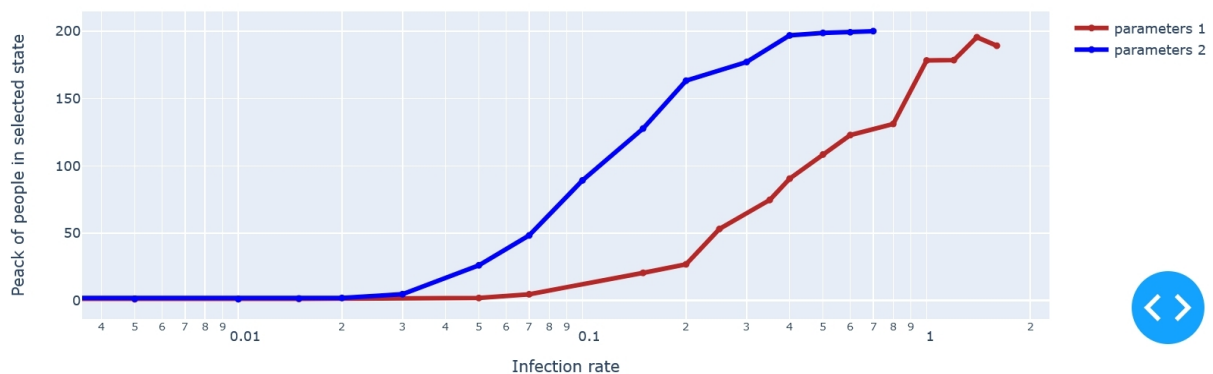


Figure 8: Graphs of the dependence of the peak number of infected on the infection rate, with the x-axis in a logarithmic scale

Graphs are interactive, so the user can choose the desired scale or move it on the field.

5 Conclusions

The developed software, unlike other tools that provide the ability to simulate an epidemic outbreak, is more versatile because it has a large number of input parameters. In addition, this software collects various epidemiological models (SI, SIS, SIR, SEIR) and models of the Erdős-Rényi, Watts-Strogatz, Barabási-Albert networks, and a complete graph for simulating the epidemic outbreak. Each of the parameters a user has to set by themselves using the dashboard.

Statistics of various diseases for different populations can be used for simulations. The simulations themselves are performed using the Gillespie algorithm, which provides a realistic simulation of events in time in contrast to other methods. The simulations are fairly fast but require a lot of RAM to simulate an epidemic outbreak in large networks.

This software can be valuable to researchers (the [8] paper studied the impact of network topology

and its parameters on the example of COVID-19 data using the software), but at the same time is quite simple and has a clear graphical interface that promotes the popularization of epidemiology.

Link to the source code: <https://github.com/YulianKuryliak/EpidemicOutbreakPredictor>

6 Future developments

We are looking for data on real social contact networks, to make our simulation closer to the real world. We are planning on adding the following functionality:

- setting a custom network;
- nodes immunization (its importance and effect are shown in the articles [4, 2]);
- nonpharmaceutical measures;
- detection of the links through which the infection occurs;
- concentration on reproduction value (R) what gives the ability to compare our results with other works.

Acknowledgment: This work was supervised by Michael Emmerich, from Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands and Dmytro Dosyn from Institute of Computer Science and Information Technologies, Lviv Polytechnic National University, Ukraine.

References

- [1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [2] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1–26, 2008.
- [3] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu. A time-dependent sir model for covid-19 with undetectable infected persons. *IEEE Transactions on Network Science and Engineering*, 7(4):3279–3294, 2020.
- [4] M. Emmerich, J. Nibbeling, M. Kefalas, and A. Plaat. Multiple node immunisation for preventing epidemics on networks by exact multiobjective optimisation of cost and shield-value. *arXiv*, 2020. doi:2010.06488.
- [5] P. Erdős and A Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
- [6] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.
- [7] H. W. Hethcote. Three basic epidemiological models. In *Applied mathematical ecology*, pages 119–144. Springer, 1989.
- [8] Y. Kuryliak, M. Emmerich, and D. Dosyn. On the effect of complex network topology in managing epidemic outbreaks. 05 2021.
- [9] A. Lipowski and D. Lipowska. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*, 391(6):2193–2196, 2012.
- [10] Y. M. Marwa, I. S. Mbalawata, S. Mwalili, et al. Continuous time markov chain model for cholera epidemic transmission dynamics. *International Journal of Statistics and Probability*, 8(3):1–32, 2019.
- [11] J. R. Norris. *Continuous-time Markov chains I*, page 60–107. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- [12] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

- [13] J. Romeu. A markov chain model for covid-19 survival analysis. *Technical Report*, 2020. doi:10.13140/RG.2.2.36349.18408.
- [14] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [15] G. Xie. A novel monte carlo simulation procedure for modelling covid-19 spread over time. *Scientific reports*, 10(1):1–9, 2020.

Yulian KURYLIAK,
Institute of Computer Science and
Information Technologies,
ICSIT of Lviv Polytechnic National University,
12 Stepan Bandera street, Lviv, 79000
UKRAINE
Email: yulian.kuryliak.mnsam.2021@lpnu.ua

Keep it Safe

Matei Marin

Abstract

KEEP IT SAFE! is a tower defense strategy game, whose main purpose is to defend the base at the end of the road. While the base is being attacked by enemy tanks or planes, the player must find the best strategic positions in which to place the towers that can provide protection against the waves of attacking enemies. The game is powered by AI, so that the gameplay enhances the fun and becomes more interactive. The game is created in Unity and is based on their Tower Defense Template which includes all of the 3D models and audios and some of the scripts. KEEP IT SAFE! is available on WebGL, Windows and MacOS. Thus, it is available online and the players can play in any browser or download it. These actions can be carried out free of charge on the online platform itch.io. For a better quality of the graphics and the game, it is recommended to download the game.

1 Introduction

This paper is a presentation of the game KEEP IT SAFE! covering its functionalities and appearance, the technologies used and, last but not least, it gives an overview of its most important features. Therefore, the article starts with a brief illustration of the game interface and the variety of game elements. As the game is based on artificial intelligence, it is explained to the reader the way it has been implemented and its utility. Next, the most important functionalities of the game and their mode of operation and implementation are presented for a better understanding of the game process. The paper ends with a short conclusion pointing out the noteworthiness of the game.

2 Game Overview

2.1 Description

As a tower defense game, the main goal of KEEP IT SAFE! is to protect the base at the end of the path and prevent the enemies from reaching it, by strategically placing defensive structures on grids. At the start of the level the player gets a total of ten lives. If an enemy manages to reach the base, one life will be subtracted and the consumption of all lives will lead to the loss of the game. Based on the remaining lives at the end of the level there will be calculated a star score. The player also gets some energy for the start. The amount of energy differs from one level to another.

Energy is used for placing towers or upgrading them, each tower having its own price and is generated by destroying enemies or by energy pylons.

The game has currently twenty levels with five different themes, six models of defense towers with various uses (shooting towers, slowing down towers, energy generators) and eight types of enemies, both ground and air. After buying a tower it can be upgraded two times or be sold. The 3D models and the themes were imported from the Unity Tower Defense Template.

The game is powered by AI which is used for the enemies to find the right path. Thus, attackers are able to always find the shorter and faster way to the base and avoid obstacles or tower barriers. If the path is blocked by pylons, AI will detect it and the enemy tanks will destroy the towers that stand in their way.

2.2 User Experience and Interface

The interface has a minimalist theme with a total of just six scenes: the menu (Fig.1), the settings, the level list (Fig.2), the pause- and upgrade-screen and the actual game screen.

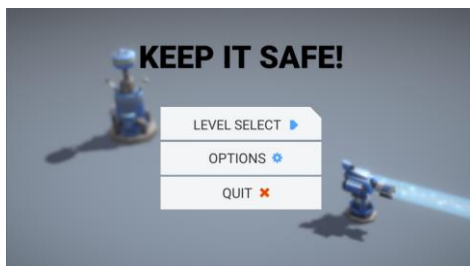


Figure 1: Menu

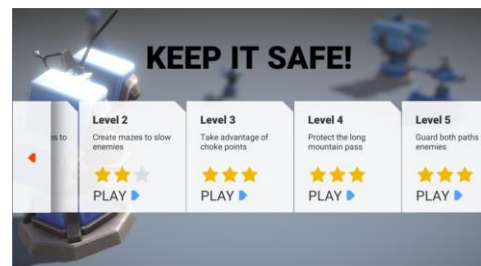


Figure 2: Level List

The Level Select button (Fig.1) links the user to the Level List (Fig.2).

In the Options section, players can adjust the volume for the game Fig.3. There are two types of volumes, for music and sound effects. The Master Volume option is to change the total volume.



Figure 3: Options Screen

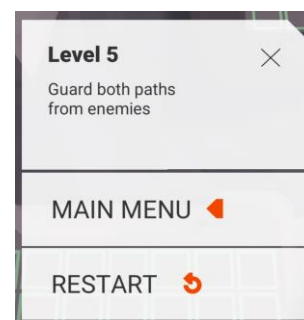


Figure 4: Pause Pop-Up

During the game the user can pause the game and restart it, or just return to the main menu. (Fig.4) The game is easy to understand due of its simplicity but also thanks to its explanations.

2.3 Towers

The game includes several types of defense towers from which players can choose to defend their base (Fig.5). They have different uses, some being able to attack only ground enemies, others only

air ones, but the most useful can attack both ground enemies and those from the airspace. These are (in order from left to right in Fig.5): the assault cannon, the rocket launcher, the plasma lance, the energy pylon, the EMP generator (slows down enemies) and the missile array.



Figure 5: Game Weapons

Features:

- Assault Cannon – shoots at both ground and air enemies, has a high fire rate, but causes low damage
- Rocket Launcher – only shoots at ground enemies, but causes high damage
- Plasma Lance – shoots at both ground and air enemies on a high range, causes very high damage, but has a slow fire rate
- Energy Pylon – generates energy (Fig.8)
- EMP Generator – slows down nearby enemies (Fig. 7)
- Missile Array – shoots on both ground and in the air, causes very high damage, but self-destructs after firing a missile barrage

The game gets even more interactive as each tower except for the Missile Array can be upgraded up to Level 3 or sold. After an upgrade the aspect of the tower changes too. (Fig. 6)

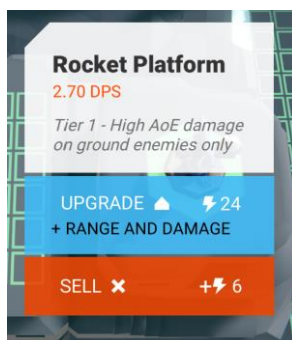


Figure 6: Tower Actions

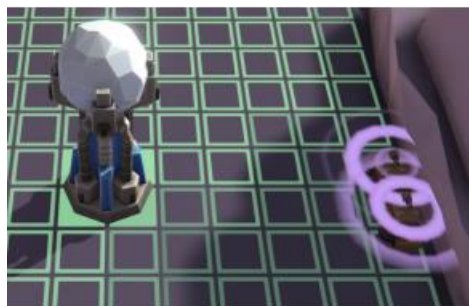


Figure 7: EMP Generator

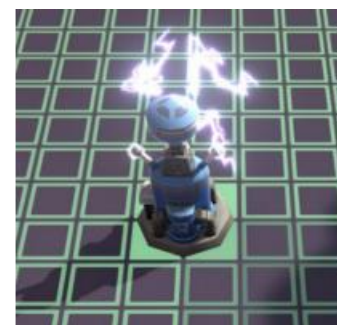


Figure 8: Energy Pylon

The player can also choose between two methods of positioning the defense towers: Grid Placement Areas (Fig.9), which allow multiple towers to be placed in the same place, and Single Placement Areas (Fig.10) which, as the name suggests, allow the placement of a single tower in a certain area. This of course comes with advantages. The areas where the player can only place one tower are higher, which offers a longer shooting range.



Figure 9: Grid Placement Area

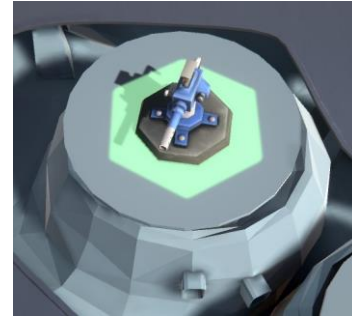


Figure 10: Single Placement Area

3 Artificial intelligence

3.1 Use

Artificial Intelligence has an important role in this game, because with its help enemies are always able to find their way to the base. Players can create mazes to slow down enemies or block one of the ways, while leaving the other open. Agents have to analyze the map and find the shortest path to the base.

If the enemies would have a predetermined path, and the player would build an obstacle, the attackers would hit that one and remain blocked. The power to detect the obstacle and to bypass it is conferred by artificial intelligence and is absolutely necessary for the smooth running of the game.

Also, enemies have the ability of destroying towers, but this feature isn't always on. It enables just when the script discovers that there is no possible way to the base and it has to create one.

3.2 Implementation

Enemies can follow the path to the base thanks to a node-system made up of multiple nodes (usually located at intersections or corners). The enemies' goal is to get to the next node – this loop continues until it gets at the End node. All nodes are interconnected and have an order number, so that they can be followed in the right order. The first node is called Start node and the last one End node. Placing nodes at crossroads resulted in the creation of a rough map of the level, which enemies can follow to get to the home base.

Fig.11 and Fig.12 represent possible ways an enemy can follow in order to get to the base in two different themes of the game. In Fig.11 there are two possible paths: In case one of them is blocked, AI detects it and takes the other. Nodes are the white spheres or quadrilaterals in the figures and the yellow lines are the connections between them.

The script calculates which is the next node to reach. Even if the player creates a maze, the agent is able to find the shortest way to the node, by avoiding the colliders of the turrets. As it

follows the pattern built by the colliders of the tower on the NavMesh (represents the accessible surface), the agent can find its way to the next destination.

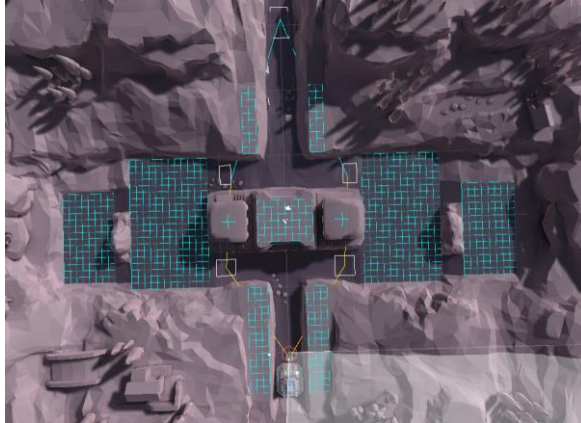


Figure 11: Nodes in Level 5



Figure 12: Nodes in Level 4

3.3 Code

3.3.1 Node Editor

When building in Unity, users can customize the Game Engine through code, which gets really helpful in order to create multiple GameObjects of the same kind. For creating the nodes inside the Unity editor it had to be modified so that all the things necessary to create a node are integrated in the new editor.

A very important part of a node is the AreaMeshCreator, which is a GameObject introduced as a child of the node (Fig.13), and which represents the node. The following code structure allows the Unity Editor to directly create a new AreaMeshCreator every time a new node is created. To avoid duplication of this GameObject, the script first checks with the help of the if statement whether there is already a AreaMeshCreator or not.

```
protected void AddMeshCreator()
{
    var meshObject = m_Node.GetComponentInChildren<AreaMeshCreator>();

    /// AreaMeshCreatorObject already exists so no need to create it again
    if (meshObject != null)
    {
        Selection.activeGameObject = meshObject.gameObject;
        return;
    }

    GameObject newGameObject = new GameObject("Node Mesh");
    newGameObject.transform.SetParent(m_Node.transform, false);
    meshObject = newGameObject.AddComponent<AreaMeshCreator>();

    Selection.activeGameObject = meshObject.gameObject;
    Undo.RegisterCreatedObjectUndo(meshObject.gameObject, "Created AreaMeshCreator");
}
```

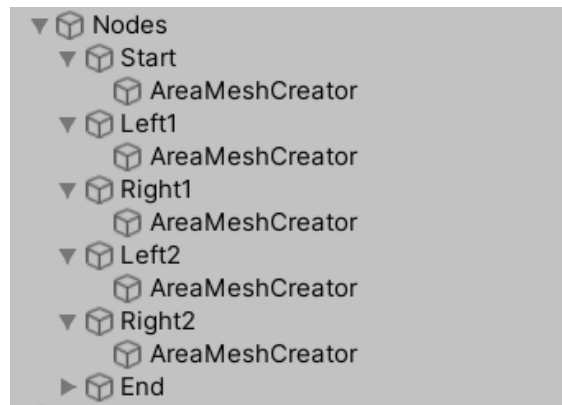



Figure 13: Node Hierarchy

Visualizing what the code processes, can also be really helpful. That's why the yellow lines were added to the editor to track the way which the attackers will follow. For displaying the connections, it was necessary to go through a `for` loop to take all the nodes in turn. Then the program was instructed to draw a line between the current node and the past node.

```
#if UNITY_EDITOR
protected virtual void OnDrawGizmos()
{
    if (linkedNodes == null)
    {
        return;
    }
    int count = linkedNodes.Count;
    for (int i = 0; i < count; i++)
    {
        Node node = linkedNodes[i];
        if (node != null)
        {
            Gizmos.DrawLine(transform.position, node.transform.position);
        }
    }
}
#endif
```

3.3.2 Agent Configuration

The logic behind the agents behaviour when taking decisions is easy to understand, as there are just some steps.

The enemy is spawned on the Start Node and then sets its next destination which is the next node in the level. This action is done while running the function:

```
public virtual void GetNextNode(Node currentlyEnteredNode)
```

This function identifies the next node to move to by adding 1 to the current node. After the target is set, the script runs the next function, which starts the enemy to the next node:

```
MoveToNode();
```

These steps happen repeatedly until the agent reaches the final node, the End node, and runs the functions presented in 4.5, or until it gets destroyed on its way.

But during these steps, the script runs other functions in the background in order to check if the way is valid.

One of this functions is verifying if the way is blocked. In order to do so, there was declared a new boolean to store whether the way is blocked or not. By analyzing the NavMesh (which represents the accessible surface), the script knows if the previous statement is true or false and saves it.

```
protected virtual bool isPathBlocked
{
    get { return m_NavMeshAgent.pathStatus == NavMeshPathStatus.PathPartial; }
}
```

Similar to the last function is another one, which calculates the remaining distance to the base and then tells if the agent is close enough to it. This one too uses a protected boolean to store the result.

```
protected virtual bool isAtDestination
{
    get { return navMeshNavMeshAgent.remainingDistance <=
navMeshNavMeshAgent.stoppingDistance; }
}
```

4 Code Implementation

4.1 Code overview

The game was mainly written in C# using the Visual Studio Code editor. All scripts were imported to Unity and in order to do so, the programs are using the library:

```
using UnityEngine;
```

After importing them to Unity in different folders they could be used for different actions by adding a new script component and selecting the needed one.

4.2 Game Manager

The Game Manager is in short terms the screen that handles the level list, the main screen, the buttons etc. and it differs from scene to scene. It's program uses four libraries for importing all game data and for managing the scenes.

```
using Core.Data; /// imports game data
using Core.Game; /// imports game data
using UnityEngine; /// main library for developing Unity games
using UnityEngine.SceneManagement; /// for detecting the scene
```

The level list consists of several rectangles containing the name of the level, a short description of it, but also the number of stars obtained by the player. By identifying the level, the program can easily display its name and its description, but for the star rating it has to run more commands.

At first, the program has to get the id for the current level in order to access the rest of the information, which is done by the following function.

```
public LevelItem GetLevelForCurrentScene ()
{
    string sceneName = SceneManager.GetActiveScene (). name;
    return levelList.GetLevelByScene (sceneName);
}
```

In the next step it will be checked if the identified level is completed or not. This is done by creating a `public bool`.

```
public bool IsLevelCompleted (string levelId)
{
    if (! levelList.ContainsKey (levelId))
    {
        Debug.LogWarningFormat ("[GAME] Cannot check if level with id = {0} is completed. Not in level list",
            levelId);
        // displays an warning if the requested level id is not in the level list
        return false;
    }
    return m_DataStore.IsLevelCompleted (levelId);
}
```

In case the level isn't completed, the program will simply display zero stars, otherwise it will have to get the star number for the current level. The process for doing so is similar to the last code sequence but the `public bool` will be replaced by `public int GetStarForLevel(string levelId)`, as the program has to return an integer not a state.

4.3 Towers

4.3.1 Target detection

To target an enemy, it must first be detected. To detect an attacker, the program will have to scan the environment for colliders (spheres or capsules, representing the enemies) and create a list of all valid targets it has to destroy. Once an enemy gets destroyed, it disappears from the list which constantly gets updated and adds the new ones.

For the start, the program has to return the enemies data, which is done by the following code structure.

```
public Targetable GetTarget ()
{
    return m_CurrrentTargetable;
}
```

The list of targets in range is created by adding all Targetable elements together and then returning their IDs.

```
public List <Targetable> GetAllTargets ()
{
    return m_TargetsInRange;
}
```

After creating the list and fetching all the data related to the attackers, towers are prepared for aiming.

4.3.2 Aiming

For the most efficient use of the towers, they always shoot at the nearest target in their range. This is done by analyzing the data of all the enemies displayed in the created list (see 4.3.1).

At the beginning of the code block two variables are declared.

```
Targetable nearest = null;  
float distance = float.MaxValue;
```

The first one is a `Targetable` one, which means it can save an enemy (in this case, the nearest enemy), and the second one represents the minimum distance ever recorded from the whole list (each time the program measures a smaller distance, this variable will be updated to the new distance).

The program runs an `for` loop in order to go through all of the enemy records. Using vectors, the program is able to calculate the distance between the tower and the current enemy. The distance will be saved in the variable name `currentDistance`, just like in the following code structure.

```
float currentDistance = Vector3.Distance (transform.position, targetable.position);
```

After this instruction is finished, the program verifies whether the value of `currentDistance` is greater than the value of `distance`. In case it is, it means that there is a nearer enemy than that one detected before, and the `distance` variable gets the value of the `currentDistance` variable. Also, the variable `nearest` gets updated and saves the current enemy as the nearest one. These instructions are done using an “if statement”, as it is shown below.

```
if (currentDistance < distance)  
{  
    distance = currentDistance;  
    nearest = targetable;  
}
```

After the `for` loop ends, which means that the program checked all the enemies on the list, the variable `nearest` will have saved the nearest to the tower of all the enemies. The target aiming animation is described in 4.3.4.

4.3.3 Shooting

For the shooting action, the towers have to send particles representing bullets to the enemies and then simulate an explosion. After the effect, the code has to subtract health from the targeted agent. The visual effects were created in Unity Particle System. The following script runs the created effects:

```
/// effects  
ParticleSystem pfxPrefab = m_Damager.collisionParticles;  
var attackEffect = Poolable.TryGetPoolable<ParticleSystem>(pfxPrefab.gameObject);  
attackEffect.transform.position = m_Enemy.position;  
attackEffect.Play();
```

After, playing the effects damage will be applied to the attacker, using the following component:

```
protected virtual void Awake()
{
    m_Damager = GetComponent<Damager>();
    m_Timer = new Timer(delay, DealDamage);
}
```

4.3.4 Mechanics

The main tower mechanic is the idle rotation, which is done while the turret is out of activity. The tower repeats a basic 360 degrees rotation movement over and over again, until it discovers an enemy in range. At first, the script checks if there is no enemy in its range:

```
if (m_CurrentTargetable == null) // do idle rotation
```

If the condition is met, the program can run the following code structure which will constantly rotate the turret:

```
m_WaitTimer -= Time.deltaTime;
if (m_WaitTimer <= 0)
{
    m_CurrentRotationSpeed = (Random.value * 2 - 1) * idleRotationSpeed;
}
```

But if the condition isn't met, the turret will have to stop its rotation and aim the target. In order to give the towers the opportunity to fire, the firing angle, the distance and the force with which the projectile is propelled must be taken into account. That's why different mathematical constants and laws of physics have a very important role in this game. Among them are the Euler angles (Fig.14) and the quaternions (Fig.15) that were used for the tower shooting system. According to Wikipedia, both are applied in three-dimensional space, which is why most 3D games use them. In the presented game, Euler's angles are used to position rigid objects in three-dimensional space, and quaternions, which are hypercomplex numbers (numbers obtained by expanding the set of complex numbers), have the role of coordinating rotations in the 3D environment. As the towers can be placed on different levels and as some of them shoot both ground and air, which implies rotation of the turret, quaternions are absolute necessary for smooth animations. The equation below is the formula for calculating a quaternion, which was used in the game as well.

$$H=a+b*i+c*j+d*k, \tag{1}$$

where a , b , c , and d are real numbers; and i , j , and k are the basic quaternions

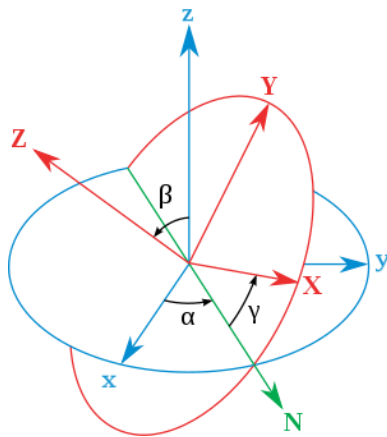


Figure 14: Euler Angles

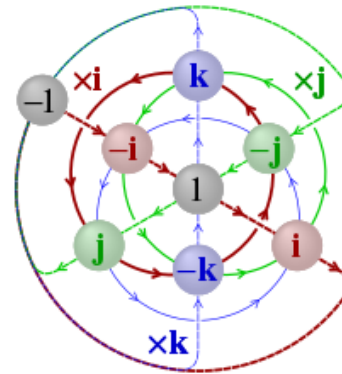


Figure 15: Quaternions

4.4 Wave Manager

The Wave Manager is a part of the game which coordinates the number and type of enemies which attack the base. Each level is formed out of ten enemy waves (Fig.16), each wave containing some attackers which will enter the environment based on an specified delay. For an easy creation of the level, a script helps creating a no-code method for the configuration of the waves. Thus, waves can be created inside the Unity editor just by selecting the type of enemy, introducing a delay to spawn (the program runs a timer and when the delay to spawn passed, the specified enemy will be created in the scene), and a place where to spawn (in this case at the beginning of the path named *Start*). The size of the wave (Spawn Instructions >> Size) and the time which has to pass until the next wave starts must also be specified. (Fig. 17).

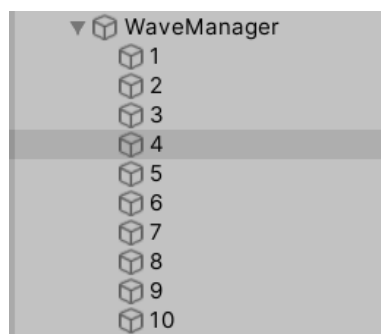


Figure 16: Wave Manager

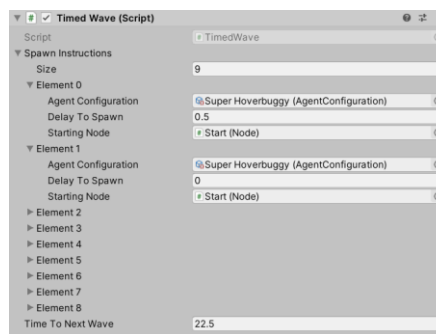


Figure 17: Wave Configuration

As there are multiple waves on each level, the script has to know which wave is currently active, and this is done by creating an `int` variable where to store this data in:

```
protected int m_CurrentIndex;
```

There must be a list of all waves from the current level for the script to run them in order:

```
public List <Wave> waves = new List <Wave> (); /// creates the list
```

Other important data regarding the Wave Manager is the total number of waves which is easily calculated by the following code, which creates an integer where it stores the requested number (obtained by counting the waves):

```
public int totalWaves
{
    get {return waves.Count; }
}
```

After all this data is set, the waves can be started. In case there is no data (0 waves), the script will return an error:

```
public virtual void StartWaves()
{
    if (waves.Count > 0)
    {
        InitCurrentWave();
    }
    else
    {
        Debug.LogWarning("[LEVEL] No Waves on wave manager. Calling spawningCompleted");
        SafelyCallSpawningCompleted();
    }
}
```

After a wave has ended, the script will run a function which checks whether all waves were completed. If this state is false, the program will increase the current wave number with 1 and continue, otherwise it will run another function, which confirms that all the enemies were spawned and ends the level.

4.5 Base Attack

4.5.1 Summary

The Base Attack action will take place when an enemy reaches the base to be defended. This is represented by a red hexagon. After an enemy reaches the hex, the program will run a 2-second countdown, after which it will destroy the enemy and subtract one point from the player's life.

4.5.2 Home Base Configuration

In order for enemies to be able to attack the base, it must first be configured in the script like it is done below:

```
public class PlayerHomeBase: DamageableBehavior
```

The `DamageableBehavior` component, which was declared in another script (its function was coded there), tells the script that the base can be attacked and destroyed. The obtained component will be used in the following actions.

4.5.3 Attackers

Similar to the process described in 4.3.1, the program has to identify all the enemies which reached the base and create a list of them, in order to be able then to carry other actions. This is done by the following code structure, which basically creates a list of agents (enemies):

```
protected List <Agent> m_CurrentAgentsInside = new List <Agent> ();
/// The current Agents within the home base attack zone
```

After the enemy destroys the base, it has to disappear from the level. Therefore there was created the following function:

```
agent.Remove();
```

This function is being used in other code structures in order to destroy an enemy (see 4.5.5).

4.5.4 Timer

The timer is declared as a `public float` with the value of 0.5 f.

```
public float homeBaseAttackChargeTime = 0.5f;
```

Then all what remains to do is to start the timer, but first the script has to check when the agent reaches the home base (the final node). Once it reaches it, the program verifies if a timer already exists (from a previous agent). If it doesn't, there will be a new timer created, but if it does, the previous timer will just be reset. All the instructions above are executed by the following script.

```
void OnDestinationReached (Node homeBase)
{
    m_FinalDestinationDamageableBehaviour =
    homeBase.GetComponent<DamageableBehaviour>();
    if (m_HomeBaseAttackTimer == null)
    {
        m_HomeBaseAttackTimer = new Timer(homeBaseAttackChargeTime,
        AttackHomeBase);
    }
    else
    {
        m_HomeBaseAttackTimer.Reset();
    }
    m_IsChargingHomeBaseAttack = true;
}
```

4.5.5 Player Damage

After the timer runs out, the agent should then apply damage to the home base. This action is fired on completion of `m_HomeBaseAttackTimer` (the timer). A new function is created and its content takes damage on the Base, but it also destroys the enemy at the end through the command `agent.Remove()` (see 3.5.3).

```
protected void AttackHomeBase()
{
    m_IsChargingHomeBaseAttack = false;
    var damager = GetComponent<Damager>();
    if (damager != null)
    {
        m_FinalDestinationDamageableBehaviour.TakeDamage(damager.damage,
        transform.position, agent.configuration.alignmentProvider);
        /// The homebase gets damaged
    }
    agent.Remove(); /// The enemy gets removed
}
```


4.5.6 Effects

Last but not least, special effects, both visual and sound effects, are very important when it comes to video games. The visual ones are created with Unity's particle system and the sound ones are based on sound files. The base attack has four different effects, two for the charging period (the time span during which the enemies are attacking) and the rest for the moment when the attack is completed and the base is damaged.

```
public ParticleSystem chargePfx; /// The particle system when an attack is charging
public ParticleSystem attackPfx; /// The particle system for an attack
public RandomAudioSource chargeSound; /// The sound when an attack is charging
public RandomAudioSource attackSound; /// The sound for an attack
```

5 Conclusion

Although KEEP IT SAFE! is built on a simple idea and doesn't present any extravagant functionalities, it offers variety with each level, due to the fact that the user has to find the logical solution to defeat the AI. Therefore, KEEP IT SAFE! might be a good option for gamers in search of a new adventure and can be downloaded from the platform itch.io: <https://jamglass.itch.io/keep-it-safe>.

Acknowledgement: This work was supervised by Professor *Monica Oancea*, from *Gheorghe Lazăr National College, Sibiu, Romania*.

References

- [1] Quaternion - <https://en.wikipedia.org/wiki/Quaternion> (12/02/2021)
- [2] Quaternions in Unity - <https://docs.unity3d.com/ScriptReference/Quaternion.html> (12/02/2021)
- [3] Random.Range Function - <https://docs.unity3d.com/ScriptReference/Random.Range.html> (12/02/2021)
- [4] Tower Defense Template - <https://learn.unity.com/project/tower-defense-template?uv=2017.2> (12/02/2021)
- [5] Euler Angles - https://en.wikipedia.org/wiki/Euler_angles (12/02/2021)
- [6] Unity Docs - <https://docs.unity3d.com/Manual/index.html> (12/02/2021)
- [7] Tower Defense Games - https://en.wikipedia.org/wiki/Tower_defense (12/02/2021)
- [8] Asset Store - <https://assetstore.unity.com/> (12/02/2021)

Marin MATEI
Gheorghe Lazăr National College
1-3 Gheorghe Lazăr str, Sibiu, 500035
ROMANIA
E-mail: matei.i.marin@gmail.com

Improving the overall user satisfaction for applications generating traffic mix over 4G networks

Vasile Horia Muntean

Abstract

Regardless of the transmission environment used, when improving the overall experience perceived by the final user with regards to an application generating rich media content, it is important to take into consideration the users' particular set of characteristics. If the transmission environment is a wireless medium, then the challenge of delivering satisfactory media content is even harder. To achieve this goal by addressing the user satisfaction problem, an adaptive algorithm centred on the user requirements was developed, DQOAS – Dynamic Quality Oriented Adaptation Scheme [1]. The present article shows that the good performances observed for DQOAS algorithm in IEEE 802.11 standard and in 4G LTE environment by applying a special mapping scheme for Quality-of-Service parameters, can be replicated even in the scenarios where the users connected to an LTE network are mobile, moving with speeds higher than average pedestrian speed. Using the proposed mapping scheme will provide the possibility to modify stream priorities, and by doing so, avoid any changes done by the LTE scheduler on the decisions made by the above-mentioned algorithm. Since LTE system scheduler is a highly important part in the LTE quality assurance process, multiple LTE scheduling strategies are also analysed in the mobility context. Results presented in the final section are showing that in a mobility scenario, using DQOAS and the proposed mapping scheme together with the Proportional Fair scheduling strategy yields better results than the Round Robin or Max Throughput schedulers, all this while reaching the highest number of satisfied users.

1 Introduction

Improving the perceived quality of media rich content delivery is synonym to providing solutions for multiple technological challenges, that directly affects the final user satisfaction, and its quality of experience (QoE). This QoE variable impacts in consequence the quality of any used application. It is also important to recognize that the expectations of the end-users today in terms of delivered multimedia quality are extremely high. Supporting media rich content streaming while maintaining a high quality becomes a more complex task when the delivery medium is a wireless one (e.g. wireless LANs), because these technologies generally tend to provide a lower available bandwidth. Also, their overall performance is greatly influenced by external factors and by instant network load (in terms of resource requests and number of clients connected). The user location and their individual mobility pattern is also an important factor that directly influences the quality of delivered media stream.

3GPP Long Term Evolution (LTE) technology is a natural evolution from the previous cellular technology, UMTS/HSDPA, required by the subscriber demands for higher quality delivery (high

data rates) while also increasing their mobility. The Radio Access part of LTE technology employs the use of OFDM technique (Orthogonal Frequency Division Multiplexing), offering support for multiple channel bandwidths, ranging from 1.4MHz up to 20MHz in both frequency and time division duplex operating modes [2,3]. For uplink (UL) communication, using OFDMA (Orthogonal Frequency Division Multiple Access) implies an increased PAPR (Peak-to-Average Power Ratio). Using Single Carrier FDMA (SC-FDMA) reduces PAPR, thus improving one key parameter of today's handheld devices: battery life.

Regarding the peak data rates obtained in 3GPP LTE Release 10 (LTE Advanced), they go up to 1 Gbps in downlink (DL) and 500 Mbps in uplink, with a peak spectral efficiency (bps/Hz) of 30 for downlink (8x8 MIMO) and 15 for uplink (4x4 MIMO). Compared to other older technologies, LTE is specifically designed to provide flexible spectrum options (1.4 to 20 MHz bandwidth), to increase the use time of all connected user equipment by employing SC-FDMA in UL and to offer a simple network architecture with open interfaces. In terms of Quality of Service (QoS) support, LTE technology has performed a big leap forward compared to older technologies. But current content quality demands and the high amount of traffic load can lead to unsatisfactory user experience. The QoS concept as it is defined in the 3GPP specifications, is centred on prioritizing different traffic types and data flow classes. This is done with the help of a simple QoS mechanism that uses profiles based on QCI (QoS Class Identifiers), sending them on specific bearers that are created by the LTE network.

This paper presents a QoS mapping alternative to original LTE QoS parameter mapping, with good results when deployed in scenarios where the traffic is mixed. The purpose of the proposed mapping scheme is to increase the end-user satisfaction levels, by using the same QoS class for different traffic flows. This mapping scheme is analysed in a scenario where users are not static, having speeds of 10km/h.

The paper is structured as follows. Section 2 presents general aspects and studies regarding LTE scheduling algorithms and QoS parametrization in case of traffic mix scenarios, while the 3rd section presents the proposed QoS mapping scheme. Section 4 details the obtained results, followed by conclusions in section 5.

2 Scheduling algorithms used in LTE

As LTE technology is already a proven solution and has been adopted by most telecommunication operators, multiple research works are now focused on improving network delivery for both UL and DL. One way of doing this is by developing scheduling algorithms capable to improve the delivered quality in various load conditions, while considering data flows belonging to different traffic classes. These solutions try to improve the overall network capacity in terms of QoS flows supported, while reducing the utilized resources.

The scheduling mechanism dynamically determines every 1ms (once per Subframe) which User Equipment (UEs) should use the Uplink / Downlink shared channel to transmit or receive. Even more, it also allocates a fixed number of resources to be used. To be able to perform this type of dynamic allocation, the scheduler requires as input continuous measurements of both uplink and downlink channels. For the UL channel, the eNodeB itself can perform these measurements, while the UE is transmitting data. Based on these measurements, the suited power and MCS (Modulation and Coding Scheme) can be selected. For the DL channel, the UE is responsible for providing feedback, and this is done by the UE through periodical reporting of specific information grouped in so called CQI (Channel Quality Indicators), that tell the eNodeB the instantaneous channel quality. Based on these reports, the eNodeB can choose the appropriate coding and modulation, and the associated transmission power, can implement interference management techniques, and can perform time or frequency selective scheduling. These reporting mechanisms for both uplink and downlink are presented in Figure 1 and Figure 2 below.

Generally, each type of data flow has specific QoS requirements. In the process of ensuring that these demands are met, a round-robin scheduling algorithm might not provide the expected results. This is because a user is uniquely defined at any point in time by specific network conditions, that impact directly his average and maximum link capacity. Hence, more complex scheduling algorithms that also consider the channel conditions, QoS requirements or retransmissions were developed, in order to perform an optimal allocation of the scarce radio resources to LTE users.

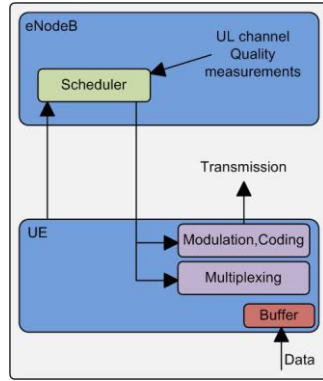


Figure 1: Uplink feedback process

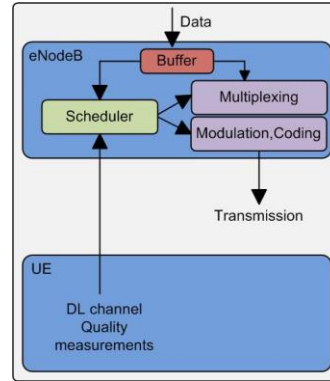


Figure 2: Downlink feedback process

Schedulers can be grouped in two main categories according to [4], based on the traffic type: schedulers dedicated for non real-time [5] and real-time [6] flows. Also, other parameters can be considered in order to perform a different schedulers classification: queue or channel aware [7] and queue and channel-aware schedulers [8].

3 QoS parameters mapping

In case an application is generating data flows organized in distinct QoS queues and a QoS mechanism that applies the RR (Round Robin) algorithm, according to [9] it is possible to describe the satisfaction of user i using the following relation:

$$\frac{(f_1 + \frac{\alpha}{\rho} f_2) \cdot T \cdot \lceil \frac{N}{n} \rceil \cdot \Delta}{(T + d^{max}) \cdot \beta} \leq \frac{1}{1 - \varepsilon}, \quad (1)$$

where f_1 and f_2 are the average transmission ratios for the 2 different traffic classes and ρ represents the priority between the two services. The time interval required for the communication is represented by T , n is the number of users scheduled each TTI (Transmission Time Interval) with a length of Δ , while N is the max cell load required to satisfy user i 's quality criteria. d^{max} denotes the maximum possible delay and finally ε is the maximum ratio between the sum of delayed/lost packages and the total number of packages that can still keep the perceived service quality as satisfactory. Further, α and β are defined as follows: $\alpha = s_2 / s_1$ and $\beta = (s_i^{max}) / s_i$. Here s_1 and s_2 represent the service average packet size and s_i^{max} is the maximum data size transmitted in a single transport block to user i .

In a scenario with multiple users, ε differs between users since they have different network conditions. Based on Dynamic QOAS description, M_i (min quality level accepted) is defined for each user for the incoming media stream. Since M_i represents the minimum level that Dynamic QOAS algorithm will deliver for any user i , the definition for ε as described previously has no

relevance because packets that will be delayed or lost will trigger a quality decrease below the minimum accepted level. One solution for this problem is to increase M_i with the max accepted ratio between the sum of delayed/lost packages and the total number of packages, ϵ : $M_i^{new} = M_i(1 + \epsilon)$.

Prioritization done based on service types yields impressive results in scenarios where data sources are independent, because the flow that is classified as high priority will experience a capacity gain while the second service will have a capacity loss. Considering scenarios where the data source is identical for different flows, it is possible for a user to experience an increased QoE per-application if the considered flows are identically classified, using only one QoS class ($\rho=1$). In this case since multiple flows will have the same priority for a user, the queue sorting and scheduling algorithm will treat the packets from the same application (even delivered via non-identical bearers) in a similar manner, thus keeping the same queuing delay and in the end improving the Quality of Experience for the application as a complete product.

4 Test Results

The solution described above was implemented and assessed with the help of LTE System Level Simulator [10], capable of simulating different network types (SISO, MIMO) and different transmit modes (TxM, OLSM). The mapping scheme proposed in section 3 of this paper and DQOAS algorithm's basic principle were integrated to test the possible benefits that can be obtained by using an adaptation algorithm based on user requirements, over LTE networks.

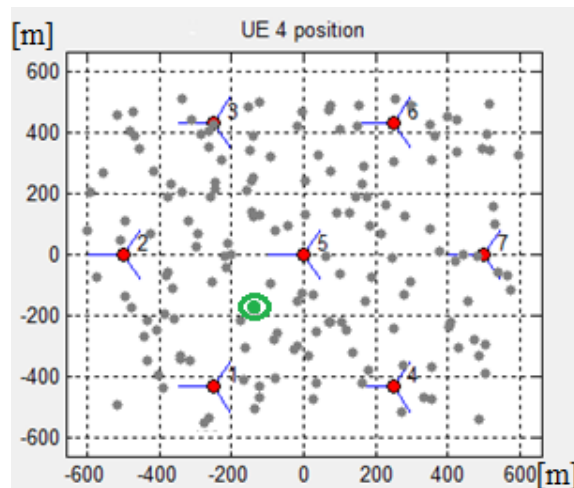


Figure 3: LTE simulated environment

As depicted in Figure 3, the test scenarios were tested using 7 eNodeB stations, each having 10 active users attached. To emulate a traffic mix, two streams classified with distinct priorities are transmitted to each UE from the base station on which the UE is attached. A particular user i is labelled as satisfied by the quality received only if its instant quality level is maintained above $M_i^{new} = M_i(1 + \epsilon)$. In the first scenario, minimum accepted quality level was set at 1Mbps for all users.

To create a more realistic representation of plausible scenarios, user mobility was introduced as a factor, with a speed of 10km/h, so that the users are randomly moving in the coverage area of their assigned eNodeB. Table 1 presents the most important parameters used for simulations performed.

The algorithm together with the described mapping scheme were tested using two scenarios with a duration of 25s each. Two different data streams are sent in parallel to each of the 70 users that are constantly moving on the coverage area. One of the streams is considered to have

multimedia content (e.g. video) while the other stream mimics the throughput pattern of a web browsing connection. Out of the 70 connected users, User53, connected to eNodeB5 and highlighted in green in Figure3, was selected for analysis and its throughput and BLER (Block Error Rate) values were analysed. The following downlink schedulers were tested in both scenarios: Round Robin (RR), Maximum Throughput (MT) and Proportional Fair (PF).

Table 1: Simulation parameters

Parameter	Value
Frequency	2.0 GHz
Bandwidth	5 MHz
Thermal noise density	-174 dBm/Hz
Receiver noise figure	9 dB
nTX x nRX	2 x 2
TTI length	1e-3 s
Simulation length	50000 TTIs
UE speed	10 Km/h

The setup for the first test scenario employs the default QoS mapping definition together with standard LTE QoS mechanism for delivering the two streams with different priorities for all users. In the second test setup, the delivery algorithm is changed to DQOAS. To assess also the quality of the second data stream (considered to be a web browsing session, simulated using a TCP download), a minimum quality level was also set, to a value of 0.250 Mbps. Only the last 20 seconds of each simulation are considered in computing the average BLER and throughput values for all users, to avoid any instabilities due to time required by the simulator to stabilize the traffic and connect all users.

Test Scenario 1

The average throughput calculated for stream 1 in case the used scheduler is Proportional Fair is 1.32Mbps, with no drop below 1Mbps during the last 20 seconds of the simulation, and this value satisfies the quality requirements defined previously for User53. In case of the web browsing stream, the throughput averages at a value of 0.368Mbps on the last 20 seconds of the simulation. But in this case, the instant data rate dropped below the required level in 2 intervals (a total of 5 seconds). From a quality of experience perspective, this 5 second interval is more than enough for a user to initiate the session termination in case of this second stream. Table 2 presents the average block error rate values obtained for User53 in this test scenario, on both streams.

Table 2: Average BLER [%] values for User53

	PF scheduler	RR scheduler	MT scheduler
Stream 1	5.8	7.8	7.9
Stream 2	7.9	7.9	6.2

The results obtained in this simulation scenario are showing exactly what was expected: in case of a default mapping scheme and the standard LTE QoS mechanism, PF scheduler has performed better in comparison with RR and MT.

Test Scenario 2

Same as in the previous setup, all three schedulers were used to analyse the performance of the proposed solution in case of a mobility scenario. Considering first the Proportional Fair scheduler, the average value obtained over the last 20 seconds of the simulation is 1.211Mbps in case of the first stream, and 0.291Mbps for the web browsing, so higher than the minimum quality thresholds defined for User53. One thing easily noticeable is that the avg throughput of the media-rich stream is lower than the one obtained in the first test scenario. This is happening since DQOAS algorithm, together with the proposed mapping scheme, can perform an optimal resource allocation by limiting the throughput for certain users, to release resources for other users and increase their perceived quality over the minimum required threshold. DQOAS realizes this optimization because while performing the adaptation process, it considers the quality requirements of all users connected to one eNodeB. And since the two streams are mapped with identical priority (same QoS class), the LTE scheduler can no longer separate the data coming from the same application and prioritize it differently based on the default mechanism used in the LTE networks (QCI based). Using the proposed mapping scheme, the two individual streams are treated with the same priority by the system scheduler, and this creates the opportunity for the used adaptation algorithm to increase the general QoE of the application, as perceived by the final user.

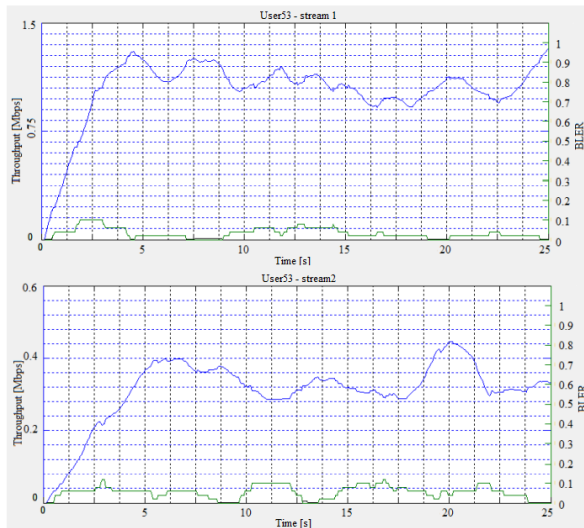


Figure 4: User53 throughput and BLER (PF scheduler)

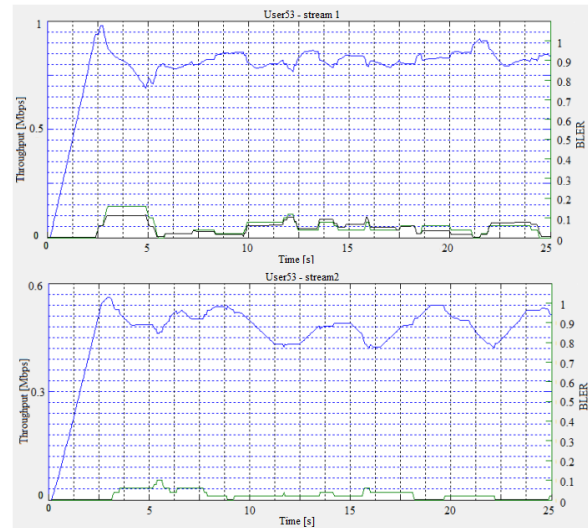


Figure 5: User53 throughput and BLER (RR scheduler)

Figure 4 shows the measurements recorded for User53 in case PF scheduler is used, while Figure 5 presents the same measurements, but recorded while RR scheduling strategy is used. The average throughput values obtained for Round Robin scheduler are 0.822Mbps and 0.481Mbps for the first and second data streams. As it was the case for PF scheduler in this test scenario, both of values are enough to satisfy the quality requirements for the selected user, even if they are lower in comparison with the values obtained in the first simulation. As expected, Maximum Throughput scheduler's performance is below PF and RR, but it is worth mentioning that it is still able to maintain the throughput average above M_i^{new} . A drop below the quality threshold is present in case of the second stream, but the duration is approximately 2 seconds, so the overall quality can be considered satisfying.

Average BLER measurements for all three schedulers during the second scenarios can be observed in Figure 6. As it was the case also in the first simulation scenario, the best BLER results for both data streams (combined) are obtained when the used scheduler is Proportional Fair.

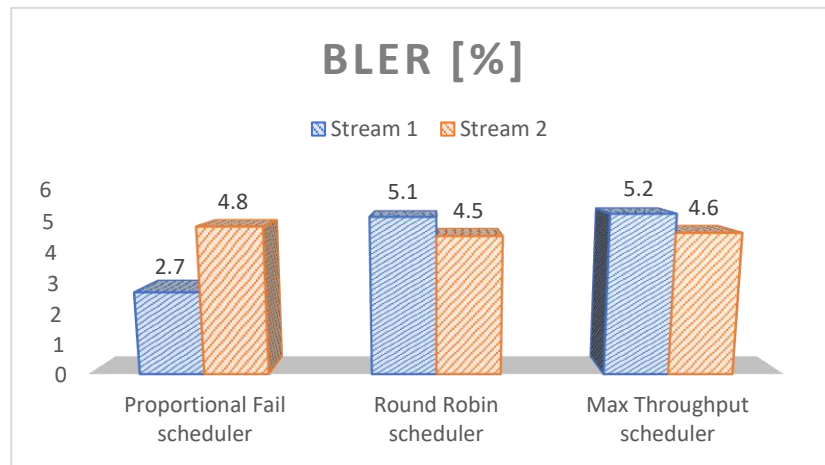


Figure 6: Average BLER values

Comparing the results with those from the previous test scenario, it is noticed that in this case the algorithm was able to reduce the average BLER values, that will lead to an improved data transmission and in the end to an improved user perceived quality.

Table 3: Percentage of satisfied users by the quality of both data streams

Scheduler	% of satisfied users		
	PF	RR	MT
Test 1	47	46	24
Test 2	62	51	32

5 Conclusions

Based on the measurements presented in section 4 of this paper, a first conclusion to be considered is that the DQOAS algorithm in conjunction with the proposed mapping scheme can obtain good results when deployed over an LTE network where users are moving with speeds higher than the average pedestrian speed (5km/h). The prioritization scheme is necessary, as the alteration of the original QCI based prioritization present in LTE will prevent for the decisions taken by the adaptation algorithm to be overridden by the LTE scheduler, giving DQOAS a way to manage the radio resourced directly. Another observation based on the obtained results is that by using the adaptation algorithm, link efficiency has improved, and the number of satisfied users increased, compared to default approach used in LTE (number of satisfied users - out of the 70 considered - can be seen in Table 3). DQOAS algorithm also improves the data rate stability, by reducing the throughput fluctuations and BLER, which is a very important parameter in defining the overall system performance.

References

- [1] V. H. Muntean and G.-M. Muntean, "A Novel Adaptive Multimedia Delivery Algorithm for Increasing User Quality of Experience during Wireless and Mobile Learning", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Bilbao Spain, May 2009.
- [2] K. Bogineni et co., "LTE Part I: Core Network – Guest Editorial", IEEE Communications Magazine, vol. 47, no. 2, February 2009.
- [3] 3GPP TS 36.104 version 15.3.0 Release 15, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception", July 2018.
- [4] R. Madan, A. Sampath B. Sadiq, "Downlink scheduling for multiclass traffic in LTE", EURASIP Journal on Wireless Communications and Networking, vol. 2009, July 2009.
- [5] M. O. Kayali et co., "Downlink scheduling in LTE: Challenges, improvement, and analysis", 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, June 2017.
- [6] N. Chen and S. Jordan, "Enhanced exponential rule scheduling algorithm for real-time traffic in LTE network", International Journal of Electrical and Computer Engineering (IJECE), April 2020.
- [7] G. Berardinelli et co., L. A. de Temino, "Channel aware optimized proportional fair scheduler for LTE downlink", Peer-to-Peer Networking and Applications, January 2020.
- [8] M. M. Nasralla, "A Hybrid Downlink Scheduling Approach for Multi-Traffic Classes in LTE Wireless Systems", IEEE Access (Volume 8), April 2020.
- [9] I. Siomina and S. Wanstedt, "The Impact of QoS Support on the End User Satisfaction in LTE Networks with Mixed Traffic", IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France, September 2008.
- [10] M. Wrulich and M. Rupp J. C. Ikuno, "System level simulator of LTE networks", Proceedings of IEEE 71st Vehicular Technology Conference, Taipei, Taiwan, May 2010.

Vasile Horia Muntean
Lucian Blaga University
Computers and Electrical Engineering
Emil Cioran 4, 550025, Sibiu
ROMANIA
E-mail: horia.v.muntean@ulbsibiu.ro

4G LTE and 5G NR - An overview; Technical aspects of the Air interface

Vasile Horia Muntean

Abstract

During the last years, we have witnessed and taken part in the shaping of the new connectivity paradigm: instant data access, reduced delays, high throughput, maximized user experience, increased mobility and rich media content consumption. The fast shift in the way network-based services evolved based on the above users' requirements is mainly possible due to the quick development of wireless access technologies. But providing high throughput, low delay connectivity on an anywhere-anytime basis is a challenging task that can only be achieved through a global deployment of real mobile broadband connectivity. Some of the first technologies developed to accommodate these new user behaviours – provisioning of high-quality voice, data and multimedia services on mobile devices while keeping user costs accessible – were WiMAX (Worldwide Interoperability for Microwave Access) and LTE (3GPP Long Term Evolution). If ten years ago WiMAX technology had the advantage over LTE due to a larger real-world deployment and more mature specifications, nowadays it is clear that LTE was the winner of that battle, most likely due to the reduced migration costs for the operators, increased flexibility in terms of frequency band options and adaptive antenna systems, higher peak data rates and uplink/downlink performances plus the high advantage offered by it being able to support higher mobility scenarios. But new technologies like IoT are emerging and user demands keep adding more and more pressure on the existing LTE infrastructure, so therefore new wireless networks are developing, and the most important is the 5G NR (New Radio) mobile broadband network. This paper investigates 4G LTE and 5G NR technologies with regards to their architectural differences, focusing mainly on the performance of their first two layers. It also tries to determine if 5G will eventually replace the 4G LTE networks or if they will work together, complementing each other in such a way that all current and future user demands will be carried through.

1 Introduction

Mobile broadband networks developed at a very high pace during last fifteen years because used handheld devices became affordable for the masses, leading to a constant increase in the demands for high-quality mobile communication services, for instant data access, reduced delays, high throughput, maximized user experience, increased mobility and rich media content consumption. Accommodating all these demands required constant redefinition and improvement of the air interface used for mobile communications, increasing system performance and capacity [1].

Radio bandwidth as a resource is very limited, and the high increase in the number of wireless network users together with their requirements is pushing towards the adoption of novel

technologies to improve the spectrum efficiency. LTE was one of the first technologies to adopt the usage of such technologies, like Multiple Input Multiple Output antennas (MIMO) or orthogonal frequency division multiplexing (OFDM), boosting the network performance by increasing the spectrum efficiency [2,3]. For 5G technology, multiple types of modulation techniques and waveforms were considered [4], but finally Cyclic Prefix (CP) OFDM was selected, with a variation called Discrete Fourier Transform spread OFDM (DFT-s-OFDM or as it is also known, SC-FDMA) used for specific scenarios.

From an economical perspective, 3GPP LTE and 5G technologies are different, because they are currently at different maturity stages. 4G LTE is the more mature technology, with standardization efforts starting in 2005 and having its first commercial deployment more than ten years ago, in Sweden. Works on 5G 3GPP standards started just in 2015, with first specifications being ready in late 2017. First commercial deployment on a large scale took place in 2019 in South Korea, and it was for the 5G NSA (non-standalone), meaning the 5G RAN Radio Access Network (RAN) is deployed on top of the 4G network core. The 5G SA (standalone) relies on a completely new Packet Core Network, designed specifically for 5G technology. As of June 2021, at least 12 network operators have commercially available 5G SA networks [5].

In this paper a comparative study between 3GPP LTE and 5G technologies is presented, focused on Physical and MAC layers. Some important aspects related to overall system architecture are discussed, highlighting the similarities and the differences with regards to the radio and protocol aspects of the air interface.

The structure of the paper is as follows: Part 2 introduces some general concepts about the architecture of the two technologies, while in Part 3 and Part 4 some radio and protocol aspects of the air interface are discussed. In the last part, conclusions are presented.

2 4G LTE and 5G NR – General aspects

A. 4G LTE overview

Long term evolution technology came as a natural development of previous cellular technologies, UMTS/HSDPA, to be able to support highly increasing user demands in terms of throughput, mobility, and Quality of Service (QoS). Compared to its 3G predecessors that used a fixed channel bandwidth of 5MHz, LTE offers multiple channel bandwidths, between 1.4MHz and 20MHz. Considering the radio interface, 3G technology employed the use of CDMA, while radio access part of 3GPP LTE standard is using OFDMA / SC-FDMA. Selection of SC-FDMA modulation for the uplink communication had the main purpose of reducing the PAPR (peak-to-average power ratio) which was significant in case of OFDMA. The reason for this is to extend the battery life of handheld devices, maximizing the network on-time for the users, with a proportional effect in the network operators' revenue. Regarding the duplexing scheme, support for both FDD (Frequency Division Duplex) and TDD (Time Division Duplex) [6] is introduced, increasing the overall configuration flexibility and adaptability of this technology.

The theoretical peak downlink speed achievable in a 4G network vary from 100Mbps in case a SISO (Single Input Single Output) antenna configuration is used, to 326Mbps in case of a maximal MIMO configuration (4x4). For uplink, peak data rates go as high as 86Mbps, when 64QAM modulation is used. The introduction of all these new technologies had the purpose of enabling LTE to operate in a mode efficient manner in terms of radio resource usage, to be a more power efficient technology and to provide a much higher theoretical throughput to accommodate user demands. The architecture of LTE network was designed to be simple with open interfaces, so data from a user equipment passes through only two nodes until core network. LTE network architecture can be seen in Figure 1 [7,8], where the two nodes involved in the communication between the user and the core network are easily visible. The first node is the base station or eNodeB, after which the differentiation between user plane and control plane is taking place.

User data is routed through serving gateway (S-GW) while control data is directed towards the mobility management entity (MME) [9].

LTE standard was designed considering the backwards compatibility with previous 3GPP standards from the architecture point of view, making sure the network operators can integrate the new technology gradually, still using their current legacy networks to provide user services. The way LTE standard was created, makes it possible to separate the entire architecture in two main parts: the Core Network (CN) or the Evolved Packet Core (EPC), and the Access Network (AN), or the E-UTRAN. Together, CN and AN, they create the Evolved Packet System (EPS), which in turn has the role of connecting the mobile equipment to PDN (Packet Data Network), providing access to services powered by the internet (e.g. Voice over IP).

In the control plane, MME is the part of the EPS that is responsible of the following tasks: user authentication, bearer management, Non-Access Stratum signalling, mobility between different access networks, roaming and S-GW selection. In the user plane, the central role is played by the Serving GW. For all users connected to the EPS, there is only one S-GW associated, with role in transporting user data packets, and also acting as a mobility anchor in case of an inter base station handover.

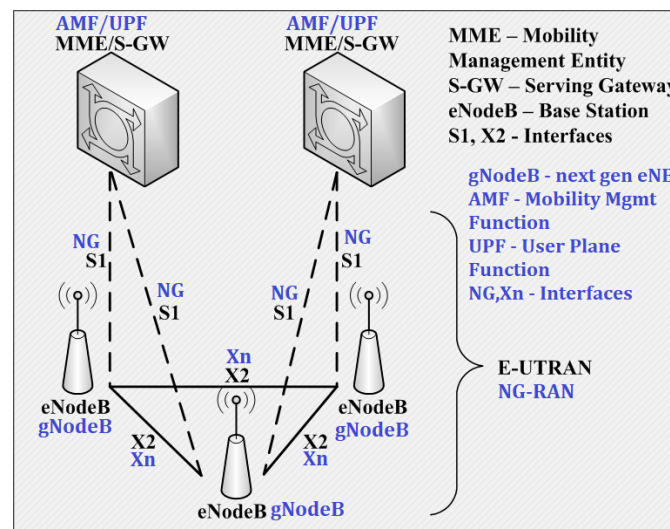


Figure 1: LTE (black) and 5G (blue) Overall Architecture

B. 5G NR overview

Ubiquitous connectivity we are witnessing today led to the development of a completely new generation of mobile communications, the 5G mobile communications system, where the approach to develop this new technology also changed, compared to previous generations. 5G precursors were mainly focused on using the state-of-the-art technology to achieve as much as possible for the operators and for the users, while 5G technology was designed with a focus on specific use cases and particular applications like vehicular communications or Internet of Things (IoT) applications consisting of huge number of sensors requiring low data rates and possibly also very low latency. This technology is capable of providing a high degree of flexibility, supporting at the same time low data rate applications like IoT's remote sensors, vehicular communications requiring extremely low latency and applications with huge data rate requirements.

Some very specific requirements were set when this standard was developed. In terms of peak data rates, it was targeted to achieve an increase of at least 20 times of current LTE data rates, meaning that 20Gbps in DL and 10Gbps in UL per gNodeB (5G compliant base station) should be possible. Mobility and connection density was also factored in from the beginning, with

connectivity support for speeds up to 500km/h and with a possibility to accommodate more than one million devices per square kilometre. Latency is also considered, with a reduction of 5 times as compared to LTE (4ms in 5G vs 20ms in LTE).

Considering the proven advantages of OFDMA waveforms used in LTE and other wireless standards like IEEE 802.11 family, the 5G standardisation body selected it as the waveform to be used also in the new technology. The reasons for selecting of OFDMA for downlink transmissions and SC-FDMA for uplink communication were identical between 4G and 5G, but 5G NR came with some important additions. For example, the subcarrier spacing that was fixed in LTE to 15kHz, is now becoming flexible and can be increased to 240kHz, still preserving the carriers' orthogonality. The new modes for subcarrier spacing are introduced to support new channel bandwidths, that can go up to 400MHz in case 5G is operating in mmWave bands. Another important difference compared to LTE is the support of mmWave bands, and 5G standard even considers using frequencies up to 86GHz for specific applications. Massive MIMO and active beam steering are other technical additions to the standard, in its attempt to achieve all goals that were set up when the development began, in 2015.

Regarding the system architecture, this is very similar with LTE architecture, as it can be observed in Figure 1. It is also a two-node architecture, because only 2 nodes are standing between the user and the 5G core network. Because this new technology was designed to accommodate many different applications, some special techniques had to be used to increase the scalability and efficiency of these 5G networks, and they were implemented in the new 5G core network. Techniques like Software Defined Networking (SDN), Network Functions Virtualisation (NFV) or Network slicing found their place in 5G core network and they all work together to support the predicted 1.2 billion 5G active connections by 2025 [10].

3 Air Interface - Radio Aspects

The size and actual frequency of available radio spectrum for a specific technology are some of the key factors in accommodating current wireless data usage, since spectrum efficiency solutions are proven insufficient to keep pace with the huge increase in data traffic.

4G LTE technology uses licensed frequency channels starting from 451MHz (band 72) up until 3800MHz (band 43), most used frequency being 1800MHz (band 3). In unlicensed spectrum, LTE uses the 5GHz band, especially for carrier aggregation. In terms of channel bandwidth, the 3GPP standard defines specific sizes, between 1.4MHz up to 20Mhz. Channel bandwidth directly affects the number of resource blocks (RBs) supported, and this varies from 6 resource blocks for a 1.4MHz bandwidth to 100 RBs for a 20MHz bandwidth. The definition of LTE-Advanced introduced the concept of carrier aggregation, increasing the maximum available bandwidth (in contiguous or non-contiguous spectrum) to 100MHz [9].

5G technology supports two different frequency ranges, known as FR1 (frequency range 1) and FR2 (frequency range 2). Frequency bands included in FR1 are the same as the bands supported by 4G LTE technology, while FR2 includes only high frequency bands, above 7.125GHz, not used until now by traditional mobile operators. Channel bandwidth is also different between FR1 and FR2, with FR1 supporting channel bandwidths between 5MHz and 100MHz, while in FR2 channel bandwidth can vary from 50MHz to 400MHz. The introduction of millimetre waves (mmWaves) support (FR2) in 5G creates a big advantage in delivering very high data rates, even if new challenges must be considered, due to specific propagation characteristics for this type of frequencies [11].

Regarding the deployment of these two technologies in paired and unpaired spectrum, both have support for FDD (separate bands are used for uplink and downlink) and TDD transmissions (same frequency band is used for both uplink and downlink). In terms of peak data rates, 5G can obtain a much higher theoretical throughput, 20 times higher than the one obtained in LTE. But as

mentioned before, achieving this throughput is possible only in special conditions, due to propagation characteristics of mmWaves (much smaller coverage, signal levels affected by natural factors, etc).

Advanced antenna solutions are implemented in both LTE and 5G, with the main purpose of increasing capacity or to create spatial diversity. These technically complex antenna techniques are used in addition to available spectrum bands described above, to offer higher data rates to all users in a cell coverage area and to extend this coverage in specific use cases. The antenna techniques implemented in LTE standard are MIMO, Beamforming, and Spatial Division Multiple Access (SDMA) with either open or closed loop. Common MIMO antenna configurations for downlink used in LTE implementations are 2x2, 4x4 or 4x2, with 2/4 antennas on the base station side and 2/4 antennas on the mobile equipment. For uplink communications, LTE supports 2x2 MIMO, but with every new release of the standard, the number of MIMO schemes supported also increases. Beamforming technique is used in LTE to reduce interferences and to temporarily increase the link capacity in a certain area of a cell, by focusing the antenna on a specific zone [12]. In case of 5G technology, besides the solutions implemented and supported by LTE standard, some new antenna strategies are available: 5G massive MIMO and Beam Steering. Massive MIMO implementation in 5G means usage of a very large number of antenna elements to simultaneously transmit and receive data, increasing the number of active users that can receive/transmit a high throughput [13]. From a size perspective, the antenna system used for massive MIMO will have similar dimensions as the one used in LTE but the antenna elements will have a smaller size, increasing the total number of elements in the same physical space (antenna systems with 64, 96 or even 128 antennas are available). Beam Steering technique is similar to beamforming but applied to massive MIMO systems. Using specific algorithms for signal processing, best radio path is determined, and specific antenna elements can be set up to send the radio signal on that specific path to reduce unwanted interferences and to increase the efficiency of the communication.

In terms of waveforms used, both standards are using SC-FDMA in the uplink connection, while OFDMA is selected for downlink.

Based on the radio aspects of the air interface for the two protocols, 5G provides more flexibility in terms of available channel bandwidths and channel frequencies, offering also a much higher theoretical throughput. Multiple antenna techniques are also more advanced in 5G, but LTE new releases will most likely introduce support for new antenna configurations. But all these advantages that are brought by 5G also come with a much higher system complexity, especially for massive MIMO deployment and for mmWaves, where frequencies have different propagation characteristics compared to lower frequencies used in LTE and require more sophisticated network planning to ensure constant conditions for all users.

4 Air Interface - Protocol Aspects

As it can be observed in Figure 2, 4G LTE and 5G NR air interface protocol architectures are very similar, except one additional sublayer added in 5G User Plane L2, SDAP (Service Data Adaptation Protocol) [14]. This new protocol implements a new mechanism for QoS, being responsible for Quality-of-Service provisioning on the air interface of 5G technology, mapping specific QoS flows in a PDU session to a dedicated Data Radio Bearer, created with the correct QoS level. This protocol layer also classifies the outgoing packets with a specific QoS Flow ID (QFI), so that these packets are treated by the 5G Core Network according to their QoS level. In LTE, the QoS concept revolves around the Evolved Packet System (EPS) bearer, and it can have two different configurations in terms of bitrate: non-guaranteed bitrate bearer (non-GBR) and guaranteed bitrate bearer (GBR), uniquely identifying from a QoS point of view all data packets exchanged between a mobile equipment and the core network.

The radio protocol stack for the two technologies has two different architectures, depending on the type of data that is transmitted or received. If the data going through the stack is a signalling message, then it will use the Control Plane side of the architecture, while if it is user data, the User Plane side will be used. There is a common structure for both planes, similar to OSI model, where a Physical Layer is defined as L1, and where L2 layer structure consists of multiple sublayers: MAC, PLC (Radio Link Control) and PDCP (Packet Data Convergence Protocol) sublayers. Additionally, for 5G, we also have the SDAP sublayer as part of L2. Physical layer L1 is responsible of creating the transport channels and making them available for MAC sublayer, who in turn creates the Logical channels required by the RLC. Next step is the creation of Radio bearers, and these will be the actual bearers used for data transmission over the air interface, as the user plane terminates directly into eNodeB/gNodeB. On the control plane, RRC is the last protocol used on radio side and it enables the exchange of control messages between the mobile equipment and base station, also forwarding the non-access stratum (NAS) messages from core network to ensure that critical functionalities like authentication, mobility management or security control can work properly.

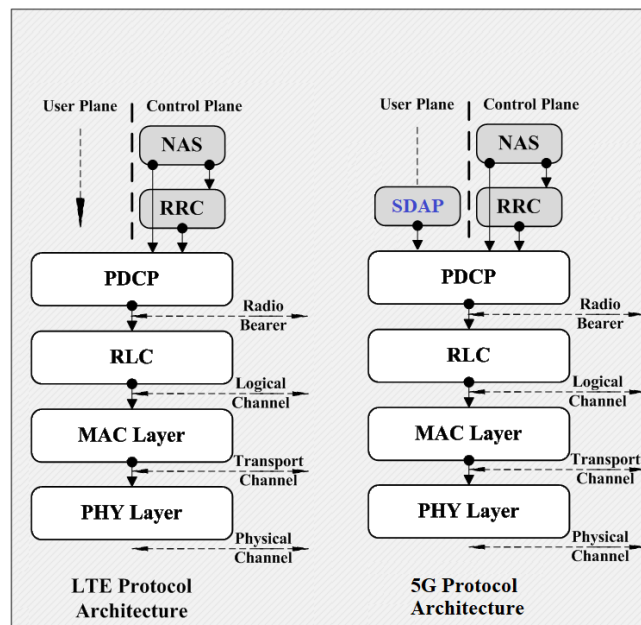


Figure 2: 4G LTE and 5G NR protocol stack

PDCP sublayer receives from below layer the PDCP SDU (Service Data Unit) packets and is responsible of PDCP SDUs retransmission triggering, PDCP PDU (Protocol Data Unit) forming and routing, duplicate deletion, header compression and (de)ciphering. Next, the RLC performs data transfer to MAC sublayer, and this can be done in three different transmission modes, for both 4G LTE and 5G NR: TM or Transparent mode, used generally for real time data transfer, AM or Acknowledged mode, used for non-real-time data transfer, and UM or Unacknowledged mode, used for transmission of signalling information. Finally, MAC sublayer has the role of mapping the logical channels on the transport channels from Physical layer and performing error correction using HARQ (Hybrid Automatic Repeat Requests) [8,14]. Another important operation realized by MAC sublayer is the continuous reporting of scheduling information, ensuring that dynamic scheduling of radio resources is possible.

It is easy to observe that both standards are technically similar, but considering the present market, LTE has the advantage as being a proven technology with a very high adoption rate by the mobile operators and with a clear advantage in terms of compatible mobile devices. 5G NR,

despite being in the first stages of deployment for mobile communication operators, is already receiving support from mobile device manufacturers that are creating affordable 5G terminals. And considering the current demand for more connectivity and for deployment and support of special applications like vehicular communications or IoT, the path towards a global adoption for 5G seems to be set.

5 Conclusions

In this paper a comparative study between 4G LTE and 5G NR technologies was proposed, to determine if differences between the two technologies could provide some indications about their future. Table 1 summarises a comparison between the two considered technologies, taking into consideration some important technological parameters for network operators and for the final users. These parameters were selected based on the comparison done in each section, to be able to better highlight the differences and similarities between 4G and 5G technologies.

Table 1: Comparison between 4G LTE and 5G NR

Parameter/Criterion	4G LTE (+/-)	5G NR (+/-)
Availability	++	-
Migration costs	++	-
Frequency band options	+	++
Peak data rates	+	++
Adaptive antenna systems	+	++
UL performances	+	++
DL performances	+	++
Mobility	+	++
Radio access modes (TDD&FDD)	++	++
QoS provisioning	++	++

Similar studies were also performed a decade ago, analysing the similarities and differences between WiMAX, that was already a proven technology, and 3GPP LTE, which was in its first trial stages. Many have given the edge to WiMAX, but the time has shown that 4G LTE emerged as the clear winner. But in this case, same as 4G didn't replaced completely 3G, it is expected that 5G will not fully replace LTE. As noted in the paper, 5G non-standalone (5G NSA) mode requires LTE infrastructure, and more than that, there are multiple scenarios where LTE is required to work together with the new standard (e.g. 4G LTE offering coverage to less populated areas while 5G takes care of the highly dense urban areas). A possible scenario is the one in which 4G takes the role of a bridge connecting major cities, where 5G will take over the demand and handle user connectivity in highly populated areas.

References

- [1] M. Lauridsen et co., “*From LTE to 5G for Connected Mobility*“, IEEE Communications Magazine (Volume 55, Issue 3), March 2017.
- [2] C. Deniz et co., “*On the performance of LTE downlink scheduling algorithms: A case study on edge throughput*“, Elsevier Computer Standards & Interfaces Journal Volume 59, August 2018.
- [3] S. R. Kondamuri and A. Sundru, “*Performance analysis of hybrid PAPR reduction technique for LTE uplink communications*“, Elsevier Physical Communication Journal Volume 29, August 2018.
- [4] S. Parkvall et co., “*NR: The New 5G Radio Access Technology*“, IEEE Communications Standards Magazine (Volume 1, Issue 4), December 2017.
- [5] *5G Standalone Update: Member Report*, Global Mobile Supplier Association, <https://gsacom.com>, June 2021
- [6] K. Bogineni, R. Ludwig et co. “*LTE Part I: Core Network – Guest Editorial*” in IEEE Communications Magazine, vol. 47, No. 2, February 2009.
- [7] S. Sesia, I. Toufik and M. Baker, “*LTE – The UMTS Long Term Evolution: From Theory to Practice*“, John Wiley&Sons Ltd, 2009.
- [8] 3GPP TS 36.300 v15.2.0, “*LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description*“, Release 15, July 2018.
- [9] 3GPP TS 36.104 v15.3.0 Release 15, “*LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*“, Release 15, July 2018.
- [10] Van-Giang Nguyen et co. “*SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey*“, IEEE Communications Surveys & Tutorials, Vol. 19, No. 3, Third Quarter 2017
- [11] 3GPP TR 38.912 v14.1.0, “*5G; Study on new radio access technology*“, October 2017.
- [12] I. Kim, J Um and S. Park, “*Performance analysis of the key aspects affecting capacity of 4G LTE networks*“, International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), October 2017.
- [13] H Asplund et co., “*Advanced Antenna Systems for 5G Network Deployments: Bridging the Gap Between Theory and Practice*“, Academic Press, ISBN 978-0-12-820046-9, 2020.
- [14] 3GPP TS 38.300 v16.7.0, “*5G; NR; NR and NG-RAN Overall description*“, Release 16, October 2021.

Vasile Horia Muntean
Lucian Blaga University
Computers and Electrical Engineering
Emil Cioran 4, 550025, Sibiu
ROMANIA
E-mail: horia.v.muntean@ulbsibiu.ro

Apparent personality analysis based on robust estimation

Milić Vukojičić, Mladen Veinović

Abstract

Apparent trait prediction is one of the hardest tasks in the domain of computer science. Multimodal approaches where the traits are extracted from various sources are giving the best prediction when we compare results to the Revised NEO Personality Inventory (NEO-PI-R model). Multimodal apparent trait personality based on text, images, audio, and handwriting can give us a wide range of estimation for each of the Big Five properties (extraversion, agreeableness, openness to experience, conscientiousness, neuroticism). This problem can be solved by using the aggregation function at the end of the model. The problem that this paper tries to overcome is to reduce the influence of the outliers on the final prediction of the model, by introducing nonlinearity to the final prediction of the model. Aggregation functions such as Max and Min are linear functions and they are creating outliers at the end of the model. We will introduce the nonlinear Huber function to get better results from the outliers and minor deviations from the output of NEO-PI-R model.

KEYWORDS: apparent personality analysis, personality classification, feature classification, robust loss function, aggregation function

1 Introduction

Detecting first impressions is one of the most advanced problems in the domain of computer science, machine learning, and artificial intelligence. The first impression can be seen as the initial perception of another person, where the observer can get a positive or negative evaluation of the observed person and estimate their psychological and physical characteristics. Personality Computing is the domain of computing that has developed greatly in recent years where detecting, predicting, and understanding human behavior is mainly based on soft computing methods. The main aim is to have traits that are measurable and stable over time [1]. Vinciarelli & Mohammadi showed in their paper [2] that interest for apparent personality detection from the year 2000 is in constant growth. While detecting Big Five personal traits is mostly done by individual input parameters such as text, image, audio, or handwriting in this paper we want to see a multimodal approach and combine these four inputs into one output that can predict each trait extraversion, agreeableness, openness to experience, conscientiousness and neuroticism so-called OCEAN or Big Five model [3]. Traits are suitable for statistical analysis [4] they also can have slight variation over the years, but we will not try to monitor them in the long run. Our goal is to detect traits in the context of behavior which is put in the situation where apparent personality is detected by an observer [5]. We will show the difference in the prediction between NEO-PI-R [6] tests taken by

subjects and traits detected from the model represented in this paper. Linear aggregation models can be very useful for minimizing different problems like slight differences between different outputs of the neural networks [7]. The main idea of this paper is to use nonlinear Huber function to deal with outliers from the aggregation layer [8].

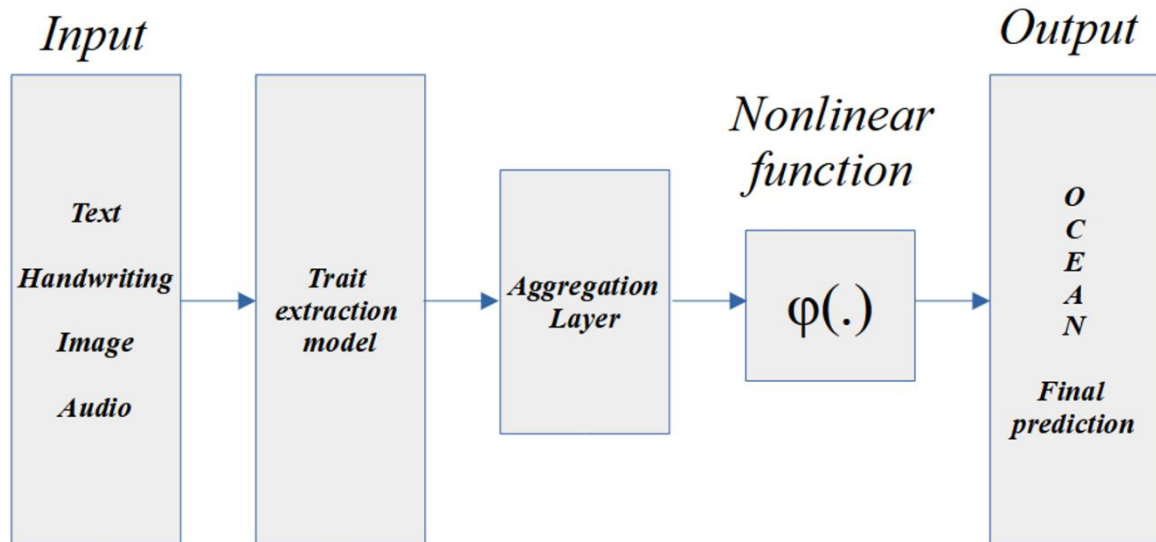


Fig. 1: The framework of the proposed model with nonlinear function, where input is text, handwriting, image, and audio, and extracted traits and output is represented as five (OCEAN) parameters between 0 and 1

The rest of this paper is organized as follows: Section 2 will introduce a multimodal approach of trait extraction, and also show some of the state of the art models for individual trait extraction from different inputs. Individual models have problems with trait prediction because of the uniformity of inputs. Section 3 will introduce the proposed nonlinear method which strives to overcome problems of the aggregation functions and introduce better fitting for individual traits at the end. Section 4 will show the experiment which is conducted on the subjects where we can see a comparison between the model proposed in this paper and earlier models.

2 Related work

In this section, we will review different models used to detect traits from subjects based on the different inputs like text, image, handwriting, and audio. We will also represent an aggregation model which can give us better results than individual models.

2.1 Detecting traits from different inputs

In the trait detection process, we can deal with the different inputs. Almost all work done is in the domain of the different neural networks which are trained for the specific set of inputs. For personality detection from text, we can see that models based on word vectorization and convolutional neural network [9] or Mairesse [10] and combinations of them are giving the best output for Big Five trait prediction from the text (O: 62.68, C: 57.30, E: 58.09, A: 56.71, N: 59.38). Traits extracted from the text are very close to the fifty-fifty chance prediction. The problem is that some papers are giving us the output of the model as 1 or 0 for each of the traits, we need to be

careful which activation function we can use in the final layers [11]. This is not a good representation of traits itself because we need to see even the slight variation in traits or their facets [12].

Visual-based personality extraction and traits extracted from images and models used for this purpose are giving better results than models based on the text itself. Most of the models are based on convolutional neural networks and long-short term memory [13,14]. The best accuracy of prediction is shown in the NJU-LAMBDA [15] model with an accuracy of 91.1. Traits like neuroticism and openness to experience are very hard to detect.

Audio-based personality extraction is mainly based on time domain and frequency domain features. Models like Linear Prediction Cepstral Coefficient (LPCC) [16] and Mel Frequency Cepstral Coefficients (MFCC) [17] are giving the best accuracy between 70 and 89 percent. The traits detected from the audio-based features are sometimes misleading because of the quality of the signal.

Personality detection from handwriting is maybe the hardest task because of the input parameters. This is also one of the oldest methods for personality trait prediction. Models proposed in papers Chen & Lin [18] and Hidden Markov Models and neural networks Fallah and Khotanlou [19] are showing the accuracy of 72 - 78 percent. On the first look, this can be seen as a very low accuracy rate for prediction but those models can be used to give the observer more information about certain traits and their facets.

	Accuracy of prediction
Personality detection from text [9]	62.68
Personality detection from image [13]	91.1
Personality detection from audio [16]	70 - 80
Personality detection from handwriting [18]	72 - 78

Table 1: Comparison of accuracy between the different models

2.2 Aggregation model

The Aggregation model shown in the paper Apparent Personality Analysis Based on Aggregation Model [7] is based on the different aggregation functions which are being used to have a better representation value out of multiple inputs in the multimodal trait prediction models. In the mentioned paper there are different aggregation functions shown like Min, Max, Mean and Median [20]. The overall prediction of the individual models based on text, images, audio, or handwriting can be improved when output of these models are combined, or so called multimodal trait prediction models. Aggregation functions are improving results of prediction of consciousness and extraversion when we use Median and Mean aggregation functions, but we still have problems with openness to experience. The problem that can be seen in the proposed aggregation model is that we have outliers that we cannot trust. Aggregation layers based on Min and Max functions are creating

the biggest deviation when we compare the output of the aggregation model with the NEO-PI-R test results. The main problem is the linearity of the model.

We can define an array of traits as $T_o = \{O_t, O_i, O_a, O_h\}$, $T_c = \{C_t, C_i, C_a, C_h\}$, $T_e = \{E_t, E_i, E_a, E_h\}$, $T_a = \{A_t, A_i, A_a, A_h\}$, $T_n = \{N_t, N_i, N_a, N_h\}$, where T_o represents OCEAN traits extracted from text O_t , image O_i , audio O_a and handwriting O_h . We need to create arrays for all of the OCEAN traits. Next step is to create sorted arrays, we will use $sorted(T_o)$, where $T_o = \{x_1, x_2, x_3, x_4\}$ and $x_1 \leq x_2 \leq x_3 \leq x_4$, this means that we can obtain $\min(T_o)$ and $\max(T_o)$, because of the sorted array as $\min(T_o) = x_1$ and $\max(T_o) = x_4$, $sorted(T_o) = \{x_{min}, x_2, x_3, x_{max}\}$. The same process can be used to obtain min and max aggregation functions on T_c, T_e, T_a and T_n .

Mean and Median aggregate function (1,2) will give a better estimation of traits. The main problem here is that some traits in some cases are better represented with outliers like Min and Max. Because of the even number of elements in array, Median must be defined with n which represent the length of the array, in our case each array will have $n = 4$, we will define $n_A = \frac{n}{2}$ and $n_B = \frac{n}{2} + 1$, where n_A represents the second number in the sorted array and n_B represent third number in the array.

$$mean(x) = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

$$median(x) = \frac{x_{n_A} + x_{n_B}}{2} \quad (2)$$

3 The proposed nonlinear method

The problem that we want to overcome with the newly proposed method in this paper is that Min and Max aggregation models are outputting a lot of outliers. To overcome this problem, we propose the model based on the Huber loss function [21] which is used in robust regression and we will make our model less sensitive to outliers created from the Min and Max aggregation layers. In our model we will use the Huber function over the square error loss function. In this paper, we choose the Huber function because is less sensitive to outliers than the square error function, and we can get a smoother approximation of our traits than simply using the Min and Max aggregation function.

Huber loss function (5) will be defined as:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta, \\ \delta(|a| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases} \quad (3)$$

If we have small values of a function will be quadratic and for large values of a function will be linear. The δ value that gives the best result is 1.35, where $\delta > 3$ gives results like Mean Squared Error and values for $\delta < 1.3$ gives results like Mean Absolute Error [22]. The aggregation model will give us four different parameters for each value of the Big Five model (O, C, E, A, N). This means that each trait will have value from Min, Max, Mean and Median aggregation (Omin, Omax,

Omean, Omedian ... Nmin, Nmax, Nmean, Nmedian). To use the Huber function, we will sort the values from smallest to largest.

In Figure 2 we will display the difference between the Huber loss function and square error loss and fit the function on four random sample values.

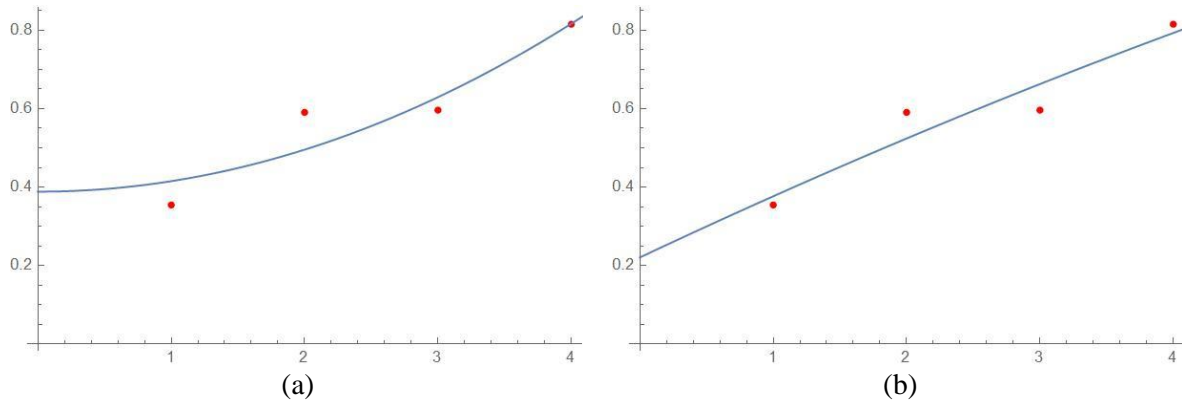


Fig. 2: The difference between Huber (a) and square loss (b) function showed on a random sample

We can see that if we use the Huber function for fitting in Figure 2 on the set of values $\{0.354752, 0.590692, 0.596393, 0.81523\}$ we will get the better fitting than square error loss function and our model will be less sensitive to the outliers. Let's take a random person from our database. This person's traits are represented with four different models, models based on text, image, audio, and handwriting. The output of the model is shown in Figure 3.

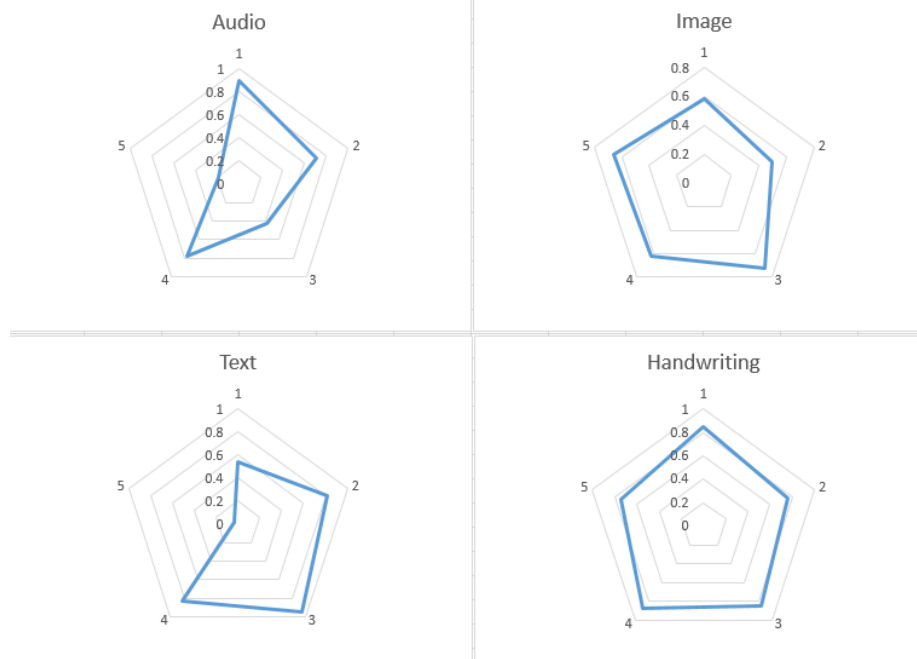


Fig. 3: Output of audio, image, text and handwriting model (1 - O, 2 - C, 3 - E, 4 - A, 5 - N)
In the Figure 4 we can see the output of the Min and Max aggregation model, as well as the output of the Huber function and NEO-PI-R test.

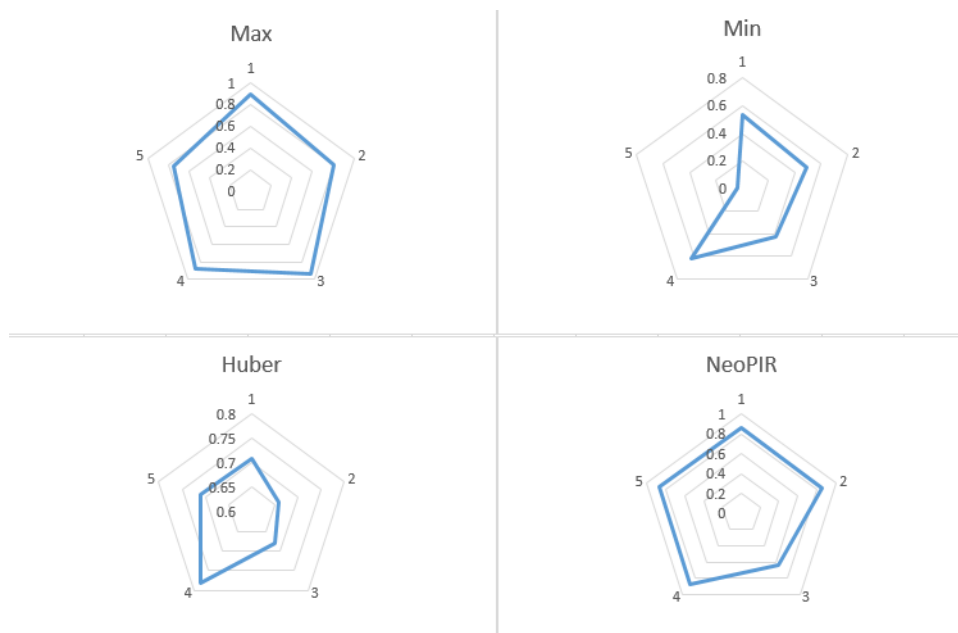


Fig. 4: Output of Min and Max aggregation functions, Huber and NEO-PI-R results

As we can see the model based on Huber function is less sensitive to the outliers that are made from Min and Max aggregation model. This also make model based on Huber function more accurate than models based on Max and Min aggregation with we compare it to the NEO-PI-R test results.

4 Experiments

Our experiment is based on the database of the 64 subjects between age 18 and 70 with 31 female and 33 male participants. Among the participants, we have different education levels and different job titles. We will compare the result of the NEO-PI-R test of each participant with the result of the proposed model. The distribution among the subjects is represented in Figure 5.

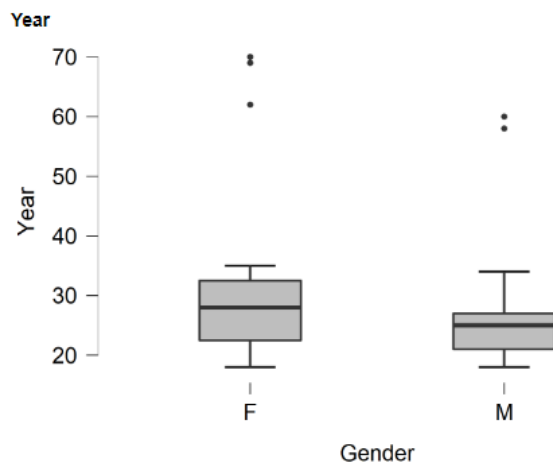


Fig 5: Distribution of male and female subjects

The subject did the NEO-PI-R test and the distribution of their test results are represented in Figure 6.

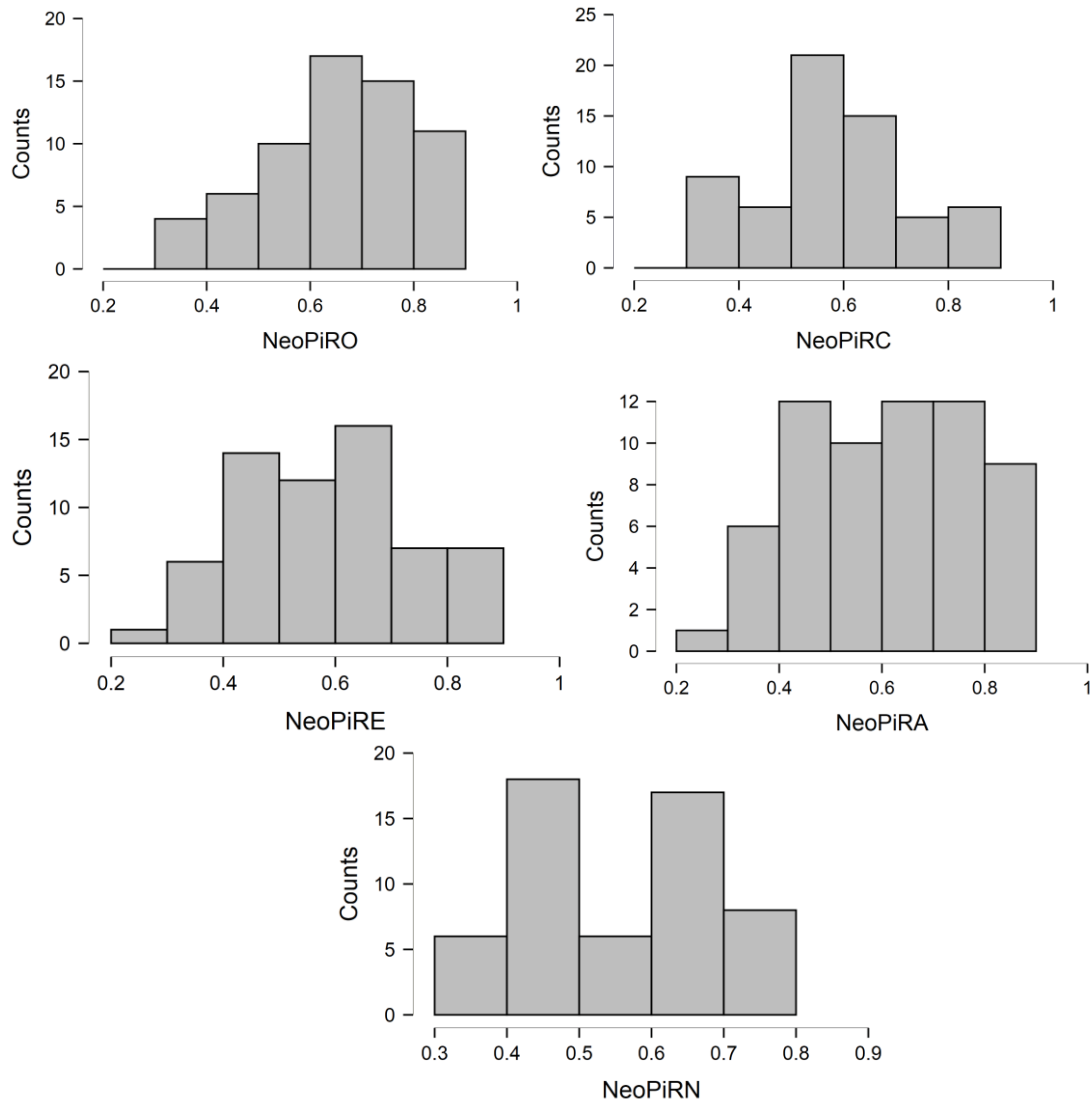


Fig. 6: Distribution of NEO-PI-R test results

The main problem shown in the previous research [7] based on the aggregation function Min, Max, Mean and Median is that Min and Max can be seen as outliers from the NEO-PI-R results. In Figure 7 We are showing the difference between NEO-PI-R tests and Min and Max aggregation models on openness to experience and extraversion.

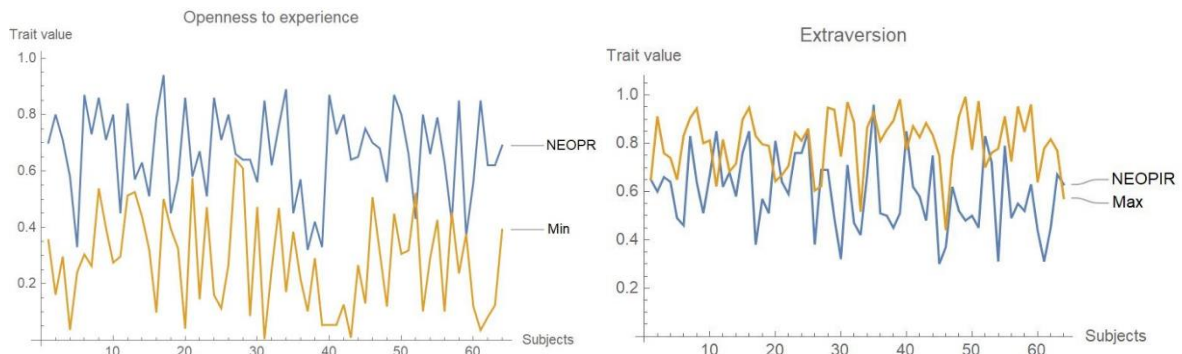


Fig. 7: Results of Min and Max aggregation models compared to NEO-PI-R test results, Extraversion NEO-PI-R model and Max aggregation and Openness to experience NEO-PI-R model and Min aggregation

The proposed model of using nonlinearity and the Huber function on the final aggregations (Min, Max, Mean, Median) can deal with outliers better and give more precise results when we compare it to the NEO-PI-R results. In Figure 8 We can see the nonlinear model based on the Huber function.

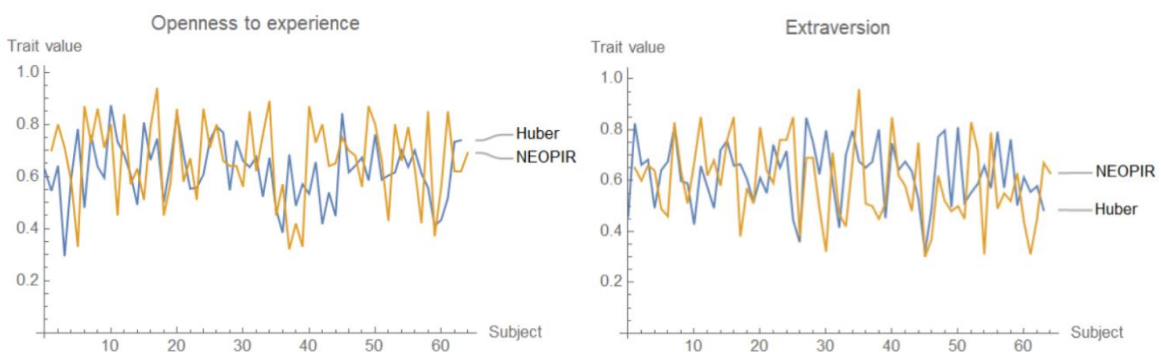


Fig. 8: Results of nonlinear models based on Huber function compared to NEO-PI-R test results, Extraversion NEO-PI-R model and nonlinear model and Openness to experience NEO-PI-R model and nonlinear model

Results of absolute and relative error and comparison of the previous models based on aggregation are shown in Table 2. We can see that the Nonlinear Huber function can be used to improve overall results in the case where we have outliers created from the Max and Min aggregation function. While Mean and Median aggregation functions are also giving us a very close approximation of the subject's traits [7], nonlinear models based on the Huber function can improve the overall prediction of the models when we use it for detection and to make our model less sensitive to outliers from the aggregation layers.

	Absolute error[Δx]					Relative error[δx]				
	O	C	E	A	N	O	C	E	A	N
Max aggregation	0.27	0.29	0.35	0.32	0.34	81%	79%	75%	71%	77%
Min aggregation	0.55	0.46	0.46	0.45	0.48	52%	60%	72%	63%	68%
Nonlinear Huber function	0.17	0.16	0.13	0.14	0.14	24%	28%	25%	24%	23%

Table 2: Results of proposed nonlinear model based on Huber function compared to Max and Min aggregation models

5 Conclusion and future development

Apparent personality analysis based on the different inputs can be seen as a challenging problem in the domain of psychology and computer science. While NEO-PI-R can be one of the best tools for detecting personal traits, models based on handwriting, images, audio, and text can come very close to the result of the NEO-PI-R tests. Also, we have seen that models based on the linear aggregation function can give some improvements but they can also create outliers. The model proposed in this paper can overcome the issue created with outliers and give an even better prediction in the term of apparent personality traits detection. The concrete application of the model can be seen in the domain of when companies recruit and select candidates, psychological assessment, schools can use it during student selection, and many more real-world applications.

Future work can be based on soft computing methods like metaheuristic and swarm intelligence algorithms which can give a more flexible prediction of the traits. In the domain of apparent personality analysis and detection not only of Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism but also their facets which can represent more accurate traits of the subjects.

References

- [1] Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). Personality traits. Cambridge University Press.
- [2] Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3), 273-291.
- [3] Raad, B. D. E., & Perugini, M. E. (2002). *Big five factor assessment: Introduction*. Hogrefe & Huber Publishers.
- [4] Vinciarelli, A. (2016). Social perception in machines: The case of personality and the Big-Five traits. In *Toward Robotic Socially Believable Behaving Systems-Volume II* (pp. 151-164). Springer, Cham.
- [5] Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177-182.
- [6] Costa, P. T., & McCrae, R. R. (1992). *Neo personality inventory-revised (NEO PI-R)*. Odessa, FL: Psychological Assessment Resources.

- [7] Vukojičić, M., Veinović, M. (2021). Apparent Personality Analysis Based on Aggregation Model. Paper presented at Sinteza 2021 - *International Scientific Conference on Information Technology and Data Related Research*. doi:10.15308/Sinteza-2021-220-225
- [8] Barron, J. T. (2019). A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4331-4339).
- [9] Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2), 74-79.
- [10] Mikolov, T., Yih, W. T., & Zweig, G. (2013, June). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 746-751).
- [11] Sibi, P., Jones, S. A., & Siddarth, P. (2013). Analysis of different activation functions using back propagation neural networks. *Journal of theoretical and applied information technology*, 47(3), 1264-1268.
- [12] DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of personality and social psychology*, 93(5), 880.
- [13] Chen, M., Wang, S., Liang, P. P., Baltrušaitis, T., Zadeh, A., & Morency, L. P. (2017, November). Multimodal sentiment analysis with word-level fusion and reinforcement learning. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (pp. 163-171).
- [14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [15] Zhang, C. L., Zhang, H., Wei, X. S., & Wu, J. (2016, October). Deep bimodal regression for apparent personality analysis. In *European conference on computer vision* (pp. 311-324). Springer, Cham.
- [16] Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4), 357-366.
- [17] Costa Jr, P. T., & McCrae, R. R. (2008). *The Revised Neo Personality Inventory (neo-pi-r)*. Sage Publications, Inc.
- [18] Chen, Z., & Lin, T. (2017). Automatic personality identification using writing behaviours: an exploratory study. *Behaviour & Information Technology*, 36(8), 839-845.
- [19] Fallah, B., & Khotanlou, H. (2016, April). Identify human personality parameters based on handwriting using neural network. In *2016 Artificial Intelligence and Robotics (IRANOPEN)* (pp. 120-126). IEEE.
- [20] Grabisch, M., Marichal, J. L., Mesiar, R., & Pap, E. (2009). *Aggregation functions* (No. 127). Cambridge University Press.
- [21] Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics* (pp. 492-518). Springer, New York, NY.
- [22] Kovačević, B., Milosavljević, M. M., Veinović, M., & Marković, M. (2017). Robust digital processing of speech signals. Springer.

Milić Vukojičić
Singidunum University
Napredni sistemi zaštite
Danijelova 32, 11000 Beograd
Serbia
E-mail: vukojcic.milic@gmail.com

Mladen Veinović
Singidunum University
Napredni sistemi zaštite
Danijelova 32, 11000 Beograd
Serbia
E-mail: mveinovic@singidunum.ac.rs

LIST OF AUTHORS

- Kowshik BHOWMIK** College of Wooster
Mathematical and Computational Sciences
1189 Beall Ave, Wooster, OH-44691
USA
E-mail: kbhowmik@wooster.edu
- Adoris-Eliau DORAN** Lucian Blaga University of Sibiu
Faculty of Sciences, Dept. of Informatics
Str. Dr. I. Rațiu, No. 5-7, Sibiu 550012
ROMANIA
E-mail: elian.doran@ulbsibiu.ro
- Martin S. DZHUROV** University of Rouse "Angel Kanchev"
Department of Natural Sciences
8 Studentska str., POB 7017, Ruse
BULGARIA
E-mail: martindzhurov2000@gmail.com
- Răzvan Gheorghe FILEA** Babes-Bolyai University
Informatics
Str. Mihail Kogălniceanu, nr. 1, Cluj-Napoca,
ROMANIA
E-mail: razvan.filea@gmail.com
- Jana GAWRILOW** University of Applied Sciences Würzburg-Schweinfurt
Faculty of Computer Science and Business Information
Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: jana.gawrilow@student.fhws.de
- Susanna GÖTZ** University of Applied Sciences Würzburg-Schweinfurt
Faculty of Design
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: susanna.goetz@fhws.de
- Patrick HAASE** University of Applied Sciences Würzburg-Schweinfurt
Faculty of Computer Science and Business Information
Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
Email: patrick.haase@student.fhws.de
- Felix HUSAC** Lucian Blaga University of Sibiu
Sciences Department
Str. Doctor Ion Rațiu, Nr.5-7, Sibiu, 550012,
ROMANIA
E-mail: felix.husac@ulbsibiu.ro

Srianuradha KANDAVEL MLCI Laboratory,
EECS Department, University of Cincinnati
OH 45220
USA
E-mail: kandavsa@mail.uc.edu

Matthias KECKL University of Applied Sciences
Würzburg-Schweinfurt
Faculty of Computer Science and
Business Information Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: matthias.keckl@outlook.de

Yulian KURYLIAK Institute of Computer Science and
Information Technologies,
ICSIT of Lviv Polytechnic National University,
12 Stepan Bandera street, Lviv, 79000
UKRAINE
Email: yulian.kuryliak.mnsam.2021@lpnu.ua

Marin MATEI Gheorghe Lazăr National College
1-3 Gheorghe Lazăr str, Sibiu, 500035
ROMANIA
E-mail: matei.i.marin@gmail.com

Vasile Horia MUNTEAN Lucian Blaga University
Computers and Electrical Engineering
Emil Cioran 4, 550025, Sibiu
ROMANIA
E-mail: horia.v.muntean@ulbsibiu.ro

Anca L. RALESCU University of Cincinnati
EECS Department
2600 Clifton Ave, Cincinnati, OH-45220
USA
E-mail: ralescal@ucmail.uc.edu

Lui RUCK University of Applied Sciences Würzburg-Schweinfurt
Faculty of Computer Science and Business Information
Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: lui.ruck@student.fhws.de

Jule SCHUMANN University of Applied Sciences Würzburg-Schweinfurt
Faculty of Computer Science and Business Information
Systems
Sanderheinrichsleitenweg 20, 97074 Würzburg
GERMANY
E-mail: jule.schumann@student.fhws.de

Mladen VEINOVIĆ

Singidunum University
Napredni sistemi zaštite
Danijelova 32, 11000 Beograd
Serbia
E-mail: mveinovic@singidunum.ac.rs

Milic VUKOJIĆIĆ

Singidunum University
Napredni sistemi zaštite
Danijelova 32, 11000 Beograd
Serbia
E-mail: vukojicic.milic@gmail.com

SPONSORS (in alphabetical order)



AUSY Technologies Romania



Asociația BIT



CodexWorks technologies



Fundația Academia Ardeleană



**Global Solutions for
Development**



Keep Calling



Nagarro



NTT Data



PAN FOOD



Omeron Technologies, Romania



ProIT



ROPARDO



Top Tech



VISMA