EMERGING 2019

The Eleventh International Conference on Emerging Networks and Systems Intelligence

ISBN: 978-1-61208-740-5

September 22 - 26, 2019

Porto, Portugal

EMERGING 2019 Editors

Jaime Lloret Mauri, Polytechnic University of Valencia, Spain

EMERGING 2019

Forward

The Eleventh International Conference on Emerging Networks and Systems Intelligence (EMERGING 2019), held between September 22-26, 2019 in Porto, Portugal, constituted a stage to present and evaluate the advances in emerging solutions for next-generation architectures, devices, and communications protocols. Particular focus was aimed at optimization, quality, discovery, protection, and user profile requirements supported by special approaches such as network coding, configurable protocols, context-aware optimization, ambient systems, anomaly discovery, and adaptive mechanisms.

Next-generation large distributed networks and systems require substantial reconsideration of exiting 'de facto' approaches and mechanisms to sustain an increasing demand on speed, scale, bandwidth, topology and flow changes, user complex behavior, security threats, and service and user ubiquity. As a result, growing research and industrial forces are focusing on new approaches for advanced communications considering new devices and protocols, advanced discovery mechanisms, and programmability techniques to express, measure and control the service quality, security, environmental and user requirements.

The conference had the following tracks:

- Technology and networking trends
- Quality and optimization

We take here the opportunity to warmly thank all the members of the EMERGING 2019 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to EMERGING 2019. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the EMERGING 2019 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that EMERGING 2019 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of emerging networks and systems intelligence. We also hope that Porto provided a pleasant environment during the conference and everyone saved some time to enjoy the historic charm of the city.

EMERGING 2019 Chairs

EMERGING Steering Committee

Carl James Debono, University of Malta, Malta
Raj Jain, Washington University in St. Louis, USA
Ioannis Moscholios, University of Peloponnese, Greece
Henrik Karstoft, Aarhus University, Denmark
Samuel Fosso Wamba, Toulouse Business School, France
Masakazu Soshi, Hiroshima City University, Japan
Robert Bestak, Czech Technical University in Prague, Czech Republic
Dong Ho Cho, KAIST, Republic of Korea
Tamás Matuszka, INDE R&D, Budapest, Hungary
Emilio Insfran, Universitat Politecnica de Valencia, Spain
Mark Leeson, University of Warwick, UK

EMERGING Industry/Research Advisory Committee

Euthimios (Thimios) Panagos, Perspecta Labs Inc, USA
Dimitris Kanellopoulos, University of Patras, Greece
Linda R. Elliott, Army Research Laboratory, USA
Jason Zurawski, Lawrence Berkeley National Laboratory / Energy Sciences Network, USA
Rawa Adla, University of Detroit Mercy | Ford Motor Company, USA
Arun Das, Visa Inc., USA
Shaohan Hu, IBM Research, USA
Xiaochuan Fan, HERE North America LLC, USA

EMERGING 2019

COMMITTEE

EMERGING Steering Committee

Carl James Debono, University of Malta, Malta
Raj Jain, Washington University in St. Louis, USA
Ioannis Moscholios, University of Peloponnese, Greece
Henrik Karstoft, Aarhus University, Denmark
Samuel Fosso Wamba, Toulouse Business School, France
Masakazu Soshi, Hiroshima City University, Japan
Robert Bestak, Czech Technical University in Prague, Czech Republic
Dong Ho Cho, KAIST, Republic of Korea
Tamás Matuszka, INDE R&D, Budapest, Hungary
Emilio Insfran, Universitat Politecnica de Valencia, Spain
Mark Leeson, University of Warwick, UK

EMERGING Industry/Research Advisory Committee

Euthimios (Thimios) Panagos, Perspecta Labs Inc, USA
Dimitris Kanellopoulos, University of Patras, Greece
Linda R. Elliott, Army Research Laboratory, USA
Jason Zurawski, Lawrence Berkeley National Laboratory / Energy Sciences Network, USA
Rawa Adla, University of Detroit Mercy | Ford Motor Company, USA
Arun Das, Visa Inc., USA
Shaohan Hu, IBM Research, USA
Xiaochuan Fan, HERE North America LLC, USA

EMERGING 2019 Technical Program Committee

Andrea F. Abate, University of Salerno, Italy Vision and image related trends
Mohd Helmy Abd Wahab, Universiti Tun Hussein Onn Malaysia, Malaysia
Manal Abdullah, King Abdulaziz University, KAU Jeddah, Saudi Arabia
Rawa Adla, University of Detroit Mercy | Ford Motor Company, USA
Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Shakeel Ahmad, Southampton Solent University, UK
Mohammed GH. I. AL Zamil, Yarmouk University, Jordan
Firkhan Ali Bin Hamid Ali, Universiti Tun Hussein Onn Malaysia, Malaysia
Adel Al-Jumaily, University of Technology, Sydney, Australia
Cesar Analide, Universidade do Minho, Portugal
Antonio Arena, University of Pisa, Italy
Tulin Atmaca, Telecom SudParis, France
Mozhgan Azimpourkivi, Florida International University, USA
Sabur Hassan Baidya, University of California Irvine, USA
Uldanay Bairam, Wageningen Food & Biobased Research, Netherlands

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

An RSU Placement Framework for V2I Scenarios 1
Baris Kara and Atay Ozgovde

Dynamic Path Discovery for In-band Control Plane Communication in a Tactical SDN Network 9
Anders Fongen

Semi-Automated Footwear Print Retrieval Using Hierarchical Features 16
Tim vor der Bruck and Thomas Stadelmann

The Use of E- Portfolio to Develop Student's Self-reflection in Pre-school - The case of a Private Saida (Lebanon) High School 23
Rouaa Chahine and Hassan M. Khachfe

Vibration Analysis with Application in Predictive Maintenance of Rolling Element Bearings 28
Theodor D. Popescu, Dorel Aiordachioaie, and Anisia Culea-Florescu

A Method of Feature Extraction from Time-Frequency Images of Vibration Signals in Faulty Bearings for Classification Purposes 34
Dorel Aiordachioaie, Theodor Popescu, and Bogdan Dumitrascu

Fault Detection using NLMS Adaptive Filtering for a Wastewater Treatment Process 40
Mihaela Miron, Anisia Culea-Florescu, and Mihai Culea

A Battery Charging Smart System Using a Power Management Algorithm and Adaptive Impedance 46
Nicusor Nistor, Laurentiu Baicu, and Bogdan Dumitrascu

An RSU Placement Framework for V2I Scenarios

Baris Kara
Department of Computer Engineering
Galatasaray University
Istanbul, Turkey
e-mail: bariskara35@gmail.com

B. Atay Ozgovde
Department of Computer Engineering
Galatasaray University
Istanbul, Turkey
e-mail: aozgovde@gsu.edu.tr

***Abstract*— Edge computing has become a prominent computing strategy when mobile devices and Internet of Things (IoT) became popular in the last decade and cloud computing could not meet the computational requirements of some of these devices/applications. What edge computing can provide differently from cloud computing is low latency in communication, high quality of service, and support for high mobility. Connected and autonomous vehicles scenarios can be considered as an important application field for edge computing as these are the key requirements to implement a vehicular network. In this study, we aim to present a solution to one of the high level problems in vehicular networks: efficient Road Side Unit (RSU) placement by addressing network coverage and computational demand. We propose an RSU placement framework for generating RSU placement models based on traffic characteristics of a target area. Moreover, our work includes extending capabilities of a simulation framework designed for edge computing scenarios. Therefore, we can evaluate the performance of the generated models and validate their functionality by running simulations on this environment.**

Keywords-edge computing; V2I; connected vehicles; roadside units.

I. INTRODUCTION

With increasing popularity of mobile devices and Internet of Things (IoT) in the last decade, cloud computing had been leveraged to solve the problem of making complex computations with limited device resources by provisioning remote computing and storage resources. Edge computing, on the other hand, was suggested as a new computing paradigm when the limitations of the centralized data centers started to emerge. Satyanarayanan et al. [1] describe these limitations as long Wide Area Network (WAN) latencies and bandwidth-induced delays. Because of these limitations, cloud computing is not a suitable computing strategy for scenarios which require real-time data processing and relies on fast feedback.

Edge computing is a good candidate to solve these problems by bringing computing resources to the edge of the network, usually one hop away from the user. The features of low latency in communication, high quality of service and support for high mobility makes edge computing an optimal solution for the computational requirements of a wide range of applications in different domains. Connected and autonomous vehicles scenarios are considered as a good application field for edge computing [2].

The components of the Vehicle-to-Infrastructure (V2I) scenarios can be mapped to edge computing elements as follows:

- Road Side Units (RSU) are the edge computing units in vehicular networks because of their proximity to the vehicles, providing computational, storage resources and high bandwidth link, and transfer data with minimum latency.
- Vehicles are the resource poor clients as they have limited computation and storage resources due to the requirements of small-size and low-cost hardware systems [3].
- Vehicular applications are edge applications as they demand complex computation and large storage.

Applications deployed into RSUs receive data from vehicular applications such as trajectory, speed, destination coordinates, etc. in short intervals, aggregate and process them in real time and send response back to senders or to the relevant vehicles within the network range. Here again, low latency and high quality of service are the key factors to build this ecosystem.

Deploying a limited number of RSUs into a smart city is a challenging work since satisfying two requirements at the same time brings us to a trade-off problem. RSUs should be placed in an area in a way that satisfies both network coverage for vehicles and computational demand for the edge applications at maximum level considering the traffic density on the road network.

The objective of this study is to implement an RSU placement framework for generating RSU placement models based on traffic characteristics of an area. We aim to provide a flexible tool that can be configured for designing a placement model in favour of network coverage or computational demand. Additionally, our work includes extending capabilities of an open source simulation framework, EdgeCloudSim, proposed by Sonmez et al. [4]. By adding new modules to support simulations for V2I scenarios and designing realistic traffic scenarios for a target area in London city centre, we validate the functionality of the proposed RSU placement framework and evaluate the performances of the generated RSU placement models

The rest of the paper is organised as follows: Section II reviews the related work. Section III introduces the simulation environment, V2ISIM. In Section IV, the

 ISBN: 978-1-61208-740-5

reference scenario is described, and Section V explains RSU placement models. In Section VI, we address the RSU placement results, in Section VII, we analyse the simulation results, and finally, Section VIII outlines the concluding remarks.

II. RELATED WORK

Previous research addressing edge computing in vehicular networks mostly suggest new frameworks and architectures in which cloud and edge processing units, and mobile devices/vehicles are integrated into a new ecosystem. Their main focus is to provide solutions for computational challenges, such as resource allocation and Virtual Machine (VM) migration [3] [5] [6]. Our work can be considered as a complementary study which is built on top of an existing architecture addressed by these studies.

On the other hand, RSU placement problem has been addressed by several studies in both highway and smart city scenarios. Highway scenarios have different traffic characteristics than smart cities (e.g. fast moving vehicles, sparse traffic, etc.), therefore, the communication infrastructure should be designed considering these requirements. Studies focusing on RSU deployment to the highways [7] [8] [9] differ from our study from this aspect.

There are also studies addressing RSU placement in smart cities. These studies mostly approach the problem from network coverage aspect without taking computational demand into account. As a result, the placement models presented in these works do not guarantee fulfilling computational requirements of the edge applications deployed to the RSUs. In their study, Liang et al. [10] formulate optimal RSU deployment problem as an integer linear program (ILP). In their model, V2I communication is extended with multi-hop Vehicle-to-Vehicle (V2V) communication. Also, RSUs in their model can have different configuration settings. In our scenario, we don't consider multi-hop communication in order to minimize the latency in V2I communication, and all the RSUs have same capabilities. Chi et al. [11] propose an RSU allocation algorithm with a concept of intersection priority. They aim to maximize the intersection coverage by deploying RSUs to the important intersections. Similarly, Gomi et al. [12] propose an RSU placement method by calculating placement priority for each intersection. In their work, they also consider road elements that affect radio wave spreading such as buildings and aim for a better communication performance. These methods consider the intersections as deployment points of RSUs, whereas in our model, we divide the target area into cells and build our deployment logic on top of these cells. Trullols et al. [13] propose a maximum coverage approach for modelling the problem of RSU deployment. Their model is based on deploying RSUs as Dissemination Points (DPs) and maximizing the number of vehicles that contact the DPs. In the study of Balouchzahi et al. [14] the problem of RSU placement formulated to binary integer programming. Unlike from the other studies, their work address highway and urban scenarios at the same time. Similarly, Premsankar et al., [15] use mixed linear integer programming formulation for the problem, but their formulation focuses on minimizing the deployment cost of edge computing devices by jointly satisfying a target level of network coverage and computational demand.

III. V2ISIM

We needed a simulation environment for our study in order to validate the functionality of the proposed RSU placement framework and compare the performances of the generated RSU placement models. For this purpose, we used EdgeCloudSim, which is an open source tool designed for simulating edge computing scenarios where it is possible to conduct experiments that consider both computational and networking resources [4]. We extended the capabilities of the framework by defining components and modules specific to V2I scenarios and referred to this extended simulation environment as V2ISim.

EdgeCloudSim is also extended from another simulation environment, CloudSim, which allows modelling of cloud computing infrastructures and application services [16]. While EdgeCloudSim implemented edge processing units and modelled edge computing network, we introduced RSUs as computing units for V2I scenarios and extended the network model for vehicular network. We also implemented a mobility module to integrate traffic scenarios into the environment.

The key components we implemented in V2ISim are as follows:

- **RSUManager:** This component is responsible for creating RSU instances in the system based on the configuration provided. The configuration should include RSU resource definition as well as Global Positioning System (GPS) coordinates in decimal degrees.
- **TrafficLoadGenerator:** A traffic input file, which includes vehicle trajectory data, should also be provided to the simulation environment. *TrafficLoadGenerator* is responsible for creating tasks using task characteristics received from task configuration file. When the simulation starts running, these tasks are scheduled for processing in due course.
- **TrafficTaskBroker:** This component is responsible for managing the lifecycle of a task. After the task is created, there are 3 stages it has to follow until it is completed: vehicular application submits task to the RSU, task is processed in the RSU and finally, the response is sent back to the vehicle. When the task reaches to one stage, it is rescheduled by *TrafficTaskBroker* for the next one.
- **RSUOrchestrator:** Its responsibility is to find the RSU that the task will be submitted. To achieve this, first, nearest RSU to the vehicle is detected. Then, if the vehicle is within the range of the RSU, task is submitted to it by *TrafficTaskBroker*.

To find the nearest RSU for a given vehicle position efficiently, all RSU coordinates are saved in a K-D tree (K-Dimensional Tree) when the application starts. A K-D Tree is a data structure for efficient search and nearest-neighbour(s) computation of points in K-dimensional space.

 ISBN: 978-1-61208-740-5

We used an open source K-D tree implementation in our application [17].

- **RSUMM1Queue:** We use M/M/1 queue model to simulate the network delay. This component is responsible for calculating task upload and download delays.
- **SimLogger:** Lastly, all the important task data, RSU data and as well as calculated system metrics are logged by *SimLogger* in different logging levels.

At the end of the simulation, 3 output files are generated for each traffic input file provided:

- **Generic logs:** this file includes most important simulation results such as number of successfully processed tasks, number of failed tasks, average service time, average network delay and average RSU utilization rate. The values logged in this file are used as metrics while comparing system performances for different RSU placement models.
- **RSU utilization logs:** this file keeps the utilization rates for each RSU logged for each simulation second. These values are used as metrics while comparing system performances from utilization aspect for different RSU placement models.
- **Task assignment logs**: It keeps the logs of number of assigned and failed tasks for each RSU.

IV. REFERENCE SCENARIO

This section outlines the reference scenario that we considered for V2I communication and the target area we used for the case study, and explains our approach on generating traffic dataset to run our experiments.

A. Scenario and Parameters

In our reference scenario, we consider a smart city equipped with V2I communication infrastructure. All vehicles are smart or connected with the ability of running vehicular applications. Vehicular applications send one task to the nearest RSU per second in case the vehicle is in the network coverage of any RSU and the data is processed by edge applications deployed to the RSUs. When the task is successfully processed, RSU sends a response back to the vehicle. There are 4 cases a task can fail:

- **Coverage:** Vehicle is not in range of any RSU's network
- **Capacity:** RSU is out of capacity and cannot process incoming task
- **Bandwidth:** Task cannot be sent through network due to congestion
- **Mobility:** Vehicle leaves the RSU network coverage after sending the task

We assume all RSUs have same hardware capacity and the tasks sent by the applications are identical. In our scenario, each RSU has 1 Mbps bandwidth. Average task payload size is 1024 bytes for both upload and download operations. We also assume that each RSU has an equipped server with 600Mhz Central Processing Unit (CPU) and 500MB Random Access Memory (RAM), and average task length is 300 Machine Instructions (MI). Table I shows the parameters for RSU and task configurations. All the simulations run as part of this study are based on these values.

TABLE I. RSU AND TASK PARAMETERS AND VALUES

Parameter	Value
RSU Network Range	250m
RSU Bandwidth	1 Mbps
CPU	600 Mhz
Memory	500 MB
Average Task Payload Size	1024 byte
Average Task Length	300 MI
Task arrival rate	1 Hz

B. Target Area

We chose an area of 3 x 3 kilometres in London city centre as the target area for deploying RSUs. To be able to run traffic simulations and calculate RSU locations, we needed to extract the road network of the target area. To obtain the road network, we outlined the target area on *OpenStreetMap* [18], which is a free collaborative map application, then we exported it in *xml* format. Since the map data includes a variety of information such as buildings, parks, restaurants, etc. we processed the file to only include road network elements such as motorways, intersections, and traffic lights.

C. Traffic Dataset

Due to the lack of publicly available vehicle trajectory dataset for the target area, we used Simulation of Urban Mobility (SUMO) framework to generate realistic traffic dataset. SUMO is an open source, microscopic and continuous road traffic simulation framework designed to handle large road networks [19]. Apart from its simulation capabilities, SUMO includes several scripts for traffic and road network operations. We used *randomTrips* script in SUMO library to generate random vehicle routes on the road network. The output route file, along with the network file should be provided to SUMO to run a traffic simulation.

In our study, traffic density plays an important role on RSU placement process as the computational demand depends on number of vehicles in the system. Thus, to cover scenarios with different traffic volumes, we generated 8 trajectory files for 500, 1000, 1500, 2000, 2500, 3000, 3500, and 4000 vehicles in the target area. Each file contains trajectory data logged for each simulation second, such as vehicle id, type, coordinates, speed, angle, lane, etc. As a result, more than 8 million logs were produced in total for the traffic dataset. Figure 1 shows heat maps of the generated vehicle trajectories for number of vehicles 500 and 4000. Traffic congestions can be observed in the centre of the map when higher number of vehicles used in the scenario.

 ISBN: 978-1-61208-740-5

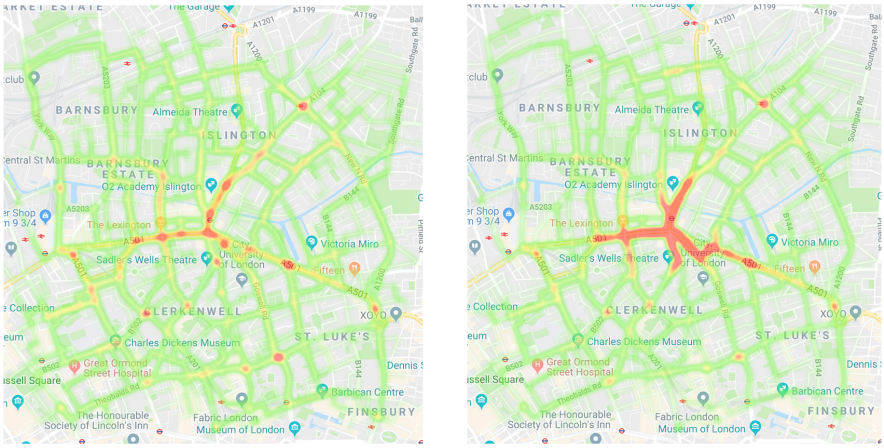


Figure 1. Heat maps of generated vehicle trajectories for the target area for number of vehicles (a) 500 and (b) 4000

V. RSU PLACEMENT MODELS

We generated 2 RSU distribution models addressing RSU placement problem in a smart city: Uniform RSU distribution and Weighted RSU distribution. This section outlines the algorithms we used for each distribution models.

A. *Uniform RSU Distribution*

Uniform RSU distribution serves for two purposes in our study: First, we used it as the base model which we made optimizations on the RSU locations in the next steps. Second, we used performance results of this model to compare with the results of generated placement models. The algorithm of this model is quite basic; after calculating the optimal distance between two RSUs, the number of RSUs required for full network coverage on the target area is calculated. Then, target area is divided into cells and RSUs are placed into these cells equidistant from each other. We referred to these cells as territories. Therefore, full network coverage is enabled in the target area, whereas computational demand is ignored.

B. *Weighted RSU Distribution*

Weighted RSU distribution model redistributes some RSUs in uniform distribution model by taking computational demand into account. In the uniform distribution model, despite of full network coverage, high task failure rates might be observed since RSUs might not meet high computational demand using their limited resources. It is especially expected to experience this problem in the territories with higher volumes of vehicle traffic, i.e., traffic congestions. An external parameter, θ, is the relocation factor, and it determines the number of the RSUs to be relocated. Relocation step addresses selecting θ% least utilized RSUs and move them to the territories where more computational resources are needed.

Therefore, we aim to decrease capacity related task failures by bringing additional computational resources to meet the higher demand. On the other hand, relocated RSUs will result in network related task failures as no RSUs will serve to vehicles at these territories. Value of θ should be assigned considering the difference of traffic volumes in different territories as this trade-off is only reasonable if total number of task failures decreases after the relocation.

The algorithm for this placement model consists of 4 steps:

- **RSU Selection:** This step addresses finding the RSUs placed at the territories with lower traffic volume, thus have low utilization rates. To detect these RSUs, we calculate task assignment rates for each RSUs in the uniform distribution. RSUs with less task assignment rates are marked to be moved in
- the territories with higher resource demand. We select θ% of least utilized RSUs in this step.
- **Territory Selection:** To detect territories that need additional resources to meet high computational demand, we analyse the performance of the RSUs in uniform distribution model under a heavy load. The territories containing the RSUs with higher capacity related task failure rates are the candidates to support with additional RSUs.
- **RSU Distribution:** In this step, we first calculate a weight factor using task failure rates for each candidate territory. Then using the weight factor, we calculate number of RSUs to be assigned into each territory. Finally, we distribute the selected RSUs into these territories.

  RSU Placement: This step addresses placing selected RSUs into the candidate territories. The first RSU is placed in the middle of territory centre and neighbour territory centre with the highest computational demand among all neighbours. The second RSU is placed between the territory centre and neighbour territory centre with the second highest computational demand, and so on.

VI. PLACEMENT RESULTS

We developed a Java application as the implementation of the suggested placement algorithms and referred to it as *RSU Distributor*.

A. *Uniform RSU Distribution*

Network range of RSUs can reach up to 1000 meters if there are no obstructions, and 250-350 meters in cluttered urban areas [20]. In our scenario, we assumed that each RSU works best with a coverage of 150 meters due to the shadowing effect of the buildings and we decided to place RSUs 300 meters far from each other. Therefore, to cover an area of 9 km² with RSUs working in their best performances, we needed to have 100 RSUs in total.

After assessing the number of RSUs to place, we processed the area map by dividing it into territories each with the size of 300 by 300 meters, and we assigned each territory an id sequentially. Then, we placed one RSU into the centre of each territory, therefore 100 RSUs were evenly distributed on the area. Figure 2 shows distribution of the RSUs into the territories based on the uniform distribution model.

B. *Weighted RSU Distribution*

To calculate the RSU coordinates for weighted RSU distribution model, we started by running V2ISIM for uniform distribution model, therefore, we generated the inputs required for weighted distribution algorithm: task assignment rates and task failure rates for each RSU. The simulation tool requires two input files: vehicle trajectory data and RSU coordinates. As traffic input data, we provided the traffic dataset we generated using SUMO and configured RSU and task characteristics by providing parameters listed in Table I. Some important simulation properties can be seen in Table II.

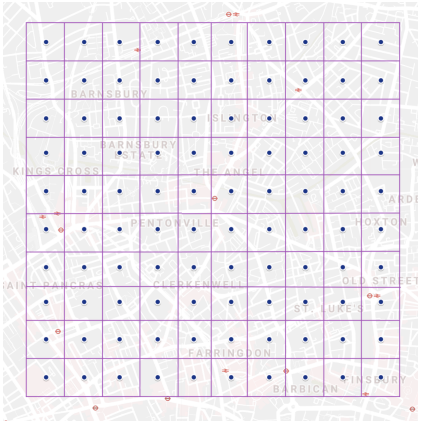


Figure 2. RSU Locations on Uniform Distribution Model

TABLE II. SIMULATION PROPERTIES

Parameter	Value
Total number of traffic logs	8 147 468
Total number of RSUs	100
RSU placement model	Uniform
Simulation time	1 hour

TABLE III. RSU IDS SELECTED FOR RELOCATION

θ	RSU ids
10	3, 11, 39, 4, 9, 49, 5, 90, 88, 2
20	3, 11, 39, 4, 9, 49, 5, 90, 88, 2, 74, 89, 79, 69, 1, 91, 6, 70, 93, 12
30	3, 11, 39, 4, 9, 49, 5, 90, 88, 2, 74, 89, 79, 69, 1, 91, 6, 70, 93, 12, 84, 98, 92, 87, 8, 99, 14, 59, 80, 19

TABLE IV. TERRITORY IDS AND NUMBER OF RSUS TO ASSIGN

θ	RSU ids
10	55(2), 54(1), 45(1), 35(1), 48(1), 33(1), 34(1), 65(1), 53(1)
20	55(3), 54(2), 45(2), 35(2), 48(2), 33(1), 34(1), 65(1), 53(1), 58(1), 47(1), 46(1), 75(1), 36(1)
30	55(4), 54(3), 45(3), 35(3), 48(3), 33(2), 34(1), 65(1), 53(1), 58(1), 47(1), 46(1), 75(1), 36(1), 25(1), 71(1), 38(1), 63(1)

In *RSU Distributor*, simulation logs were aggregated and processed to calculate the values of task assignment rates and task failure rates of the RSUs. By assigning 10, 20, and 30 to θ, we run the application and generated 3 different RSU placement models. For each value of the θ, Table III shows the selected RSUs for relocation and Table IV shows the number of RSUs to be assigned to each territory.

After running RSU distributor with these inputs, 3 different distribution models were produced based on weighted distribution model algorithm. Figure 3 shows RSU placements for θ=10, 20, and 30 respectively.

VII. SIMULATION RESULTS

For generated RSU placement models, we run a set of simulations on V2ISim using a laptop with Intel Core i7-8850H CPU and 16GB RAM. Table V shows the time spent to run each simulation.

We classify traffic densities of the traffic input files we used for the simulations into 3 categories:

- Number of vehicles below 1500 as low traffic volume
- Number of vehicles between 1500 and 3000 as medium traffic volume
- Number of vehicles more than 3000 as high traffic volume

The graph in Figure 4 shows comparison of task failure rates for uniform distribution and weighted distribution for θ=10, 20, and 30.

TABLE V. SIMULATION PROPERTIES

Simulation	Duration
Uniform RSU Distribution	6 hours 10 minutes
Weighted RSU Distribution θ=10	9 hours 6 minutes
Weighted RSU Distribution θ=20	6 hours 36 minutes
Weighted RSU Distribution θ=30	6 hours 19 minutes

 ISBN: 978-1-61208-740-5

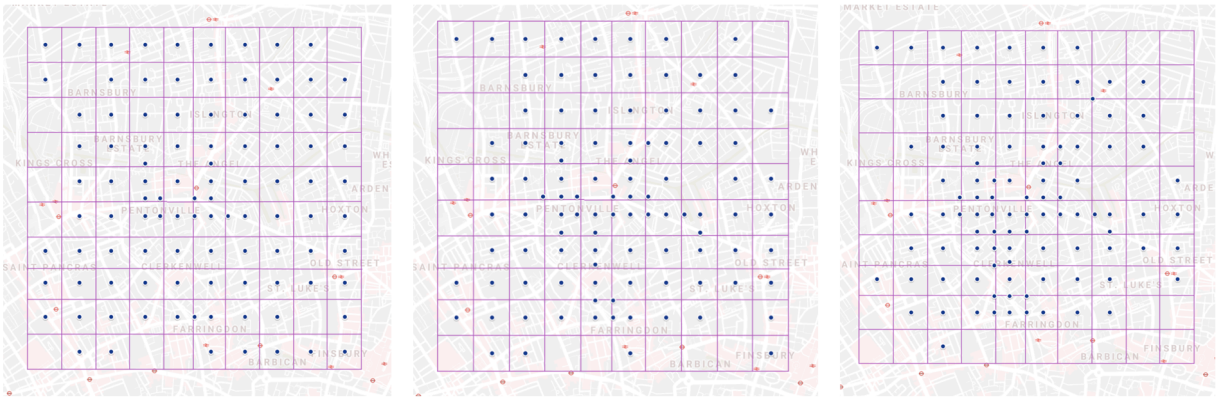


Figure 3. RSU Locations on Weighted Distribution Model for (a) θ=10, (b) θ=20, (c) θ=30

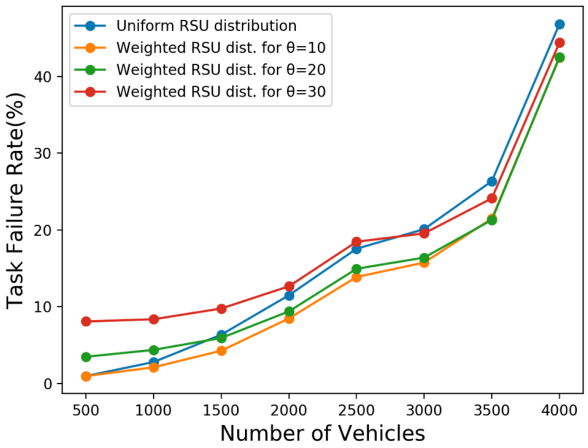


Figure 4. Task Failure Rates

Task failure rates can be considered as our most important metric while evaluating system performance. A system with low task failure rates is more reliable and functions better.

We can observe that the system functions best for weighted distribution model for θ=10 under any traffic volumes. The graph also shows that when the number of vehicles in the system increases, task failure rates also increase for all RSU distribution models. Considering the sharp increase between 3500 and 4000 vehicles for all models, we can claim that if the traffic density is over a threshold, RSUs will not handle the load and the system will crash. Below 1000 vehicles, there is no significant gap between weighted distribution model for θ=10 and uniform distribution model, however, after this point, we can observe an increase on this gap.

On the other hand, while uniform distribution model performs better than the weighted distribution models for θ=20 and 30 under low traffic volume, weighted distribution model for θ=20 outperforms it for medium traffic volume and weighted distribution model for θ=30 outperforms it for high traffic volume. This is because while network coverage is a more important factor for the low traffic volume, resource capacity becomes more critical than the other factors when traffic density increases.

Lastly, the graph shows that relocating less utilized RSUs to the territories with higher load improves the system to a certain point. Weighted distribution model for θ=10 outperforms uniform model for low, medium and high traffic volumes and it is the most optimal relocation factor among all the others. However, for θ=20, weighted model only performs better for medium and high traffic volumes, and for θ=30, it only functions better for high traffic volume. The reason for this is the trade-off between network coverage and resource capacity. When a less demanded RSU is relocated into a position to share the load in a busy area, capacity originated failure rates will decrease for the RSUs in the target territory, however coverage originated failure rates will increase for the original source territory.

As a result, by evaluating the results of Task Failure Rate graph, we can conclude that:

- θ=10 outperforms all others under any traffic load.
- uniform distribution model can be used for low traffic volume
- weighted model for θ=20 can be used for medium and high traffic volumes
- weighted model for θ=30 does not perform well under any traffic load

Figure 5 shows the comparison of average service time of the RSUs in the unit of seconds. The service time is the sum of download and upload delays and task processing time. As can be seen on the graph, increasing load had a similar impact on RSU service times for all distribution models, and all weighted distribution models performed better than the uniform model for all traffic volumes. The reason is, both download and upload delays and processing time depend on the RSU demand in that particular time. When an RSU needs to serve to higher number vehicles, they experience more delays on network and processing time. And as a result of sharing the high load with relocated RSUs, all weighted models provide better results in terms of service time.

While measuring system performance, another important metric is average utilizations of the RSUs. A system in which RSUs run with a low capacity is less efficient than another system with higher RSU utilization. On the other hand, a system with RSUs running in full capacity for a certain level of computational demand is not able to sustain higher loads. Since the simulations we run with low and medium traffic volumes do not create significant load on

 ISBN: 978-1-61208-740-5

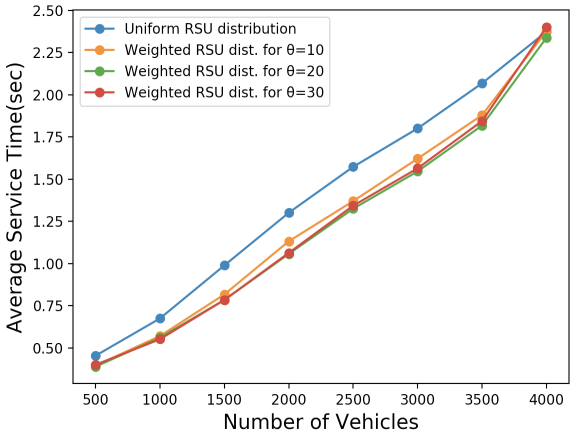


Figure 5. Average Service Time

majority of the RSUs, we compared utilization of RSUs using only the results of the simulations run with 3500 vehicles. 3500 is the number which creates the highest traffic volume without breaking the system.

Figure 6 shows histogram of average RSU utilization for uniform distribution model and weighted distribution models for θ=10. The histogram shows a significant improvement for weighted model in terms of RSU utilization because of two reasons: first, number of RSUs running in the lowest capacity (<10%) is lower than the uniform model, therefore RSU resources were used more efficiently. Second, number of RSUs running in high capacity (>%80) is also lower, therefore the load is distributed more evenly among the RSUs. This shows that relocating a less utilized RSU to share the high load in a territory provides good results in terms of utilization and can serve as a good optimization technique.

Figure 7(a) and 7(b) shows task failure reasons and breakdowns for uniform distribution model and weighted distribution model for θ=10 respectively. In uniform distribution, no task failure due to network coverage can be observed since it was specifically designed by addressing full network coverage. When the traffic volume is low, vehicle mobility is the reason for the majority of the task failures.

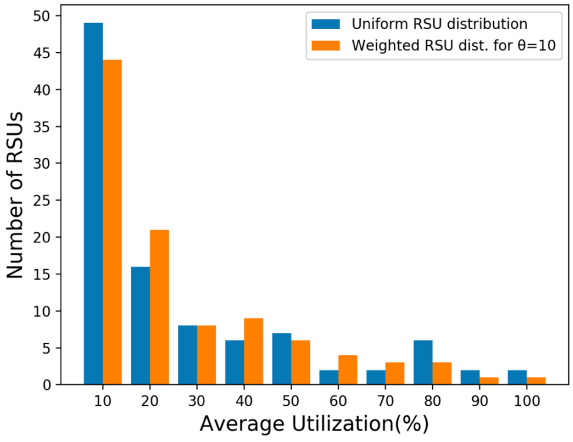


Figure 6. Average Utilization Histogram (3500 vehicles)

However, when traffic density increases, mobility failure rate decreases and RSU capacity failure becomes the main reason of the task failures. Also a high number of task failures can be observed due to exceeding bandwidth capacity when the number of vehicles is 4000 in the simulation.

On the other hand, when vehicle number is 500, network coverage is the main reason of the task failures for weighted distribution model for θ=10 since the vehicles within the range of relocated RSUs cannot connect to any RSU to assign their tasks. When traffic density increases, coverage and mobility failure rates decrease and RSU capacity failure becomes the main reason of the task failures.

VIII. Conclusions And Future Work

In this study, we proposed an RSU placement framework to be used for generating optimal RSU placement models based on traffic characteristics of a target area. Our solution addresses satisfying two criteria for RSU placement problem: network coverage and resource demand. Our work also includes extending capabilities of EdgeCloudSim, a simulation framework designed for edge scenarios, by introducing V2I components and modules. We referred to this extended simulation environment as V2I framework.

In order to validate the functionality of the proposed RSU placement framework, we generated a set of RSU distributions for a target area in London city centre based on uniform and weighted RSU distribution models. Then, we conducted experiments using V2I framework to compare their performances under different traffic loads. The experiments showed that weighted distribution model with replacement factor (θ) 10 performs best under any traffic load. Also we observed that weighted distribution models provided better results in terms of service time and resource utilization.

As future work, we plan further optimisations on weighted distribution model by eliminating input θ from the system and calculate optimal number of RSUs only based on given traffic input data. Also, in this study, we had our main focus on the communication between vehicle and RSU, however inter-RSU communication is an accepted form of communication in Vehicular ad-hoc network (VANET) in which RSUs can exchange data with each other [21]. By implementing this in V2ISim, task transfers between RSUs will be possible and task failures due to vehicle mobility will be prevented. Moreover, some technical factors that can impact the communication between vehicles and RSUs should be studied and findings should be reflected to the study. These can be determining the noise level for the RSUs in close proximity and shadowing effect of the buildings. Finally, in the next phases of the study, we might still have the requirement of working with simulation based traffic datasets as finding real traffic datasets is not always possible. In that case, in order to validate the results and prove the consistency, we will have an approach to generate multiple datasets using different traffic simulation environments.

 ISBN: 978-1-61208-740-5

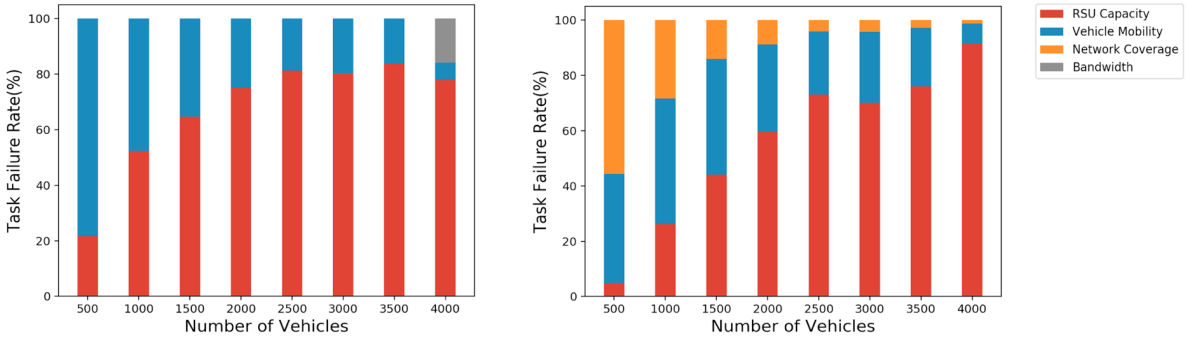


Figure 7. Task Failure Breakdown (a) Uniform Distribution Model (b) Weighted Distribution Model (θ=10)

ACKNOWLEDGEMENT

This work is supported by the Galatasaray University Research Foundation under the Grant No. 18.401.003.

REFERENCES

[1] M. Satyanarayanan, P. Bahl, R Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," IEEE Pervasive Computing, vol. 8, no.4, pp. 14-23, Oct-Dec 2009.

[2] P. Corcoran and S. K. Datta, "Mobile-Edge Computing and the Internet of Things for Consumers: Extending cloud computing and services to the edge of the network," IEEE Consumer Electronics Magazine, vol. 5, no.4, pp. 73-74, Oct 2016.

[3] R. Yu, Y. Zhang, S. Gjessing, W. Xia, and K. Yang, "Toward cloud-based vehicular networks with efficient resource management," IEEE Network, vol. 27, no.5, pp. 48-55, Oct 2013.

[4] C. Sonmez, A. Ozgovde, and C. Ersoy, "EdgeCloudSim: An Environment for Performance Evaluation of Edge Computing Systems," International Conference on Fog and Mobile Edge Computing (FMEC), pp. 39-44, May 2017.

[5] S. K. Datta, R. D. F. Da Costa, J. Härri, and C. Bonnet, "Integrating connected vehicles in Internet of Things ecosystems: Challenges and solutions," IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1-6, Jul 2016.

[6] M. A. Salahuddin, A. Al-Fuqaha, M. Guizani, and S. Cherkaoui, "RSU cloud and its resource management in support of enhanced vehicular applications," IEEE Globecom Workshops (GC Wkshps), pp. 127-132, Dec 2014.

[7] T. J. Wu, W. Liao, and C. Chang, "A Cost-Effective Strategy for Road-Side Unit Placement in Vehicular Networks," IEEE Transactions on Communications, vol. 60, no.8, pp. 2295 – 2303, Jul 2012.

[8] X. Liya, H. Chuanhe, L. Peng, and Z. Junyu, "A randomized algorithm for roadside units placement in vehicular ad hoc network," IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks, pp. 193–197, Dec 2013.

[9] A. O'Driscoll and D. Pesch, "Hybrid geo-routing in urban vehicular networks," IEEE Vehicular Networking Conference, pp. 63–70, Dec 2013.

[10] Y. Liang, H. Liu, and D. Rajan, "Optimal placement and configuration of roadside units in vehicular networks," IEEE 75th Vehicular Technology Conference, pp. 1-6, May 2012.

[11] J. Chi, Y. Jo, H. Park, T. Hwang, and S. Park, "An Effective RSU Allocation Strategy for Maximizing Vehicular Network Connectivity," International Journal of Control and Automation, vol. 6, no. 4, pp. 259-270, Aug 2013.

[12] K. Gomi, Y. Okabe, and H. Shigeno, "RSU Placement Method Considering Road Elements for Information Dissemination," The Sixth International Conference on Advances in Vehicular Systems, Technologies and Applications, pp. 68-73, Jul 2017.

[13] O. Trullols, M. Fiore, C. Casetti, C. F. Chiasserini, and J. M. Barcelo Ordinas, "Planning roadside infrastructure for information dissemination in intelligent transportation systems," Computer Communications, vol. 33, no. 4, pp. 432–442, Dec 2010.

[14] N. M. Balouchzahi, M. Fathy, and A. Akbari, "Optimal road side units placement model based on binary integer programming for efficient traffic information advertisement and discovery in vehicular environment," IET Intelligent Transport Systems, vol. 9, no.9, pp. 851-861, Nov 2015.

[15] G. Premsankar, B. Ghaddar, M. Di Francesco, and R. Verago, "Efficient Placement of Edge Computing Devices for Vehicular Applications in Smart Cities," IEEE/IFIP Network Operations and Management Symposium, pp. 1-9, Apr 2018.

[16] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software Practice and Experience, vol. 41, no.1, pp. 23-50, Jan 2011.

[17] S. D. Levy, "KD-Tree Implementation in Java and C#," https://simondlevy.academic.wlu.edu/software/kd, retrieved Aug 11, 2019

[18] https://www.openstreetmap.org, retrieved Aug 11, 2019

[19] P. A. Lopez et al., "Microscopic Traffic Simulation using SUMO," 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2575-2582, Nov 2018.

[20] A. K. Ligo, J. M. Peha, P. Ferreira, and J. Barros, "Comparison between Benefits and Costs of Offload of Mobile Internet Traffic Via Vehicular Networks," 43rd Research Conference on Communications, Information and Internet Policy, pp. 1-39, Nov 2015.

[21] R. Barskar and M. Chawla, "Vehicular Ad hoc Networks and its Applications in Diversified Fields," International Journal of Computer Applications, vol. 123, no.10, pp. 7-11, Aug 2015.

 ISBN: 978-1-61208-740-5

Dynamic Path Discovery for In-band Control Plane Communication in a Tactical SDN Network

Anders Fongen
Norwegian Cyber Defence Academy (FHS/CIS)
Lillehammer, Norway
email:anders@fongen.no

***Abstract*—Software Defined Networking (SDN) offers promising improvements in operational control of tactical networks, in terms of traffic prioritization, topology management, infrastructure protection, resource monitoring, configuration and deployment. Military networks are characterized by relatively slow radio links which are vulnerable to detection and intrusion, so novel technologies like SDN are highly relevant and actively researched for these purposes. However, the SDN architectural blueprint needs several modifications to meet the typical requirements of a tactical mobile network. This paper addresses the need for reliable in-band control plane traffic across the southbound interface, and suggests two different algorithms to obtain adaptive forwarding decisions for southbound protocol traffic across the data plane links. Among the challenges related to the implementation of adaptive forwarding mechanisms in SDN equipment is the limited expressiveness in the OpenFlow language. Based on expiry mechanisms in flow rules, the SDN switches were able to choose alternative forwarding ports in case of link or switch failure in the grid. The conclusion of the study is that it is possible to make adaptive mechanisms with recovery times comparable to the Spanning Tree Protocol (STP), but with better utilization of link resources since link loops are allowed to exist for load distribution (contrary to the STP protocol).**

***Keywords*—software defined networks; tactical networks; adaptive forwarding; resilience**

I. INTRODUCTION

In a tactical military network, the links are the resource of greatest scarcity and need to be utilized as efficiently as possible. Also, the links are exposed to a range of threats to security, integrity and availability. The SDN blueprint, developed with centralized data centers in mind, assumes a separate set of links for the control plane, which is not an affordable luxury in a tactical network [1]. The use of so-called *in-band control plane communication* has been pursued in this paper for this particular reason.

The term *In-band control plane* describes an SDN configuration were the links between the Network Elements or switches (NEs) carry both user data and southbound protocol traffic between the NEs and the SDN controller (SDNC). In-band control plane have been widely discussed and is implemented in a basic manner in OpenVswitch [2]. However, the OpenVswitch implementation does not offer adaptive paths for the southbound traffic in case of link or NE failure.

The need for reliable control plane connectivity must be combined with the desire to control the links with regard to traffic and security policies. The Spanning Tree protocol (STP, IEEE 802.1D) offers a form for adaptive paths since it may reconstruct the forwarding path in case of link or node failure, but the STP protocol does not employ redundant links for load-balancing purposes and does not offer the flexible traffic policing for which the SDN is popular.

Redundancy management, both for the data plane and control plane traffic, should employ all available link resources both for resilience and load balancing purposes. The presence of link loops requires that received frames cannot be broadcast in order to avoid traffic loops. This restriction affects both discovery protocols like Address Resolution Protocol (ARP), etc., and the bridge link-address learning process.

The construction of the distributed logic necessary for the implementation must take into account the primitive nature of the NE; They are not programmable in the ordinary sense, only through rule-oriented protocols like OpenFlow. The lack of a transaction context in OpenFlow processing generates race conditions and inconsistent intermediate states [3], and failure detection must be handled through the flow expiry mechanism [4].

A. Related Research

The application of SDN in tactical military networks has been discussed from perspectives of robust flow separation [1], security [5] and in-band control plane [1]. Still there are few proposals on how to implement a robust and resilient control plane tunneled through the data plane links. Schiff et al. [3] and later Canini et al. [6] provides good analysis of the problem space and a solution outline, but no detailed and tested proof of concept. Both papers propose the same "hard-timeout"-based mechanism for failure detection, but do not provide a convincing fail-over mechanism.

Sharma et al. [7] offers a comprehensive solution to the connectivity problem as well as congestion in the control plane. They base their solution on modification of the OVS switch code, which is avoided by the work presented in this paper, since the solution should be able to run on commercial and unmodified OF-switches. A fully implemented and tested algorithm for dynamic path discovery in an in-band control plane has, to the author's best knowledge, not been presented before.

B. Contribution of the paper

The research question being pursued in the presented paper can be expressed as *is it possible to implement adaptive routing in the control plane using OpenFlow protocol mechanisms?* The challenges related to this question is that link failure in the control plane will isolate NEs from the SDN controller and NEs will have to detect and recover from link failure autonomously.

 ISBN: 978-1-61208-740-5

The contribution of this paper is the investigation of discovery and fail-over mechanisms in the control plane. It will report on algorithms, design patterns and the experimental evaluation. This paper does not address the obvious security problems related to SDN in a tactical environment, since those problems have been addressed in other recent publications [5].

The remainder of the paper is organized as follows: Section II discusses the benefits of in-band control plane and two alternative design methods. Section III presents the chosen technology components and Section IV shows the configuration of the experimental network used in the experiment. Section V discusses how the underlying SDN protocols can support the operations necessary for the purpose of the experiment. Sections VI and VII provide details of the two proposed algorithm for redundancy management and contains the main contributions of the paper. Furthermore, Section VIII describes the mechanisms necessary for operation of a network without using broadcast frames. Section X summarizes the paper with a few conclusive remarks.

II. IN-BAND CONTROL PLANE

The initial SDN architecture presumes the existence of a separate link infrastructure where every NE is directly connected to the SDNC. Through this infrastructure (the control plane) the SDNC will manage the topology control and traffic policing in the data plane. The control plane is silently expected to be reliable, and any fail-over mechanisms are kept outside the SDN scope.

> *"An SDN controller may use an SDN data plane for some or all of its internal or external interfaces, as long as the SDN controller does not rely for its connectivity on the operability of the data plane that it controls; otherwise, the SDN controller may find itself stranded or irrecoverably fragmented."* - Sect 6.4 of [8]

In a military network, the resource of greatest scarcity is the set of communication links, and a separate control plane infrastructure is an unaffordable luxury. Intuitively, there are resource benefits by merging the two planes into a shared set of communication links. Multiplexing and switching units external to the NE can allow the two types of traffic to be logically separated yet sharing a link. This solution is expensive and cumbersome in terms of hardware resources though, and offers no fail-over mechanisms.

In a more straightforward manner, the OpenVswitch [2] offers a mode whereby it at startup time allows Dynamic Host Configuation Protocol (DHCP), ARP and TCP traffic to and from the SDNC to pass through its switching fabric. The trick is to assign the control plane IP address of the NE to the switch pseudo interface, not to a physical port. OpenVswitch will then use its switching fabric as a MAC learning switch to find a path to the SDNC, using the OpenFlow `NORMAL` output port.

Despite a low cost solution to the in-band control plane problem, it still offers no fail-over mechanism. Besides, the presence of redundant links in the data plane will result in traffic loops.

OpenVswitch offers redundant link management through the STP. However, STP will simply disable links that result in loops and prune the link structure into a tree. Nor will STP support the tight traffic policing and priority mechanisms that are the hallmark of the SDN architecture.

Therefore, there is a need for a mechanism to allow an in-band control plane to be dynamically overlaid on the data plane link structure, without reserving valuable links for fail-over purposes only. The load-sharing capabilities of the redundancy must be utilized and the traffic policing enforced by the SDNC must not be hindered. This paper investigates two approaches to meet these requirements:

1) A reactive method, where the individual NEs discover the path to the SDNC and connects to it.
2) A proactive method, where the SDNC keeps a catalogue of the switches and the links, and actively constructs a tree structure and connects the switches accordingly.

Both methods allow the NEs to be indirectly connected to the SDNC, and both allow for fail-over operations to take place if a link or an NE falls out of service. However, they differ in their programming structure and how well they fit into the software environment.

III. TECHNOLOGY PLATFORM

In this section, the choice of technology components will be described. The components are all software, including operating system, hypervisor and system-level components.

The study of a medium sized networks with more than 10 nodes is best conducted in a virtualized environment. The hypervisor of choice is Oracle's VirtualBox, which is free, easily configured, and offers the right degree of scalability. The limit of four ports per VM is the most limiting factor during the experiment.

For the NEs, complete instances of Linux were chosen. The reason for this choice is that the experimental network is used for testing several services and protocols auxiliary to the OpenFlow protocol, and a general computing platform offers the necessary flexibility and software availability, contrary to Mininet. The Linux instances do not need a GUI and was installed with a text console interface only for the sake of saving memory.

The chosen OpenFlow switch (the NE) implementation is OpenVswitch [2], which is easy to install, relatively easy to configure, and offers the necessary inspection and logging mechanisms for testing and debugging purposes.

As the network controller (SDNC), the Ryu framework was used [9]. Ryu is very popular as an experimental platform with a relatively low abstraction level: OpenFlow statements are generally not automatically generated, but individually constructed through Python programming code. For the experimentation at hand, Ryu performs well and with good stability, although the API and the required design patterns takes some time to learn.

For all the chosen technology components, an important property is the community support offered. Most problems are easily solved through these support resources.

 ISBN: 978-1-61208-740-5

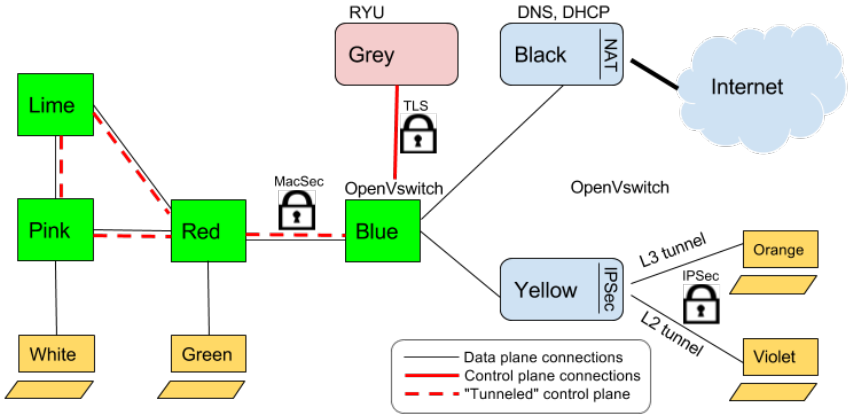


Fig. 1: Current SDN laboratory configuration

IV. Experimental network

The network used in the experiment is shown in Figure 1. The network consists of a number of green switching nodes (NEs), a number of yellow general clients and a number of blue nodes for serving IPSec, OpenVPN, DHCP, DNS, VXLAN etc.

The links between the NEs are somewhat redundant and consequently form loops. The redundant links are essential for the study of fail-over mechanisms and load balancing services. The experimental network was used also for investigating security mechanisms in an SDN based environment [5], thus the presence of IPSec, MacSec and TLS is indicated in the figure.

V. Software design patterns

Adaptive forwarding would intuitively need general programmable logic in the switches, in order to test environmental conditions and make decisions accordingly. This is particularly likely when the data plane and the control plane use the same links, and an NE would need to find an alternative forwarding path in a situation where it is isolated from the SDNC.

Another matter is that the NE does not have any direct mechanism to reveal the physical port it uses for the control plane traffic. The controller needs that information in order to install flows that avoid traffic loops (the use of the OpenFlow `NORMAL` output port must be avoided for the same reason).

The OpenFlow mechanisms that were used for building the necessary distributed logic were:

1) The output port `CONTROLLER` which allows traffic that matches a flow to be handled over to the SDNC, with information about the ingress port of the switch. This mechanism is part of the Port Discovery procedure which will be discussed in Section VI-A.
2) The use of flows associated with expiry mechanisms and high priority in combination with low priority flows without expiration. In case the high priority flows fail to be renewed, they will disappear and the low priority flow will be set in effect. The necessary fail-over mechanisms are built on this design pattern, which is also proposed by Canini et al. in [6].
3) Timestamps on certain connection requests. Where there are several paths from an NE to the SDNC, the first node common to these paths will experience connection requests from the same node over a short period of time. The timestamps will serve to recognize the first request and discard the others.
4) Proxy operation of broadcasts. To avoid traffic loops, NEs cannot broadcast received data, only data which is locally originated. Broadcast packets are passed on to the SDNC which may locally resolve the situation or pass them on to each NEs with the instruction to flood them.
5) Data plane forwarding based on MPLS labels. Flooding as seen in MAC-learning bridges cannot be used, but the SDNC keeps track of all client MAC addresses and their associated NE. The SDNC will install flows (on-demand) to attach MPLS labels to frames signifying the egress NE in the path to the destination. Forwarding in the data plane is based on MPLS labels only.

In the following sections the two different algorithms will be described in more details. The two alternatives are designated as reactive and proactive, respectively.

VI. Reactive adaptive forwarding algorithm

This algorithm does not presume any knowledge about the collection of links and switches. The knowledge is built through a discovery process. The switches can be in one of three states: *Disconnected, Connected* and *Operating*:

Disconnected: The switch is not under control of the SDNC and contains only flows installed during the bootstrap process. These flows floods TCP and ARP packets originated from the switch (in-port: `LOCAL`) destined to the SDNC, and sends ARP and TCP traffic from the SDNC to the `LOCAL` port for processing by the MAC-learning switch fabric.

Connected: The switch is under control of the SDNC through the OpenFlow protocol, but is unable to bridge connections from lower-tier switches. In this mode, it still floods locally generated packets to the SDNC. The SDNC will initiate

 ISBN: 978-1-61208-740-5

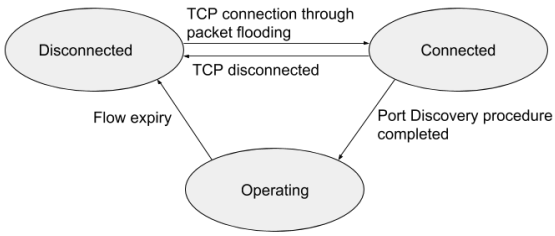


Fig. 2: State diagram of a switch/NE

the *Port Discovery procedure* when the switch enters the connected mode.

Operating: The Port Discovery procedure is completed, and the switch now communicates unicast (called "unicast flows") with the SDNC through its *command-port* (c-port). The switch is able to receive connection requests (TCP SYN) from lower-tier switches and pass them on towards the SDNC, which happens later as explained in Section VI-B.

The state transitions are shown in Figure 2. Note that the transition from Operating to Disconnected happens as the installed flows expires due to lost communication from the SDNC. The SDNC has to refresh the "unicast flows" on a regular basis to avoid them from expiring. If the communication is lost, the flows that floods packets to the SDNC will again come into effect and let the NE look for other paths to the SDNC. *This is the fail-over mechanism available in OpenFlow when the NE becomes disconnected from the SDNC.*

A. The Port Discovery procedure

The port on a NE currently used for the southbound interface is called the *command port* (c-port). The c-port value is important to the SDNC, since it need to set up flows in the NE for unicast communication with the SDNC. Unicast communication is also necessary for forwarding of TCP connections from lower-tier NEs without creating traffic loops.

The procedure to identify the c-port and for subsequent flow installation is called the *Port Discovery procedure* and is shown in Figure 3. It consists of the following steps:

1) The SDNC sends a UDP packet to the NEs IP address. It will be trapped by a flow (installed during the Connected state) and sent back to the SDNC as an OpenFlow PacketIn message, revealing the ingress port of the UDP packet. This port will be used also for the southbound traffic, called the c-port.
2) Flows will be installed for the NE-SDNC communication to take place over the (c-port,`LOCAL`) pair of ports.
3) Flows will be installed on-demand to handle connection from NEs in the lower tiers of the link tree. The individual flows will be installed as the lower-tier NEs floods their TCP SYN packets, since their MAC address and the connected NE port is not known until then. The detail of lower-tier connections will be presented in Section VI-B.

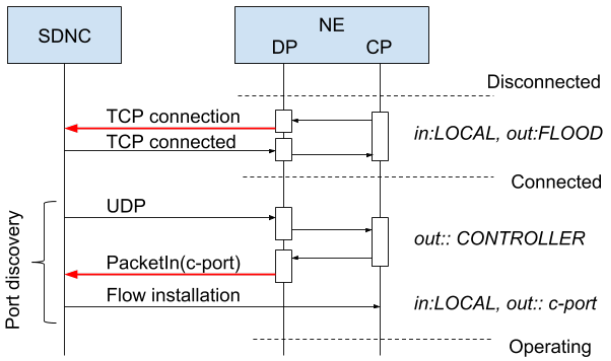


Fig. 3: Protocol elements of the Port Discovery procedure. Red arrows indicate flooded traffic. DP=data plane, CP=control plane

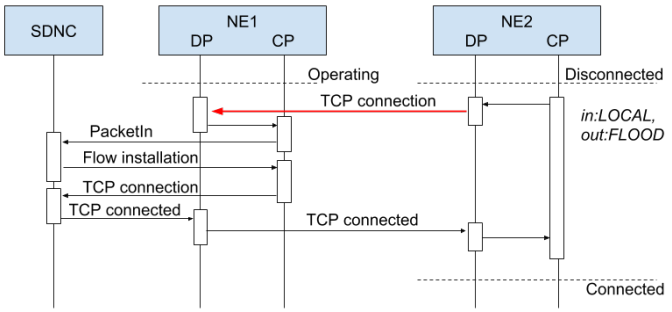


Fig. 4: Protocol elements for lower-tier NEs connecting to the SDN via a higher-tier NE

B. Connection of lower-tier NEs

NEs do not know if they are directly or indirectly connected to the SDNC, and they do not know which port that leads in that direction. The state diagram and the Port Discovery procedure just described are valid for all NEs regardless their position in the link tree.

The following paragraphs describe the process whereby an NE establishes an indirect connection to the SDNC. The protocol elements of this process is shown in Figure 4.

ARP and TCP SYN packets flooded from an NE in the Disconnected state will be received by one or more of its link neighbors. Each of them will, provided that they are in Operating state, pass the frame towards the SDNC, which will respond by an installation of the flows for passing southbound traffic between the ingress port and the c-port, i.e., connecting the lower-tier NE to the control plane link tree so that it will enter the Connected state.

Three details should be mentioned: (1) This process will be repeated for every tier of NEs towards the SDNC, since they all need these flows to be installed to serve the new connecting NE, (2) the flows are given an `idle_timeout` expiration time, to sanitize stale flows if forwarding paths change, (3) several of the NEs neighbors may try to create a forwarding path to SDNC, and a timestamp based mechanism will reject a connection attempt if the same lower-tier NE has connected through the same higher-tier NE recently (the last 10 seconds), applying the heuristic that the first received connection request has chosen the best path. This suppression mechanism will

 ISBN: 978-1-61208-740-5

have to keep rejecting until the lower-tier NE has completed the Port Discovery procedure and started to use c-port (rather than flooding) for communication with the SDNC, i.e., it will have entered the Operating state.

VII. Proactive adaptive forwarding algorithm

Another algorithm for dynamic path discovery and management was investigated, based on a proactive design principle. The NEs, the clients, the communication links and the ports they are connected to are known in advance in the form of a *catalogue*, and the spanning tree calculation can be done by the controller and the resulting structure be imposed on the NE structure proactively.

Having a detailed catalogue of the communication resources and the connected clients may sound cumbersome, but it is frequently seen in military application that detailed information about the system configuration is a prerequisite for the planning of an operation. For this reason, the presence of a catalogue is not regarded as an unreasonable requirement. The following sections will present the essential characteristics of the proactive algorithm and implementation details.

A. SDNC-initiated connections

The Ryu framework does not offer any blocking operations for connecting to NEs. The TCP connection always originates from the NE and results in an event in the controller code. An unconnected NE will continuously make connection attempts, so once the SDNC has created a path between it and the NE, a connection is expected to take place within a few seconds. The failure of a connection is therefore indicated in the SDNC code by a timeout event, while a successful connection is indicated by a connection event.

Once connected, the liveness of the connection is monitored by the SDNC through heartbeat messages. Connection loss is signalled in the SDNC by a *dpset* event.

B. No port discovery needed

Port details are recorded in the catalogue so the Port Discovery procedure described in Section VI-A is no longer necessary. The c-port is now used differently than in the reactive algorithm; the c-port is used only for transit traffic to and from lower-tier NEs. The control plane traffic originated in the NE (ARP and TCP port 6633 from the `LOCAL` port) is flooded to all ports, like in the Connected state in Section VI. Similar incoming control plane traffic from any port is forwarded to the `LOCAL` port. This choice was made to simplify the procedure whereby the SDNC will operate an NE through a different port, as a part of a fail-over procedure. The flooded traffic will not be forwarded by the neighbor NEs except for the parent node in the spanning tree, which is given explicit flow instructions for that purpose.

Figure 5 shows how connections are being made by NEs in successively lower tiers in the spanning tree. The tree structure is shown on the right side, and the interaction diagram shows the details of the protocol. The spanning tree is traversed in-order and connections are accepted from successively lower tiers of the tree. Dotted lines are links that are not part of the spanning tree but still employed by the data plane. The grey boxes on the interaction diagram on the left indicates operations that are given a deadline and watched by a timer mechanism. Transactions shown with a red arrow are flooded to all output ports of the NE.

C. Fail-over procedure

The failure of a NE or a link, indicated by the loss of a TCP connection or failure to create one, will cause a fail-over procedure to be executed. The loss of connection to a node can be caused by failure in any link or node along the path from the SDNC to the NE, and a link failure will cause loss in all nodes connected along that link, but the resulting software events may be generated in any order. It is therefore a complicated task to identify the exact failed link, and a heuristic has been chosen to simplify the algorithm: Communication loss to a node is taken as an indication of a failure in the link closest to it, i.e., the link connecting to the nearest parent node. This simplification may generate unnecessary fail-over actions, but the spanning tree will still be operating correctly and the optimized tree structure will be constructed later as a result of other procedures soon to be described.

If a disconnected node has not alternative paths to it, the parent node will keep its existing flow, expecting the node to resume operation later. This rule alleviates the consequences of a mistaken fault detection, as discussed in the previous paragraph.

Figure 6 shows state transitions on a spanning tree during a link failure. Since node E is first reported as disconnected, the fail-over action in State 2 is taken. Later, when node C also is reported as lost, the fail-over action shown in State 3 is taken. All nodes are now connected, although node E is one step deeper than necessary in the tree. On a later instance, the SDNC can use the Peer Discovery information (cf. Section VIII-B) to learn that the link C-E is in operation, and optimize the spanning tree accordingly.

VIII. Broadcast-free operation

The suggested network design is an L2 switched network with topology loops, and must therefore refrain from ordinary broadcast-operation in order to avoid endless traffic loops. The switches cannot broadcast received frames as a part of the switches' MAC-learning process. For this purpose four services have been implemented:

A. Proxy ARP

ARP requests from a client (connected to an NE) are trapped by the NE and passed to the SDNC for resolution. For this purpose, the SDNC maintains an ARP table (MACaddr,IPaddr) which learns from all packets delivered to the SDNC, not only ARP replies. In case of a table miss, the SDNC will instruct every individual NE to flood the request, and the resulting ARP reply is trapped in the first switch and sent to the SDNC, which will update its ARP table. Since NEs do not forward the flooded request, traffic loops will not be formed.

Actually, the ARP table also contains columns for the identifier of the NE connecting to the client, and the port number of the NE used for the connection. This information is

 ISBN: 978-1-61208-740-5

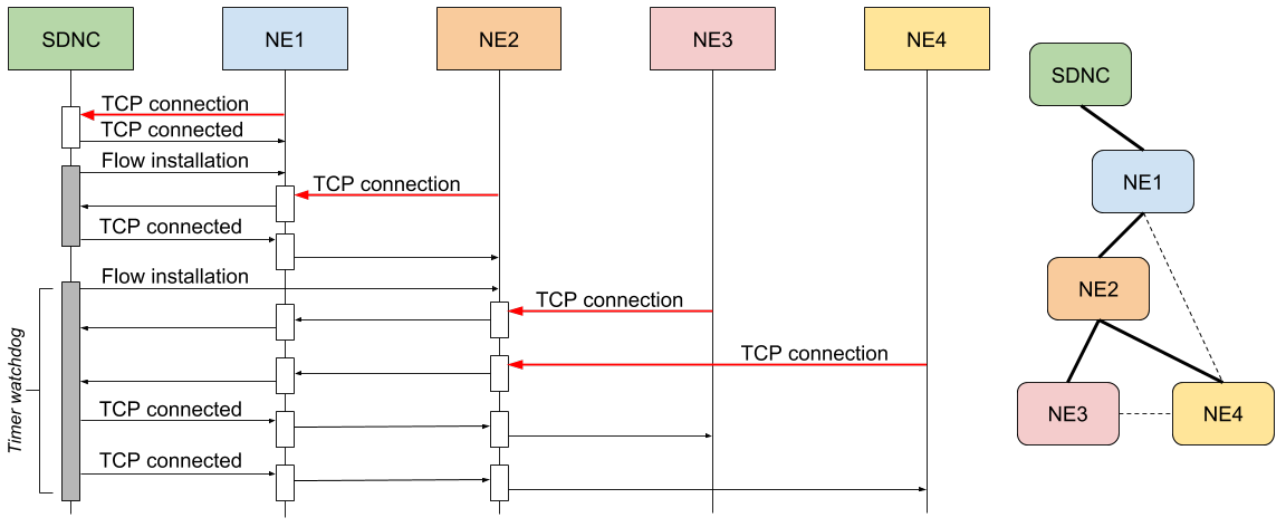


Fig. 5: Connection creation in a spanning tree of NEs using the proactive algorithm. See the text for detailed explanations.

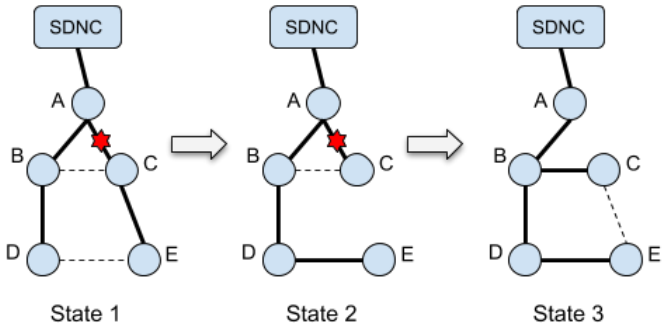


Fig. 6: A possible set of state transitions in the spanning tree as a result of superficial fault detection. See text for full details.

used for MPLS based forwarding, explained in section VIII-C. The bond between L2-mechanisms like MPLS to an IPv4 protocol like ARP is not a good design, and Section VIII-D will describe a cleaner alternative to broadcast operation.

B. Peer Discovery

To identify all links in the data plane, and to detect link failure, regular messages are sent from SDNC to every NE with instructions to flood the data through all ports. At the other end of the link, the received frame is passed to the SDNC which adds a timestamp and updates its link database. This method is quite similar to what is found in OpenFlow Discovery Protocol [10] and Bidirectional Forward Detection (RFC 5880).

In the case of the proactive algorithm (cf. Section VII), explicit messages for this purpose may not be necessary, since every NE will reply to heartbeat polling from the SDNC with a flooded frame, and these frames can be used for peer discovery. This protocol variant has not been tested.

C. MPLS based forwarding

In the NEs, traditional MAC-learning is not used, since that involves broadcast operations. Instead, every NE is associated with an MPLS label value, and every frame sent from a client will be given an MPLS label indicating the egress NE on the path to the destination. The egress NE will strip off the MPLS label and deliver the frame to the destination host.

This process relies on a number of information sources: The ingress switch needs a flow to attach the MPLS label, and the egress switch will need a flow to strip off the MPLS label and deliver the frame through the correct port. Intermediate NEs will need flows to associate MPLS labels with an output port. This forwarding information is derived from the peer discovery protocol link database (using a shortest path calculation) and installed as flows in NEs as needed.

D. Multicast trees

The ARP proxy described in Section VIII-A binds L2-mechanisms to the IPv4 protocol, which is not a desirable design, since the infrastructure should not make any assumptions with regard to the network-layer protocol in use. Therefore, at a later iteration of the design, the peer discovery link database was used to build multicast trees for every NE, so that broadcast frames initiated from one NE will propagate to every NE in a loop-free manner, and the NE will again deliver the frame to every connected client. MPLS labels are used to identify the originating NE of a broadcast frame, so that intermediate NEs may make the correct forward decisions.

This arrangement permits the use of, e.g., IPv6 Neighbor Discovery Protocol (NDP), so that IPv6 and IPv4 traffic can use the same mechanisms for broadcast frames. However, it is not established if the ARP proxy described in Section VIII-A is still more effective in terms of link usage.

IX. Performance and Scalability

The performance and the scalability of the proactive and reactive algorithms will be discussed in this section. The lab design shown in Figure 1 was used to develop the algorithms and test the functional correctness of the reactive algorithm (the proactive design has not been tested). However, this design

 ISBN: 978-1-61208-740-5

was not sufficient for scalability experiments or comprehensive performance measurements.

The chosen baseline for the performance discussion is the Spanning Tree Protocol in the default configuration. A number of measurements indicated that STP recovers the network in **30 seconds after** a link failure. The reactive algorithm, with our chosen parameters, exhibits an average recovery time of **14 seconds**.

A link discovery mechanism based on polling (in the form of renewed flows) will introduce a trade-off between failure detection time and generated traffic volume. A halved detection time will require the doubled number of liveness control messages. Besides, once a link connection has been established, an unconnected switch will make connection attempt with regular intervals and thus introduce a mean delay after a path to the SDNC has been established.

With regards to scalability, both the reactive and proactive algorithms need to build up a path between NE and SDN step by step. As the number N of NEs grow, the average number of links D between an NE and the SDN is expected to be growing like the depth of a tree:

$$D = O(logN) \tag{1}$$

The establishment of a single step in the path from an NE to the SDN will generate a constant number of messages, so the total volume M of messages associated with path discovery from every NE is expected to be

$$M = O(NlogN) \tag{2}$$

The required time T for re-connection is assumed to grow with the number of links in the path and with the same order as D:

$$T = O(logN) \tag{3}$$

During the experiments, it was observed that a connection from an NE before the SDNC had closed the previous SSL connection resulted in a "duplicate connection attempt". A re-connection attempt from an NE should not happen earlier than this timeout value in the SDNC, which is a configurable parameter value in Ryu.

X. Conclusion and future research

The contribution of this paper is a detailed analysis and a proof-of-concept implementation of an in-band control plane with failure detection and dynamic path discovery. These properties allow for a control plane that survives link and node failure, since it will employ alternative paths to connect the NEs to SDNC. An important contribution is that these mechanisms are offered in the presence of link loops. Link loops represent redundant communication resources which should be employed for load balancing and traffic separation, not only for resilience, which is why the Spanning Tree Protocol was abandoned.

Future research and development on this topic will include better testing of the separation between L2 and L3 protocols, so that the ARP proxies are replaced by mechanisms to distribute broadcast frames along multicast trees. We also are in the process to include multi-tenancy separation in the L2 forwarding mechanisms based on the clients' X.509 certificate information. IPv6 support will also be added to the prototype.

References

[1] J. Spencer and T. J. Willink, "SDN in coalition tactical networks," in *2016 IEEE Military Communications Conference, MILCOM 2016, Baltimore, MD, USA, November 1-3, 2016*, pp. 1053–1058, 2016.

[2] "Open vSwitch." http://openvswitch.org. Online, Accessed Aug 2019.

[3] L. Schiff, S. Schmid, and M. Canini, "Ground control to major faults: Towards a fault tolerant and adaptive SDN control network," in *46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, DSN Workshops 2016, Toulouse, France, June 28 - July 1, 2016*, pp. 90–96, 2016.

[4] L. Schiff, S. Schmid, and P. Kuznetsov, "In-band synchronization for distributed sdn control planes," *SIGCOMM Comput. Commun. Rev.*, vol. 46, pp. 37–43, Jan. 2016.

[5] A. Fongen and G. Køien, "Trust management in tactical coalition software defined networks," in *2018 International Conference on Military Communications and Information Systems, ICMCIS 2018*, pp. 1–8, Institute of Electrical and Electronics Engineers Inc., 5 2018.

[6] M. Canini, I. Salem, L. Schiff, E. M. Schiller, and S. Schmid, "A self-organizing distributed and in-band SDN control plane," in *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, GA, USA, June 5-8, 2017*, pp. 2656–2657, 2017.

[7] Y.-L. Su, I.-C. Wang, Y.-T. Hsu, and C. H.-P. Wen, "FASIC: A fast-recovery, adaptively spanning in-band control plane in software-defined network," in *IEEE GLOBECOM 2017*, pp. 1–6, Institute of Electrical and Electronics Engineers Inc., 12 2017.

[8] O. N. Foundation, "SDN architecture, issue 1.0." https://www.opennetworking.org/, 2014. Online, Accessed July 2019.

[9] "Ryu SDN Framework." https://osrg.github.io/ryu/. Online, Accessed Aug 2019.

[10] "OpenFlow Discovery Protocol." http://groups.geni.net/geni/wiki/OpenFlowDiscoveryProtocol. Online, Accessed Aug 2019.

 ISBN: 978-1-61208-740-5

Semi-Automated Footwear Print Retrieval Using Hierarchical Features

Tim vor der Brück

School of Information Technology
Lucerne University of Applied Sciences and Arts
Rotkreuz, Switzerland
Email: tim.vorderbrueck@hslu.ch

Thomas Stadelmann

Forensity AG
Technopark
Root-D4, Switzerland
Email: thomas.stadelmann@forensity.com

***Abstract*—Footwear prints are one of the most commonly found pieces of evidence on crime scenes. They can be used both to connect different crimes and to give important clues to the identity of the culprit. In many cases, these footwear prints are distorted or incomplete, which makes fully automated approaches for their identification and comparison unreliable. Hence, we propose a semi-automated approach, where a representation of key features is obtained manually by forensic experts, the comparison with outsole models from a database is done by a computer. To account for potentially poor quality footwear print pictures, we introduce a hierarchical fuzzy search that ranks the outsole models according to their degree of correspondence with the features of the footwear print. Furthermore, we conducted an evaluation that demonstrated the usefulness of the proposed approach.**

Keywords–Footwear print; retrieval; tree edit distance.

I. INTRODUCTION

Footwear prints can be secured on almost every crime scene [1] and are daily used by law enforcement authorities for the following purposes [2]:

1) Crime scenes where footwear prints belonging to the same shoe have been secured can be linked and deliver valuable information for forensic intelligence and crime analysis.
2) The outsole design of shoes from arrestees can be compared with the footwear prints from the crime scenes and suspects can be linked to open crimes.
3) Crime scene footwear impressions can be associated with a specific outsole model and deliver helpful information for investigations or support purpose 1 or 2.

Police investigators collect for all three purposes images of footwear prints. In case of a newly committed offense, the police investigators manually compare the footwear prints found at the new crime scene with footwear print images originating from earlier offenses, often collected in cardboard files or binders. For this, they are looking for certain striking patterns or characteristics that the footwear prints have in common. In particular, the following two steps have to be performed (see Figure 1). Step one is feature extraction, where the sole patterns are recognized and encoded into an abstract representation. The second step is the search process, where an abstract representation of an image is compared with other abstract representations from a data collection. Up until now, humans are still more effective than computer algorithms in interpreting images with footwear prints from crime scenes since they can better distinguish sole patterns from the noisy background. Therefore, established computerized footwear systems work with images that are encoded by humans yet [3]. This means human forensic footwear examiners assign predefined codes to the identified features on the image. Afterward, images can be searched by the assigned codes, which build an abstract representation of the sole pattern on the image. With this approach, every image interpretable by an expert can be processed. Albeit, assigning the right code is not an easy task because there is an infeasible amount of pattern designs that are constantly being changed but the predefined codes remain fixed. Thus, it is quite possible that two experts encode the same pattern differently. However, an unambiguous representation is typically an important prerequisite for finding corresponding patterns. Using traditional exact search methods, only footwear models are determined that completely matches the given input. Therefore, police departments have limited the number of people coding the images to assure a common standard. This is feasible if the footwear database only belongs to a small department or the administration of multiple departments is centralized.

The benefit of a common data exchange of footwear print information between different police departments has been recognized already in the early nineties [2]. Different initiatives have taken up this issue during the last couple of years [2] as offenders get more mobile and can only effectively be opposed by cooperation. Besides political and organizational issues [4], one important limitation of sharing information is the unavailability of an efficient search system to find relevant information with reasonable effort in big data collections containing images and outsole patterns nationwide or even internationally.

In this paper, we present a search engine with an advantageous division of labor between human and machine. The feature extraction is accomplished by human experts. This way, also very noisy and distorted images, which are currently only interpretable by humans, can be encoded. To ensure a homogenous classification by different users, the number of features is restricted to a rather small standardized set. The retrieval is accomplished by a fault-tolerant search engine, which ranks all outsole models stored in our database according to their degree of correspondence with the input shoe track. See Figure 2 for the architecture of our proposed approach called *Fast*.

 ISBN: 978-1-61208-740-5

The remainder of this paper is organized as follows. In Section II we give an overview of the current state of the art regarding automated footwear print retrieval. In Section III we describe our proposed fuzzy search, whereas its evaluation is contained in Section IV. Finally, we give a conclusion and an outlook to possible further work in Section V.

II. Related Work

From the late seventies on [5], police departments in different countries introduced computer-assisted footwear systems. What they all have in common is that the footwear print images are encoded by humans. This means the users assign predefined codes to the detected geometric forms on the footwear print image. For most of the systems, self-developed coding schemes are employed [3]. The amount of predefined codes varies between the different systems and most systems work with more than 40 different codes subdivided into different groups. To achieve a better selectivity, several additional coding elements can be found among the different systems [6]:

- Different zones on the outsole that can be encoded separately
- Additional properties for some codes to specify the geometric form they represent (e.g., horizontal/vertical)

The existing coding systems work with filters based on simple string matching. As it is possible that different experts encode the same pattern differently, relevant footwear images can easily be missed by the search [7]. To reduce mistakes during the retrieval process some systems work only with a few basic codes [8]. Gross et al. [9] could easily distinguish 99% of the impressing using only nine design elements types in combination with size-relationship.

In the current research, there are several fully automated approaches for footwear print comparison that make use of image processing. Usually, they aim to detect certain geometrical patterns on the images (for instance by conducting a Hough transform) [4] [10] . These patterns are then leveraged to obtain an abstract footwear representation, which allows for accurate similarity estimation. Quite lately, deep learning methods that require no explicit feature engineering become popular in computer vision and also for footwear comparison. Neal Khosla and Vignesh Venkataraman, for instance, [11] construct neural networks for footwear images by means of the VGGNet classifier (VGG stands for Visual Geometry Group) and estimates their similarity by applying the vector norm on activation value differences of neurons belonging to certain layers of the network.

So far, it is still uncertain if there is a fully automated method that provides good recognition rates on real crime scene data. Many published methods has been evaluated on synthetic data or the dataset of used crime scene marks has not been published. Therefore, it is often not possible to repeat experiments or compare methods using benchmarks. [7] assumes that published results are not reliable because of missing gold standard datasets and evaluation standards.

The main drawbacks of fully-automated methods are the runtime, which can amount up to several seconds for a single comparison, and the difficulty for a forensic footwear expert

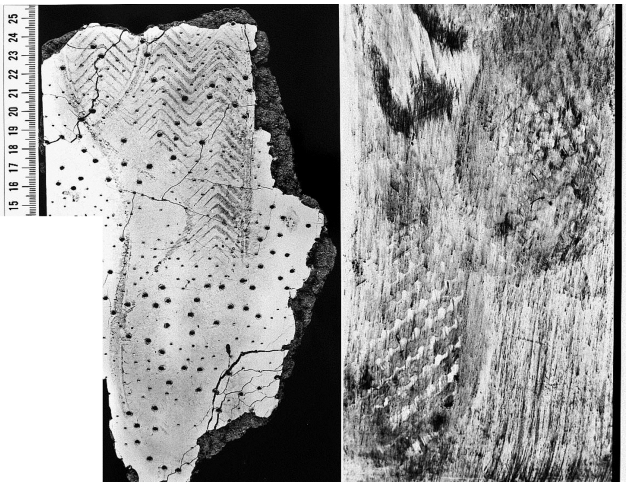


Figure 1. Footwear print from a crime scene.

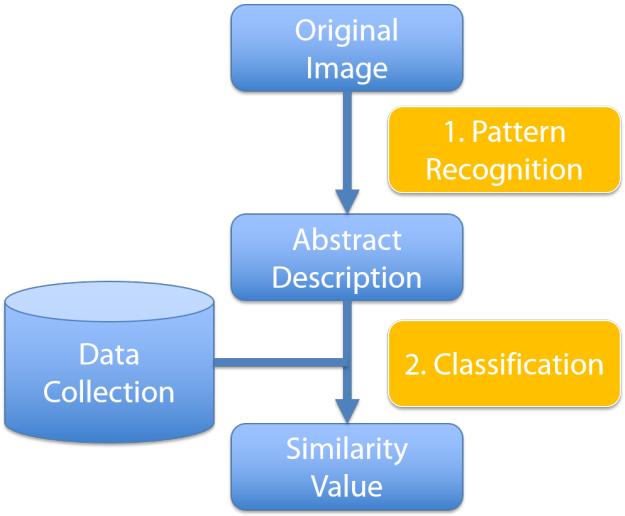


Figure 2. Overview of our proposed methodology for footwear print retrieval.

to take influence in the obtained results. Deep learning-based approaches in contrast are rather fast but behave like black box models and lack interpretability. Furthermore, a neural network has a vast amount of free parameters, which have to be optimized preferably automatically. Such a parameter optimization consumes a large amount of runtime and also needs a lot of training data. This is different with semi-automated approaches like our proposed method, which are however rather rare in academic research. One example in the literature is the method of Gird [2], in which all footwear features values are manually specified and only the search is conducted automatically. Albeit, the proposed method is not error-tolerant and can only establish perfect matches, which considerably hinders its practical usage.

III. Our Approach

In the following, we will first give a rough overview on how our approach works and go into detail afterwards.

 ISBN: 978-1-61208-740-5

A. Overview

For a footwear print found at a crime scene, the following two tasks are of interest:

- Identifying the correct outsole model that is associated with this footwear print
- Comparing two footwear prints from different crime scenes

We will first describe scenario one and cover afterward how to deal with the case that footwear prints should be compared directly with each other.

Basically, a footwear print and its associated outsole model can be represented by a noisy channel model. The informational content of the outsole model is transmitted through a channel that is influenced by environmental conditions and results in a usually distorted and incomplete footwear print. The task of the retrieval system is to obtain the most likely outsole model for the observed print. In practice, this can be a very tedious and difficult task for the following reasons. First, the footwear print could be several hours old already and was potentially affected by weather conditions like rain or wind. Second, only a part of the shoe sole might have had enough contact with the soil to produce a visible mark, which can be particularly the case for hard and dry surfaces. Finally, the sizes of the outsole model and print might be different, which can cause the attribute values not being identically but only proportionally to each other.

Therefore, our approach can establish approximate (fuzzy) matches between shoe tracks and associated outsole models and is error-tolerant in the following ways:

- Different attribute values: Attribute values like *line with* or *circle radius* of a shoe track and its associated outsole model can exhibit minor deviations.
- Different Granularities: The outsole model and the shoe track could be described by different levels of abstraction. Consider for example the case that the outsole model clearly contains an ellipse while the associated shoe track picture is strongly blurred and noisy. The forensic footwear examiner might be unsure whether the picture conveys an ellipse or rather a circle and selects the feature *round shape* instead, which is a hypernym (supertype) of both ellipse and circle.
- Missing features: Not all features contained in the outsole model might be visible at an associated footwear print since the latter potentially conveys only a part of the entire shoe sole. However, if a certain feature shows up in the footwear print, it should also be represented in the outsole model for establishing a match.

B. Tree representation

We represent both the footwear print as well as the outsole models by a tree that describes the observed outsole patterns in a hierarchical way and contains the following type of nodes:

- Feature type: type of a visual shoe sole pattern like lines, round shape, etc. All feature types are ordered hierarchically in a taxonomy tree.
- Feature: an instance of a feature type
- Attribute: property of a feature. Each attribute has a name and an associated value. For example, the attributes for the feature *circle* are *radius* and *line width*. We discern between the following attribute types:
  - Nominal: values of a nominally scaled attribute have no natural ordering and can therefore only be compared for equality.
  - Ordinal scaled: attribute values of an ordinal scaled feature can be enumerated in a natural order. If we compare two different values, one can always decide, which value is smaller and which is larger. However, there might not be a natural origin.
  - Ratio scaled: attribute values of a ratio scale both have a natural ordering and an origin. In addition, the ratio of two attribute values can be interpreted in a natural way, e.g., if this ratio assumes the value of two, then we can conclude that the attribute value belonging to the dividend is twice as high/good/large than the one belonging to the divisor.

An example of such a tree is depicted in Figure 3. A list of all features and associated attributes are given in Table III, the full feature taxonomy is specified in Figure 7. Our employed feature, feature types, and attributes were devised in cooperation with our industry partner, a company specialized in forensics, by a profound study of available outsole models and of existing literature (see for instance [10] for an extensive analysis of outsole patterns) and competing systems.

Note that we provide a browser-based tool to specify the features perceived from the footwear images and which also constructs the associated feature tree automatically. The comparison trees of outsole models are created on the fly for every user query from associated relational database entries. To reduce processing time, we use some kind of prefiltering that rules out entries having low feature agreements with the query tree. In this way, the number of database entries, for which feature trees actually need to be constructed can be considerably reduced and the entire tree creation stage usually takes no more than one second.

C. Obtaining a similarity estimate

To obtain a numerical similarity estimate between a footwear print and an outsole model, we compare their associated trees with each other. There are several tree comparison methods mentioned in the literature, where a selection of them is described below.

Tree Kernel: A tree kernel is a positive-semidefinite similarity measure often employed as SVM (short for support vector machine) kernel (see [12]). Most popular is the common subtree tree kernel, which is based on the assumption that two trees are similar if they contain a lot of common subtrees.

 ISBN: 978-1-61208-740-5

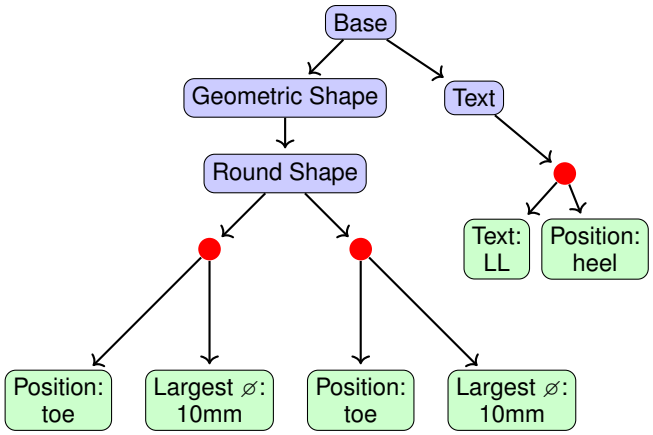


Figure 3. Example tree representing a footwear print / outsole model. The red circles represent unnamed feature nodes.

pg-gram-Distance: The pg-gram-Distance is determined akin to the Jaccard Index from set union/intersection of fixed size subtrees [13].

Tree Edit distance: The tree edit distance is a generalization of the Levenshtein string distance to trees. It counts the minimal number of required removal, insert and edit operations to transform the first argument tree into the second [14] .

We opted to use the tree edit distance as our similarity measure for several reasons. First, it can be computed completely unsupervised and requires no training. Secondly, there are quite efficient computation methods for determining the tree edit distance of ordered trees that employ dynamic programming and have quadratic time complexity in respect to the number of tree nodes [14]. Finally, most implementations allow the insert, edit and delete operations to be weighted differently, which is an advantage over the tree kernel and a necessity for our scenario. This is because a certain feature found at a footwear print should also show up in the outsole model, while an outsole model feature can be missing at the footwear print. Hence, we want a deletion operation on nodes belonging to footwear prints to be more expensive than an insertion. However, if we compare two footwear prints wich each other directly, then the deletion and insertion weights should actually be identical.

The computational complexity of the tree edit distance is polynomial for ordered and NP-hard for unordered trees. We can impose an ordering on the attribute and feature type nodes by comparing their names alphabetically. However, there is no meaningful way to compare two feature nodes other than for equality, so they are actually unordered. Thus, if we want to compare two trees, where the first argument tree contains a feature type node that is the parent of several feature nodes, we compute the ordered tree edit distance for all possible permutations of such feature nodes and take its maximum value as overall tree edit distance. Formally, this distance is given by:

$$sim_u(t_1, t_2) = \max\{sim_o(u, t_2) | u \in perm(t_1)\} \quad (1)$$

where $perm(t_1)$ is the set of all permuted trees. Consider for example a tree, in which two feature type nodes have more than one child, namely the first node two and the second one three, then the set $perm(t_1)$ consists of $2 \times 3 = 6$ elements in total.

Note that we use fixed weights for insertion and deletion operations. These weights must be specified in advance in a configuration file and can be adjusted by the forensic footwear examiner for each node type individually. For node modifications, the total weight is given by the product of a node type-specific weight, which is constant and can be adjusted in a configuration file as well, and the value of a similarity function applied on the compared nodes. For non-attributable and nominal-scaled attribute nodes, this function always assumes the value of zero for identical nodes and one otherwise. However, if the attribute values are ordinally or cardinally scaled, we would like small deviation of attribute values to be less penalized than large differences. Hence, we define the node similarity function as follows: $sim(n_1, n_2) = e^{-\frac{(val(n_1)-val(n_2))^2}{\delta}}$ where the parameter δ determines the slope or decay rate of the function and $val : Attr \rightarrow \mathbb{R}$ denotes a function that is defined on the set of all attributes $Attr$ and retrieves the attribute's current numerical value.

IV. Evaluation

In this section, we will describe the evaluation we conducted to demonstrate the usefulness of our proposed approach.

A. Goals of the evaluation

The goal of the evaluation is to get answers to the following questions:

1) How accurate is the fuzzy search?
2) Are the results reproducible with different users?
3) Does the accuracy change if user search a reference or a crime scene print

B. Description of the used test images

Kortylewski [4] published labeled images with real prints from crime scenes and corresponding references. These images were provided from different German police departments for a standardized evaluation of search algorithms. This data set allows anyone to reproduce results of published algorithms and check them for reproducibility. The images show different types of quality (c.f. image 1) and ensure that tested performances have an informative value concerning the later daily use in the field. In addition, thanks to publically available data, results from different publications can be compared. For this evaluation all 1 500 references and all the 300 prints from crime scenes provided by [4] were integrated into the testing system. To get a reasonable workload when searching for prints from crime scenes additional 350 images where introduced to the testing system. To make the images searchable, all of them were codified by the same person.

Our method was tested by a forensic scientist, experienced with the algorithm (P1), an ordinary person, already experienced on how to use our fuzzy search (P2), and another ordinary person not having any such experience (P3).

 ISBN: 978-1-61208-740-5

TABLE I. RANKING WITH 650 FOOTWEAR PRINTS FROM CRIME SCENES.

No.	1	3	4	5	6	8	9	10	14	15	16	18	19	23	24	27	28	29	31	33
Rank User 1	1	3	1	2	4	1	33	1	2	12	3	2	1	27	1	11	15	1	1	1
Rank User 2	3	7	1	5	1	4	1	4	1	4	2	18	1	6	12	4	12	4	1	4
Rank User 3	5	5	2	1	1	30	1	1	1	1	5	15	4	16	1	3	3	1	2	4

TABLE II. RANKING WITH 1500 FOOTWEAR PRINTS FROM CRIME SCENES.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank User 1	8	8	31	26	1	7	13	1	1	1	10	4	3	1	22	14	27	48	21	1
Rank User 2	1	17	1	11	1	15	27	11	7	12	22	2	10	2	11	25	4	6	16	3
Rank User 3	15	10	27	1	5	14	32	1	7	12	14	3	1	1	18	40	8	48	1	1

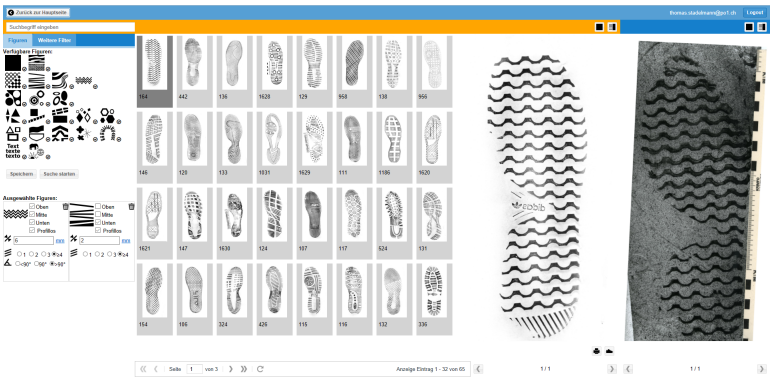


Figure 4. Screenshot of the Fast user interface.

C. Test system and test procedure

We conducted an online experiment, in which people annotated footwear print images via a browser-based web application. Once logged in, the participants first had to select a print from a crime scene and decide afterward whether they want to search for corresponding outsole models or rather for corresponding prints from other crime scenes. The workflow for processing an arbitrary footwear print image consisted of the following five steps:

1) Selecting a feature to describe the sole pattern on the image
2) Selecting attributes to specify the selected feature
3) Launch search request
4) Check results on the preview screen
5) Recording the position of the corresponding image

The participants could repeat the steps 1 to 4 as often as desired. In this way, they were able to assign multiple features and could, therefore, describe the sole pattern as precise as possible. The users were told to repeat the steps 1 to 4 until they found the corresponding images but not to codify the image as much as possible.

Our system was running on a web server with 8 vCores and 32 GB RAM. For every requested footwear print, the system retrieved the 32 most similar entries (footwear prints or outsole models) from the database and displayed them on the screen as thumbnails. The participants could click on the thumbnails to obtain a larger preview image. If the corresponding result could not be found on the first page, the participants were able to switch to the next 32 results.

All test participants received at first a short introduction of about 30 minutes. In this tutorial, they were introduced to the overall functionality of the system and the exact procedure of the test. Afterward, they were asked to independently identify the corresponding reference and prints from crime scenes for the 20 prints.

D. Results

The results for both search types (footwear print - outsole pattern and footwear print - footwear print) are presented in tables I and II (cf. Figure 5 and 6). For every request, the final position of the corresponding result is indicated. Yellow fields shoe results between position 33 and 64. This means, user preferred to scroll to the second page to get the results instead of undertaking a better coding.

For both search processes, the tables demonstrate that the difference between the rankings conducted by individual users is rather small and the system is easily applicable for both - forensic scientists and unexperienced users. In our experiment, a single user needed up to 2 hours to process all the 40 test samples. They reported to us that to the end of the experiment they were able to speed up considerably because they got meanwhile familiar to the system.

 ISBN: 978-1-61208-740-5

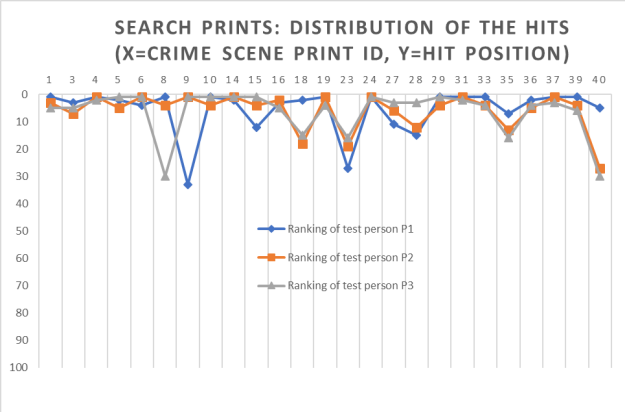


Figure 5. Ranking with 650 prints from crime scenes.

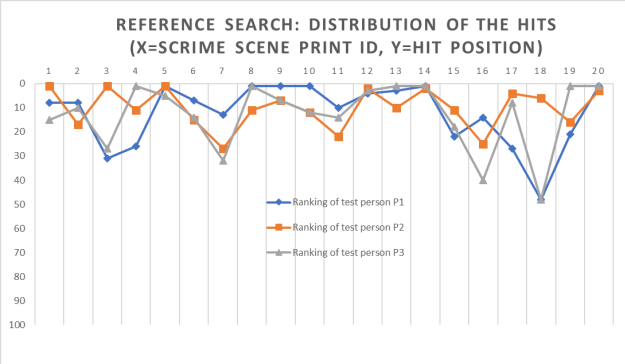


Figure 6. Ranking with 1500 prints from crime scenes.

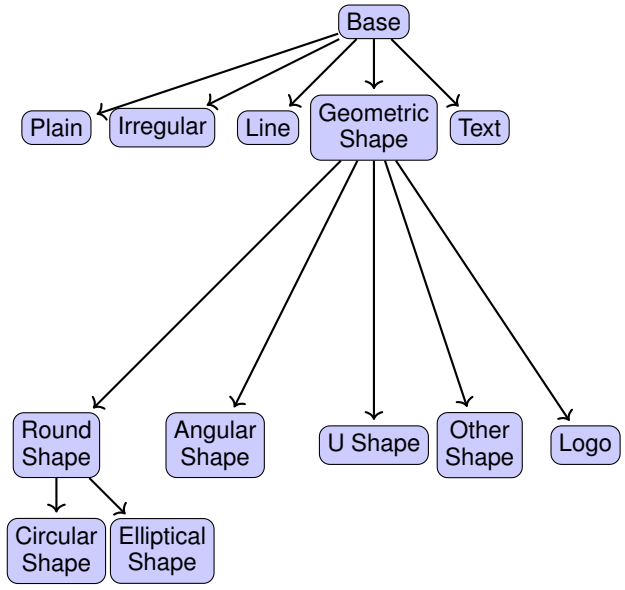


Figure 7. Feature Hierarchy

TABLE III. ALL EMPLOYED FEATURE TYPES AND ASSOCIATED ATTRIBUTES.

Base	
Attribute	Type
Position Tip	Boolean
Position Middle	Boolean
Position Heel	Boolean
Position Border Tip / Middle	Boolean
Position Border Heel	Boolean
Plain	
Attribute	Type
-	-
Irregular	
Attribute	Type
Texture Type	Enumeration (Crepe, Spots)
Spots Thickness	Floating point
Line	
Attribute	Type
Width	Floating point
Shape	Enumeration (round, straight segments)
Number of connected segments	Integer
Angle between segments	Floating point
Number of parallel lines	Integer
Distance between parallel lines	Double
Amount of crossed lines	Integer
Angle between crossed lines	Floating point
Geometric Shape	
Attribute	Type
Amount	Integer
Distance	Floating point
Round Shape	
Attribute	Type
Largest diameter	Floating point
Number concentric forms	Integer
Distance between concentric forms	Floating point
Filled	Boolean
Ring	Boolean
Ring width	Floating point
Circular Shape	
Attribute	Type
-	-
Elliptical Shape	
Attribute	Type
Shorter radius	Floating point
Angular Shape	
Attribute	Type
Number corners	Integer
Largest length	Floating point
Filled	Boolean
Regular	Boolean
U-Shape	
Attribute	Type
-	-
Other Shape	
Attribute	Value Type
Army cross	Boolean
Army rand	Boolean
Logo	
Attribute	Type
Trademark	String
Text	
Attribute	Value Type
Font size	Floating point
Text length	Floating point
Text	String

 ISBN: 978-1-61208-740-5

V. Conclusion and Future Work

We presented a novel approach for footwear print retrieval based on a hierarchical tree representation of footwear prints and outsole models consisting of features, feature types and attribute nodes. The proposed tree representation allows for establishing matches between entries of different feature/attribute orderings or levels of abstraction. As similarity measure, we opted for the tree edit distance due to its ability for incorporating weights and its polynomial runtime complexity. We applied our approach on a given set of footwear prints and compared the automatic assignments with that of human annotators. The evaluation showed the usefulness of our system and a high degree of correspondence between human and automated search results, even though we did not spend much time on parameter tuning yet. We identified possible future work in the following areas: Search process, GUI, and evaluation.

A. Search process

Our approach is highly customizable by a large number of weights, which can be adjusted separately for the individual type of modification (insertion, deletion, and replacement), node type (feature, feature type and attributes) and comparison mode (footwear prints against each other or footwear print vs. outsole model). Since a manual specification of so many parameters is quite tedious and typically involves a lot of trial and error cycles, we plan for future work to implement a genetic algorithm that adjusts them automatically.

B. GUI

Currently there is no automated check if the features and attributes, the forensic expert enters, actually matches to patterns in the footwear image. However, such a cross-check would be very beneficial but requires advanced image processing techniques. One could even go a step further and generate a first guess for the perceived features and attributes.

Besides, our system only lists similar outsole patterns but does not make any statement, whether the similarity is actually close enough so that the footwear print actually could belong to one of the identified outsole models or not. To accomplish this, one would have to derive some kind of threshold value for our similarity score that discerns the two possible outcomes *Match found* and *There is not match*.

C. Evaluation

It would be interesting to investigate, how the number of features and the accuracy of the system correlates. Finally, possible future work also includes an evaluation of other state-of-the-art systems on our dataset, which would allow for a quantitative comparison.

Acknowledgments

We thank our working collegues for their support regarding this paper and the InnoSuisse organization for funding this research.

References

[1] V. S. S. Mikkonen and P.Heinonen, “Use of footwear impressions in crime scene investigations assisted by computerized footwear collection system,” Forensic Science International, vol. 82, no. 1, 1996, pp. 67–79.

[2] A. Girod, “Computerized classification of the shoeprints of burglars’ shoes,” Forensic Science International, vol. 82, no. 1, 1996, pp. 59–65.

[3] A. Girod, C. Champod, and O. Ribaux, Trace de Souliers. Lausanne, Switzerland: Presse polytechniques et universitaires romandes, 2008.

[4] A. Kortylewski, “Model-based image analysis for forensic shoe print recognition,” Ph.D. dissertation, Basel University, 2017.

[5] W. Ashley, “What shoe was that? the use of compuerized image database to assist in identification,” Forensic Science International, vol. 82, no. 1, 1996, pp. 7–20.

[6] R. Davis, “An intelligence approach to footwear marks and toolmarks,” Journal of the Forensic Science Society, vol. 21, no. 3, 1981, pp. 183–193.

[7] H. Majammaa, “Footwear databases used in police and forensic laboratories,” Information Bulletin for Shueprint/Toolmark Examiners, 2000, pp. 133–157.

[8] A. Girod, “Efficiency of computerised database of burlarsoles’ standards,” Information Bulletin for Shoeprint/Toolmark Examiners, vol. 6, no. 1, 2000, pp. 125–132.

[9] S. Gross, D. Jeppesen, and C. Neumann, “The variability and significance of class characteristics in footware impressions,” Journal of Forensic Identification, vol. 63, no. 3, 2013, pp. 332–351.

[10] S. N. Srihari, “Analysis of footware impression evidence,” U.S. Department of Justice, Tech. Rep. 233981, 2011.

[11] N. Khosla and V. Venkataraman, “Building image-based shoe search using convolutional neural networks,” Stanford University, Tech. Rep. CS23 / Course Project Reports, 2015.

[12] A. Moschitti, “Efficient convolution kernels for depencency and constituent synatactic trees,” in Proceedings of the European Conference on Machine Learning, 2006, pp. 18–22.

[13] N. Augsten, M. Böhlen, and J. Gamper, “Approximate matching of hierarchical data using pq-grams,” in Proceedings of the 31st International Conference on Very Large Databases, 2005.

[14] M. Pawlik and N. Augsten, “Tree edit distance: Robust and memory-efficient,” Information Systems, vol. 56, 2016, pp. 157–173.

 ISBN: 978-1-61208-740-5

The Use of E-Portfolio to Develop Student's Self-reflection in Pre-school

The case of a Private Saida (Lebanon) High School

Rouaa Chahine and Hassan M. Khachfe
School of Education, and Business, Educational, and Medical Optimization Research Institute (BE-MORE)
Lebanese International University
Beirut, Lebanon
E-mail: 31730701@students.liu.edu.lb; hassan.khachfe@liu.edu.lb

***Abstract*—This study presents electronic portfolio (E-portfolio) as a tool for reflection, and appraises in great depth its impact on students' metacognition. The suggested simulation was empirically examined by investigating the nature, and composition of its framework using descriptive evidence from 24 participants of students in pre-school along with their parents and 36 preschool teachers, 4 admission people and 1 learning support assistant. Findings revealed students who used E-portfolio as tool of reflection had experienced a deeper understanding in learning by the action of realizing the point of strength and areas of improvement that played a role in student's intrinsic motivation. They also showed that digital reflection is more effective in simple decision-making skills than traditional methods.**

Keywords- E-portfolio; interactive learning environment; self- reflection.

I. INTRODUCTION

As we are moving in time, developing teaching methods, strategies and tools have become a must. An Electronic portfolio (E-portfolio), perhaps one of the latest trends in education, is becoming essential for both teachers and students, since it carries within its pages a preview of the student's or teacher's challenges, strengths, and areas of improvement, and foremost, it is considered a main tool for reflection. For this, including E-portfolios in schools, as a tool for reflection became something that Houssam Edeen Hariri High School (HHHS) is seeking to have to accomplish their vision, mission and goals.

In an overview of the school, HHHS is a private school that follows a holistic learning program, and places significant efforts in recruiting highly qualified teaching staff. It has been accredited as a Primary Years Program (PYP) in Saida, Lebanon. Since the topic of E-portfolio for students at schools is still gaining momentum and in the process of development in Lebanese schools, there is a need to investigate more about the relation between E-portfolio and self-reflection to find its efficacy. This study aims to analyze the impact of the use of E- Portfolio to develop student's self-reflection in early years in the above-mentioned school on the process of changes from traditional to E-portfolio. The rest of this paper is organized as follows. Section II presents a review of related literature about self-reflection and E-portfolio in schools. Section III describes the methodology of the study. Section IV provides the results of the study, and the analyses of the results. Section V proposes recommendations based on the findings of the study. The acknowledgement and conclusions close the article.

II. LITERATURE REVIEW

A review of self-reflection in the light of technology will be presented, and the following will be included: (1) definition of technology, portfolio as a mean of self-reflection and E-portfolio, (2) process of change into technology, (3) Theory of Connectivism, (4) summary. There are many types as of portfolios: Showcase, cumulative, goal based, process, active, evaluation, electronic, and mini portfolios. Many facilitators find the portfolio an accurate tool for assessing students especially those with disabilities since it shows the learning progress for each student. Portfolios also give students the opportunity to watch their progress and reflect on it, for example a learner would write "in this page you can find my areas of strength or the areas I need to improve." This involves learners in the process of evaluating their progress [1].

Other researchers suggested that traditional portfolios are not reflective of disabilities, and E-portfolios are way easier for all types of students equally [2]. E-portfolios allow the process of self-reflection to be more efficient. Using animations in E-portfolios has a significant effect in the process of self-reflection. Use of animations together with assessment methods and techniques will result in having the students actively participating in the process evaluation [3]. Also, it engages students and motivates them to use online communication tools [4]. Providing game-based learning engages the learners in an interactive-authentic problem-solving situation that provides enjoyable and motivating learning experiences [5].

Connectivism mentions learning as the ability to construct networks among education [6]. It is focused on connecting with people in a network in order to share experiences and specialized knowledge [6]. To support this process, current education has the goal of combining two effective tools that are meant by this change. The first tool is using Web 3.0 technologies for the learning platform in order to create user-centered learning and people [1]. The second tool is using well-defined website that is concerned in education objectives and outcomes between materials [8]. When using both tools, Connectivists will deliver communication networks consisting of experts for direct help and communication among learners.

 ISBN: 978-1-61208-740-5

The challenge of such an approach is revealed by the evaluation of the effectiveness of this tool in the light of change [8].

In summary, considering the elements for change, the future effective goals, and the learners in the educational institution who need this reflection, the school will be able to achieve the expected success. Thus, checking the current situation for the school and checking on students', teachers' and parents' needs and capacities will help on knowing how to move accordingly in less time that results in efficient change.

III. METHODOLOGY

The methodology is utilized to check the effectiveness of E-portfolios as intervention or enhancement instrument in pre classes at private school at KG2 classroom in Lebanon.

A. Research Design

The study applied an experimental design. Reference [9], "in an experiment the investigator controls the application of the treatment". The experimental strategy permitted examination of the effectiveness and impact of using E-portfolio to enhance the students' self-reflection skills knowing that self-reflection is an important key concept according to the PYP and in the enhanced PYP self-reflection must be integrated into all units of inquiry. The dependent variable in this study is: Utilizing E-portfolio as an intervention to prove students' self reflection at pre-school and develop reflection skills at this age. However, the independent variables are: education goal, social learning and environment, students' perception toward instruction, and evaluation. In this design, triangulation method offers the integration matter [8]. A descriptive quantitative research design was used.

B. Sample Size and Population

The target population of the current study contained students at pre-school, in a private school in KG2C (3-4 years old) from HHHS. One hundred fifteen participants were involved in the study (twenty four students, forty eight parents, forty one teachers, one learning support assistants and two school principals).

The selected sample of study followed the stratified random sampling technique. Stratified random sampling is suitable methodology in order to make balanced, meaningful, comparisons between sub-groups in the population [10].

C. Data Collection Tools

In this study, the researcher prepared questionnaire to be filled by students, principals, teachers and parents investigating their perceptions. The researcher used questionnaire for teachers examining effectiveness of using E-portfolio as a teaching method in KG2 based and taken from the International Journal of E-portfolio. The questionnaire is divided into two sections. The first section consists of 8 questions about the impact of digital portfolio on teachers, spread among 5 action scale with numerical value for each answer as 1= Developing, 2= Basic, 3= Satisfactory, 4= Efficient, and 5= Proficient. The second section consists of 6 questions that form a reflection for teachers about the impact of E-portfolio on the process of teaching and learning, the teachers and learning support assistant are asked to choose between yes and no. This questionnaire helped in targeting the main points that the researcher has to focus on and made data collection process go smoother.

Unlike teachers and learning support assistants' questionnaire, the students' questionnaire is made up of scale divided into 4 levels by which students will reflect on their journey of using E-portfolio derived from Marzano Scale on Teachers Pay Teachers website. This questionnaire was based on visuals which made learners engaged in the self-assessment process.

As for parents, they are asked to fill a questionnaire that the school always asks them to fill after each student led-conference.

All statistical analyses were carried out using the Microsoft Excel. The analyses that were examined in the study included: (1) Frequency Analysis to analyze the number of participants; (2) Descriptive Analysis to analyze the perceptions of participants towards the use of E-portfolios and its effects on self-reflection and metacognition.

The questionnaires selected were depending on how much each question is relevant to the case study itself and how much it reflects each teacher's, parents' and student's concerns when thinking about technology and E-portfolio.

IV. RESULTS AND DISCUSSION

This section introduces data analysis. It analyzes teachers' questionnaires that show their perceptions regarding the impact of digital portfolios on their teachings and on students as well. Moreover, it analyzes parents' questionnaires concerning their children learning, and it analyzes students' questionnaires to assess their perceived self-reflection. Afterward, the research questions and hypotheses are shown to correlate with the effectiveness of the E-portfolio in enhancing students' reflection skills. Finally, correlations between results and two UN Sustainable Development Goals (SDGs): Goal 4 (Quality Education) and Goal 13 (Climate Action) are discussed [11].

A. Results of Teachers' Questionnaires

When teachers developed their own digital portfolios, they learned more about efficiently using technology, rethought of their existing teaching practices, and enhanced their lesson planning. In addition, teachers gained the ability to teach their students how to create a digital portfolio. The use of technology by students was a new interesting experience for them. This result demonstrates a shift in pedagogical practice through incorporating technology to a higher degree and enhancing teachers' practices in having more timely communications with their students. The creation of digital portfolios engaged both teachers and students in a give-and-take process of learning.

 ISBN: 978-1-61208-740-5

These results significantly indicate that teachers were able to express the efficient and motivational use of digital portfolios on students learning, academic standards, self-assessment, and reflection skills. Also, our findings demonstrate that the development of digital portfolios by teachers positively impacted their students' way and amount of learning as well as their relation with them.

TABLE I. RESULTS OF TEACHERS' QUESTIONNAIRE

Criteria	Scale				
	1	2	3	4	5
Knowledge about digital portfolios	0%	4.9%	58.5%	24.4%	12.2%
Ability to create and use a digital portfolio	26.8%	24.4%	48.8%	0%	0%
Ability to teach your students how to create a digital portfolio	12.2%	63.4%	22%	2.4%	0%
Ability to use and integrate technology	0	24.4%	36.6%	36.6%	2.4%
Attitude towards using technology in the classroom	0	12.2%	12.2%	73.2%	2.4%
Collaboration with other teachers in or outside your school in the use of educational technology	0	0	0	0	100%
Ability to coach/ support colleagues in the use of educational technology	0	12.2%	36.6%	48.8%	2.4%

The relation between students and teachers is very critical since it is part of the reflection process that teachers use in order to guide curriculum and to assess individual and group understanding of concepts.

TABLE II. RESULTS OF PARENTS' QUESTIONNAIRE

Questions	*Yes*		*No*	
	Frequency	%	Frequency	%
Did using a digital portfolio with your son/daughter reflect more knowledge on their taught materials?	38	79.2%	10	20.8%
Did using a digital portfolio reflect how much your students learned?	40	83.3%	8	16.4%
Do you think your son/daughter learned academic content standards differently through reflecting while using digital portfolios?	38	79.2%	10	20.8%
Was using digital portfolios with your son/daughter important?	41	85.4%	7	14.6%

Upon comparing parents' answers to teachers' answers of the same set of questions, it can be noticed that their answers were approximately matching. Hence, these results validated how the use of technology in digital portfolio made parent engagement efforts even more effective.

Digital portfolios have been shown to push updates to parents rather than expecting them to check, and they were also proven to connect parents with what is actually happening in the classroom. This clearly emphasizes the positive impact of using a digital portfolio with their children on reflecting knowledge and learning.

B. Results of Students' Questionnaire

This is critical as self-reflection process does not only encourage students to think about their own thinking but also help them in developing their ability to know how to think. Portfolios give a mean for the students to monitor their achievements and in turn they become confident and motivated to take risks in the future [12]. Hence, there is a statistically significant relation for E-portfolio on the process of teaching and learning among preschoolers. Also, E-portfolio correlates with students' self-reflection skills like metacognition.

TABLE III. RESULTS OF STUDENTS' QUESTIONNAIRE

Self- reflection		Frequency	%
I know what I am good at very well. I feel like I could talk about it for someone else.		1	4.2%
I know my strength points pretty well. I remember every situation right after the question.		12	50%
I feel like I am still discovering what I am good at. I still have some questions and am unsure sometimes.		8	33.3%
I need to do Lots of tasks to know what do I know and what I don't. I am not sure what to do most of the time.		3	12.5%

 ISBN: 978-1-61208-740-5

C. Discussion and Interpretation

Most parents referred to E-portfolios as means of communication with their children in a way that connects home and school more deeply. Parents' perceptions were divided; many parents found using E-portfolio is essential especially that they found how their children were leading in the student led conference and talking about each and every single activity independently. On the other hand, some parents argued the idea of having their children learning through technology with mentioning that it would be harmful of learners at that age. Throughout the survey, most parents reported many benefits of E-portfolios and the strongest benefits included being able to document progress and the promising digital aspect of portfolios in a reflective manner. Most parents also indicated that E-portfolios provided a glimpse into their child's classroom, and they enjoyed the digital features of being able to hear and/or see as part of their perception toward E-portfolios implementation in the class.

As for teachers, their perceptions differed as the researcher is conducting the study; before starting to use the E-portfolio many teachers' responses were negative and they mentioned that it wouldn't affect the learning process for each learner then after applying E-portfolio they found how interactive, reflective and independent the students were and they found out how using E-portfolio reduces time teachers need to prepare and finish students' portfolios so they were motivated to start applying E-portfolios in their classrooms with their students. Foremost, many teachers realized that learners are capable to do self-reflection at that age which made them motivated to start the procedure with their learners.

Data collected from the teachers' questionnaire showed that teachers clearly expressed subjective satisfaction from E-portfolios. A critical theme from teachers was conducting an authentic E-portfolio learning process using "real-life application of their learning" and "watching children take their work seriously". As mentioned earlier, the collected data further show that E-portfolios impacted the teaching methods of the involved teachers. One important thing was that the teachers were able to have more insights toward each child through their selection of artifacts and their reflections on their work. Finally, through E-portfolios, teachers were able to have a strong view of children as capable learners whereby they described students as protagonists who have high engagement, reflection and ownership. Obviously, teachers viewed the creation and use of E-portfolios as positive teaching tools for their students and themselves.

Learners were extremely excited; they asked about every single detail concerning their E-portfolio and they were responsible to show the best image about each learner's profile in his/her E-portfolio. Implementing E-portfolio for students was a privilege that motivated the students in the process of learning and teaching. In addition, students showed a deeper thinking concerning the provided content and themselves as learners. In particular, this finding is helpful because most empirical studies are limited to students at elementary schools rather than preschoolers.

Overall, students, teachers and parents expressed subjective satisfaction to the use of E-portfolios. Hence, most parents, teachers, and students found E-portfolios effective for reflection.

These results allowed the researcher to accept the following hypothesis:

- H_A1: There is a statistically significant relation for E-portfolio on the process of teaching and learning among preschoolers.
- HA2: E-portfolio correlate with students' self-reflection skills like metacognition.
- Parents, teachers, and students find E-portfolio effective for reflection.

Correlation between Results and the UN Sustainable Development Goals (SDGs), The UN Sustainable Development Goals (SDGs) are considered as blueprints that lead to better achievements and more sustainable future for everyone.

Two major SDGs that could be related to the current study are goals 4 and 13 as discussed below:

- Goal 4 "Quality Education" [11]: Our findings reveal that children at HHHS are receiving a well-qualified care and pre-primary education that develops their deep understanding about subject. Moreover, they learn how to reflect upon their learning experiences and progress which increases their self-confidence, responsibility and metacognition.
- Goal 13 "Climate Action" [11]: Since every single and even the smallest action can make a difference and lead to a change, we decided to undertake a correlation between the creation and use of digital portfolios and this SDG. Using digital portfolios will reduce the use of papers, and thus their production and their unfavorable effects on the environment.

V. CONCLUSIONS AND RECOMMENDATIONS

The major focus of this study is to enhance the learning process in a digital context that requires new tools and methods, and E-portfolio is one of them. The current study discussed the potentials and merits of using digital portfolios. It is concluded that digital portfolios act as potent tools for assessing students' work and progression, structuring learning and teaching processes, enhancing communication and collaboration, sharing experiences and resources, and finally for supporting the self-reflection and metacognition of students. Moreover, the use of E-portfolios led to an enhanced quality of teaching, played a key role toward a sustainable development, reduced the use of papers, their production and their unfavorable effects on the environment, which correlates with the sustainable development goal that combats climate change and its impacts. Moreover, the use of digital portfolios helps in overcoming the technical problems associated with storing and accessing information while creating hardcopy portfolios.

 ISBN: 978-1-61208-740-5

A. Recommendations

Based on the review of the literature and the obtained findings of the current study, the accompanying pedagogical recommendations on some critical issues are suggested for future studies in order to bridge the existing research gaps:.

a) More training and integration of activities regarding structural reflection, social interaction, and self-awareness could be involved in E-learning materials in order to increase students' performance in self-monitor, self-modification, and self-evaluation.

b) Urge teachers to begin with introducing metacognitive strategies into the classroom .

c) Embed E-portfolios as "long-term professional development tools" for both inservice and preservice teachers are recommended.

d) Both private and public education departments should ensure the implementation of these tools within the teaching and learning processes.

e) Increase the time apportioned in class for the use if E-portfolios and some sort of re-organization could be done to enable this new teaching method to take up enough time in the class

f) Future research should examine and assess the differences between male and female students toward self-reflection and metacognition.

g) Articulate the concept of self-reflection and metacognition and encourage them among our students via E-portfolios. Ongoing efforts must ensure a sustainable dedication to effectively integrate E-portfolios to enhance student learning.

B. Limitations

This study suffers from some limitations. First, the sample was limited to only one school (115 participants: 24 students, 48 parents, 41 teachers, 1 learning support assistants, and 2 school principals). Thus, different results might be obtained if more teachers from different grade levels, schools, and areas were involved and studied. Secondly, our data is represented in the form of quantitative measures and relies only on descriptive statistics, which cannot confirm any statistical significance between data. In addition, the study sets self-report data, which is a type of data that in spite of speaking to the teachers' and parents' perceptions as participants, there was a limitation regarding the reliability of the collected data. It is also essential to report that the participants' perceptions may be subjected to internal bias due to several factors. However, obtaining similar findings from teachers, parents and students minimizes the possibility of having unreliable data.

C. Implications for Further Study in Lebanon

There is an absence of similar studies in Lebanon and a deficiency in other Arab countries, which have close pedagogical environment. It was also noted that the implementation of E-portfolios was more concentrated in universities more than schools. However, there is a possibility of inequitable comparison due to differences in the studied students' educational stages (preschoolers, primary education, secondary education, undergraduate, graduate), and the degree of pedagogical improvement between the other environments and cultures (USA, Malaysia, Taiwan, China, UK, and Turkey). These discrepancies would be better addressed in future research.

REFERENCES

[1] Barrett, H. "Authentic assessment with electronic portfolios using common software and Web 2.0 tools." www.electronicportfolios.org/web20.html, 2012

[2] Driessen, EW., Overeem K., van Tartwijk J., van der Vleuten C.P., and Muijtjens A.M. "Validity of portfolio assessment: which qualities determine ratings?" Journal of Medical Education. vol. 40 no.9, pp. 862-6, 2006.

[3] Lewis, M. A., and Maylor, H. R., "Game playing and operations management education." International Journal of Production Economics, vol. 105, pp. 134–149, 2007

[4] Barbera, E. "Mutual feedback in e-portfolio assessment: An approach to the netfolio system." British Journal of Educational Technology, vol. 40, no.2, pp. 342–357, 2009. DOI:10.1111/j.1467-8535.2007.00803.x

[5] Barzilai, S. and Blau, I., "Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences", Vol. 70, January 2014, pp. 65-79, 2014.

[6] Downes, S.,"Resources for distance education worldwide." International Review of Research in Open and Distance Learning, 2007.

[7] Nicolay et al., "International Journal of Industrial Ergonomics, vol. 35, nb. 7, pp. 605-618, 2015

[8] Fetters, M. D., Curry, L. A., and Creswell, J. W., "Achieving integration in mixed methods designs-principles and practices." Health Services Research, vol. 48, no. 6, pp. 2134-2156, 2015. doi:10.1111/1475-6773.12117

[9] Mosteller, F., and Youtz, C. "Quantifying Probabilistic Expressions", Rejoinder. Statist. Sci. 5 (1990), no. 1, 32--34. doi:10.1214/ss/1177012251.

[10] Gay, L.R., "Educational Research: Competencies for Analysis and Application. Columbus", OH: Merrill Publishing Company, 1987.

[11] Morton, S., Pencheon, D., and Squires, N. "Sustainable Development Goals (SDGs), and their implementation: A national global framework for health, development and equity needs a systems approach at every level." British Medical Bulletin, vol. 124, no. 1, pp. 81-90, 2017.

[12] Bransford, J., Brown, A., and Cocking, R. "How People Learn: Brain, Mind, and Experience & School." Washington, DC: National Academy Press, 2000.

 ISBN: 978-1-61208-740-5

Vibration Analysis with Application in Predictive Maintenance of Rolling Element Bearings

Theodor D. Popescu

National Institute for Research
and Development in Informatics
8-10 Averescu Avenue
011455 Bucharest
Romania
Email: Theodor.Popescu@ici.ro

Dorel Aiordachioaie
and Anisia-Culea Florescu

"Dunarea de Jos"
University of Galati
Galati, Romania
Email: Dorel.Aiordachioaie@ugal.ro
Email: Anisia.Florescu@ugal.ro

***Abstract*—The paper presents the vibration analysis problem with application in predictive maintenance of Rolling Elements Bearings (REB). After an overview of the maintenance approach, the condition monitoring in predictive maintenance is presented. A general view on change detection problem, with application in vibration monitoring, precedes some experimental results obtained in REB operating, for multiple faults and faults which gradually occur, with the conceptual description of the algorithm used. The approach proved to offer more robust detection of faults in REB, able to assure proactive actions in predictive maintenance.**

***Index Terms*—Fault detection and diagnosis; Rolling element bearings; Optimal segmentation; Vibrating signals.**

I. Introduction

Vibration analysis is one of the most effective tool used to check the health of plant machinery and diagnose the causes. The health of a machine is checked by routine or continuous vibration monitoring, giving an early indication of a possible failure and offering countermeasures to avoid a possible catastrophic event. Every machinery problem generates specific spectrum patterns, which are identified using frequency and phase analysis.

Vibration monitoring problem consists of machines condition and the change rate of its behavior. It can be ascertained by selecting a suitable parameter for deterioration measuring and recording its value for further analysis. This activity is known as condition monitoring. The great part of the defects encountered in the rotating machinery give rise to a distinct vibration pattern, or "vibration signature". Vibration monitoring has the ability to record and identify vibration "signatures" for monitoring rotating machinery. Vibration analysis is applied by using transducers to measure acceleration, velocity or displacement, depending of the frequencies making the object of the analysis. Different mechanical and electrical faults generate vibration "signatures" and careful scrutiny and deep study eliminates different possibilities and concludes to a single fault.

The problem of fault modeling and predictive health monitoring of Rolling Elements Bearings (REB) is one of great interest and made the object of many papers and books. El-Thalji et al. [1] presents such a monitoring procedure that includes detection, diagnosis and prognosis, to extract the features related to the fault occurrence. A general overview of various condition-monitoring and fault diagnosis techniques for REB in current practice is discussed in [2]. The paper of Randall and Antoni [3] offers a tutorial to guide the reader in REB diagnostics using vibrating signal analysis, and presents different case studies. An application of blind source separation method in diagnosis rolling bearing faults is presented in [4]. The study [5] presents a procedure for fault detection of roller bearings using signal processing and optimization techniques.

The matter of monitoring of REB plays a crucial role in the assessment of the overall health state of a rotating machine and is still a challenge. A new approach operating in time domain, using the optimal segmentation of vibration signals [6] occurred during REB operating, is used in the present paper. It offers new possibilities for more robust detection of changes in REB, and assures proactive actions in predictive maintenance.

The paper is organized as follows. Section II has as subject the maintenance approach, while in Section III, we present the condition monitoring problem in predictive maintenance. Section IV offers a general view on change detection problem with application in vibration monitoring. Finally, Section V presents some experimental results obtained in REB operating, for multiple faults and faults which gradually occur, and the conceptual description of the algorithm used.

II. Maintenance Approach

Usually, the maintenance is performed as *preventive maintenance*, at fixed time intervals, or as *reactive maintenance*, after the fault occurs. In the last case, it is necessary to perform immediately maintenance actions, while in the *predictive maintenance*, after a warning of a fault occurrence, the problem solving is carried out when necessary, so to avoid disruption of machine operations. A comparison of different maintenance

 ISBN: 978-1-61208-740-5

types, with disadvantages and advantages, is given in [7]. We present in the following some aspects concerning these approaches, to be taken into account, mainly in predictive maintenance of REB.

A. Reactive Maintenance

This approach refers to machine running until a fault occurs and involves fixing problems only when the fault occurs. It represents the simplest and cheapest approach in terms of maintenance costs; often it implies additional costs, usually due to unplanned downtime. It can be seen as an easy solution to many maintenance strategies.

In rotating machines, REB represent the most critical components, both in terms of initial selection, as well as in how they are maintained. Monitoring the condition of rolling bearings is essential and vibration based monitoring is frequently used to detect an early fault.

B. Preventive Maintenance

The preventive maintenance implies the scheduling of regular machine shutdowns, even if they are not required; this will increase the maintenance costs as some machine components are replaced, when this is not necessarily required. Some risks could appear due to replacing a defective machine part, incorrectly installing or reassembling parts. A frequent result of preventive maintenance consist of the fact that the maintenance is performed when there is nothing wrong in machine operating. Significant costs saving can be obtained by predictive maintenance.

C. Predictive Maintenance

The predictive maintenance refers to the process of monitoring the machine condition as it operates in order to predict which components are likely to fail and when. So, the maintenance can be planned and there is the possibility to change only those components that show failure signs in their operation. The predictive maintenance principle consists of taking additional measurements in order to predict the behavior of machine components that are susceptible of failure, and also to predict when these failures will occur. Usually, these measurements include machine vibration, and machine operating parameters: flow, temperature, pressure, etc.

The continuous monitoring detects, in advance, the onset of component problems, so the maintenance is performed when needed. By this approach, unplanned downtime is reduced, and also the risk of catastrophic failure is reduced. This will increase the efficiency and reduce the costs. By predictive maintenance strategy, applied in rolling bearings, the costs can be cut, giving in advance, a warning of a possible failure, enabling remedial action in advance.

III. Condition Monitoring

Condition monitoring consists of machine monitoring for early signs of failure so that the maintenance activity can be better planned, with reduced down time and costs.

The monitoring of vibration, temperature, voltage or power and oil analysis is frequently the most used. Vibration is the most widely used for its ability to detect and diagnose failure problems, but it offers also a prognosis on the useful life and possible failure mode of the machine. The prognosis is much more difficult to be performed and usually relies on continue monitoring of the fault to estimate the time when the machine will become unusable, taking into account the known experience in similar cases.

Vibration monitoring can be considered the most widely used predictive maintenance technique, and can be applied to a wide area of rotating machines. Machine vibration comes from many sources such as bearings, gears, unbalance, etc., each sources having its own characteristic frequencies, manifesting as a discrete frequency, or as a sum and/or difference frequency. It can generate complex vibration signals, which cause problems in vibration analysis, but some techniques, with a high sensitivity to faults, can reduce the complexity of the analysis. Bearing defects can affect higher frequencies, offering a basis for detecting incipient failure.

Usually, the detection uses the basic form of vibration measurement, where the vibration level is measured on a broadband basis (10-1000 Hz or 10-10000 Hz). The spikiness of the vibration signal, in machines with little vibration other than in the case of the bearings, is highlighted by the Crest Factor, indicating an incipient defect, and the a great value of the energy given by RMS level indicates a severe defect.

These measurements offer limited information, but they can be useful for trend evaluation; increasing vibration level highlights the machine condition deterioration. Also, a comparison of the measurement level with some vibration criteria from literature proves to be useful in practice.

Generally, rolling bearings generate very little vibration in faults absence, and present specific frequencies when a fault occurred. At the beginning of a fault, for a single defect, the vibration signals present a narrow band frequency spectrum. As the malfunction increases, an increase in the characteristic defect frequencies and sidebands can be noticed, with a drop in these amplitudes, broadband noise increasing and considerable vibration at shaft rotational frequency [7]. At very low machine speed, low energy signals are generated by the bearings, difficult to be detected. Also, bearings located within a gearbox are difficult to monitor, because of the high energy at the gear, which can mask the bearing defect frequencies.

IV. Change Detection in Vibration Monitoring

The CD problem is frequently present for continuous monitoring of systems like machinery, structure, process, equipment or plant, using data provided by the sensors. So, it is possible to anticipate the abnormal functioning or these systems, before

 ISBN: 978-1-61208-740-5

it occurs and to reduce the maintenance costs. The normal behavior of the system can be described by a parametric model, without using artificial excitation, reducing the speed of the equipment or temporary stop. If such early detections are possible, large changes of the system can be prevented, and the effects of defects, mechanical fatigue, etc., can be quickly anticipated, raising the usability of the system.

The applications in this field make use of theories based on statistics, providing theoretical instruments to solve the early detection problem. Many industrial processes are based on known physical principles, with available analytical models, and for very complicated or unknown models, semi-physical or black-box models can be used. Vibration analysis and surveillance of machinery or industrial equipments represent important cases of detection and diagnosis problems.

The CD problem refers to detection of the change (the alarm) and evaluation of the change (estimation), providing information, in some cases, for diagnosis (source isolation). The performance criterion of a change detection algorithm consists in its ability to correctly detect the changes, with minimum delay and minimum probability of false decisions. So, it must respond to the small changes (sensitivity to changes), without being affected by the disturbances, noise or modeling errors (robustness of the algorithm). The sensitivity and robustness properties are usually in conflict, a good change detection algorithm must perform a compromise between the two aspects.

Two basic approaches in CD are reported as based on quantitative models (using analytical redundancy) and qualitative models, which can be conveniently combined to improve the robustness of the generation of quantitative residuals. In the case of analytical exact models absence, learning models, such as fuzzy and neural models, can be used. Moreover, the neural networks can be used for classification of the residuals, while fuzzy logic is useful for decision making. The methods based on quantitative models are oriented to identification (parameter estimation), observers (state estimation) and parity space. Some heuristics results, obtained from the previous experience, can be used for diagnosing the origins of the failure or change, based on the dispersion of the characteristics.

Almost all CD solutions assume that the monitored system can be described with sufficient precision by a finite-dimensional linear model. In practice, if the system is more complex than the structure described by a finite-dimensional model, the parameter estimates will still converge, but their values can be strongly dependent on the experimental conditions. The algorithms will not be able to separate the changes determined by the external conditions from those occurred by the internal defect of the investigated system, so the classical tests will fail. The problems mentioned above point out the requirement of the robust CD algorithms, able to separate the changes determined by the external conditions from the changes of the internal dynamics of the system.

The first generation of CD algorithms is based on strong hypotheses, or strong assumptions, which are difficult to verify in practice. So, a second generation of solutions were required, insensitive to the uncertainty of the system's dynamics, to the operating environment, and to large noise, statistically unknown. In our opinion, the central problems to be addressed in the CD area refer to robustness, sensitivity and versatility. The lack of robustness of the classical algorithms concerns the failure of the detection, if one or more of the hypotheses assumed during the design are not verified in practice. The sensitivity relates to the ability of the algorithm to detect the change, even if there are small scale incipient changes. Finally, the versatility is linked to the ability of the methods and techniques to solve more CD problems, using the same set of algorithms.

To solve the vibration monitoring problem different techniques have been developed, one can mention: analysis of overall vibration level, frequency spectrum, envelope spectrum, cepstrum analysis, etc. [7]. The success of vibration monitoring, in many practical cases, requires specialized functions and tools. Simple application of CD techniques on original mono- or multivariate vibration signals can assure successful monitoring. Sometimes, it is necessary that some signal pre- or postprocessing procedures to be applied, to emphasize and highlight the characteristics of the vibration signals making the object of the analysis. So, some signal processing techniques can be used in conjunction with CD techniques: independent component analysis (ICA), time-frequency analysis (TFA), energy distribution (ED) evaluation in time-frequency domain. These techniques are implemented in a software toolbox, Matlab VIBROTOOL Toolbox [8], built as a set of programs that compute specific parameters and solve specialized tasks for vibration monitoring. A general approach, making use of these techniques, and a case study having as object the condition monitoring of a rotating machine, an industrial pump, with a progressed pitting in gears, is presented in [9].

The CD problem can be solved by change point estimation (mean change), change detection using one and two model approach, with different distance measures and stoping rules [10], multiple change detection [6], detection and diagnosis of model parameter and noise variance changes [11], for mono- and multivariable vibration signals. Some algorithms, making the object of [12] and [13] in CD, represented the starting points in developing these algorithms. The analysis of the vibration signals behavior reveals that most of the changes that occur are either changes in the mean level, variance, or changes in spectral characteristics.

V. Fault Detection in Rolling Elements Bearings

This section presents some experimental results, obtained in a case study, having as object fault detection in REB, as well as the conceptual description of the algorithm used.

A. Test Data

The experiments performed use a data set from [14], with three faults having different locations: $F1$ (Inner race), $F2$

 ISBN: 978-1-61208-740-5

(Ball) and $F3$ (Outer race), and four sizes of the faults; $F0$ denotes no faults; only the data for the first case (06HH) have been used (see Table I).

TABLE I. 1ST DATA TEST SET (6203 BEARING TYPE).

Fault size	$F0$ Free	$F1$ Inn. Race	$F2$ Ball	$F3$ Outer Race
0.000"	$y_0(t)$	-	-	-
0.007"	-	$y_1(t)$	$y_2(t)$	$y_3(t)$
0.014"	-	$y_4(t)$	$y_5(t)$	$y_6(t)$
0.021"	-	$y_7(t)$	$y_8(t)$	$y_9(t)$
0.028"	-	$y_{10}(t)$	$y_{11}(t)$	-

$y_0(t)$ contains 4,096 samples recorded during normal conditions operating, while $y_i(t)$, $i = 1, \dots 11$ indicate files/vectors, containing each 4,096 samples, for the cases with faults; the sampling rate was of 12,000 samples/s.

B. Preliminary Analysis

For the signals mentioned above, some statistical features in time domain [2], have been computed, and are given in Table II, offering a general view of the signal characteristics.

TABLE II. STATISTICAL FEATURES OF THE SIGNALS $y_0(t)$, $y_1(t)$, ..., $y_{11}(t)$ IN TIME DOMAIN.

Signal	RMS	Mean	Var.	Cres. fact.	Skew.	Kurt.
$y_0(t)$	0.999	-0.002	0.998	3.796	-0.094	2.890
$y_1(t)$	0.992	0.007	0.985	5.145	0.124	5.456
$y_2(t)$	1.007	0.021	1.014	3.720	0.003	2.997
$y_3(t)$	0.997	0.016	0.995	5.189	0.088	7.698
$y_4(t)$	0.997	-0.001	0.995	4.016	0.067	4.281
$y_5(t)$	1.013	0.013	1.027	5.299	0.012	7.032
$y_6(t)$	0.987	0.078	0.974	9.747	-0.144	22.505
$y_7(t)$	0.724	0.001	0.525	6.937	-0.066	5.775
$y_8(t)$	0.978	0.046	0.958	3.779	0.023	2.982
$y_9(t)$	1.018	0.011	1.037	6.495	0.315	6.868
$y_{10}(t)$	0.981	0.019	0.963	4.378	0.043	3.457
$y_{11}(t)$	0.955	0.002	0.913	9.992	-0.086	21.255

The signals, making the object of the analysis, are simultaneously characterized in time and frequency domain using their mean localizations and dispersions. So, the averaged time and the time spreading, as well as the averaged frequency and the frequency spreading [15], are given in Table III for signals analyzed.

TABLE III. TIME-FREQUENCY STATISTICAL FEATURES OF THE SIGNALS $y_0(t)$, $y_1(t)$, ..., $y_{11}(t)$.

Signal	Aver. time	Time spread	Aver. freq.	Freq. spread
$y_0(t)$	2.104e+003	4.251e+003	-8.197e-009	0.287
$y_1(t)$	2.032e+003	4.155e+003	-2.359e-008	0.850
$y_2(t)$	2.026e+003	4.103e+003	-1.035e-006	0.906
$y_3(t)$	2.090e+003	4.167e+003	-2.206e-008	0.969
$y_4(t)$	1.944e+003	4.157e+003	-5.457e-009	0.804
$y_5(t)$	2.082e+003	4.247e+003	-3.880e-008	0.983
$y_6(t)$	1.954e+003	4.099e+003	-1.229e-008	0.920
$y_7(t)$	1.993e+003	4.843e+003	-1.134e-008	0.820
$y_8(t)$	2.057e+003	4.187e+003	-1.800e-007	0.968
$y_9(t)$	2.054e+003	4.273e+003	-1.604e-007	0.857
$y_{10}(t)$	2.006e+003	4.184e+003	-1.435e-007	0.909
$y_{11}(t)$	2.085e+003	4.081e+003	-9.584e-010	0.911

C. Algorithm Description

The model used in the case study is a linear regression model with piecewise constant parameters [6],

$$y_t = \phi_t^T \theta(i) + e_t, \;\; E(e_t^2) = R_t, \tag{1}$$

where y_t is the observed signal, $\theta(i)$ is the d-dimensional parameter vector in data stationary segment i, ϕ_t is the regressor. The noise e_t is assumed to be Gaussian with variance R_t. Its important feature is that the jumps divide the vibration signals into a number of independent segments, since the parameter vectors in different segments are independent.

To solve the segmentation problem, all possible segmentation k^n are considered, estimate one linear regression model in each segment, and then choose the particular k^n that minimizes an optimality criteria of the form:

$$\widehat{k^n} = \arg \min_{n \geq 1, 0 < k_1 < \dots < k_n = N} V(k^n) \tag{2}$$

For the measurements in a i-th segment, $y_{k_{i-1}+1}, \dots y_{k_i} = y_{k_{i-1}+1}^{k_i}$, results the least square estimate and its covariance matrix:

$$\hat{\theta}(i) = P(i) \sum_{t=k_{i-1}+1}^{k_i} \phi_t R_t^{-1} y_t, \tag{3}$$

$$P(i) = \left(\sum_{t=k_{i-1}+1}^{k_i} \phi_t R_t^{-1} \phi_t^T \right)^{-1}. \tag{4}$$

The following quantities are used in optimal segmentation algorithm:

$$V(i) = \sum_{t=k_{i-1}+1}^{k_i} (y_t - \phi_t^T \hat{\theta}(i))^T R_t^{-1} (y_t - \phi_t^T \hat{\theta}(i)) \tag{5}$$

$$D(i) = -\log\det P(i) \tag{6}$$

$$N(i) = k_i - k_{i-1} \tag{7}$$

where $V(i)$ - the sum of squared residuals, $D(i)$ - $-\log\det$ of the covariance matrix $P(i)$ and $N(i)$ - the number of data in each i segment, and represent sufficient statistics for each segment. The data and quantities used in segmentation k^n, having $n-1$ degrees of freedom are given in Table IV.

TABLE IV. DATA AND QUANTITIES USED IN OPTIMAL SEGMENTATION PROCEDURE.

Data	$y_1, y_2, \dots, y_{k_1}$	...	$y_{k_{n-1}+1}, \dots, y_{k_n}$
Segment	Segment 1	...	Segment n
LS est.	$\hat{\theta}(1), P(1)$	...	$\hat{\theta}(n), P(n)$
Statistics	$V(1), D(1), N(1)$	...	$V(n), D(n), N(n)$

To solve the optimal segmentation procedure, different types of optimality criteria have been proposed [13]. In the following we will use Maximum A posteriori Probability estimate

 ISBN: 978-1-61208-740-5

(MAP) criterion [6]. The number of segmentations k^n is 2^N (can be a change or no change at each time instant), and this raises problems concerning the dimensionality.

The conceptual description of the MAP estimator [6], [13] for the data and quantities given in Table IV it is presented below, for three different assumptions on noise scaling: (i) known $\lambda(i) = \lambda_0$, (ii) unknown but constant $\lambda(i) = \lambda$ and (iii) unknown and changing $\lambda(i)$, where q is the change probability at each time instants ($0 < q < 1$).

Data: Vibration signal $y_t, \quad t = 1 \dots N$

Step 1: Examine every possible segmentation, parameterized in the number of jumps n and jump times k^n,separately.

Step 2: For each segmentation, compute the best models in each segment parameterized in the least square estimates $\hat{\theta}(i)$ and their covariance matrices $P(i)$.

Step 3: Compute in each segment:

$$\begin{aligned} V(i) &= \sum_{t=k_{i-1}+1}^{k_i} (y_t - \phi_t^T \hat{\theta}(i))^T R_t^{-1} (y_t - \phi_t^T \hat{\theta}(i)) \\ D(i) &= -\log\det P(i) \\ N(i) &= k_i - k_{i-1} \end{aligned}$$

Step 4: MAP estimate, $\widehat{k^n}$, for the three different assumptions on noise scaling

(i) known $\lambda(i) = \lambda_0$,

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} (D(i) + V(i)) + 2n \log \frac{1-q}{q}$$

(ii) unknown but constant $\lambda(i) = \lambda$,

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} D(i) + (Np - nd - 2) \times \times \log \sum_{i=1}^{n} \frac{V(i)}{Np-nd-4} + 2n \log \frac{1-q}{q}$$

(iii) unknown and changing $\lambda(i)$,

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} (D(i) + (N(i)p - d - 2) \times \times \log \frac{V(i)}{N(i)p-d-4}) + 2n \log \frac{1-q}{q}$$

Results : Number n and locations k_i , $k^n = k_1, k_2, \dots, k_n$

In a practical problem, only one of the equations from **Step 4** is evaluated, according with the assumption on noise scaling of the procedure.

For the exact likelihood evaluation, there are implemented recursive local search techniques and numerical searches based on dynamic programming or Markov Chain Monte Carlo (MCMC) techniques [6], [13].

Starting from the optimal segmentation results, it is possible to analyze the data resulted for each stationary data segment to locate and diagnose the occurred fault or change in the REB: outer race, inner race, bearing cage, ball (roller), according with the frequency area where it has occurred.

D. Multiple Fault Detection

Started from the data given in TABLE I data sequences with multiple faults have been generated, for 3 types of events: inner race faults, ball faults and outer race faults, with different fault size: 0.007", 0.014", 0.021", 0.028", for the first two cases, and 0.007", 0.014", 0.021" for the third case. The following data sets have been used in the analysis, for fault detection:

$$\begin{aligned} s_1(t) &= [y_0(t), y_1(t), y_4(t), y_7(t), y_{10}(t)] \\ s_2(t) &= [y_0(t), y_2(t), y_5(t), y_8(t), y_{11}(t)] \\ s_3(t) &= [y_0(t), y_3(t), y_6(t), y_9(t)] \end{aligned}$$

resulting data sequences of 20480 values for signals $s_1(t), s_2(t)$ and 16384 for signal $s_3(t)$. The real faults instants were 4097, 8193, 12288 and 16384. These data sets offer the possibility to fault detection of a graduate size of fault, for the cases mentioned above.

The experimental results refer to the signals $s_1(t), s_2(t), s_3(t)$ and the segmenting algorithm presented above with unknown and constant noise scaling, and MCMC algorithm, [6], with a value of jump probability, $q = 0.3$ and appropriate design parameters in search scheme, for different model orders, na. The fault instants detected for different model orders na are presented in Table V, Table VI and Table VII for $s_1(t)$, $s_2(t)$ and $s_3(t)$, respectively.

The signal $s_1(t)$, making the object of the analysis, and the estimated multiple fault times for the inner race, $na = 20$ and $q = 0.3$, are presented in Figure 1, while the signal $s_2(t)$ and the estimated multiple fault times for ball, $na = 20$ and $q = 0.3$ are given in Figure 2. The signal $s_3(t)$ and the estimated multiple fault times for the outer race, $na = 60$ and $q = 0.3$ are presented in Figure 3.

TABLE V. FAULT DETECTION IN SIGNAL $s_1(t)$ USING DIFFERENT MODEL ORDER.

Model order	Fault detection instants
$na = 10$	4096, 8687, 9501, 10684, 11322, 11500, 12570, 12627, 12967, 13068, 13961, 14527, 14627, 14777, 15964, 16384.
$na = 15$	4096, 8687, 9502, 10684, 11501, 12570, 14777, 16384.
$na = 20$	4096, 8195, 8687, 11502, 13026, 16384.

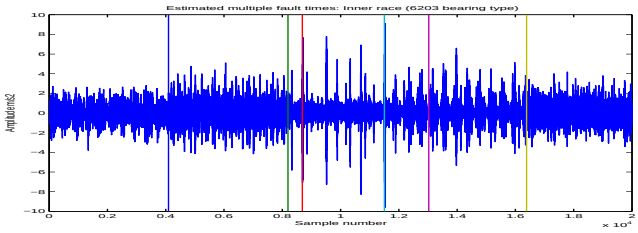


Fig. 1. The signal $s_1(t)$ and estimated multiple fault times for inner race, $na = 20$, $q = 0.3$.

The changes in signals $s_1(t)$, $s_2(t)$ and $s_3(t)$, resulted after data concatenation, are gradual, and the effect may increase, producing new changes in the signal dynamics that can be

TABLE VI. FAULT DETECTION IN SIGNAL $s_2(t)$ USING DIFFERENT MODEL ORDER.

Model order	Fault detection instants
$na = 10$	4096, 8191, 8497, 8614, 9305, 9929, 11946, 16385, 16711, 16901, 18065, 18129.
$na = 15$	4096, 8190, 11946, 16385, 16719, 18108, 18128.
$na = 20$	4096, 8190, 11945, 16385, 16751, 18233.

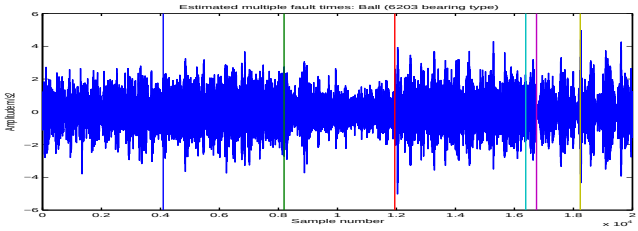


Fig. 2. The signal $s_2(t)$ and estimated multiple fault times for the ball, $na = 20$, $q = 0.3$.

detected by the algorithm. The further deterioration of the rolling element bearing during operating occurs new fault instants, different from 4096, 8192, 12288 and 16384 instants. According with data from Table V, Table VI and Table VII, one can notice that in all the cases the main faults are detected. Also, it can be noted that for the models of high order ($na = 20$, $na = 20$ and $na = 60$, respectively), only the main faults are detected at instants 4096, 8192, 12288 and 16384 or near instants. The models of high order, can increase the robustness of the optimal segmentation algorithm to gradual, or small changes in signal dynamics. Different values of q offer similar results, but a higher order of the model leads to a better fault detection, the model being able to better approximate the signal dynamics.

VI. CONCLUSIONS

The paper presents a vibration analysis approach, with application in predictive maintenance of REB. The experimental results, presented in the case study, have as object detection of the multiple faults, as well as of the faults, which gradually occur, in REB operating. The optimal segmentation method is based on maximum a posteriori probability estimator and need a minimum of design parameters, depending to a great extend of the linear regression model order. The used approach offers new possibilities for more robust detection of changes

TABLE VII. FAULT DETECTION IN SIGNAL $s_3(t)$ USING DIFFERENT MODEL ORDER.

Model order	Fault detection instants
$na = 10$	4096, 4383, 7081, 7170, 7897, 7950, 8192, 12298, 12367, 12480, 12982, 13151, 13260, 13407, 13596, 14042, 14179, 14378, 14489, 14668, 14823, 15169, 15271, 15575, 15605, 16050, 16229.
$na = 15$	4096, 8192, 12296, 12368, 12479, 12669, 12813, 13261, 13455, 13596, 14042, 14173, 14378, 15015, 15164, 15271, 15469, 15605, 16051, 16346.
$na = 20$	4096, 8192, 12293, 12367, 12479, 12669, 12813, 13261, 13460, 13594, 14042, 14189, 14378, 15271, 15473, 15604, 16051.
$na = 60$	4096, 8198, 12287, 12352, 14057.

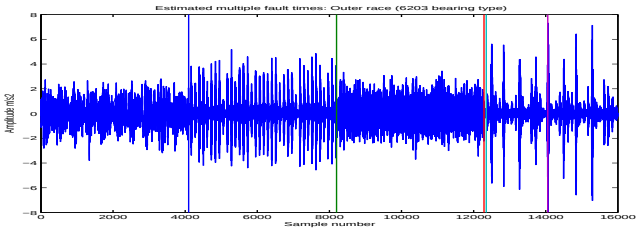


Fig. 3. The signal $s_3(t)$ and estimated multiple fault times for outer race, $na = 60$, $q = 0.3$.

in vibration signals, and assures proactive actions in predictive maintenance.

ACKNOWLEDGMENT

The authors thank the Ministry of Research and Innovation for its support under the 2019-2022 Core Program, Cod PN 301 200, Project RO-SmartAgeing.

REFERENCES

[1] I. El-Thalji and E. Jantunen, "A Summary of Fault Modelling and Predictive Health Monitoring of Rolling Element Bearings", *Mechanical Systems and Signal Processing*, vol. 60-61 pp. 252-272, 2015.

[2] T. R. Lin, K. Yu and J. Tan, "Condition Monitoring and Fault Diagnosis of Roller Element Bearing", INTECH, http://www.intechopen.com/books/bearing-technology, pp. 39-75, 2017.

[3] R. B. Randall and J. Antoni, "Rolling Element Bearing Diagnostics - A Tutorial", *Mechanical Systems and Signal Processing*, vol. 25, pp. 485-520, 2011.

[4] C. Yi, Y. Lv, H. Xiao, G. You and Z. Dang, "Research on the Blind Source Separation Method Based on Regenerated Phase-Shifted Sinusoid-Assisted EMD and Its Application in Diagnosing Rolling-Bearings Faults", *Applied Sciences*, vol. 7, pp. 1-18, 2017.

[5] D. H. Kwak, D. H. Lee, J. H. Ahn and B. H. Koh, "Fault Detection of Roller-Bearings Using Signal Processing and Optimization Algorithms", *Sensors*, vol. 14, pp. 283-298, 2014.

[6] Th. D. Popescu, "Signal Segmentation Using Changing Regression Models with Application in Seismic Engineering", *Digital Signal Processing*, vol. 24, pp. 14-26, 2014.

[7] S. .J. Lacey, "The Role of Vibration Monitoring in Predictive Maintenance", *FAG Technical Publication*, Schaeffler Limited, UK, 2010.

[8] Th. D. Popescu and D. Aiordachioaie, "VIBROTOOL - Software Tool for Change Detection and Diagnosis in Vibration Signals", *in Proceedings of the 59th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS 2016)*, Abu Dhabi, United Arab Emirates, October 16-19, pp. 640-643, 2016.

[9] Th. D. Popescu and D. Aiordachioaie, "A General Approach for Change Detection in Vibration Signals with Application in Machine Health Monitoring, *in Proceedings of the 6-th International Symposium on Electrical and Electronic Engineering (ISEEE'19)*, 18-20 October 2019, Galati, Romania.

[10] Th. D. Popescu, "Blind Separation of Vibration Signals and Source Change Detection - Application to Machine Monitoring", *Applied Mathematical Modelling*, vol. 34, pp. 3408-3421, 2010.

[11] Th. D. Popescu, "Detection and Diagnosis of Model Parameter and Noise Variance Changes with Application in Seismic Signal Processing", *Mechanical Systems and Signal Processing*, vol. 25, pp. 1598-1616, 2011.

[12] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Applications*, Prentice Hall, N.J., 1993.

[13] F. Gustafsson, *Adaptive Filtering and Change Detection*, Wiley, 2001.

[14] * * * Case Western Reserve University Bearing Data Center, *http://csegroups.case.edu/bearingdatacenter/home*, 2017.

[15] P. Flandrin, *Temps-fréquence*, Hermes, 1998.

 ISBN: 978-1-61208-740-5

A Method of Feature Extraction from Time-Frequency Images of Vibration Signals in Faulty Bearings for Classification Purposes

Dorel Aiordachioaie
"Dunarea de Jos" University of Galati
Electronics and Telecom. Department
Galati, Romania
e-mail: Dorel.Aiordachioaie@ugal.ro

Theodor D. Popescu
National Institute for Research
and Development in Informatics
Bucharest, Romania
e-mail: Theodor.Popescu@ici.ro

Bogdan Dumitrascu
"Dunarea de Jos" University of Galati
Electronics and Telecom. Department
Galati, Romania
e-mail: Bogdan.Dumitrascu@ugal.ro

***Abstract*—Time-frequency image processing is considered in the context of change detection and diagnosis purposes of bearings with faults and vibration signal processing. The analysis of these images reveals some difficulties in obtaining accurate results in classification. Some images are different compared to other images from the same set. New images are obtained by applying a criterion based on the contours generated by the main components of the analyzed time-frequency image. The transformed images are less complex, and could be with white and black only. Features based on statistical moments are considered, selected and used to define discriminant functions, which could improve the results of the classification. The features include the number of the contours, the average area defined by the contours, the variance of the areas and the Renyi entropies.**

Keywords-signal; vibration; image; time-frequency transform; signal processing; feature selection; classification.

I. INTRODUCTION

An important activity in industry, for safe work and quality of the products, is the Change Detection and Diagnosis (CDD) of various processes. These two activities are parts of wider domain, called condition-based and predictive maintenance, as described in some excellent books with theory and applications [1] [2] [3]. In the field of vibrational processes, i.e., processes which generates mechanical vibrations, with or without faults or damages, advanced signal processing algorithms are intensively used to elaborate accurate and robust algorithms for process diagnosis [4] [5] [6].

One of the more complex signal processing method is based on time-frequency transform, and next on time-frequency images, as described in [7] [8] [9]. The structure of such processing chain is presented in Figure 1. Signals from the process under study are pre-processed both on continuous and discrete time, mainly by filtering and scaling. Next, a time sliding window is considered for the computation of the time-frequency transform. The parameters of the sliding window depend on the statistical properties of the analyzed signals, to meet the condition of the statistical stationarity. The coefficients of the time-frequency transform are considered as elements of an image. From this point all processing steps are based on image processing, for various tasks, as fault detection and diagnosis.

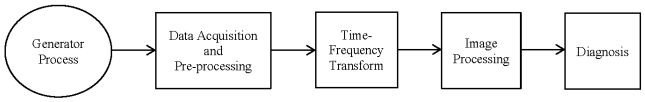


Figure 1. The block structure of signal processing for CDD

This work considers the last block before diagnosis, i.e., image processing for classification purposes. The previous processing blocks, mainly for signal processing, i.e., data acquisition and pre-processing, time-frequency transforms, are described in the Sections of the paper. Direct classification of time-frequency images do not offer always the best results in CDD activities, as described in [10] [11]. It is the main objective of this paper to define algorithms for feature selection and extraction, in order to expect better results of the classification and diagnosis purposes.

The rest of this paper is organized as follows. Section II describes the basic transforms applied to the vibrating signals, i.e., time-frequency and Renyi entropy. Section III describes the data which was used for experiments, including time-frequency images. Section IV addresses the proposed method for feature selection. Section V goes into the results of the experiments. The conclusion and acknowledgement close the article.

II. SIGNAL TRANSFORMS

Signal transforms are used to compute specific features of the analyzed signal or to change the analysis system, e.g., time-frequency transforms, or to compute and extract other relevant features, e.g., the Renyi entropy.

A. Time-Frequency Transform

Time-frequency transforms are advanced processing techniques for data processing, and especially for data coming from non-stationary signals. A general theoretical framework is presented in [12] [13]. Examples of signals and applications are audio signals [14], mechanical vibrations [15] or biomedical signals [16].

If $x(t)$ is a continuous (possible complex) signal, the time-frequency transforms can be obtained from the general formula [20]

 ISBN: 978-1-61208-740-5

$$C_x(t,\omega)=\frac{1}{4\pi^2}\iiint x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)\phi(\theta,\tau) \cdot e^{-j\theta t-j\omega\tau+j\theta u}dud\tau d\theta \quad (1)$$

where $\phi(\theta,\tau)$ is the kernel function, which imposes the properties of the distribution, and "*" denotes complex conjugation. If the kernel function is 1, then the Wigner distribution is obtained.

For the case where *x(t)* is an analytical signal, the Wigner distribution is called the Wigner-Ville distribution (WVD) [17]. This distribution satisfies a large number of desirable mathematical properties, as described in the specialized literature [18] [19]. In particular, the WVD is always real-valued; it preserves time and frequency shifts and satisfies the marginal properties. It has also some drawbacks, as the apparition of the cross-terms. This is the reason for using the Choi-Williams Distribution (CWD) where the kernel function is [12]

$$\phi(\theta,\tau)=\exp\left[-(\theta\tau)^2/\sigma^2\right] \quad (2)$$

This distribution function adopts exponential kernel to suppress the cross-term that results from the components that differ in both time and frequency centers.

The discrete WVD is defined by [13]

$$W(n,m)=\frac{1}{2N}\sum_{k=0}^{N-1}x(kT) \cdot x^*((n-k)T)\cdot\exp\left(-\frac{j\pi\cdot m\cdot(2k-n)}{N}\right) \quad (3)$$

It is informal to verify that *W(n,m)* is a periodic function of period $2N$ in both time and frequency. The last relationship shows that in the range $0\le n<2N-1, 0\le m<2N-1$, representing one period, the WVD needs only be calculated over the range $0\le n<N-1, 0\le m<N-1$, having an area of one quarter that of the complete period.

The coefficients of the time-frequency transform define an image, which will be called a Time-Frequency Image (TFI).

B. Entropy Transform

The objective of the section is to introduce a criterion to describe the time-frequency images, from content point of view, before measuring the similarities for classification purposes. The start point is the registration process of the image, as pre-processing step before classification of these images.

Common algorithms for image registration could use: translation, on *x* and *y* directions; rigid processing, which means translation plus rotation; similarity, which means translation, rotation and scaling; affine transformation, which considers translation, rotation, scaling, and shearing. The choice of one of them is based mainly on the content of the image, the sources and the number of the images which are considered for registration. Simple registration methods of the images, from the content point of view, use intensity-based registration algorithms. As complexity rises, feature-based is more indicated. Details and examples are available in many references [20] [21].

The registration time is rapidly growing from translation to affine transformation. Sometimes, for complex transforms - like affine, the registration process could diverge. This is the reason to consider new methods valid for time-frequency images – in general – and in the case of bearings, in particular. The proposed procedure considers a number of maxima which will be considered as reference, i.e., their positions remain unchanged during and after registration.

Based on the above remarks, a criterion based on the image similarity is promoted, in order to select/decide if an image of the set could be registered or not. In principle, if the similarity with the reference image is low, then the registration of that image is skipped. The similarity is evaluated based on information content with the help of Renyi entropy. Definition of the alpha order Renyi entropy from [22] is considered as

$$HR_\alpha(\mathbf{I})=-\frac{1}{1-\alpha}\log_2\iint\left(\frac{I(t,f)}{\iint I(u,v)dudv}\right)^\alpha dtdf \quad (4)$$

By discretization of this measure by setting $t=n\cdot\Delta t$ and $f=k\cdot\Delta f, n,k\in Z$ yields

$$HR_\alpha(\mathbf{I})=-\frac{1}{1-\alpha}\log_2\sum\sum\left(\frac{I[n,k]}{\sum\sum I[n',k']}\right)^\alpha + \log_2(\Delta t\cdot\Delta f) \quad (5)$$

Depending on the processed data, a number of images from the data set will be discarded, i.e., not considered for the registration step, as in [24]. The rejected images have very low similarity comparing with the prototype of the considered set and could come from some transient processes during the raw signal acquisition stage.

III. DATA DESCRIPTION AND IMAGE ANALYSIS

Data were considered for the case of faults in bearings, available from [23], and briefly described in Table I. Three types of faults are available, like F1 (Inner race), F2 (Ball) and F3 (Outer race). The case F0 means no faults. Vibration data from four sizes of the faults are available. The data set has the advantage of consistency, by considering faults from incipient/small size (0.007") to a larger one (0.028").

The sampling rate is 12,000 Hz, the motor load is 0 hp, and all data are from drive end bearing (DE). A set of four classes of patterns are considered as: C#0 – no faults, defined

 ISBN: 978-1-61208-740-5

by d0; C#1 –inner race, defined by {**d1, d6, d9, d14**}; C#2 – ball, defined by {**d2, d7, d10, d15**}; C#3-outer race defined by {**d3, d8, d11**}. The vectors **d4, d5, d12**, and **d13** correspond to other sites of the transducers.

TABLE I. DATA TEST SET

Fault size	Faults			
	F0	**F1**	**F2**	**F3**
	Free	**Inner Race**	**Ball**	**Outer Race**
0.000“	**d0**	-	-	-
0.007"	-	**d1**	**d2**	**d3, d4, d5**
0.014"	-	**d6**	**d7**	**d8**
0.021"	-	**d9**	**d10**	**d11, d12, d13**
0.028"	-	**d14**	**d15**	-

IV. THE PROPOSED METHOD

Taking into account the results obtained by other methods, e.g., [10] [11], a method to select the important features of the time-frequency image and to define a new set of features, is developed. The method considers the content of the image by taking images with a pre-defined number of peaks, e.g., 1 to 3, depending on the complexity of the image. Thus, a new image is considered and defined in terms of contours, defined by the above peaks, which will be called transformed images or contour-based images (CBI).

In the set of the four next figures, i.e., Figure 2 to Figure 5, the raw/original images and the transformed images are presented, for all four classes, i.e., *C*#1 to *C*#4. On blue background, the set of the registered images are presented. Registration is made on the set of the images obtained from the frames of each record/file, as described in [24]. These images are considered as the prototypes of the classes.

The transformed images are presented on white background. A primary analysis of these images reveals some interesting properties:

a) the common content of the images is of vertical lines, as C#1(1,6,9), C#3(3,11,12);
b) the class C#2 has a very complex patterns, for all cases (2,7,10, and 15);
c) The classes C#1 and C#3 have some strange patterns, C#1(14) and C#3(8). Keeping all these images will damage the final classification, so images with high dissimilarity are removed from processing.

In order to extract the right information/features from the transformed images, the following elements are considered in defining the necessary features for classifications:

i) the number of contours, *Nc*, as a measure of the complexity;
ii) the area of the polygons, *Ac*, as a measure of the spreading on horizontal plane;
iii) the variance of the above areas, *var(Ac)*, as a measure of the complexity;
iv) the average of the area of the polygons, $E\{A_c\}$;
v) the mean of the squared values of areas, $E\{A^2_c\}$;
vi) the Renyi entropy of transformed images, *RH*.

A vector of features could be defined by using the above variables, as

$$\mathbf{f}_i = \left[N_c \quad \sum A_c \quad \text{var}(Ac) \quad \overline{A_c} \quad \overline{A_c^2} \quad RH \right], \quad i = 1,2,3,4 \tag{6}$$

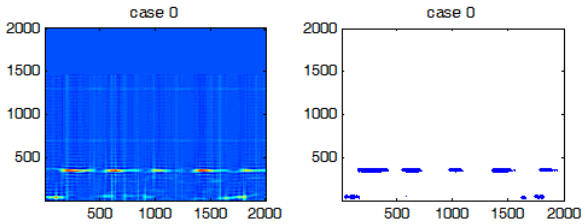


Figure 2. Class #0. Original and transformed image.

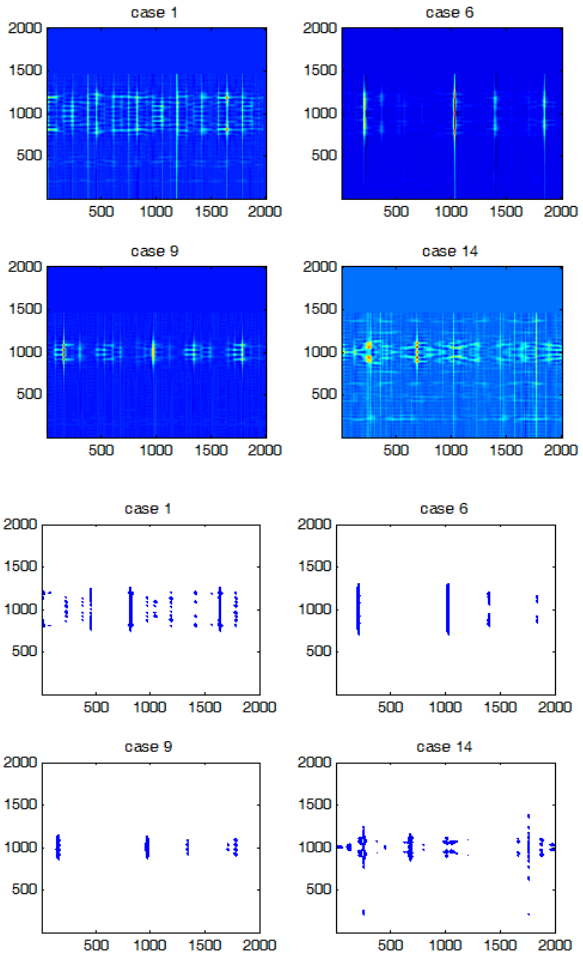


Figure 3. Class #1. Original and transformed images.

For each data vector **d** from a class, the vector $\mathbf{f}_i$ is evaluated, and matrix of features is obtained for each class, as

$$\mathbf{F}_j = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad ... \quad \mathbf{f}_4], j = \overline{1,4} \qquad (7)$$

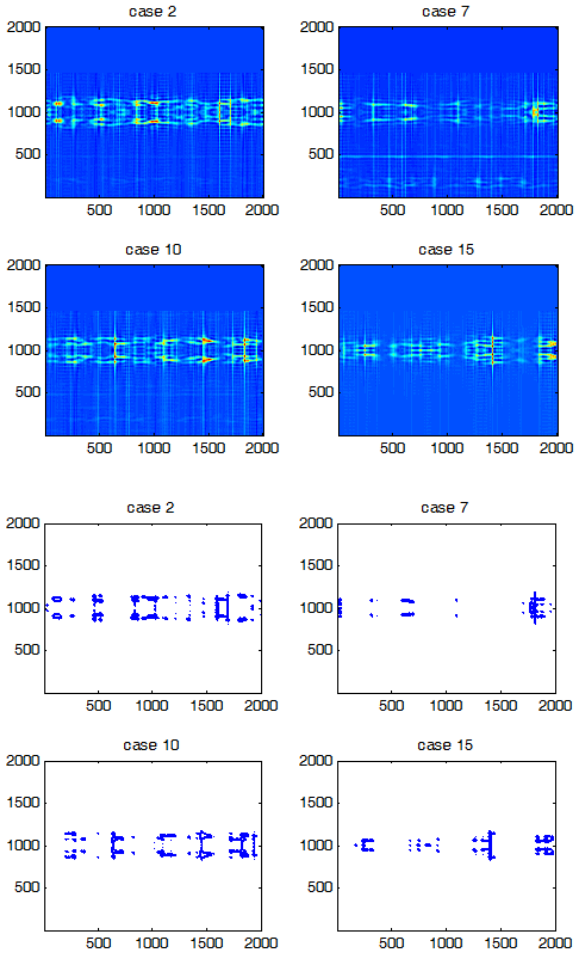


Figure 4. Class #2. Original and transformed images

The effect of the features is estimated by a general discriminant matrix

$$D(k, j) = \sum_{k=0}^{3} \sum_{j=0}^{3} (\mathbf{F}_k - \mathbf{F}_j)^2, k, j = \overline{1,4} \qquad (8)$$

or, by considering only the distinct classes, by the discriminant functions

$$D_m = E\{D(k, j) \mid k \neq j, k > j, k, j = \overline{1,4}\} \qquad (9)$$

or

$$D_v = \text{var}\{D(k, j) \mid k \neq j, k > j, k, j = \overline{1,4}\} \qquad (10)$$

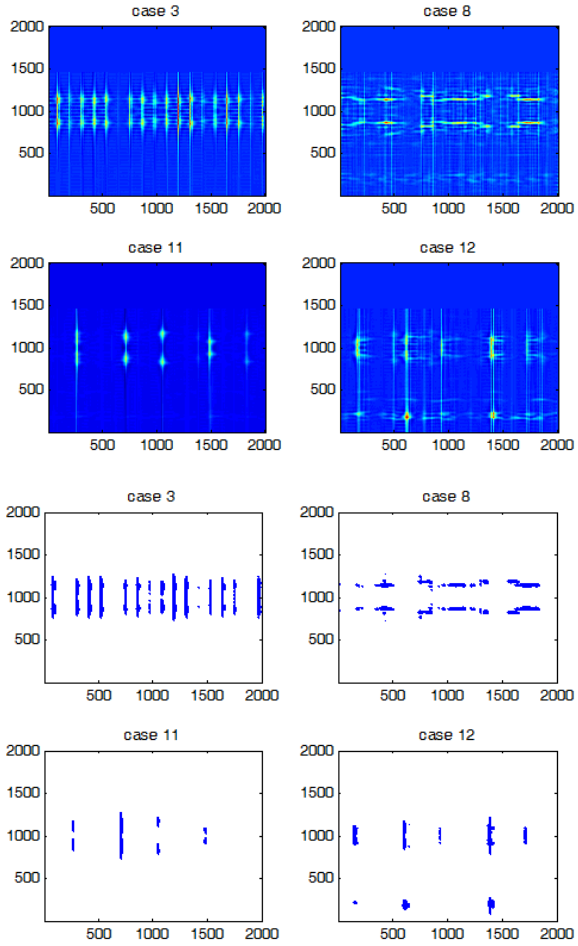


Figure 5. Class #3. Original and transformed images.

V. RESULTS OF THE EXPERIMENTS

From various experiments, the best results are presented in Table II. Figure 6 presents the features space, composed from the first three normalized features, and the values of the discriminant functions. The best structure of the feature vector, as result of the maximum variance, ***Dv***, is offered on the first line with a value of 6.23.

TABLE II. DISCRIMINANT VALUES

No	*Feature vector*	*Dm*	*Dv*
1	$[\mathbf{N_c} \quad \overline{\mathbf{A_c}} \quad \overline{\mathbf{A_c^2}} \quad \mathbf{RH}]$	3.88	6.23
2	$[\mathbf{N_c} \quad \overline{\mathbf{A_c}} \quad \text{var}(Ac) \quad \mathbf{RH}]$	3.79	5.64
3	$[\mathbf{N_c} \quad \overline{\mathbf{A_c}} \quad \text{var}(Ac) \quad \sum Ac]$	4.34	4.33

 ISBN: 978-1-61208-740-5

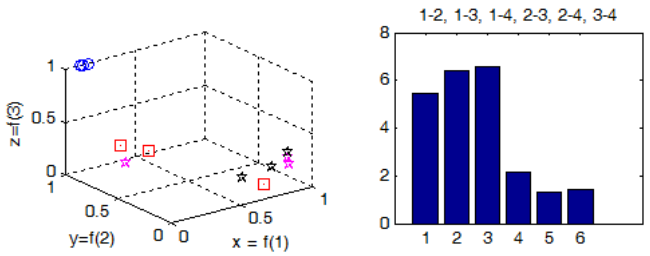


Figure 6. Feature space and the discriminant values

The unnormalized values are presented in the following. A column is used for a class. The first column is for class #0, i.e., free of faults. The next column are for classes 1, 2 and 3. For class C#0, because only one vector is available, the next three vectors are considered identically with the first one.

J1m =

23	150	179	204
23	25	86	138
23	16	173	12
23	143	48	69

J2m =

896.10	94.20	184.68	200.69
896.10	405.73	143.06	181.46
896.10	539.86	191.68	465.38
896.10	128.87	457.87	180.73

J3m = 1.0e+06 *

3.1544	0.2217	0.4132	0.5623
3.1544	1.3166	0.4640	0.3623
3.1544	1.3451	0.7012	0.9783
3.1544	0.5633	1.3616	0.4031

J4m =

4.4006	6.0824	5.9316	5.6459
4.4006	4.2532	5.8992	6.1837
4.4006	4.6008	5.7734	4.0958
4.4006	6.2462	5.3046	5.4525

Next, the features are organized in classes, one column for a parameter, and one line for a data vector of the same class. In this way, the parametric matrices from PC1 to PC4 are obtained and used in the computation of the discriminant functions.

PC1 =

0.1127	1.0000	1.0000	0.7235
0.1667	1.0000	1.0000	0.7116
0.1329	1.0000	1.0000	0.7622
0.1608	1.0000	1.0000	0.7045

PC2 =

0.7353	0.1051	0.0703	1.0000
0.1812	0.4528	0.4174	0.6878
0.0925	0.6025	0.4264	0.7969
1.0000	0.1438	0.1786	1.0000

PC3 =

0.8775	0.2061	0.1310	0.9752
0.6232	0.1596	0.1471	0.9540
1.0000	0.2139	0.2223	1.0000
0.3357	0.5110	0.4317	0.8492

PC4 =

1.0000	0.2240	0.1783	0.9282
1.0000	0.2025	0.1148	1.0000
0.0694	0.5193	0.3101	0.7094
0.4825	0.2017	0.1278	0.8729

VI. CONCLUSION

The objective of the paper was to find a solution to the classification problem, based on time-frequency images, which are quite modest with data of faulty bearings. The previous solutions have implemented feature selection and extraction based on the parameters of the first component of the processed images.

At this stage, two major improvements could be considered. The first one is based on image analysis and registration. Analysis of the images to be registered shows that it is possible to have images which are very different compared to others from the same class. These must be rejected from the registration process. The second direction, which was promoted in this paper, uses a transformed image that contains contours of the time-frequency image. The selection of the features is based on the statistical moments, as average, variance and squared average values of the areas of the contours. Information based features is also used, by promotion of the Renyi entropy.

The preliminary results show an improvement in the feature space, in the sense of clustering, and the possibility to obtain right classification results based on proposed dissimilarities functions.

In the nearest future a performance comparison of the proposed method with some classical methods based on similarity computation will be considered.

ACKNOWLEDGMENT

The work was partly supported by the Romanian Council for Research (UEFISCDI) under Grant 224/2014, “Experimental model for change detection and diagnosis of vibrational processes using advanced measuring and analysis techniques model-based (VIBROCHANGE)”.

REFERENCES

[1] J. H. Williams, A. Davies, and P. R. Drake (Eds), Condition-based Maintenance and Machine Diagnostics, Springer, 1994.

[2] L. Fedele, Methodologies and Techniques for Advanced Maintenance, Springer, 2011.

[3] M. Bengtsson, E. Olsson, P. Funk, and M. Jackson, "Technical Design of Condition Based Maintenance System", Proceedings of the 8th Congress, University of Tennessee – Maintenance and Reliability Center, Knoxville, USA, May 2nd – 5th, 2004, Paper III.

[4] K. Shin and J. K. Hammond, Fundamental of Signal Processing for Sound and Vibration Eng., John Wiley & Sons Ltd, 2008.

[5] D. M. Yang, A. F. Stronach, and P. MacConnell, "The Application of Advanced Signal Processing Techniques to Induction Motor Bearing Condition Diagnosis", P. Meccanica, Springer, 2003, vol 38(2), pp 297-308.

[6] D. Mironovs and A. Mironov, "Vibration based signal processing algorithm for modal characteristics change assessment", AIP Conference Proceedings 2029, 020043, 2018, pp.1-10.

[7] B. Boashash (Ed), Time-Frequency Signal Analysis and Processing, Elsevier, 2016.

[8] L. Stankovic, M. Dakovic, and T. Thayaparan, Time-Frequency Signal Analysis with Applications, Artech House, 2014.

[9] M. Boufenar, S. Rechak, and M. Rezig, "Time-Frequency Analysis Techniques Review and Their Application on Roller Bearings Prognostics". In: Fakhfakh T., Bartelmus W., Chaari F., Zimroz R., Haddar M. (eds) Condition Monitoring of Machinery in Non-Stationary Operations. Springer, Berlin, Heidelberg, 2012.

[10] D. Aiordachioaie and Th. D. Popescu, "Change Detection by Feature Extraction and Processing from Time-Frequency Images", IEEE 10th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp.1-7.

[11] D. Aiordachioaie, N. Nistor, and M. Andrei, "Change Detection in Time-Frequency Images by Feature Processing in Compressed Spaces", IEEE 24th Int. Symp. for Design and Technology in Electronic Packaging (SIITME), Iași, Romania, 2018, pp.181-186 .

[12] E. Sejdic, I. Djurovic, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances", Digital Signal Processing, 19, 2009, pp. 153-183.

[13] Q. Zhu, Y. Wang, and G. Shen, "Comparison and Application of Time-Frequency Analysis Methods for Nonstationary Signal Processing". In: Lin S., Huang X. (eds) Advanced Research on Computer Education, Simulation and Modeling. CESM 2011. Communications in Computer and Information Science, vol 175. Springer, Berlin, Heidelberg, 2011.

[14] M. Karjalainen and V. Pulkki, Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics, John Wiley & Sons, 2015.

[15] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques", MSSP, vol. 25 (6), 2011, pp. 2083-2101.

[16] M. Akay, Time Frequency and Wavelets in Biomedical Signal Processing, Wiley-IEEE Press, 1997.

[17] P. D. McFadden and W. Wang, Time-Frequency Domain Analysis of Vibration Signals for Machinery Diagnostics. (I) Introduction to the Wigner-Ville Distribution, University of Oxford, Report OUEL 1859/92, 1990.

[18] L. Debnath, The Wigner-Ville Distribution and Time-Frequency Signal Analysis. In: Wavelet Transforms and Their Applications, Birkhäuser, Boston, MA, 2002.

[19] L. Cohen, "Time-Frequency distributions-A review", Proc. IEEE, Vol. 77, pp. 941-981, July 1989.

[20] G. A. Ardeshir, Image Registration Principles, Tools and Methods, Springer, 2012.

[21] I. N. Bankman (Ed.), Handbook of Medical Image Processing and Analysis, Elsevier, 2009.

[22] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring Time–Frequency Information Content Using the Rényi Entropies", IEEE Transactions on Information Theory, vol. 47(4), 2001, pp. 1391-1409.

[23] Case Western Reserve University, *Bearing Data Center*, At: http://csegroups.case.edu/bearingdatacenter/home, Retrieved: 06, 2019.

[24] D. Aiordachioaie and S. M. Pavel, "Qualitative Analysis of the Time-Frequency Images of Vibrations in Faulty Bearings", IEEE 11th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI), Pitesti, Romania, 2019, pp. 1-6.

 ISBN: 978-1-61208-740-5

Fault Detection using NLMS Adaptive Filtering for a Wastewater Treatment Process

Mihaela Miron
Faculty of Automation, Computers, Electrical Engineering and Electronics
Dunarea de Jos University of Galati
Galati, Romania
Mihaela.Miron @ugal.ro

Anisia Culea-Florescu
Faculty of Automation, Computers, Electrical Engineering and Electronics
Dunarea de Jos University of Galati
Galati, Romania
Anisia.Florescu @ugal.ro

Mihai Culea
Faculty of Automation, Computers, Electrical Engineering and Electronics
Dunarea de Jos University of Galati
Galati, Romania
Mihai.Culea@ugal.ro

***Abstract*—In this paper, a method based on adaptive filtering is proposed for actuator, sensor and toxicity faults detection in a biological wastewater treatment process. Improving water quality in such treatment plants is an important and growing problem so monitoring performance and conditions, optimization and fault diagnosis for biotechnological processes are as important as fault detection. Such detection is performed here using state-parameter estimation where the detection algorithm compares the outputs of an analytical model with those estimated by the normalized least mean square adaptive filter in order to calculate the residual value for each output of the process. Numerical examples are presented in order to illustrate the performance of the proposed method.**

Keywords- wastewater treatment process; adaptive Kalman filter; fault detection; process diagnosis.

I. INTRODUCTION

Improving the quality of waters in the wastewater treatments plants has become more and more important due to the fact that the population is continuously increasing. The environment and especially general population health depends on it. As a result, monitoring performances and conditions, optimization and fault diagnosis for biotechnological processes are novelty topics in the current scientific research.

So far in literature several methods for fault detection and isolation of different types of faults have been proposed [1]-[10]; however, there are limitations in the case of monitoring complex and dynamical processes as Wastewater Treatment Processes (WWTPs).

The main challenge of this work was to propose a simple and fast method to detect different types of faults which can occur in the wastewater treatment process. Subsequently, in order to detect actuator, sensor and biological faults, this paper proposes a method based on adaptive filtering which is monitoring the changes in residuals of the model parameters.

Many adaptive algorithms have been developed over time based on two different approaches, namely the statistical approach and the deterministic approach, each with specific advantages and disadvantages [11] [12]. This paper presents a fault detection algorithm which uses one of the popular Normalized Least Mean Square (NLMS) algorithm for state estimation. This approach proves to be fit for WTTPs since adaptive filtering algorithm has to address the classical compromise between fast convergence/tracking and low misadjustment [13] [14].

The paper structure is the following: Section 2 presents the analytical model of the wastewater treatment process; Section 3 presents the NLMS adaptive filter used to estimate the WWTP outputs; Section 4 presents the fault detection approach and the residuals obtained for sensor, actuator or toxicity faults; Section 5 is dedicated to results and discussion about fault detection performance and the last one highlights the paper's conclusions.

II. THE WASTEWATER TREATMENT PROCESS

The mathematical model of the Wastewater Treatment Process (WWTP) on which this study is based is described by the following equations [7]-[10].

$$\frac{dX}{dt} = \big(\mu(t) - \mu_s(t)\big)X(t) - D(t)(1+r)X(t) + \\ + rD(t)X_r(t) \quad (1)$$

with:

$$D = \frac{F_{in}}{V} \quad (2)$$

where:
$X(t)$ – biomass concentration,
$\mu(t)$ – specific growth rate,
$\mu s(t)$ – decay coefficient for biomass,
$D(t)$ – dilution rate is : $D = \frac{F_{in}}{V}$
r – recirculating rate,
$X_r(t)$ – recirculated biomass concentration,
F_{in} – influent flow,
V – bioreactor volume,

$$\frac{dS}{dt} = -\frac{\mu(t)-\mu_s(t)}{Y}X(t) - D(t)(1+r)S(t) + D(t)S_{in} \quad (3)$$

where:
$S(t)$ – substrate concentration,
Y – yield coefficient,

 ISBN: 978-1-61208-740-5

μ_{max} – maximum specific growth rate,
S_{in} – influent substrate concentration,

$$\frac{dDO}{dt} = -\frac{(1-Y)(\mu(t)-\mu_s(t))X(t)}{Y} \cdot 10^3 - D(t)(1+r)DO(t) + \\ +60\alpha W(t)\left(DO_{sat} - DO(t)\right) + D(t)DO_{in} \quad (4)$$

where:
$DO(t)$ – dissolved oxygen concentration,
Do_{ing} – influent dissolved oxygen concentration,
DO_{sat} – saturation value of dissolved oxygen,
$W(t)$ – aeration rate,
α – oxygen transfer rate,

$$\frac{dX_r}{dt} = D_s(t)(1+r)X(t) - D_s(t)(\beta+r)X_r(t) - \\ -0.5D_s(t)(1+\beta)X_r(t) \quad (5)$$

where: D_s is the dilution rate of the sludge

$$D_s = \frac{D \cdot V}{V_s} \quad (6)$$

with V_s – sludge volume and β the rate of the sludge in excess.

$$\mu(t) = \mu_{max}\frac{S(t)}{K_s+S(t)} \cdot \frac{DO(t)}{K_{DO}+DO(t)} \quad (7)$$

where: K_s is the saturation constant of the substrate and K_{DO} is the saturation constant of dissolved oxygen.

III. Adaptive Filter Using Normalized Least Square (NLMS) Algorithm

As a general note, the adaptive filters are self-adjustable systems which adapt to various conditions and situations therefore are used in a wide range of areas. The common trait of the applications where the adaptive filters provide a good solution is based on minimizing the mean squared error between the filter output and a desired signal.

The filter's parameters are updated using a set of measured data which are used as input for the adaptive filtering algorithm. The algorithm adjusts filter's parameters so that the difference between the input and the output is minimized either statistically or deterministically and these approaches give us applications for modelling, reverse modelling, prediction or interference cancelling.

This paper uses the adaptive filter as a predictor and, in this context, it estimates the current value of the signal $\hat{x}(n)$ based on the past values $x(n-1), x(n-2), \dots, x(n-N)$. Using the linear combinations of N successive samples of the input signal, the algorithm tries to estimate the output of the desired signal $d(n)$, which is a forward version of the adaptive filter input signal.

The filter is assumed to be finite impulse response (FIR) filter of length L with coefficients $\boldsymbol{w} = (w_1 \dots w_L)^T$, input signal $\mathbf{x}(n-1) = (x(n-1) \dots x(n-N))^T$ and output defined by:

$$\hat{x}(n|\mathbf{X}_{n-1}) = \sum_{k=1}^{M} w_k^* \; x(n-1). \quad (8)$$

where $\mathbf{X}_{n-1}$ is a N-dimension for input samples.

The desired signal is:

$$d(n) = x(n) \quad (9)$$

and predicted error is defined by equation:

$$e(n) = x(n) - \hat{x}(n|\mathbf{X}_{n-1}). \quad (10)$$

By minimizing error $e(n)$, an optimal predictive signal input is made.

If NLMS algorithm is used, it can be expressed by the following equation [15]:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \text{sgn}(e(n))\, \mathbf{x}(n-1) \quad (11)$$

where μ is the adaptation [15].

IV. Fault Detection Approach

A model-based fault detection approach is proposed in this paper in order to identify anomalies which can occur in the WWTP process. The detection algorithm uses the residual values of each output and by comparing it with a threshold value it can be established when a fault occurs in the system.

As in other studies [9]-[12], the residual $R(k)$ is obtained as a difference between the estimated output $\hat{y}_p(i)$ of the process and the output of the analytical model, $y_m(i)$

$$R(k) = \frac{1}{N}\sum_{i=k-N+1}^{k}(\hat{y}_p(i) - y_m(i))^2 \quad (12)$$

where:
N = number of samples,
R(k) = the value of the residue over the last N samples,
$\hat{y}_p(i)$ - estimated output of the process by the adaptive filter,
$y_m(i)$ – output of the analytical model.

The detection efficiency is obtained through the binary signal E which is generated by using a OR function described below:

$$E(i) = \\ \begin{cases} 1 \; if R(i) > \varepsilon_X || R(i) > \varepsilon_S || R(i) > \varepsilon_{DO} || R(i) > \varepsilon_{X_r} \\ 0 \; else \end{cases} \quad (13)$$

This function detects a fault even if the selected threshold is not exceeded by the residual value on a particular output. It is necessary for the residuals to exceed at least one time one of the four thresholds in order to correctly detect the presence of a fault in the process.

A. Fault detection scheme

As presented in [7]-[10] the input parameter DO_{in} is considered to be constant. The fault detection scheme is shown in Figure 1. A model of the supervised process, in this case an analytical model, is used to provide the same

evolution of the output as the process outputs if the same values of the inputs are applied. The model outputs are compared with the estimated outputs of the NLMS adaptive filter in order to generate the residuals.

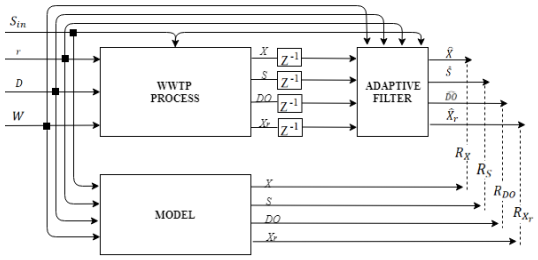


Figure 1. The fault detection scheme

B. *Fault detection parameters*

As several other studies presented [7]-[10], the algorithm's decision parameters are:

- sensibility threshold value ε which is compared with the residual value R in order to establish if the conditions of a fault occurrence are met.

- number of samples N, on which the residual value, R is obtained.

C. *Method validation by numerical simulations*

Several experiments were carried on in order to find the optimal values for the parameters N and ε and to achieve the best performances of the detection approach.

1) *Fault of the recirculation pump*

The deviations caused by the recirculation pump fault can be seen in Figure 2. When this actuator fault occurs in the process, the recirculation rate value becomes zero ($r = 0$ when $N_t \in [3500, 3899]$). The residuals are obtained for each output of the system when $N = 10, N = 20, N = 40$.

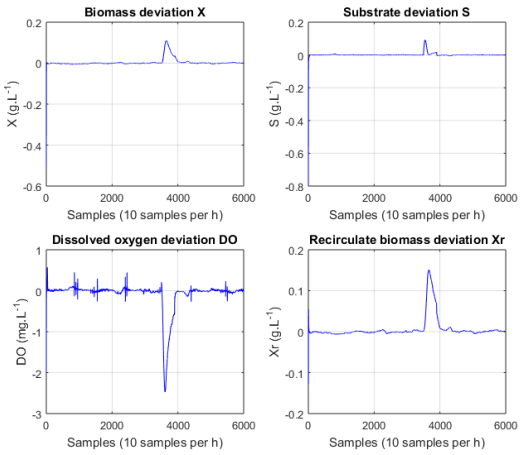


Figure 2. The outputs, X, S, DO, X_r deviations caused by the recirculation pump fault ($X = 0$ over 3500 to 3899 samples)

Figures 3 – 5 show the residuals obtained in both cases, when the system operates in normal conditions (the residual values are close to zero) and when the recirculation pump fails around sample no. 3500 (the residual value is increasing due to process output deviation caused by the pump malfunction).

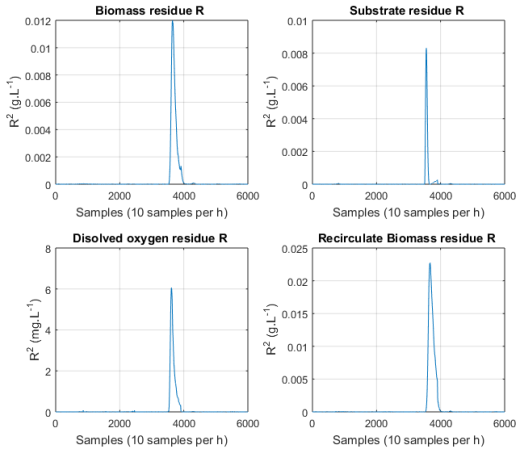


Figure 3. The residue, R for partial fault of the recirculation pump when $N = 10$

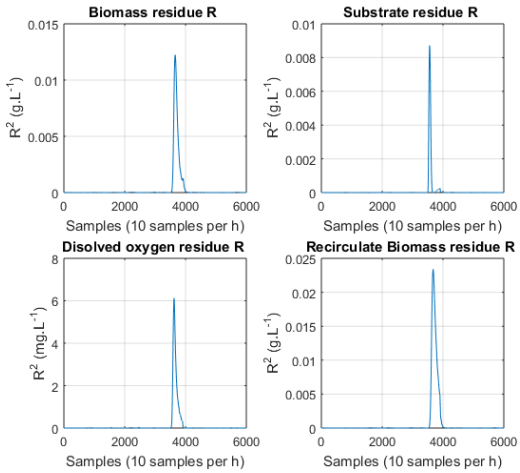


Figure 4. The residue, R for partial fault of the recirculation pump when $N = 20$

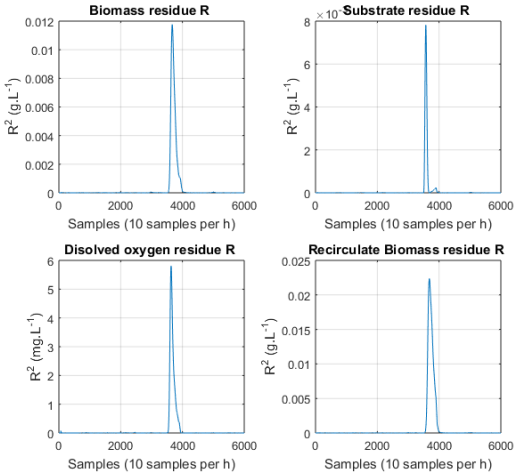


Figure 5. The residue, R for partial fault of the recirculation pump when $N = 40$

2) Fault of the biomass sensor

In general, a WWTP plant is equipped with many important sensors for monitoring the process performances and conditions [1]. Here, a fault of the biomass sensor is simulated over 400 samples. The output deviations caused by this fault are shown in Figure 6. As previously, the residuals R are generated for each output of the process when $N = 10, N = 20, N = 40$.

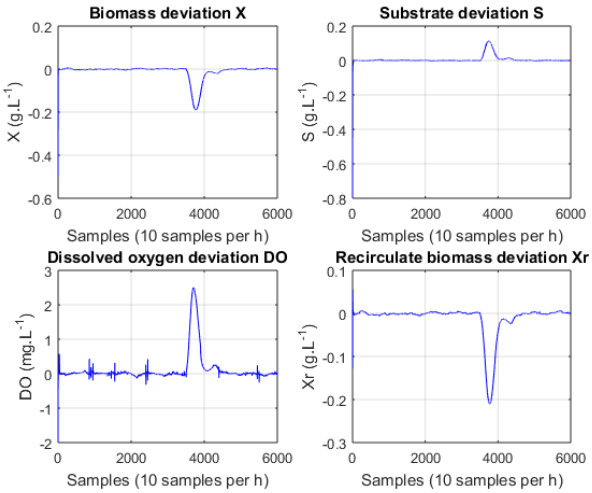


Figure 6. The outputs, X, S, DO, X_r deviations caused by the biomass sensor fault ($X = 0$ over 3500 to 3899 samples)

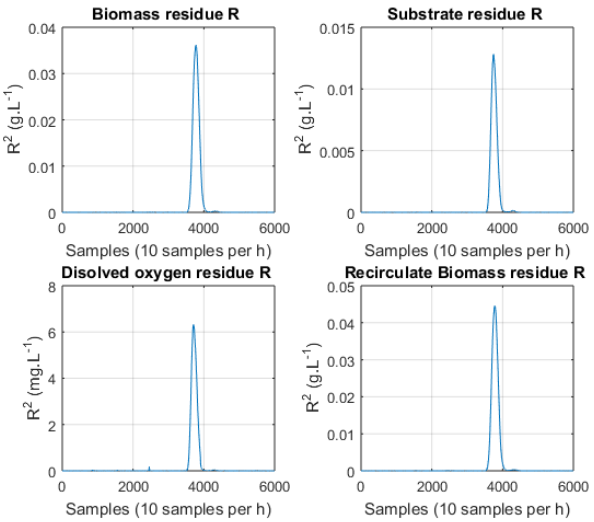


Figure 7. The residue, R for toxicity fault when $N = 10$

From Figures 7 – 9, it can be observed that the residual value R is big compared with the process and the measurement noise. So, in this case, tweaking N value to a smaller size will avoid generating false alarms.

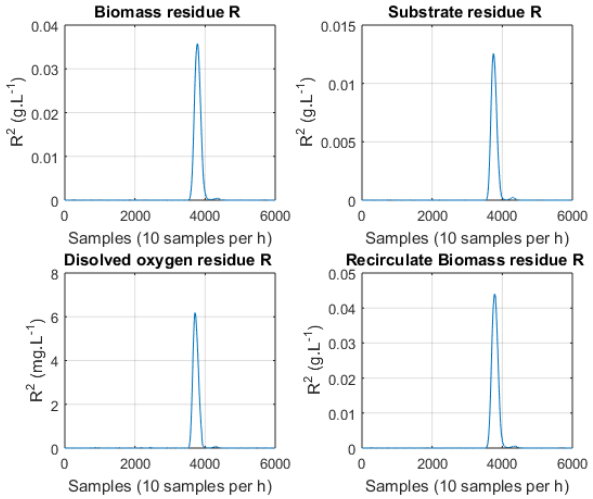


Figure 8. The residue, R for toxicity fault when $N = 20$

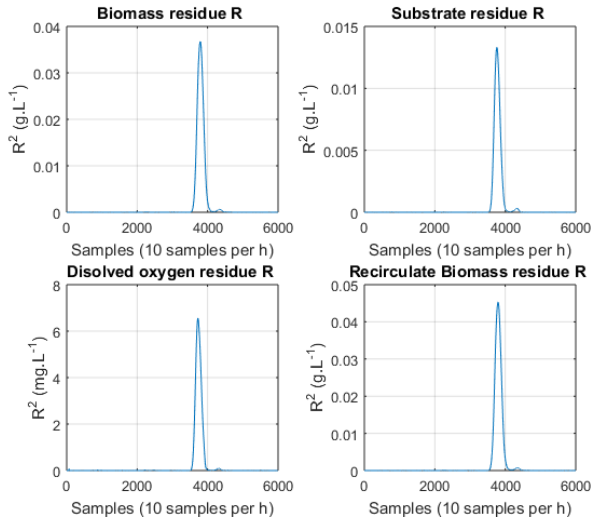


Figure 9. The residue, R for toxicity fault when $N = 40$

3) Toxicity shock fault

A fault caused by a toxic shock suffered by the microorganism's culture, presented in Figure 10, was simulated by reducing the value of the maximum specific growth rate μ_{max} by half, over $N_t \in [3500, 3899]$). Figure 11 – 13, shows the residual values at different iterations of parameter N ($N = 10, N = 20$ and $N = 40$).

Simulation results show that depending on the value of N, the measurement and process noise is reduced, which could cause a lower value of the threshold and a possible increasing of the detection time. Therefore, a compromise must be made when choosing the detection parameters in order to achieve good results.

 ISBN: 978-1-61208-740-5

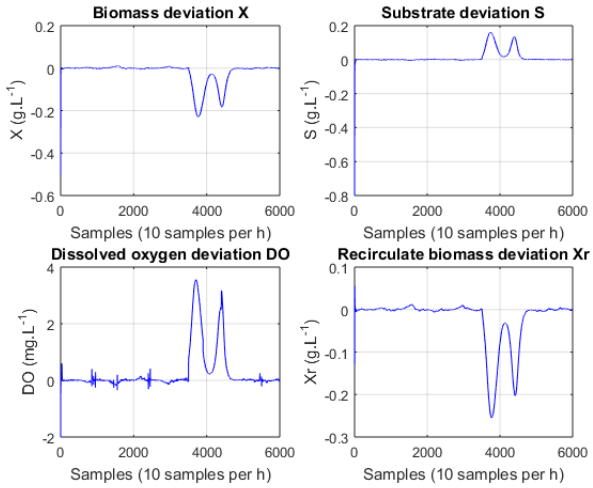


Figure 10. The outputs, X, S, DO, X_r deviations caused by the toxicity fault ($\mu_{max}/2$ over 3500 to 3899 samples)

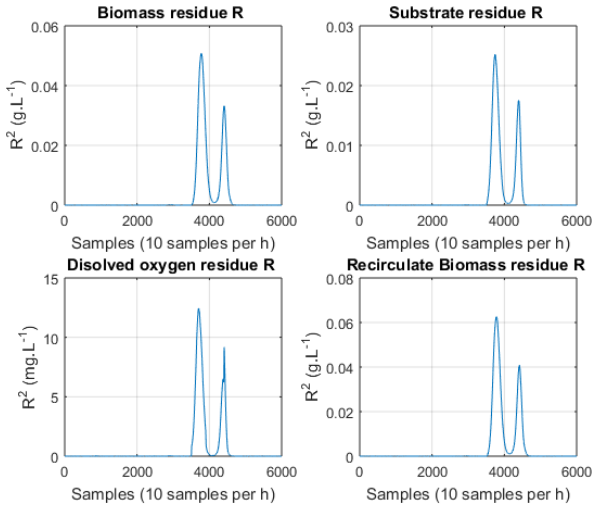


Figure 11. The residue, R for toxicity fault when $N = 10$

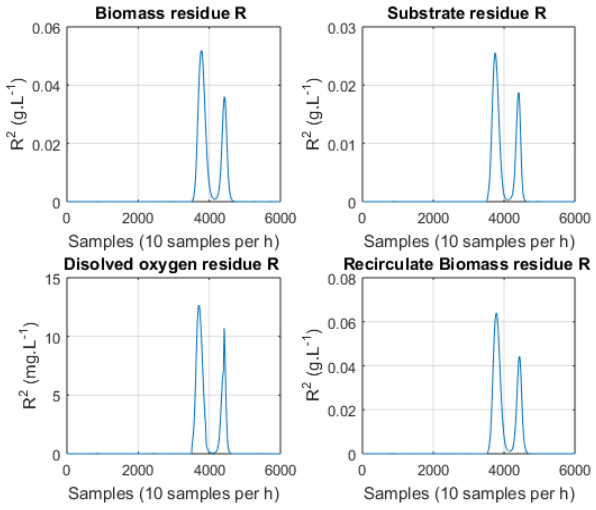


Figure 12. The residue, R for toxicity fault when $N = 20$

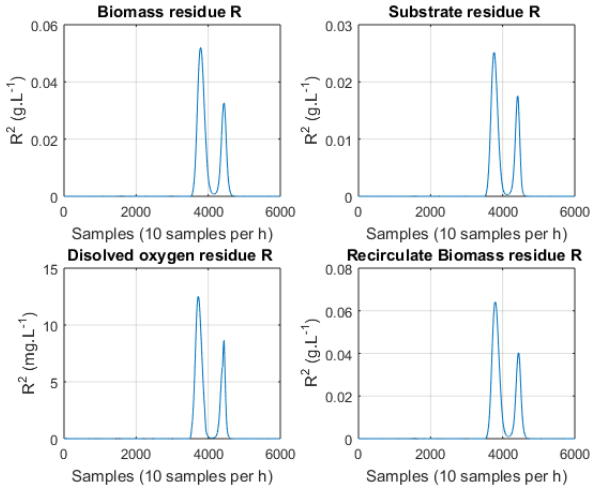


Figure 13. The residue, R for toxicity fault when $N = 40$

V. RESULTS AND DISCUSSION

In all the previously presented cases of simulated faults the value of N is set to 10, which corresponds to the criteria of choosing the detection parameters described in section IV. Also the sensibility thresholds for each output are: $\varepsilon_X = 4 \cdot 10^{-4}$, $\varepsilon_S = 2 \cdot 10^{-4}$, $\varepsilon_{DO} = 0.1$, $\varepsilon_{X_r} = 5.5 \cdot 10^{-4}$.

Further, the detection efficiency E (Figures 14 – 17) is obtained for all types of faults analyzed in the previous section: recirculation pump fault, biomass sensor fault and toxicity fault.

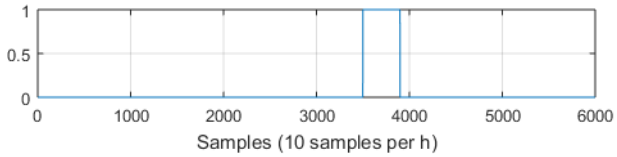


Figure 14. Fault simulated over $N_t \in [3500, 3899]$

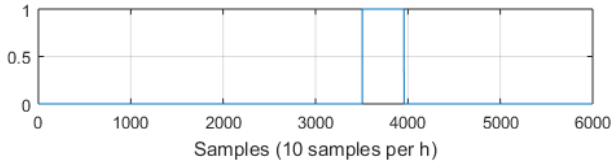


Figure 15. Alarm signal for recirculation pump fault detection (detection over $N_t \in [3506, 3958]$, after ~36 min)

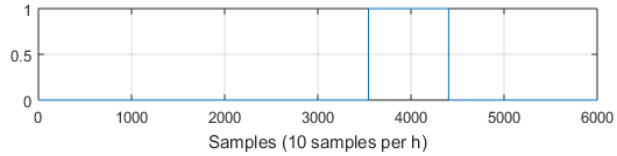


Figure 16. Alarm signal for biomass sensor fault detection (detection over $N_t \in [3543, 4405]$, after ~$4h$)

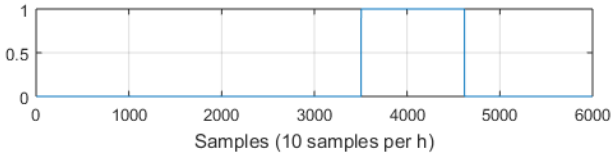


Figure 17. Alarm signal for toxicity fault detection (detection over $N_t \in [3504, 4619]$, after ~24 min)

The best detection time of the proposed algorithm was obtained in the case of toxicity fault, after approximately 24 minutes.

The fault detection algorithm simulation was run on a computer having the following specifications: Intel Core i3-6100U with 2.30GHz, 4 GB RAM memory and 500GB SSD. The detection performances are displayed in approximately 30 seconds.

VI. CONCLUSIONS

This paper proposes a model-based fault detection method for a wastewater treatment process. The detection algorithm compares the outputs of the analytical model with the ones estimated by the NLMS adaptive filter in order to calculate the residual value for each output of the process.

The alarm decision, occurring when a fault appears in the system, is enabled based on the detection parameters values (threshold ε and number of samples N). The results obtained are promising when compared with other studies [1] [2] and [8] [9] presenting aspects of sensor and actuator faults detection in WWTP.

Moreover, this paper analyses the case of a toxicity shock fault detection that could damage the microorganism's culture and cause erraticism in these types of biotechnological processes.

REFERENCES

[1] C. Lee, S. W. Choi, and I. B Lee, "Sensor fault diagnosis in a wastewater treatment process", Water Science and Technology, vol. 53, pp. 251–257, 2006.

[2] F. Nejjari, V. Puig and L. Giancristofaro, S. Koehler, "Extended Luenberger Observer-Based Fault Detection for an Activated Sludge Process", Proceedings of the 17th World Congress The International Federation of Automatic Control, pp. 9725-9730, July 6-11, 2008.

[3] I. Baklouti, M. Mansouri, H. Nounou, M. Ben Slima and A. Ben Hamida, "Fault detection in a wastewater treatment plant," 2017 International Conference on Advanced echnologies for Signal and Image Processing (ATSIP), Fez, 2017, pp. 1-5.

[4] I. Baklouti, M. Mansouri, A. Ben Hamida, H. Nounou and M. Nounou, "Novel Fault Detection Approach of Biological Wastewater Treatment Plants," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 2669-2674.

[5] L. Corominas, K. Villez, D. Aguado, L. Rieger, C. Rosén et al., "Performance evaluation of fault detection methods for Wastewater Treatment Processes", Biotechnology and bioengineering, vol. 108, no. 2, pp. 333-344, 2011.

[6] I. Baklouti, M. Mansouri, A. B. Hamida, H. Nounou and M. Nounou, "Monitoring of wastewater treatment plants using improved univariate statistical technique", Process Safety and Environmental Protection, vol. 116, pp. 287-300, 2018.

[7] M. Miron, L. Frangu, G. Ifrim and S. Caraman, "Modeling of a Wastewater Treatment Process Using Neural Networks", 20th International Conference on System Theory, Control and Computing - ICSTCC 2016, October 13-15, Sinaia, Romania, pp. 210**,** 2016.

[8] M. Miron, L. Frangu and S. Caraman „Actuator fault detection using extended Kalman filter for a wastewater treatment process", 21st International Conference on System Theory, Control and Computing, October 19 - 21, Sinaia, Romania, pp. 583-588, 2017.

[9] M. Miron, L. Frangu and S. Caraman, "Fault detection method for a wastewater treatment process based on a neural model", 5th International Symposium on Electrical and Electronics Engineering (ISEEE), October 20 – 22, Galati, Romania, pp. 1-6, 2017.

[10] M. Miron, L. Frangu and S. Caraman, "Artificial neural network approach for fault recognition in a wastewater treatment process", 22nd International Conference on System Theory, Control and Computing, October 10 - 12, Sinaia, Romania, pp. 634-639, 2018.

[11] S. Haykin, Adaptive Filter Theory , 4th edn. (Upper Saddle River, NJ, Prentice-Hall, 2002)

[12] A. H. Sayed, Fundamentals of Adaptive Filtering. Hoboken, NJ: Wiley, 2003

[13] Q. Chai, B. Furenes and B. Lie, "Comparison of state estimation techniques, applied to a biological wastewater treatment process, IFAC Proceedings Volumes, Volume 40, Issue 4, 2007,pp. 357-362.

[14] J. Dhiman, S. Ahmad and K. Gulia, "Comparison between Adaptive filter Algorithms (LMS, NLMS and RLS)", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 2, Issue 5, May 2013, pp. 1100-1103.

[15] J. Benesty, C. Paleologu and S. Ciochina, On Regularization in Adaptive Filtering. Ieee Transactions nn Audio, Speech, and Language Processing, vol. 19, no. 6, pp. 1734-1742, august 2011

 ISBN: 978-1-61208-740-5

A Battery Charging Smart System using a Power Management Algorithm and Adaptive Impedance

Nicusor Nistor, Laurentiu Baicu and Bogdan Dumitrascu.
Department of Electronics and Telecommunications,
"Dunarea de Jos" University of Galati
47 Domneasca Street, 800008 Galati, Romania
email: nicusor.nistor@ugal.ro

***Abstract*—The paper proposes an original method experimented by the authors, in order to optimize the charging process for rechargeable batteries. The method involves: testing, adapting and implementing the on board charging function, depending on each type of battery used. First of all, the method proposes an algorithm for the identification of the battery internal resistance, which depends on the wear and tear of the battery. Also, the battery charging dynamics will be created according to its wear status. This is of particular importance in the maximum power transfer especially for batteries with high specific power. Secondly, each battery must be charged taking into account the type of chemical reaction used by the manufacturer in the battery construction process. For this, we will test the dynamics of the discharging in a short time interval and adapt the load to each type of battery differently. The entire process will be implemented and controlled by a microcontroller. The proposed solution is not a new approach, the subject being of great interest especially in the automotive industry, but two of distinct solutions will be implemented cumulatively during the same optimization process.**

Keywords-smart battery charging; power management; adaptive impedance; internal resistance; dynamic charging function; microcontroller based.

I. INTRODUCTION

In this paper, a type of battery charger is proposed that uses an initial battery test to determine the charging function for the charging battery, each time a battery is connected. Efficient charging and dynamic change of the resistance introduced by the power generator during the charging process is considered.

Although the subject was closely addressed by a multitude of researchers in the field, the topic being current and interesting in the field of industrial technology and home automation [1], [2] and [4]. This research consists of proposing the realization of a battery charger made in the original variant in which the maximum power transfer and the reduction of the losses to the load is taken into account.

The steps that have been implemented in carrying out the intelligent charging process are the following: connecting a new battery to the circuit terminals, testing the nominal voltage and its internal resistance by means of a controlled discharge for a short period of time, calculating an onboard current-voltage actuation function such that in small areas the generator's resistance should be adaptive and equal to the internal resistance of the battery, and the implementation of this function in real time during charging. The role of the initial testing of battery in efficient charging process was mentioned in [3], [5].

The entire process is implemented in 8-bit process microcontroller with analog read and writes hardware capabilities and precise time interval implementation. Additionally, two functional blocks have been created, the adaptive driver circuit and discharging and testing battery, electronic blocks that realize the interface of the process with the microcontroller. The two blocks play the role in implementation of the entire algorithm. The physical and mathematical aspects of the problem were taken into account, equations describing the maximum value of the power transfer theorem and the inversion with respect to the composition of the real variable functions.

In this paper, we have explained a general case of the proposed solution with details of the software model, experimental aspects of the solution including graphs of variations of the current, voltage and resistance. In section II the solution proposed by us was described, following that in section III some experimental and graphical results will be highlighted. In section IV the conclusions are presented.

II. PROPOSED SOLUTION

A schematic block of the proposed smart battery charger is presented in Fig. 1. The proposed solution consists of a power source generator providing power to an adaptive control circuit driver that charges the battery; the microcontroller applies control signals to the adaptive control driver circuit, based on the charge feedback and the results obtained from the discharging and testing battery circuit.

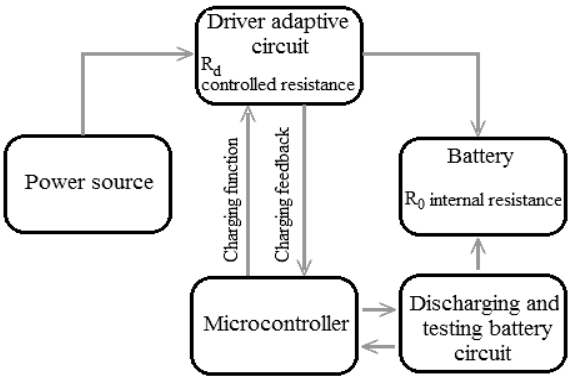


Figure 1. Block schematic of the proposed solution

From a theoretical point of view, we consider that a battery charger has to be energy efficient among other functions. We consider the case where we need to charge multiple battery elements with high energy density, like car batteries, where the energy transfer to the battery involves additional costs when the charging process has a low efficiency.

If we consider a simple circuit with generator, rechargeable battery and a measurement resistor (pull-down resistor), we can get the formula for the maximum power transfer while charging in the following form:

$$P_{Batt} = \frac{U_{R_0}^2}{R_0} = \frac{\left(U_d \frac{R_0}{R_0 + R_d}\right)^2}{R_0} \quad (1)$$

where:
-P_{Batt} is the power delivered to the battery,
-U_d is the voltage applied to the adaptive driver circuit
-U_g is the voltage applied to the battery
-R_0 is the internal resistance of the battery
-Rd is the variable resistance of adaptive driver

If we consider this function of a real variable (the internal resistance of the battery, R_0):

$$P_{Batt} = P_{Batt}(R_0) \quad (2)$$

We observe that this function has an extreme point (in this case a maximum) when the first derivative of the function is 0. After the first derivative is applied, we obtain:

$$\frac{dP_{Batt}(R_0)}{dR_0} = \frac{U_d^2\left(R_0^2 - R_d^2\right)}{\left(R_d + R_0\right)^4} \quad (3)$$

And this condition happens when:

$$\frac{dP_{Batt}(R_0)}{dR_0} = 0 \rightarrow \left(R_d = R_0\right) \quad (4)$$

In our proposed solution the internal resistance of the battery (R_0) is calculated onboard at the start of the charging process, after the tests performed by the microcontroller. The tests consist of fast discharge over very short time intervals, using the experimental circuit from Fig. 2.

The internal resistance of the battery has been calculated based on the numerical differentiation using following formula:

$$R_0 = \frac{du_{dis}}{di_{dis}} = \frac{u_n - u_{n-1}}{i_n - i_{n-1}}, \quad (5)$$

where u_{dis}, i_{dis} represents the values for the discharge voltage and discharge current during t internal resistance test of the battery.

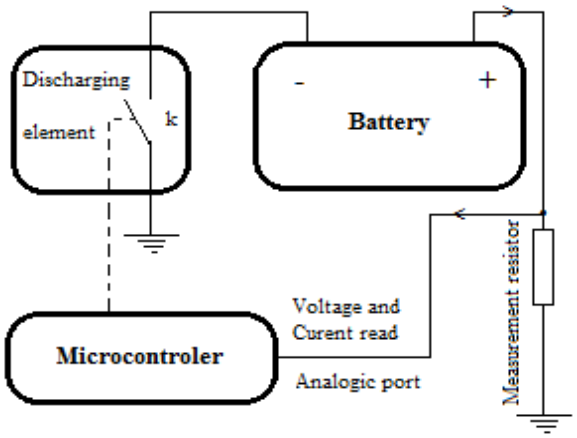


Figure 2. Discharging and testing circuit

The software used for this step was implemented in C and written on an Atmega 328p. A simplified excerpt is presented below:

```
float v1[m] = {1,2,3,…m};
float v2[n] = {1,2,3,…n};
float v3[p]= {1,2,3,…p};
void setup ()
{
float value = 0;
variable "value"
float value2 = 0;
for(int i=0; i< m; i++)
   {value=v1[i];
value2=v2[i];
value=value-(value-1)/value2-(value2-1);
v3[i]=value;
   }
```

A schematic of the section for the numerical differentiation is presented in Fig. 3.

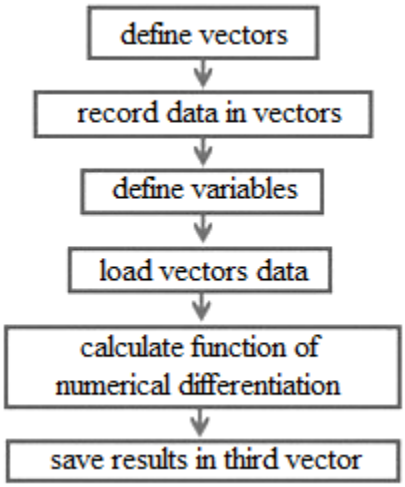
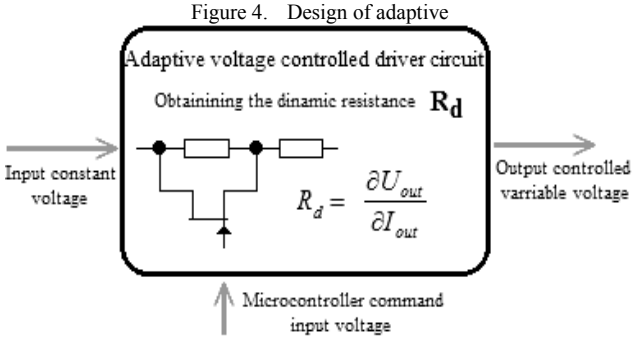


Figure 3. Schematic of the numerical differentiation section

After the value of the internal resistance of the battery is determined, the dynamic charging function can be calculated. The implementation of the charging function is done using a dynamic adaptor with variable conductance. The configuration of the adaptive controlled driver circuit is presented in Fig. 4.

Figure 4. Design of adaptive



voltage controlled driver circuit

The electric circuit contains an active element whose transconductance and dynamic resistance introduced in the circuit can be controlled by a command voltage applied to the command pin of the active element in order to achieve dynamic charging. The input of the circuit is connected to the constant voltage generator and the output provides control of the output resistance, by controlling the output voltage and current.

The formula for the adaptive impedance implementation of the battery with the power circuit is depending on forms:

$$R_d = \frac{\partial U_{out}}{\partial I_{out}} \quad (6)$$

And the differential form in the discrete case to implement is:

$$R_d = \frac{du_{out}}{di_{out}} = \frac{u_{n+1} - u_n}{i_{n+1} - i_n} \quad (7)$$

Both the resistance of battery and of the generator must be equal. For this to happen we propose the next form of charging function:

$$R_d = \frac{\partial U_{out}}{\partial I_{out}} \cong R_0 \rightarrow F_{charge} = inv[u_{dis}(i_{dis})], \quad (8)$$

where $inv[u_{dis}(i_{dis})]$ is the mathematical invert of the short-circuit resistance curve, obtained by recording the data during the controlled discharge tests.

The software used for this step was implemented in C and programmed on an Atmega 328. A simplified excerpt is presented below:

```
float v1[m] = {0,1,2,3,…m};
float v2[n];

void setup ()
{
float
value = 0;
for(int i=0; i< m; i++)
   {value=v1[i];
value=sqrt(value*value-2*value+6);
v2[i]=value;
      }
```

The schematic for the charging function is presented in Fig. 5:

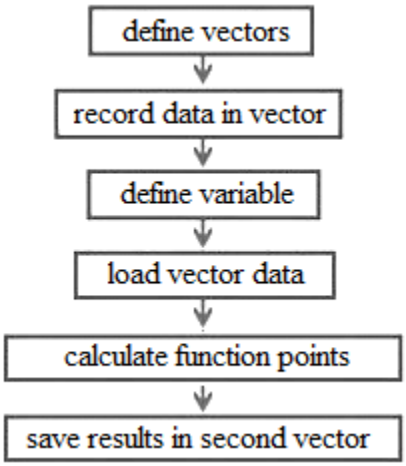


Figure 5. Schematic of the charging function

III. EXPERIMENTAL DATA AND RESULTS

In this section, experimental data and graphs are presented and the implementation of the proposed method. Using the discharging and testing block, controlled by the microcontroller, several short experiments (200-500 ms) have been performed and the values of the short-circuit voltage and current have been recorded. The Li-Ion battery of 3.7V, type 18650 was considered and experimental data obtained are presented in Table I.

Using the data from Table 1, recorded in a 300 ms long test, the graph from Fig. 6 has been generated. The graph contains three variations represented in red, green and blue. The line with a negative slope represented in blue is the current-voltage variation during the discharge and was generated using data from Table 1, consisting of 48 distinct values. It is noted that both the voltage and current drop during the discharge. The red line with zero slope represents the internal resistance of the battery and was obtained using numerical differentiation of the voltage and current.

 ISBN: 978-1-61208-740-5

TABLE I. EXPERIMENTAL DISCHARGING DATA

I[A]	U[V]	I[A]	U[V]	I[A]	U[V]
0.84	3.87	0.712	3.772	0.584	3.675
0.832	3.864	0.704	3.766	0.576	3.669
0.824	3.858	0.696	3.76	0.568	3.663
0.816	3.852	0.688	3.754	0.56	3.657
0.808	3.846	0.68	3.748	0.552	3.65
0.8	3.84	0.672	3.742	0.544	3.644
0.792	3.833	0.664	3.736	0.536	3.638
0.784	3.827	0.656	3.73	0.528	3.632
0.776	3.821	0.648	3.724	0.52	3.626
0.768	3.815	0.64	3.718	0.512	3.62
0.76	3.809	0.632	3.711	0.504	3.614
0.752	3.803	0.624	3.705	0.496	3.608
0.744	3.797	0.616	3.699	0.488	3.602
0.736	3.791	0.608	3.693	0.48	3.596
0.728	3.785	0.6	3.687	0.472	3.589
0.72	3.779	0.592	3.681	0.47	3.583

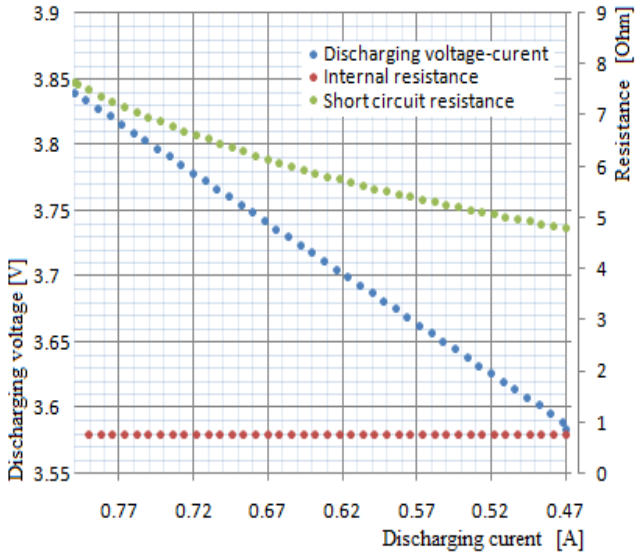


Figure 6. Discharge voltage and current

Numerical differentiation of the experimental data has led to the determination of the value of the internal resistance

$$R_0 \cong 0.76\Omega$$

The green curve with a hyperbolic variation represents the implicit variation of the current and the voltage measured at the terminals of the battery determined in short-circuits conditions and influenced by the internal resistance of the battery.

During this short test the internal resistance was constant. It implies that during the invers process, the charging process, the charging voltage and current has to follow a similar curve but with inverted variation.

The curvature and type of variation will determine the charging function. The shape of the variation of the specified mathematical function, of the current and voltage during the discharge was obtained using a second order polynomial interpolation, in this case.

$$u_{dis}(i_{dis}) = 12.47 i_{dis}^2 - 24.28 i_{dis} + 16.24 \quad (9)$$

Assigning to function the mathematical invert along the first bisector we can calculate analytical the inverse function. In this example it has the next form.

$$i_{charge}(u_{charge}) = 0.973 - 2\sqrt{0.02 u_{charge} - 0.088} \quad (10)$$

During our experiment after the completion of the test to determine the internal resistance of the battery, we will not perform onboard implementation of the analytical calculated function from (10) because it involves floating point operations and extra calculations, instead we will use the numerical invert of function, and for the experimental curve by simply inversing the experimental data obtained during the discharge process, controlled using the equation:

$$Inv[Y(X)] = X(Y), \quad (11)$$

which becomes:

$$F_{charge} = inv[u_{dis}(i_{dis})] = i_{charge}(u_{charge}) \quad (12)$$

The algorithm involves the determination of the internal resistance of the battery by performing a fast discharge test, before the start of the charging process, and the determination of the voltage-current variation curve for which the internal resistance of the battery stays constant and it is the basis for functional inversion of the charging voltage and current of the battery that does not change the internal resistance.

The difference is that the charging process will last longer than the discharge process. The time variable described in the discharge experiment is implemented using an optimized process controlled by the microcontroller.

In the following, several experimental results are presented for repeated charging of the same battery. We are presenting in Table II the experimental data of approximate 600 seconds of battery charging, in voltage and current variations. We used the same Li-Ion battery of 3.7V, 18650 types.

Fig. 7 presents the results of the voltage charging solution. Using the microcontroller controlled charging the voltage was rising with values in the [3.971-4.031V] interval. The function for the implementation of adaptive impedance was represented in red, and it represents the correction applied to the adaptive driver circuit by the microcontroller.

 ISBN: 978-1-61208-740-5

The current variation between (0.083 - 0.108) Amps, are presented in Fig. 8, during the charging process and the variation of the adaptive driver.

TABLE II. EXPERIMENTAL CHARGING DATA

U[V]	I[A]	U[V]	I[A]	U[V]	I[A]
3.971	0.108	4.000	0.098	4.016	0.090
3.974	0.106	4.001	0.097	4.017	0.090
3.977	0.105	4.002	0.097	4.018	0.090
3.979	0.104	4.003	0.096	4.019	0.090
3.981	0.103	4.004	0.096	4.020	0.089
3.983	0.103	4.005	0.095	4.021	0.089
3.985	0.102	4.005	0.095	4.022	0.089
3.987	0.102	4.006	0.094	4.023	0.088
3.988	0.101	4.007	0.094	4.024	0.088
3.990	0.101	4.008	0.093	4.025	0.087
3.991	0.101	4.009	0.093	4.025	0.087
3.992	0.100	4.010	0.092	4.026	0.086
3.993	0.100	4.011	0.092	4.026	0.086
3.994	0.100	4.012	0.091	4.027	0.085
3.995	0.099	4.013	0.091	4.028	0.085
3.996	0.099	4.014	0.091	4.029	0.084
3.998	0.099	4.015	0.091	4.030	0.084
3.999	0.098	4.015	0.090	4.031	0.083

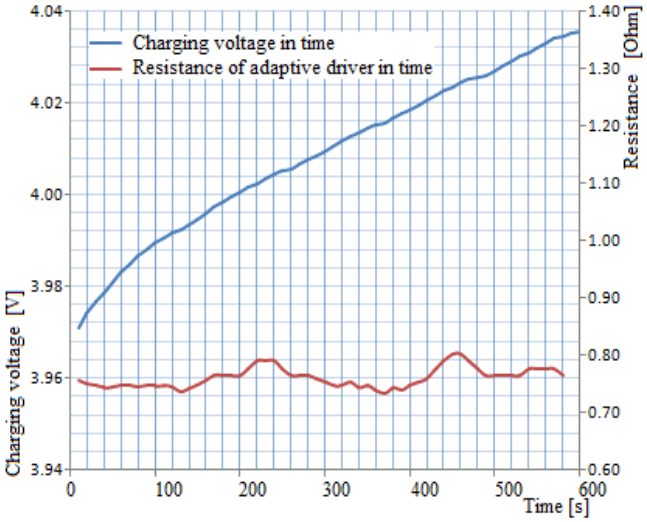


Figure 7. Charging graph of voltage in time

In Fig. 8 we are also highlighting the variation of adaptive resistance in time obtained in adaptive driver circuit. The variations of resistance of the adapter block are the answer due to the variations of internal resistance during the charging in time of the chemical element (battery).

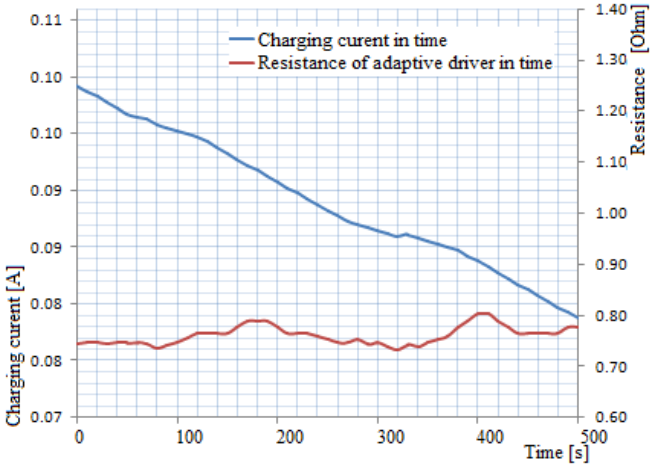


Figure 8. Charging graph of curent in time

IV. CONCLUSION AND FUTURE WORK

In this paper the possibility of using a microcontroller controlled battery charger is discussed. The method is based on theoretical considerations and numerical calculation and implementation. The experimental results show that the process can be carried out with high efficiency by controlling the conductance during the charging period using the adaptive circuit and an initial battery resistance test. Only one type of battery was tested, but the authors want to test and implement the method on several types of batteries and at various stages of wear. With the results obtained we will develop in a future project a model of functional charging in minimum time and maximum efficiency.

REFERENCES

[1] H.C. Lin, Y.J. He and C.W. Liu, "Design of an Efficient Battery Charging System Based on Ideal Multi-state Strategy", 2016 International Symposium on Computer, Consumer and Control (IS3C), Aug. 2016, pp. 956-959 , doi: 10.1109/IS3C.2016.242.

[2] D. Xu, L. Wang and J. Yang, "Research on Li-ion Battery Management System", 2010 International Conference on Electrical and Control Engineering, Wuhan, Nov. 2010, pp. 4106-4109, doi: 10.1109/iCECE.2010.998.

[3] W. Han and L. Zhang, "Charge transfer and energy transfer analysis of battery charge equalization", 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Aug. 2015, pp. 1137-1138, doi: 10.1109/CoASE.2015.7294250.

[4] J.-S. Moon, J. H. Lee, I.-Y. Ha, T.-K. Lee and C. Yuen, "An efficient battery charging algorithm based on state-of-charge estimation for electric vehicle", 2011 International Conference on Electrical Machines and Systems, Beijing, Nov. 2011, pp. 1-6, doi: 10.1109/ICEMS.2011.6073783.

[5] S.-W. Luan, J.-H. Teng, D.-J. Lee, Y.-Q. Huang and C.-L. Sung, "Charging/discharging monitoring and simulation platform for Li-ion batteries," TENCON 2011 - 2011 IEEE Region 10 Conference, Bali, Nov. 2011, pp. 868-872, doi: 10.1109/TENCON.2011.6129

 ISBN: 978-1-61208-740-5