Automation, Robotics & Communications for Industry 4.0/5.0

Proceedings

of the 3rd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI' 2023)

22-24 February 2023, Chamonix-Mont-Blanc, France

Automation, Robotics & Communications for Industry 4.0/5.0

Proceedings of the 3rd Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI' 2023)

22-24 February 2023
Chamonix-Mont-Blanc, France

Edited by
Dr., Prof. Sergey Y. Yurish
International Frequency Sensor Association,
Barcelona, Spain

IFSA Publishing

Sergey Y. Yurish, *Editor*
Automation, Robotics & Communications for Industry 4.0/5.0
ARCI' 2023 Conference Proceedings

E-mail (for orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on http://www.sensorsportal.com

ISSN: 2938-4796
ISBN: 978-84-09-47563-6
BN-20230218-XX
BIC: TJFM

Contents

Foreword

On behalf of the ARCI' 2023 Organizing Committee, I introduce with pleasure these proceedings with contributions from the *3rd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0/ 5.0(ARCI' 2023)*, 22-24 February 2023.

All submitted papers have been preliminary peer reviewed by international reviewers. The authors have revised papers after review, and resubmitted it for the second round of review by the technical committee. By this way, all mandatory corrections were made. All published papers have been presented at the event and opened for discussion.

According to the modern market study the global Industry 4.0. Market size is projected to reach US $ 377.30 Billion by 2029, at a CAGR of 16.3 % during the forecast period, 2022-2029. The Industry 4.0 means the usage of an integrated system, which consists of an automation tool, robotic control and communications. The key factors fuelling the growth of the industry 4.0 market include rapid adoption of Artificial Intelligence (AI) and Internet of Things (IoT) in manufacturing sector, increasing demand for industrial robots, rising government investments in additive manufacturing, and growing adoption of blockchain technology in manufacturing industry.

Industry 4.0 represents the 4th industrial revolution that marks the rising of new digital industry. It is defined as an integrated system that comprises numerous technologies such as advanced robotics control, automation tools, sensors, artificial intelligence, cloud computing, digital fabrication, etc. These technologies help in developing machines that will be self-optimized and self-configured. It helps in enhancing asset performance, technology usage, material usage and other industrial processes that are involves in various industries. Numerous benefits are offered by these technologies such as low operational cost, improved productivity, enhanced customer satisfaction, improved customization, and increased efficiency. The Industry 4.0 holds a lot of potentials and is expected to register a substantial growth in the near future.

The series of annual ARCI Winter IFSA conferences have been launched to fill-in this gap and provide a forum for open discussion of state-of-the-art technologies related to control, automation, robotics and communication - three main components of Industry 4.0. It will be also to discuss how to adopt the current R&D results for Industry 4.0 and to customize products under the conditions of highly flexible (mass-) production, and what we are waiting from the Industry 5.0.

The conference is organized by the International Frequency Sensor Association (IFSA) - one of the major professional, non-profit association serving for sensor industry and academy since 1999, in technical cooperation with media partners – journals: MDPI Processes, MDPI Machines and Soft Measurements and Computing.

The conference proceedings contains all peer reviewed papers, presented at the ARCI' 2023 conference. I hope that these proceedings will give readers an excellent overview of important and diversity topics discussed at the conference.

We thank all authors for submitting their latest works, thus contributing to the excellent technical contents of the Conference. Especially, we would like to thank the individuals and organizations that worked together diligently to make this Conference a success, and to the members of the International Program Committee for the thorough and careful review of the papers. It is important to point out that the great majority of the efforts in organizing the technical program of the Conference came from volunteers.

Prof., Dr. Sergey Y. Yurish,
ARCI' 2023 Conference Chairman

(0113)

Comparison of an Electric Drive with a Mechanical Drive in Mobile Machines

W. Aumer
Deggendorf Institute of Technology, Technology Campus Cham, 21 Badstraße, 93413 Cham, Deutschland
Tel.: +499971/99673-46
E-mail: wolfgang.aumer@th-deg.de

Summary: New possibilities of drive trains are given by using electrical drives. So far, process performance at self propelled harvesters can be increased by engine speed independent drive speed. Individual and variable speed and torque adjustment allows a maximum flexibility of function and design of drive trains. Results of test bench runs are shown, where an electric threshing cylinder was integrated into a combine harvester test environment. Furthermore, mechanical an electrical drives are compared towards efficiency.

Keywords: Electric drive, Automation, Functional integration, Agriculture, Mobile machines.

1. Introduction

Cost in agriculture is the motivation for permanently increasing productivity and efficiency in self-propelled harvesters. The necessity to accommodate much greater machine outputs in a limited space, requires new drive concepts. The designed space is needed to transmit the power to the various drives. In combine harvesters, for example, the distribution and transmission of drive power to the functional elements takes place via shafts, gears, chains, belts and hydraulics. Depending on the power and control requirements, the working elements are driven mechanically or hydraulically [1]. Future harvesters will have function-specific and integrated electric drives to which the electric power is supplied via a central on-board network. The integration of electric drives in the functional elements enables system simplification by reducing the number of drive and transmission elements [2].

In the ginning process, the speed stiffness of the threshing cylinder is an important characteristic of the threshing quality. The power for the threshing cylinder is transmitted mechanically and the speed is controlled by a variator. This leads to costly and complex drivelines, which can be simplified with electric drives. For research, an electric drive was designed specifically for the application of a direct drive in a threshing cylinder [3]. An air-cooled permanently excited synchronous motor in external rotor design was used. Due to the maximum motor torque and the low mass moment of inertia, the synchronous motor has a short ramp-up time, which is particularly advantageous in terms of dynamics. The objectives in the development of the electrically driven threshing cylinder are:

- Integration of the electric motor into the functional element
- Reduction of the installation space requirement due to both the integration and the simpler driveline
- Increased efficiency and speed variability (as essential cornerstones of utility value enhancement)
- Information feedback from speed and torque to incorporate dynamic processes into control strategies.

2. Testbench

For recording parameters of the electric drive, the threshing cylinder was operated on the testbench with nominal load. The load drive was directly coupled to the threshing cylinder via a chain. The electrical and thermal parameters of the threshing cylinder drive were determined. Fig. 1 shows the setup of the testbench.

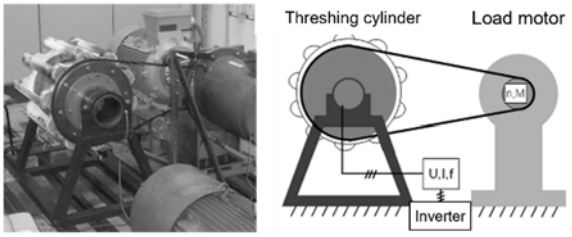


Fig. 1. Testbench setup of the electrically driven threshing cylinder.

The electrical drive of the threshing cylinder runs in speed-controlled mode. By means of torque control, the load drive provides the desired load.

The heat generated by the electric drive is dissipated via the heat sink in the axle tube on which the stator is mounted [4]. The temperature curves of the electric motor windings are shown in Fig. 2.

The abidance temperature of the electrical threshing cylinder is reached after a time of 5 h and is 130 °C. Ambient air at 16 to 18 °C was used as the cooling medium. The temperature gradient is 112 K. Thus, at an ambient temperature of 45 °C, a winding temperature of 157 °C is reached, which corresponds to the dimensioning.

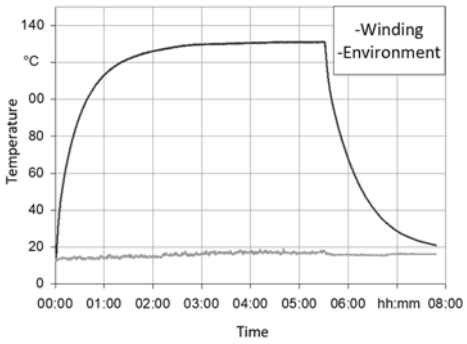


Fig. 2. Thermal characteristic of the electric drive of the threshing cylinder.

3. Laboratory Tests

For laboratory tests, the threshing cylinder with associated driveline was replaced by the electric version in a conventional threshing system, see Fig. 3.

The width corresponds to half the channel width (800 mm). As these are stationary tests, the crop is fed in defined quantities in order to reproduce realistic conditions (crop density, running speed). In this way, it was possible to determine the load behavior at different crop throughputs.

Fig. 4 shows the speed, torque and power line of the threshing cylinder for a laboratory test.

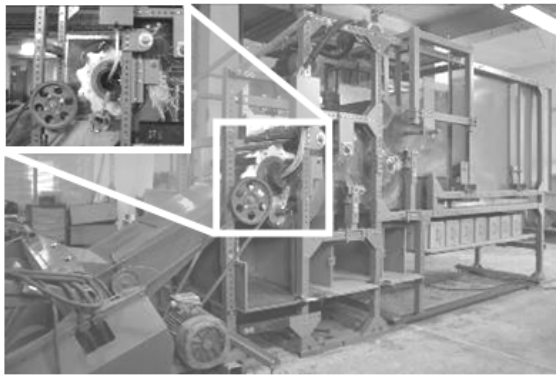


Fig. 3. Electrical threshing cylinder during laboratory tests.

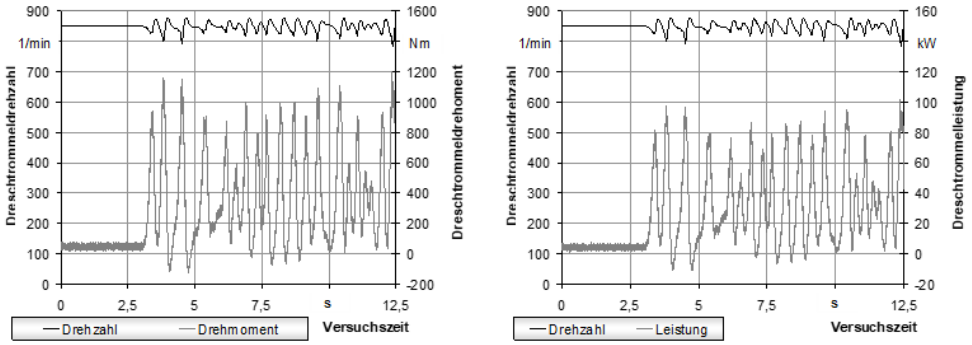


Fig. 4. Threshing cylinder speed, torque and power at 800 mm channel width and a Non-grain component (NGC) throughput of 25 t/h (based on the full channel width of 1600 mm).

At the test time of 3 s, the threshing cylinder is fed an NGC throughput of 25 t/h, based on the full channel width of 1600 mm. Non-grain components (NGC) include all components of the cut crop except the grains. The NGC throughput is a measure of the threshing drum's performance.

The drive system regulates the torque very quickly to keep the speed constant at the target speed of 850 min^{-1}. The trend of the output is similar to the torque because of the fast speed control.

For the threshing time, the effective torque is 470 Nm and the maximum torque is 1200 Nm. This results in a required overload factor of 2.55 for the design. Similar ratios result for the power. The effective power is 39 kW and the maximum power is 100 kW. The determined data meet the design, although there are still reserves.

While the high power reserves with a commercial power supply led to fast control times at the laboratory tests, higher inertia in the control system is to be expected with a diesel-electric power supply on the mobile machine.

4. Conclusions

With the functionally integrated electric drive in the threshing drum, the advantages and comparative studies with the conventional system can be proven. It can be seen that alternative drives can be an alternative for increasing productivity and efficiency in self-propelled harvesters thanks to their function-specific, decentralized and modularly designed units. In addition to the electric threshing drum, other decentralized drives in the combine harvester can usefully be driven electrically. A higher purchase price can be worthwhile for the user if it is justified by reduced specific fuel consumption, lower maintenance, increased reliability and productivity.

References

[1]. Herlitzius, Th., et. al., Integration of an Electric drive into a Tangential Threshing Cylinder, in Tagung Landtechnik, *VDI Verlag GmbH,* Hannover, Deutschland. Düsseldorf, Nov. 2009.

[2]. Bernhardt, G., Teichmann, J., Alternative Antriebskonzepte für Mähdrescher. (In German). *VDI-MEG Kolloquium Mähdrescher*, Hohenheim, 17./18. März 2005.

[3]. Aumer, W., Funktionsintegration elektrischer Antriebe in mobilen Arbeitsmaschinen, (In German), PhD Thesis, *Technische Universität Dresden*, 2018.

[4]. Miersch, S.: Erhöhung der Ausnutzung permanentmagneterregter Außenläufermaschinen durch Verbesserung der Wärmeabführung. (in German), PhD Thesis, *Technische Universität Dresden*, 2021.

(0233)

Deployment of Digital Twin in Robotics and Autonomous Driving Using ROS Architecture

Muhammad Ishfaq Hussain, Muhammad Aasim Rafique, Sayfullokh, Khurbaev, Junbom Pyo, Taweook Kim, Brain J. d' Auriol, and Moongu Jeon
School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea
Tel.: + 82627153057
E-mail: {ishfaqhussain, aasimrafique, sayfull1735, xenoluv, theater74, dauriol, mgjeon}@gist.ac.kr

Summary: In robotics and autonomous driving applications, safety and reliability are major concerns. Testing and authentication of autonomous driving and robotic applications in a simulation environment exist. Visualization and performing some operations with the real-time data of autonomous driving and robots can be dealt with Digital twin technology. Digital twin technology provides us the real-time ongoing operations on the remote machine as well. Sending the commands and performing some operations on the real autonomous robot or vehicle using the cloud machine (twin) required smooth network communication. While communication over the network the network's latency should be very low to deal with life-critical applications like robotics and autonomous driving. In this research we designed a prototype of Digital twin, using the robot operating systems (ROS) based publisher-subscriber pattern. However, the network latency is quite high while communicating over the network. Dealing with this problem, we also proposed a vehicular edge computing device to minimize network latency. We compare the results by analyzing the received timestamps in real time on the autonomous robot and on the remote machine timestamps to empower the credibility and reliability of the proposed work.

Keywords: Digitization, Digital Twin, Autonomous Vehicles, Robotics, Latency, Intelligent vehicles.

1. Introduction

The dawn of robotics and autonomous driving is quite slow due to multiple root constraints. The reason behind the scene is security, acceptance challenges, legal, and infrastructure. Moving forward towards the feet-off, eye-off, and finally, fully brain-off driving requires a reliable and safest automotive structure to bypass these challenges. Automotive manufacturers can play a vital role while selecting and in- integrating the right sensors stack, in the right place of the vehicle with evolving software stack. The existing research focuses on the design of robotics and intelligent autonomous vehicles, however, neglecting the digitization of the hierarchy of software stacks, which plays a significant role in driving. Digitize and synergize the connected robots and autonomous vehicles, Digital twin technology provides a virtual physical system, with accelerated software stack processes, faster innovation, and enhanced productivity at a lower cost. Digital twin technology overall plays a vital role in many domains like automotive, industry 4.0, smart cities, healthcare, and a lot more. The effectiveness of a digital twin is only possible if it meets the desired quality, and security at the lowest latency while communicating over the network between the twins. In this research, we proposed the robot operating system (ROS) based [1] digital twin machine which provides us with the publisher and subscriber pattern. However, there is a strong communication protocol that should be followed as robotics and autonomous systems are life-critical systems. The quality of the digital twin depends on the communication over the network between actual and physical machines affected by the latency of the network. In this research, we proposed a methodology to reduce the latency rate while communicating over the network in the robot operating system ROS environment.

2. Related Work

The digital twin is a great tool to have a glimpse into the hearts of the physical equipment. Not only it enables investigation of happenings of the present but it also helps foresee the future of the machines in the scenarios. Digital Twin technology was one of the top ten promising technologies mentioned by Gartner's report in 2018 but initially Grives, et al proposed it in 2003. Author [3] used a novel approach and worked on improving the visual guidance provided to the vehicle by working on the fusion of camera image and digital twin cloud information about other present entities in the environment. A further extension to digital twin can be seen in the [2, 4-6] where the author explored time variance, customized modeling, and multi-modal state fusion.

3. Proposed Methodology and Results

In this work, we utilized the robot operating system (ROS) publisher and subscriber Fig 1, through which we can easily receive and send the information over multiple machines. The server machine runs all the

self-driving software stack in the ROS environment. To make a digital twin machine for self-driving vehicles we proposed and utilized the ROS as the middleware for the twin and the actual machine to communicate and share information. ROS master which is running on the server machine while the twin machine is just using that ROS master as the client. In the master machine, the ROS master is running while the client machine or the twin machine is just sourcing the master over the network. As the ROS master is working on the server machine, so all the ROS nodes and topics can be easily accessed and utilized on the twin machine as well. The actual machine is running all the ROS nodes and topics which are communicating with the real car through the publisher and subscribers. The same publisher and subscribers can access and send, and receive commands respectively. In the first experiment during the communication of ROS nodes, there was a delay while sending and receiving the information over the network. The network latency was too high. To avoid mishaps and packet loss while information sharing, we added an edge device to reduce the latency. The edge device which we call, here the vehicle edge computing (VEC) device. The VEC is holding all the ROS topics and ROS nodes' information while communicating only the updated information to reduce the latency rate over the network. Some information like maps (point cloud and vector map information are the same as they are pre-defined). Only the sensor information which we are getting from the environment, needs to be updated on the twin machine. Based on that the twin can send or bypass some ROS nodes on the server machine. Using this load sharing and reducing the latency over the network through the VEC improve the proposed work as shown in Fig 2. To test the efficacy, of the proposed model, we calculate the difference between the rostopic timestamps on the server machine as well on the twin. For this mean square error MSE is utilized. Based on the results, we concluded that, for the development of robotics and the self-driving cars digital twin uses the robot operating system (ROS) publisher and subscriber. The proposed structure also incorporates an edge device to mitigate the problem of latency while communicating over the cloud between the real and the twin machine.

Fig. 1. ROS based publisher and subscriber for a digital twin.

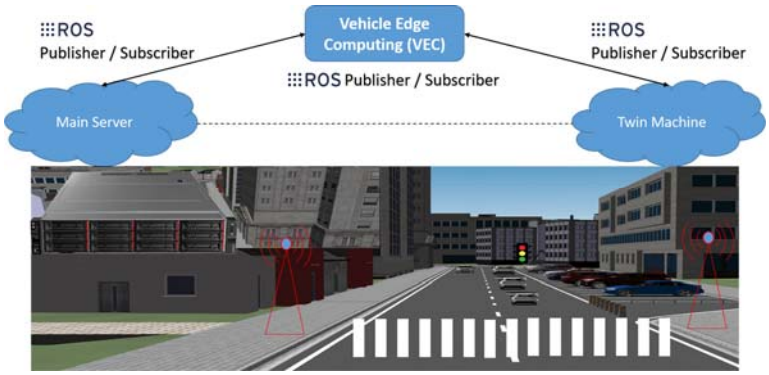


Fig 2. The Proposed Structure for the digital twin to decrease the latency using Vehicle Edge Computing (VEC).

Acknowledgement

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2014-3-00077, Development of Global Multi-target Tracking and Event Prediction Techniques Based on Real-time Large-Scale Video Analysis) and Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism (No. R2022060001, Development of service robot and contents supporting children's reading activities based on artificial intelligence).

References

[1]. M. Quigley, A. Y. Ng et al., ROS: an open-source robot operating system, in *Proceedings of the International Conference on Robotics and Automation ICRA,* Vol. 3, No. 3.2, Kobe, Japan, 2009, p. 5.

[2]. Z. Hu, S. Lou, Y. Xing, X. Wang, D. Cao, and C. Lv, Review and perspectives on driver digital twin and its enabling technologies for intelligent vehicles, *IEEE Transactions on Intelligent Vehicles*, 7, 3, 2022, pp. 417 - 440.

[3]. Y. Liu, Z. Wang, K. Han, Z. Shou, P. Tiwari, and J. H. Hansen, Sensor fusion of camera and cloud digital twin information for intelligent vehicles, in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV),* 2020, pp. 182–187.

[4]. S. Nair, S. Abbasi, A. Wong, and M. J. Shafiee, MAPLE-Edge: A Runtime Latency Predictor for Edge Devices, in *Proceedings of the Conference on Computer Vision and Pattern Recognition,* 2022, pp. 3660– 3668.

[5]. A. Niaz, M. U. Shoukat, Y. Jia, S. Khan, F. Niaz, and M. U. Raza, Autonomous driving test method based on digital twin: A survey, in *Proceedings of the IEEE International Conference on Computing, Electronic and Electrical Engineering (ICE Cube' 2021)*, 2021, pp. 1–7.

[6]. G. P. Agnusdei, V. Elia, and M. G. Gnoni, A classification proposal of digital twin applications in the safety domain, *Computers & Industrial Engineering,* Vol. 154, 2021, p. 107137.

(0570)

Design and Implementation of Time Aware Buffer for Scheduling Time Sensitive Packets in Deadline Aware Networks

Tsung-Ying Yang and Steven S. W. Lee
Department of Communications Engineering, National Chung Cheng University,
621, Chiayi, Taiwan
E-mail: ieeswl@ccu.edu.tw

Summary: Time-sensitive applications require their packets to be delivered by deadlines. Although modern switches are able to prioritize packets and store them in different queues, such buffers that do not consider the deadline of each packet are still insufficient to support time-sensitive services. In this paper, we propose a new type of buffer called Time-Aware Rotating Priority Queue (TARPQ). TARPQ consists of multiple First-In-First-Out (FIFO) queues, each associated with a different deadline period. The system organizes the FIFO queues into a logical ring and dynamically adjusts each queue's deadline period. TARPQ guarantees in-order delivery of packets. Therefore, the destination does not need to reorder packets. We modified the bmv2 emulator to enable P4 switches to perform TARPQ. Experimental results show that, compared with well-known schemes, TARPQ can reduce packet loss rate and overdue probability.

Keywords: Buffer management, Deadline aware network, Packet scheduling, Rotating priority queue, Time sensitive application.

1. Introduction

The most commonly used buffers in network switches are First-In-First-Out (FIFO) queues. A single FIFO queue only maintains packet order regardless of deadlines. Modern network switches are often equipped with multiple FIFO queues. Each queue is associated with a priority. Packets are kept in queues according to their priority. Switches can use different scheduling algorithms to deliver packets in these queues. In Strict Priority Queue (SPQ), packet delivery strictly follows queue priority. The packet at the head of queue *i* will not be delivered until all queues with priority higher than *i* have no packets left.

Some variants, such as Weighted Fair Queuing (WFQ), assign weights to different FIFO queues. A queues with higher priority is usually assigned a larger weight. Therefore, the network provides more bandwidth to transmit packets with higher priority. By assigning appropriate weights to each queue, the network can optimally control packet delay and jitter for static traffic [1, 2]. Cisco combines SPQ and WFQ and proposes Low Latency Queuing (LLQ) [3] to achieve better QoS control. Although SPQ and WFQ can provide service differentiation for packets of different priorities, they are not sufficient for time-sensitive packets, especially for bursty traffic.

The Earliest Deadline First (EDF) scheduler delivers packets according to their deadlines [4-6]. EDF provides excellent results for delivering time-sensitive packets. To implement EDF, the buffer needs to sort the packets according to the deadline. In [7, 8], the system inserts an incoming packet into the correct position in the queue according to the sorting results. Due to its high complexity, EDF is difficult to apply in high-speed networks. To reduce packet processing time, special hardware [9] or data structure [10] are required.

To reduce the time complexity, Rotating Priority Queue (RPQ) [11] is proposed to approximate EDF. RPQ uses multiple FIFO queues. Each FIFO queue is assigned a deadline period. Based on the packet deadline, the switch stores received packets in its FIFO queue. RPQ uses strict priority to deliver packets. The queue associated with the earliest deadline holds the highest priority packets. The proposed Time-Aware Rotating Priority Queue (TARPQ) is also a kind of rotating priority queue. Unlike RPQ, TARPQ can perform queue rotation even if the highest priority queue is not empty. In addition, in TARPQ, queue rotation can be triggered for load balancing purposes. Those properties enable TARPQ to achieve better performance.

In our system, every switch knows the current time. Every packet indicates whether it is time sensitive or not. Every time-sensitive packet carries a deadline assigned by its source host. The switch processes time-sensitive packets according to their deadlines.

Time-sensitive applications can be divided into two types. For the first type, every packet in the same flow has the same deadline. Time-sensitive file transfers fall under this category of services. A time-sensitive file must be delivered to the destination within a specified deadline. Each packet of the file carries the same deadline. Machine control is another time-sensitive application. The controller sends commands to its target machines. As a command is issued by the controller, it must be delivered to the target machine within the specified time. In this case, each command packet sent at a different time has a different deadline. Our system can support both types of services.

2. The Time Aware Rotating Priority Queue (TARPQ)

We use Fig. 1 to demonstrate the operation of TARPQ. A TARPQ consists of multiple FIFO queues (q_0-q_n). Each queue has a different priority. P_0 is the highest priority and P_n is the lowest priority. Packet leaving the switch follows strict priority. The left and right sides of Fig. 1 show TARPQ before and after queue rotation. Each queue is assigned a deadline period. These deadline periods for the queues do not overlap. Queues with higher priority hold packets with earlier deadlines. When a packet arrives, it will be kept in a queue according to its deadline. If the deadline carried on an incoming packet is smaller than current system time t, the packet has exceeded its deadline and the switch will discard the packet directly.

The time period for the highest priority queue is $(t,T_1]$, where t is the current time. If the queue is not the highest priority one or the lowest priority one, the time period for the queue is $(T_i,T_{i+1}]$, where $T_{i+1}-T_i$ is a constant value δ. The lowest priority queue is responsible for holding packets with a deadline greater than T_n.

In this example shown in Fig. 1, initially q_0 is the highest priority queue and q_n is the lowest priority queue. The system performs queue rotation to maintain high utilization of these FIFO queues. There are two events that can trigger queue rotation. The first event is a periodic event generated every duration δ. The second event is triggered if there is a queue with queue length larger than a threshold.

After a rotation, the priority of each queue is increased by 1 except for the highest priority queue. The highest priority queue becomes the lowest priority queue, and the time period for each queue changes to a new range.

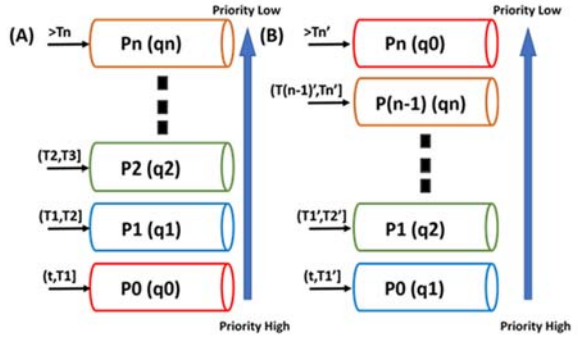


Fig. 1. FIFO queue priorities before and after rotation.

The period to perform periodic queue rotation is δ. We set $\delta = d/m$, where m is the total number of queues. In the system, the maximum allowed delay for time sensitive service is d so the maximum deadline carried by a time sensitive packet is $t+d$. The time period assigned for highest priority queue is $(t, T_1]$. The initial value of T_1 is $t+2\delta$. As time goes by, the system time t gradually increases, and the range T_1-t of the highest priority gradually shrinks. Periodic rotation is triggered when t reaches $T_1-\delta$. After rotation, the new T_i' is updated to $T_i+\delta$.

Because we rotate the queues at $t= T_1-\delta$ instead of $t= T_1$, there may still be packets left in the highest priority queue after the rotation. TARPQ records the number of highest priority packets remaining in the new lowest priority queue. If the number of remaining packets is non-zero, the system enters a transition state. In the transient state, the highest priority packets remaining in the lowest priority queue are transmitted first. The system does not deliver any packets from other queues until the remaining highest priority packets are cleared. Once all remaining highest priority packets leave the switch, TARPQ returns to a normal state. Even in the transient state, each queue still accepts packets. Note that our system guarantees that packets in each flow are delivered in order. Since the time period of each queue, except the highest priority queue, is updated by adding δ for each T_i when queue rotates, the rotation doesn't cause out-of-order problems.

In addition to periodic rotation, the system also performs load balancing rotation. A queue can overflow when a large number of packets with the same priority enter the switch. The system rotates queues to avoid load imbalance between queues. Different from periodic rotation, after load balancing rotation, the time period of each priority remains unchanged. Although two consecutive data packets belonging to a flow will be stored in different queues before and after load balancing rotation, there is no out-of-order problem. TARPQ ensures that the destination host always receives packets in order.

3. Experimental Results

We conduct three sets of experiments. The first two sets of experiments are performed in the network shown in Fig. 2 and the third set of experiments are performed in the network shown in Fig. 5. Each switch is a PC running the bmv2 emulator [13]. We modified the bmv2 program to include the proposed TARPQ queues. Each link capacity is 10 Mbps in the networks.

A. Experiment 1: Time sensitive control

In the first set of experiments, we examine the performance of the proposed TARPQ when used to deliver machine control packets. A source sends control packets to its destination. We observe five flows between hosts in the network. In addition, we also generate five background flows between the switches. All of the flows are UDP flows. We denote each flow by (s, d, t), where s is the start node, d is the destination node, and t ms is end-to-end delay requirement for each packet. These five background flows are (S4, S3, 240), (S4, S3, 180), (S1, S3, 330), (S1, S2, 200), and (S2, S3, 180).

We observed 5 time sensitive traffic flows. These flows are (H1, H6, 80), (H2, H6, 180), (H3, H6, 340), (H4, H6, 360), (H5, H6, 240). We follow a long range

dependent model to generate the flows. The α value and burst size used in the traffic generator were 1.4 and 4000, respectively. Each flow is an aggregate from 256 on/off packet generators [12].

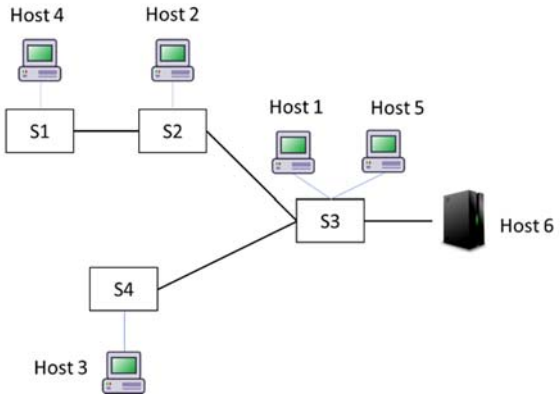


Fig. 2. Experimental network.

In the experiments, we used four FIFO queues in the TARPQ. Each queue can hold 64 packets. For TARPQ, the δ is set to 90 ms and the threshold to trigger load balancing rotation is 36 packets. We performed a performance comparison of a single FIFO, SPQ, RPQ, and TARPQ. SPQ, RPQ, and TARPQ use four queues, and each queue can hold 64 data packets. For a single FIFO queue, there are two buffer sizes. One can hold 64 packets and the other can hold 256 packets. For SPR, the time periods for the four queues are (0,100], (100,200], (200,300], and (300,400] ms. For RPQ, the Δ is set to 80 ms.

We conducted four sets of experiments. Table 1 includes the data rates used in the experiments. Fig. 3 shows the experimental results. We include packet loss rate and packet expiration rate. The former comes from not having enough space in the switches to buffer incoming packets. The latter indicates that the packet arrived at its destination, but the arrival time missed the deadline. Experimental results show that the packet success rate of TARPQ is higher than that of FIFO, SPQ and RPQ.

B. Experiment 2: Packets in a flow have random deadline

In the second set of experiments, we use the same settings as in Experiment 1. However, the deadline for each packet in a flow is randomly generated. The deadline ranges from 80 ms to 360 ms. The experimental results are shown in Fig. 4 and the TARPQ has the best performance.

C. Experiment 3: Time sensitive file transfer

In the final set of experiments, we examine the performance of transferring time-sensitive files. The experiment was performed in the network shown in Fig. 5. There are five source hosts, H1-H5. Each host uses TCP to send five types of flows to the receiver computer. The five types of flows consist of 10, 20, 30, 40, and 50 packets, and the five types of flows have deadlines of 90 ms, 135 ms, 180 ms, 320 ms, and 400 ms, respectively. Every packet from the same flow has the same deadline. In this experiment, each sending host sends a total of 5000 flows to the receiving host, 1000 of each type. The receiver is responsible for collecting packets from a total of 25000 flows and analyzing the percentage of packets that missed the deadline.

Table 1. Data rates of each flows.

Test Cases	Data rate
1	H1-H5 mean rate : 0.4 Mbps Background 1-5 mean rate : 0.8 Mbps
2	H1-H5 mean rate : 0.5 Mbps Background 1-5 mean rate : 0.1 Mbps
3	H1-H5 mean rate : 0.6 Mbps Background 1-5 mean rate : 1.2 Mbps
4	H1~H5 mean rate : 0.7 Mbps Background 1-5 mean rate : 1.6 Mbps

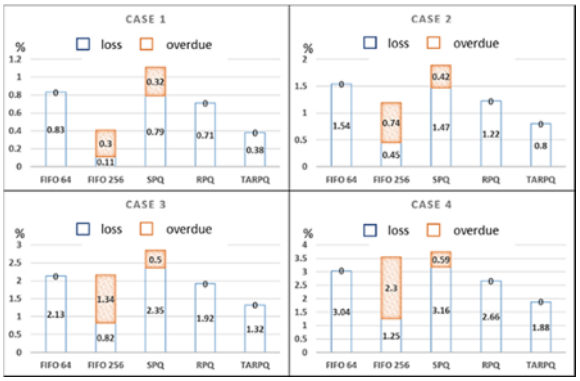


Fig. 3. Percentage of packet loss and overdue in experiment 1.

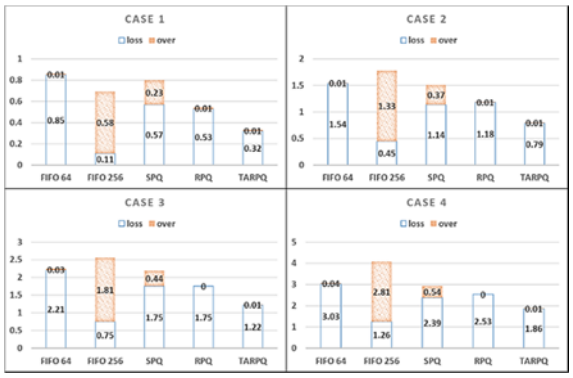


Fig. 4. Percentage of packet loss and overdue in experiment 2.

The total number of queues used in this experiment is four. TARPQ and RPQ handle packets according to their deadline. Because the maximum deadline is 400 ms, the δ of TARPQ is set to 100 ms. The Δ parameter of RPQ is 80 ms. In SPQ, all packets from type 1 files are statically assigned to the highest

priority queue. Packets from Type 2 and Type 3 flows are assigned to the second priority queue. Packets from Type 4 flow and Type 5 flow are statically assigned to the third and fourth priority queues, respectively.

Table 2 shows the ratio of traffic that fails to reach the receiver by the file deadline. Experimental results show that TARPQ has the highest success rate in meeting flow deadlines.

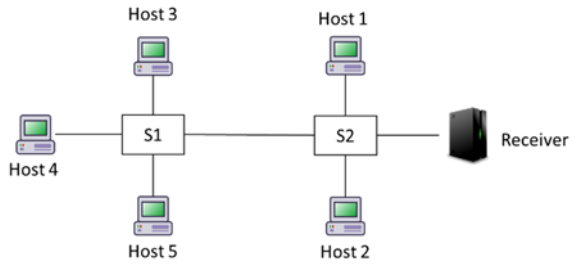


Fig. 5. Experimental network for file transfer.

Table 2. Ratio of files missed deadline.

Test Cases	Missed deadline ratio
FIFO 64	23.68 %
FIFO 256	21.83 %
SPQ	18.55 %
RPQ	20.66 %
TARPQ	11.20 %

4. Conclusions

In this paper, we present the design of the proposed TARPQ. The goal of TARPQ is to increase the success rate of delivering time-sensitive network traffic. We have implemented the proposed TARPQ on top of bmv2 emulator. In addition to TARPQ, we also evaluate the results from FIFO, SPQ, and RPQ. Experimental results show that the packet success rate of TARPQ is higher than that of FIFO, SPQ and RPQ.

References

[1]. M. Li and B. Song, Design and implementation of a new queue scheduling scheme in DiffServ networks, in *Proceedings of the International Conference on Advanced Computer Control,* Shenyang, 2010, pp. 117-122.

[2]. A. M. Elnaka, Q. H. Mahmoud, and Xining Li, Simulation based comparative performance analysis of QoS traffic scheduling using fair and delay adaptive scheduler (FDAS) versus WFQ and EDF, in *Proc. IEEE CCNC*, 2016, pp. 916-923.

[3] Cisco LLQ (Low Latency Queueing), (https://networklessons.com/quality-of-service/qos-llq-low-latency-queueing-cisco-ios).

[4] D. Liu and Y.-H. Lee, An efficient scheduling discipline for packet switching networks using Earliest Deadline First Round Robin, in *Proc. IEEE ICCCN*, Oct. 2003, pp. 5-10.

[5] P. Bommannavar, J. Apostolopoulos and N. Bambos, Deadline aware packet scheduling in switches for multimedia streaming applications, in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2013, pp. 3797-3802.

[6] M. Saleh, Z. Othman and M. Saleh, Efficient hierarchical Diff-EDF schedulability for QoS packet networks, in *Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications*, 2008, pp. 815-819.

[7] A. Martinez Vicente, G. Apostolopoulos, F. J. Alfaro, J. L. Sanchez, and J. Duato, Efficient deadline-based QoS algorithms for high-performance networks, *IEEE Transactions on Computers*, Vol. 57, No. 7, July 2008, pp. 928-939.

[8] Y. Xie, X. Tu, H. Liu, J. Li, T. Li, and J. Xie, Insertion based packets scheduling for providing QoS guarantee in switch systems, in *Proceedings of the International Conference on Communications, Circuits and Systems*, 2008, pp. 453-456.

[9] M. Alizadeh, S. Yang, M. Sharif, S. Katti, N. McKeown, B. Prabhakar, and S. Shenker. pFabric: Minimal near-optimal datacenter transport, in *Proceedings of the ACM SIGCOMM'13 Conference,* 2013.

[10] A. R. Ashok Kumar, K. L. Anusha, and S. G. Hegde, Packet scheduler for meeting deadline in 4-4; 1-4 architecture, in *Proceedings of the IEEE International Conference on Advanced Networks and Telecommunications Systems* (*ANTS*), 2018, pp. 1-6.

[11]. J. Liebeherr and D. E. Wrege, Priority queue schedulers with approximate sorting in output-buffered switches, *IEEE Journal on Selected Areas in Communications* , Vol. 17, 1999, pp. 1127-1144.

[12] Synthetic self-similar traffic generation (http://research.glenkramer.com).

[13] Behavioral Model version 2 (BMV2), (https://github.com/p4lang/behavioral-model).

(0880)

Monitoring Weld Penetration by Training A Deep Learning Model Using Inaccurate Labels

Rui Yu [1], Yize Chen [1], Jin Zhang [1], Qiang Ye [2] and YuMing Zhang [1*]
[1] Department of Electrical and Computer Engineering and Institute for Sustainable Manufacturing
[2] Department of Mathematics,
University of Kentucky, Lexington, KY 40506, USA
E-mail: yuming.zhang@uky.edu

Summary: Welding joins workpieces together by melting their facing edges. Such melting is quantified by the weld penetration state. It determines the weld integrity and is the most important critical weld quality parameter. However, it occurs underneath the workpiece and is not observable during manufacturing. The penetration state thus requires to be monitored using easily measurable surface phenomena but how they are correlated to the penetration is unknown. Deep learning provides an effective solution to automatically extract the critical features and correlate them to the penetration labels. However, to train deep learning models, the needed amount of labels is typically large. As penetration occurs underneath the workpieces, its labels are not easily obtained in many cases. This paper proposes to use inaccurate labels that may be easily obtained but have large errors with the actual labels. The key to the success of this approach is the labeling error has zero mean and the data size increases with the labeling error. To verify this idea, experiments have been conducted using gas tungsten arc welding on mild steel plates. A high dynamic range (HDR) camera is used to image the weld pool as the measurable phenomena for being used as the input of a convolutional neural network (CNN). A regular camera is used to image the back-side of the workpiece and such captured image is used to calculate the accurate label. The accurate labels are added with large random errors to generate the inaccurate labels. The output of the CNN is compared with the inaccurate labels to train the CNN. It was found that excellent accuracy has been achieved by the trained CNN to predict the accurate label. Methods that may obtain inaccurate labels without obtaining accurate penetration labels have also be proposed.

Keywords: Weld penetration, Deep learning, CNN, Ground truth, Labeling

1. Introduction

The weld penetration occurs on the bottom of the weld pool and the weld pool surface is probably the most directly related observable that may be correlated to the penetration. However, because of the extraordinarily strong radiation from the welding arc, imaging the weld pool surface has not been easy [1]. The recent progress in HDR cameras has made clearly observing the weld pool easier despite the arc radiation [2]. As such, most of the observable phenomena that are used to monitor weld penetration are based on weld pool images in recent studies. In these studies, weld pool images are directly used as the inputs of deep learning models to predict the penetration. To train such deep learning models, weld penetration labels are needed. Unfortunately, the needed number of labels to train deep learning models is extraordinarily large. Further, the weld penetration occurs underneath the workpiece so that its measurement is typically difficult. Effectively using deep learning models to revolutionize intelligent welding becomes challenging.

2. Background and Analysis

To solve this fundamental issue, this paper proposes to use inaccurate labels that can be conveniently obtained. Denote the set of the weld pool images (observable phenomena) as Ξ and the set of the penetration state labels as X. Deep learning is to fit a model that correlates Ξ to X from the dataset (X, Ξ). As X is not easy to obtain, we propose to use an alternative dataset (X^0, Ξ) to train the deep learning model. Here X^0 is an easily obtained "alternative label" that is not accurate and is related to X by $X(k) - X^0(k) = \varepsilon(k)$ $(k = 1, \dots, N)$ where N is the number of samples in the data set. The requirement for X^0 is that the labeling error ε has zero mean while its variance σ_ε^2 can be large.

For an easy understanding of why the proposed method may be effective, we use a simple example $y(k) = a_1 x_1(k-1) + a_2 x_2(k-1)$ to illustrate. We use a set of known $(a_1, a_2) = (0.5, 0.4)$ to calculate $y(k)$ from $(x_1(k), x_2(k))$ $(k = 1, \dots, N)$. To be convenient, $-1 \le x_1(k) \le 1$ & $-1 \le x_2(k) \le 1$ $(k = 1, \dots, N)$ are independent uniformly distributed random numbers. For the given (a_1, a_2), $y(k)'s$ $(k = 1, \dots, N)$ can be calculated and these calculated $y(k)'s$ are the accurate labels. However, we use $y^0(k) = y(k) - \varepsilon(k)$, where $-1 \le \varepsilon(k) \le 1$ is a uniformly distributed random number independent of x_1 and x_2, to generate $(y^0(1), \dots, y^0(N))$ as the inaccurate labels to estimate the model. The relative inaccuracy in the label is measured by $\sigma_\varepsilon^2/\sigma_y^2$. In the simulations, the mean of $\sigma_\varepsilon^2/\sigma_y^2$ is approximately 0.7. The mean ratio of the labeling error/real label is 84% which is very large. However, the modeling error over labeling error reduces with the sample number N. Simulations show $\mathrm{E}(\sigma_\varepsilon^2/\sigma_y^2)\mathrm{E}(\sigma_e^2/\sigma_\varepsilon^2) = \mathrm{E}(\sigma_e^2/\sigma_y^2) = 0.0126$, i.e., the estimation error variance is only 1.26% of that of the

real label. This is the result when N=100. When N=1000, this reduces to 0.7091*0.0019=0.13%. The large error in the labeling is compensated by increasing the number of samples.

To verify the effectiveness of the proposed idea in monitoring weld penetration using a trained deep learning model, study needs to be conducted using experimental data and a deep learning model.

3. Data Generated

The penetration is controlled by the welding current $i(k)$ and speed $v(k)$. Ideally, we wish to use the heat input $h(k) \propto i(k)/v(k)$ as the inaccurate label. However, this requires other welding conditions for each experiment to be maintained almost the same or the experiments to be extensive to assure the zero mean requirement for labeling error. In this preliminary study to propose and first verify the proposed novel concept, we conducted 3 experiments. We found that the weld penetration does not change accurately per the heat input from experiment to experiment as other welding conditions are not controlled and have affected the welding process. As such, we still measure the accurate label X and add errors to generate the inaccurate label X^0.

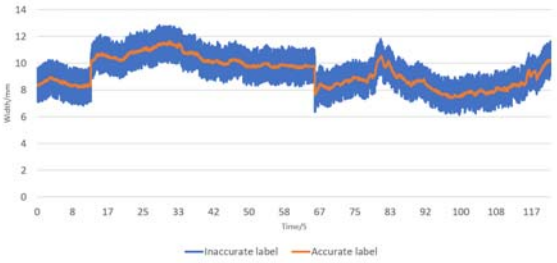
To be simple, each of the three experiments used a constant current, 85 A, 90 A, and 95 A respectively in the three experiments, and welding speed is randomly changed about 2 mm/s during the experiment. The experiments are conducted on mild steel workpieces with dimension 3mmX50mmX250mm using gas tungsten arc welding process. An HDR camera, Xiris XVC-1100, is used to image the weld pool at 60 Hz as Ξ. A regular PointGray camera is used to capture the bottom image of the weld pool. The ground truth X is calculated from the bottom images using an image processing algorithm [2]. The resultant accurate label can be seen in Fig. 1. The dataset $(X(k), \Xi(k))$ $(k = 1, \dots, 7281)$ is formed corresponding to 121.35 seconds. To form the dataset to train a deep learning model, a random error $\varepsilon(k)$ in [-20, 20] pixel ([-1.4, 1.4] mm) is subtracted from (same as added to) $X(k)$. The dataset (X^0, Ξ) for training using inaccurate labels can thus be obtained.

4. Network Model

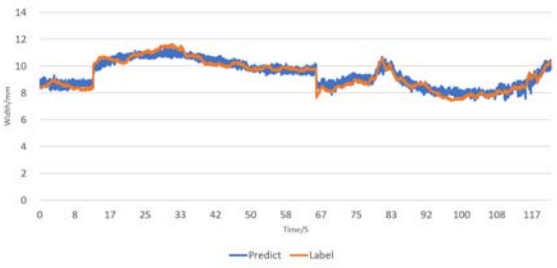
A typical CNN (convolutional neural network), which repeats convolution and a following pooling a few times, is used to covert one image $\Xi(k), 256 \times 256$, into one 1×64 feature vector $V(k)$. This repeats 2 times with parameters of the convolution layers to be (1, 32, 5, 4, 2), (32, 64, 5, 4, 2) respectively. Batch normalization and ReLU activation are conducted between each convolution and pooling layer. Finally, the CNN inputs $V(k)$ to a fully connected layer to calculate the output $O(k)$ of the CNN to compare with the inaccurate label $X^0(k)$.

5. Results

Fig. 1 shows the accurate label (in both (a) and (b)), inaccurate label (in (a)) used for training, and training/validation results (in (b)). The first 15 % in (b) is the validation result (prediction) and the rest 85 % of the plot in (b) is the fitting of the trained model. It is apparent that the model error (fitting and prediction) $X - O$ (Fig. 1(b)) is much reduced from the labeling error $X - X^0$ (Fig. 1(a)). In fact, the variance of the labeling error $(X - X^0)^T(X - X^0)/N = 130.74$. However, the variance of the actual fitting/prediction $(X - O)^T(X - O)/N = 5.53$. The fitting/prediction accuracy is acceptable by any reasonable standards. As such, we verify that a deep learning model may be trained using inaccurate labels. However, we note that this is verified under the condition that the labeling error meets the zero mean condition.



(a) Accurate and inaccurate labels.



(b) Deep learning model fitted/prediction vs actual.

Fig. 1. Training a deep learning-based weld penetration monitoring model using an inaccurate label.

6. Conclusion

An innovative concept has been proposed to address the fundamental challenge in deep learning-based weld penetration monitoring. This uses inaccurate labels whose error meets the zero mean property. Experiments verified this effectiveness.

Acknowledgements

This work is partially funded by the National Science Foundation under CMMI-2024614.

References

[1]. Zhang, Y., Wang, Q. and Liu, Y., Adaptive intelligent welding manufacturing, *Weld J.,* 100, 1, 2021, pp. 63-83.

[2]. Yu, R., et al., How to Accurately Monitor the Weld Penetration From Dynamic Weld Pool Serial Images Using CNN-LSTM Deep Learning Model? *IEEE Robotics and Automation Letters*, 7, 3, 2022, pp. 6519-6525.

(0889)

Triangular Observer-based Integral Sliding-mode Control for Wind System Energy

S. El Bouassi, Z. Chalh and E. M. Mellouli
Laboratory of Engineering, systems and applications, Sidi Mohamed Ben Abdellah University, Fez, Morocco
E-mail: sanae.elbouassi@usmba.ac.ma

Summary: In this work, in order to build a dependable controller, we offer a modified integrated sliding mode control technique, it is used in order to fix the problem of reaching phase in systems with matching disturbances. We recommend modeling the wind turbine system as a starting point. We will apply after the integral sliding mode control methodology. Only the rotor speed is available in this model , for that We use an asymptotic observer triangular ; that causes the observing errors to go toward zero over a finite period of time ; to roughly estimate the unknown rotor acceleration needed to build the control law. The control law has its roots in the stability research in the Lyapunov sense. The simulation results show that the proposed method works better than the conventional sliding mode control in terms of settling time, tracking precision, energy usage, and input smoothness.

Keywords: Integral sliding mode control, Nonlinear control, Lyapunov stability, Wind turbine, Triangular observer.

1. Introduction

The concept known as integrated sliding mode control (ISMC) was initially put out in [4, 6] as a remedy for the reaching phase issue for systems with matched disturbances. It can be viewed as a method of integrating the sliding mode controller with another controller.

Practical control systems always have nonlinear plants, making the task of designing controls for nonlinear systems challenging. Thus, a variety of nonlinear control techniques, including integral sliding mode control, have been developed for nonlinear systems to get around the challenge of designing a controller for a real system.

In this work, the first section cover the modelling of the wind energy system [7] then we present the suggested strategy then we applied it on our system ,we add the triangular observer to make an approximation of the rotor acceleration needed to build the control law, and we present some simulations. Finally, a conclusion is reached from the results.

2. Wind Turbine Modeling

The variable speed wind turbine's aerodynamic torque is given by [8]:

$$\mathrm{Ta} = \frac{1}{2}\rho\pi R^3 \frac{Cp(\lambda,\beta)}{\lambda(t)} v(t)^2, \qquad (1)$$

where ρ is the density of air, R is the rotor radius, V is the wind speed , Cp is the power coefficient and $\boldsymbol{\lambda}$ is defined as:

$$\lambda\,(\mathrm{t}) = \mathrm{R}.\dot{w}_t \,/\, v\,(\mathrm{t}) \qquad (2)$$

The rotor's dynamics is determined by:

$$J_r \dot{w}_t = T_a - T_{ls} - k_r w_t \qquad (3)$$

J_r, k_r and w_t are sequentially, the rotor's inertia, external damping, and rotational speed.

The speed w_g of the generator is provided by the following expression:

$$J_g \dot{w}_g = T_{hs} \dot{-} k_g w_g - T_{em} \qquad (4)$$

With Jg, kg are the generator inertia and friction coefficient and T_{em} is the control law ofthe system. Expecting a perfect transmission, we define n_g :

$$\mathrm{n_g} = \frac{\mathrm{Tls}}{\mathrm{Ths}} \qquad (5)$$

3. Integral Sliding Mode Control

The maximum point on the power coefficient curve corresponds to the ideal amount of wind energy which corresponds to w_{topt}, where w_{topt} is given by :

$$y_r = w_{topt} = \frac{v(t).\lambda_{opt}}{R} \qquad (6)$$

The system in (Fig. 1) and (Fig. 2) is represented by:

$$\begin{cases} \dot{w}_t = w_g \\ \dot{w}_g = f(x) * \mathrm{w_g} + \mathrm{g}\ .u(x,t) \\ y_1 = w_t \end{cases} \qquad (7)$$

The control law u=T_{em}, y=w_t.

The errors of the system are provided by:

$$e_1 = w_{topt} - w_t \ , e_2 = \dot{w_{topt}} - \dot{w_t} \quad (8)$$

The sliding surface is defined as:

$$\sigma(t) = \lambda e_1(t) + e_2(t) + ki * \int_0^{\infty} e_1(t)dt \quad (9)$$

Such as ki is the sliding gain.
Assuming the function of stability given by [7]

$$v(t) = \frac{1}{2} * \sigma^2 \quad (10)$$

By using the following equation's derivative we have

$$\dot{V} = \sigma\dot{\sigma} = \sigma(\lambda(\dot{w_{topt}} - \dot{w_t}) + ki * e_1(t)) \quad (11)$$

The condition of stability

$$\dot{V} \leq 0 \quad (12)$$

This condition satisfies the following requirement

$$\lambda(\dot{w_{topt}} - \dot{w_t}) + ki * e_1 = \begin{cases} < 0 \text{ for } \sigma > 0 \\ = 0 \text{ for } \sigma = 0 \\ > 0 \text{ for } \sigma < 0 \end{cases} \quad (13)$$

$$Tem = Jg\left(-\lambda e_2(t) - \ddot{w_{topt}} + \frac{1}{Jg}(Ths - kgw_g) - kie_1(t)\right) - k\,sign(\sigma). \quad (14)$$

K is the designer's chosen sliding control positive gain.

3. Triangular Observer Structure

For the system model described in (7), the triangle observer that causes the observing errors to tend to zero in a finite amount of time (Fig. 3) can be represented as follows:

$$\begin{cases} \dot{\hat{w}}_t = \hat{w}_g + \lambda_1 sign(w_t - \hat{w}_t) \\ \dot{\hat{w}}_g = \dot{w}_g + \lambda_2 sign(\tilde{w}_g - \hat{w}_g) \end{cases},$$

where

$$\tilde{w}_g = \hat{w}_g + \lambda_1 sign_1(w_t - \hat{w}_t)$$

and

$$\tilde{w}_t = w_t$$

Such as $sign_i$ is characterized by:

$$sign_i(..) = \begin{cases} 0 & \text{if } \tilde{x}_j - \hat{x}_j \neq 0 \\ & j \in [1, i] \\ sign(..) & \text{elsewhere} \end{cases}$$

4. Simulations

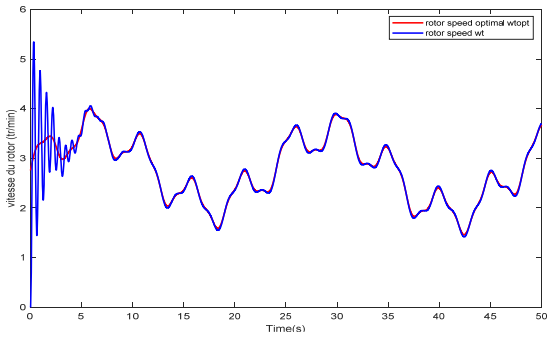


Fig. 1. Responses of the calculate and desired (optimal) rotor speed.

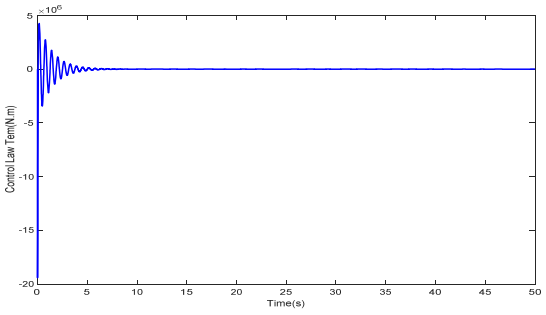


Fig. 2. The response of the control law with integral sliding mode control.

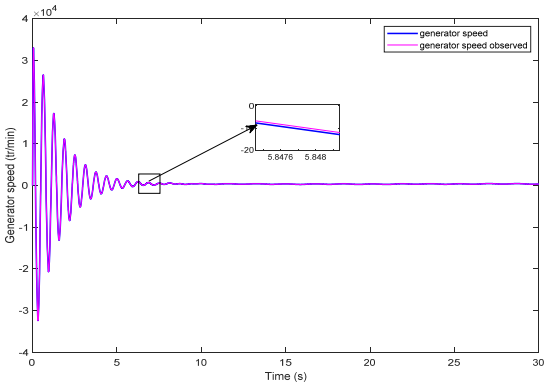


Fig. 3. The generator speed and the generator speed observed.

5. Conclusion

In this article, we have discussed the operating mode which is related to the integral sliding mode control (ISMC) for the two mass model system. The results of the simulations show the difference between the two methods. The stability of the system is guaranteed by using Lyapunov approach. In this optics, we may make and search for some approximations to certain parameters and gains in order to enhance the performance and efficiency of our system.

References

[1]. A. Sellami, D. Arzelier, R. M'hiri, J. Zrida, A Sliding Mode Control Approach for Systems Subjected to a Norm-Bounded Uncertainty, *International Journal of Robust and Nonlinear Control,* Vol. 17, Issue 4, 2007, pp. 327-346.

[2]. A. Aissaoui, A. Tahour, M. Abid, N. Essounbouli, F. Nollet, Power Control of Wind Turbine based on Fuzzy Controllers. *The Mediterranean Green Energy Forum (MGEF-13),* 2013, pp. 163-172.

[3]. C. Chatri, M. Ouassaid, M. Labbadi, Y. Errami, Integral-type terminal sliding mode control approach for wind energy conversion system with uncertainties, *Computers and Electrical Engineering*, 2022, p. 107775.

[4]. R. Naoual, E. M. Mellouli, S. Sefriti, I. Boumhidi, Fuzzy sliding mode control for the two-link robot, *International Journal of Systems, Control and Communications*, 2014, pp. 84-96.

[5]. E. M. Mellouli, M. Alfidi, I. Boumhidi, Fuzzy sliding mode control for three-tank system based on linear matrix inequality, *International Journal of Automation and Control*, 2018, pp. 237-250.

[6]. E. M. Mellouli, S. Sefriti, I Boumhidi, Combined fuzzy logic and sliding mode approach's for modelling and control of the two link robot, *IEEE International Conference on Complex Systems (ICCS)1-6,* 2012, pp. 1-6.

[7]. E. M. Mellouli, Z. Chalh, M. Alfidi, A New Robust Adaptive Fuzzy Sliding Mode Controller for a Variable Speed Wind Turbine, *International Review of Automatic Control,* 8, 5, 2015, pp. 338-345.

[8]. J. Mérida, Luis T. Aguilar, J. Dávila, Analysis and synthesis of sliding mode control for large scale variable speed wind turbine for power optimization, *Renewable Energy*, 2014, pp. 715-728.

(0912)

Collaborative Environment based on Strategic Engineering to Improve Efficiency and Reduce Vulnerabilities of Marine and Underwater Critical Infrastructures

Agostino G. Bruzzone [1,2], Marina Massei [1,2] and Antonio Giovannetti [1,3]
[1] STRATEGOS, Genoa University, via Opera Pia 15, 16145 Genova, Italy
[2] Sim4Future, Simulation Team, Trento 34, 16145 Genova, Italy
[3] Simulation Team, Savona Campus, Genoa University, via Cadorna 2, 17100 Savona, Italy
Tel.: +39 010 353 2275
E-mail: agostino@itim.unige.it, URL: www.itim.unige.it/strategos

Summary: This paper proposes an innovative architecture based on the innovative discipline called Strategic Engineering devoted to improve Performance and reduce Vulnerabilities within Marine Critical Infrastructures and Off Shore Facilities. The approach relies on Simulation Models and AI (Artificial Intelligence) to reproduce autonomous systems, UxVs (Unmanned any domain Vehicle), operating over multiple domains (i.e. UAV, USV, ROV & AUV) in joint cooperation with Traditional Assets, Sensor Grids and Industrial Process Systems & Components. The simulation involve extensive use of Intelligent Agents (IA) driving actions and reactions of all the Units, Elements and Process Models to reproduce and react to potential Accidents, Operational Problems, Menaces both on physical and cyber layers, including Threats represented by unmanned and manned Antagonists (i.e. underwater, surface and aerial threats) as well as Friendly, Suspect, Neutral and Hostile Units. The Synthetic Environment is created as Interoperable, Immersive and Intuitively thanks to innovative eXtended Reality (XR) allowing also to interact with 3D Printed Models representing some of the Assets to guarantee immediate understanding of the scenario by Decision Makers. The paper proposes a case study of using this solution for a dual use scenario addressing underwater cable and pipelines as well as offshore facilities.

Keywords: Strategic engineering, Modeling, Interoperable Simulation, AI, Intelligent agents, UAV, USV, AUV, Collaborative autonomous systems, Off shore Infrastructures, Underwater critical infrastructures.

1. Introduction

The current geopolitical situation, due to international tensions, highlights the circumstance of some infrastructures. These infrastructures are now fundamental pillars of our energy/information intensive society and represent plants and facilities to be maintained efficiently even in hard conditions as well as to protect by reducing their vulnerabilities [1-3]. For what the sea in concerned, these aspects turn in even more critical elements considering the challenging environment that the Ocean represents, due to adverse possible weather conditions and difficulties related to big Deep Sea Operations for Oil & Gas, Mining and UW Cables (Underwater) [4]. The Off Shore Plats and Infrastructures require high Efficiency, Safety and Resilience. Therefore, Strategic Infrastructures are concentrated also on coastal areas: Ports & Terminals, On Shore LNG services (Liquid Natural Gas), Power Plants, Desalination Plants, Refineries and Chemical Plants, Iron & Steel Facilities; this is due by the need to be connected to Logistics Services, Shipping Lines for Natural Resource Supplies [5, 6]. In addition to accidents and natural disasters, other major challenges rise due to the current political situation, in terms of Threats both by terrorists and individuals (e.g. hackers, piracy) and by major governmental organizations acting within different war paradigms such as conventional, cyber or hybrid warfare [7]. While previously the most devastating threats were represented by physical attacks, today cyber-attacks and autonomous systems emerge as new major threats that could generate both severe direct damages to critical infrastructures as well as indirect damages compromizing control systems & availability (e.g. hijacking ships or other assets in kamikaze actions, cyber block of operations) [8]. The economical convenience/flexibility of new UxV (Unmanned any domain Vehicles) as well as the necessity to operate in very critical conditions for humans (e.g. deep sea operation under 300 meters) push forward the use of autonomous and robotic systems as well as the digitalization of infrastructures with automatic and remote controls (e.g. off shore wind farms); all these systems are potentially vulnerable to the new threats. These aspects suggest the necessity to increase resilience on multiple domains, physical, cyber, underwater, sea surface, air, space and coastal areas by creating synthetic environment where to experiment the interoperability among different digital twins of different systems and platforms respect the potential threats and addressing new needs [3]. This paper proposes an innovative solution based on Strategic Engineering devoted to create a comprehensive virtual scenario to increase operational efficiency and create resilience within Marine Critical Infrastructure and Off Shore Facilities.

2. Strategic Engineering and Application Case

The Architecture of this solution it is based on the combined use of different methodologies such as

Modeling & Simualtion (M&S), Artificial Intelligence (AI) and Extend Reality (XR) in order to protect Critical Infrastructures such as UW Fields, UW cables and Pipelines, Off-shore platforms respect to different assets (e.g. Autonomous, Cyber and Conventional Asstes) both as Threats and Defenders operating over Multiple Domains. The general architecture include multiple models of different assets as well as IA (Intelligent Agents) to drive them within a Stochastic Discrete Event Simulation to investigate different configurations, decisions, tactics and strategies on a Virtual World. This system, titled CAPIAS (CAbles, Pipelines, marine Infrastructures & Autonomous system: protection & Simulation),, is created as Interoperable Simulator ready to be combined with other Models and even real systems in order to support different applications (e.g. Decision Making, Capability Assessment, Engineering Education, Training) respect different Scenarios in which to develop and test tactics, strategies and new system to develop an overall Resilence. The strategic importance of marine and underwater infrastructures both in the industrial and defence sector, have always made them subject to different types of attack. These structures provide a range of essential services, from energy production to transport and communications, so they play a key role in supporting Economic Development, Country Competitiveness and National Security. Offshore Wind Farms, Oil & Gas Platforms, and Underwater Pipelines are essential components of the energy production and distribution system, providing reliable sources of Energy. Underwater Data Cables result even more strategic considering their importance, confirmed by the fact that more than 95% of international data and voice traffic, including nearly 100% of transoceanic Internet traffic, is carried by submarine cables [9]; this obviously is due to the gigantic volume of data and the much more broad bandwidth of cable communications respect Satellite capabilities. In the military field, marine and underwater infrastructures play a key role in supporting National Security, including Surveillance and Defense operations. These structures are vulnerable to a wide range of threats, including natural disasters, technical failures as well as malicious attacks. However, the vulnerabilities of these structures are no longer covered just by physical attacks, but also those on the cyber layer and soft kill approaches. For these reasons it is necessary to find innovative solutions that are effective both on the physical and cyber layers. Just thinking of the networks for Oil Pipelines and Submarine Cables, it emerges that these lines represent strategic assets of vital importance for World Economies. Marine and submarine critical infrastructures represent Complex Systems integrating many different sub systems and components that the vulnerability of the entire system is affected by direct and indirect vulnerabilities of its individual elements. Off Shore Platforms, Ports, Coastal Refineries and Gas Facilities are critical infrastructures which over the years have been subject to attacks such terrorism and as piracy [10]. Although these infrastructures are protected by international legal standards [11], it is necessary to develop new capabilities and solutions able to increase resilience and decrease vulnerabilities in order to mitigate the possible damages that would affect the economy of entire Countries. The simulation of attacks and countermeasures turns out to be important in order to prepare operators and decision-makers to manage critical situations. Through the digital twins of real critical assets, in facts, it is possible to carry out a threat assessment and evaluate the impact of different course of actions (COA) [12]. Obviously this requires to introduce in the models not only Technological Systems and Platforms, but even their behaviors in order to develop the possibility to carry out several virtual experiments without relying just on humans; obviously Decision Makers choices are important and CAPIAS allows them to interact intuitively during the simulations, therefore to simulate these scenarios it is necessary to address low level decisions and actions that it could be better to delegate to IA and AI to make it easy and fast to conduct complex tests. Indeed, in this context, it becomes necessary to adopt innovative solutions that integrate information extracted from real data from the field, by processing them through modern technologies of Data Analytics and Artificial Intelligence. By this approach is possible to carry out much reliable Agent Driven Simulation based on realistic models, while analysis of the results as well as the feedback from the field of new implemented strategies allows, by Machine Learning (ML) to correct the Algorithms and to tune the models. This innovative methodological approach is called Strategic Engineering and it turns in a critical approach for supporting Decision Makers in critical situations [13].

3. Strategic Engineering at Work

Indeed, Strategic Engineering is a new discipline that use in combined way Modeling and Simulation, Data Analytics, AI & IA in closed loop with big data arriving from Digitalized Systems and Sensor Grids in order to dynamically refine the models and algorithms while supporting the Decision Makers [14]. This system and systematic approach allow us to know what happened in the past and what is on going now in order to get a holistic picture of the present. The models created are used to simulate the situation and the impacts of our decisions on the future as well as to consider the possible alternatives in terms of scenario changes and other players' reactions and moves. The real impacts of the decisions on the field, measured after their implementation, allows to dynamically refine the models and hypotheses by using Machine Learning techniques. The Simulation Team is nowadays active in developing R&D Projects (e.g. ALACRES2, CAPIAS, ACTA EX MORE) and Educational & Training Initiatives (e.g. STRATEGOS at Genoa University, International PhD Program in Strategic Engineering) in this framework by leading cooperation with Institutions, Agencies and Companies at International level [13].

In this way, it is possible to create advanced digital twin enabled to interoperate among themselves as well as to be simulated into a Federation of models driven by IA to reproduce realistic behaviors reactive to situation evolutions and other actor actions; it is very important to outline also the necessity to develop young engineers and scientists as well as to introduce managers, officers and executives in this new mind-set that enable the use of these new capabilities.
The Strategic Engineering approach is based on the idea to use Models and AI to create synergies among Decision Makers, Commanders, Executives and Engineers [7, 13]. In CAPIAS due to Strategic Engineering Approach, it is possible to move from Data, simulating Heterogeneous Networks, Sensor Grids & Digital Systems, with all their Inconsistencies & Problems, up to extract Valuable Information and create a Clear View and a Valid Scenario Awareness (Fig. 1). Therefore this is not enough to finalize decisions that should consider impact of our moves as well as possible reactions of other players and Systems. This turn possible thanks to use of Modeling & Simulation (M&S) by analyzing the different COAs (Course of Actions), considering impacts of boundary conditions, reactions of all actors as well as influence of stochastic factors; the evaluation of the KPIs (Key Performance Indicators) it turns possible to choose most promising decisions and to develop Knowledge on the on going situation. Indeed, the Artificial Intelligence and M&S create a valuable Knowledge that, by interacting with Humans, lead to Wisdom necessary for achieving Strategic Victory, real Resilience and supporting prevention & Mitigation of new threats.

4. CAPIAS Solution

In this paper we propose CAPIAS (CAbles, Pipelines, marine Infrastructure & Autonomous system: protection & Simulation) Solution, developed by SIM4Future, Simulation Team in collaboration with STRATEGOS, Genoa University. Indeed. CAPIAS is an innovative solution to address activities Off Shore and Underwater related to critical infrastructure and strategic assets based on collaborative use of traditional assets and UxVs (Unmanned any domain Vehicle) including UAVs (Unmanned Aerial Vehicles), USVs (Unmanned Surface Vehicles), AUVs (Autonomous Underwater Vehicles) as well as ROV (Underwater Remotely Operated), UW Gliders, etc. CAPIAS use extensively innovative XR (eXtended Reality) to propose to the decision makers an immersive intuitive whole interactive simulated world with Industrial Plant Operations, Accidental Failures as well as restoring activities, defensive and hostile threats at sea, underwater, in air, space and cyberspace. XR, combining AR & VR, makes easy to capture the whole scene and develop awareness on the elaborations provided by Data Fusion, Data Analytics, AI and Models.

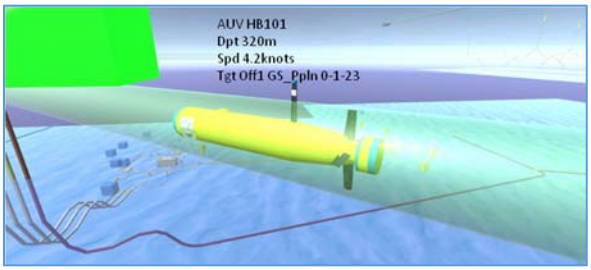


Fig. 1. Virtual AUV over an Off Shore Oil & Gas Field.

CAPIAS provides to Decision Makers even the possibility to interact with 3D Printed Models of major assets by Augmented Reality (AR); this solution allows to get a more intuitive and immediate understanding of the scenario, a full awareness of operational space and time scale. Indeed the 3D Printed Models are in same scale dimensions of the Virtual World so it allows to overlap the systems and to learn quickly intuitive commands by AR. XR solution enables to create a collaborative environment among users; indeed, CAPIAS is currently operating over Laptop, Hololens, VR Headsets, SPIDERs (Simulation Practical Immersive Dynamic Environment for Reengineering), Caves (Cave Automatic Virtual Environment) as well as on Smart phones, while even other approaches could be activated. Vice versa interoperability with real systems and other models is enabled by the adoption of High Level Architecture (HLA) Standard [14]. Obviously to make this approach effective and intuitive is necessary an extensive use of Models and AI that take care of detailed operations; for instance if a failure affect a Subsea Trees, the CAPIAS IA directing operations of an Off Shore Platform could decide autonomously to direct over there a support boat with a ready team of scuba divers in a special hyperbaric chamber or to send another boat with Underwater ROV; these AI could check possibility to operate a subsea resident ROV in the area or to send some AUV and then assign tasks and orders to the different IA of the selected system to carry out operations. CAPIAS demonstrative scenario addresses dual use (i.e. industrial and military operations) aspects respect Critical Infrastructures involving the Oil & Gas Operations of two Offshore Platforms, their Subsea Fields, some UW Pipelines, UW cables. The whole scenario is Unclassified and approved for unlimited public release, therefore it could be easily adapted to include data, behaviors and models subjected to restrictions for more realistic specific analyses. In CAPIAS, the different Models and AI guarantee the capability to reproduce collaborative and competitive interactions among different assets (traditional and autonomous systems) over the different domains In figure 3, it is proposed the use of XR, Simulation and AI with 3D Printing to make it easy, intuitive and effective to deal with the Challenges of the Multidomain Marine Scenarios to protect Strategic Assets such as underwater Cables, Pipelines and Off Shore Resources.

CAPIAS has been verified, validated and tested over a specific demo scenarios including 2 Offshore Platforms, 1 Oil Pipeline, 1 Gas Pipeline, 1 UW Data Cable and an Offshore Wind Farm; in addition to these strategic assets there hundreds of entities including boats, ships, vessels, submarines, divers, UxV, Helicopters including public demonstration. CAPIAS experimentation involved hundred test users and experts by interfacing them with the virtual world by Smart Phones, Workstations as well as by Hololens and 3D Printed models (i.e. Thaon di Revel Class, Todaro Class Submarine and an one of the two Off-Shore Platforms, courtesy of Crowd 3D). In figures 2a/2b it is proposed the synthetic environment of the area around the two Off Shore Platforms.

4. From Data to Wisdom

CAPIAS operates within Extended Maritime Frameworks (EMS, including Sea Surface, Underwater World, Air, Coastal Areas, Space & Cyber Space) and it includes:

- Off-Shore Platforms, FPSO and UW Fields with their Service Requests, Failures as well as support Assets and could be affected by Natural & Human Hazards. The Simulated Platforms include communications, Cyber Layer, Security Resources, Diving Bells, Divers, On board RHIB and Boats, ROV, AUV, UUW, UAV, Helicopters, etc.
- Underwater Assets such as Oil & Gas Pipelines, UW Fields with their requests to carry out Inspections, Service and Maintenance Plans as well as to simulate impact of Accidents and Antagonist Actions
- UW Cables for Communications and the possibility to consider different Threats, Inspections and other Actions
- Dynamic Models of Underwater communications, acoustic modem, cable data transfer, data package exchanges, sensor grids, heterogenous Networks, service operations and failures
- Marine Energy Sources such as Off-Shore Wind Farms to be considered as Assets to Protect
- Different Support Boats and Vessels including carriers for Diving Bell, ROV and other supports are part of the scenario as well as other surface entities external to Off-Shore Operations, but crossing the area
- Ships are part of the Scenario just as background traffic, but even as potential threats due to cyber actions such as GPS spoofing, hijacking or simple accidents due to failures or improper anchoring
- Innovative Solutions and Proposals for guarantee persistent surveillance and proactive readiness are part of the CAPIAS simulation. For instance it is reproduced a collaborative use of an Autonomous Systems using an USV SWATH (Small Waterplane Area Twin Hull) with a *revolver* carrying up to 3 AUV and a Catcher to recover them on a bay on its belly even during critical weather conditions,
- Navy Vessels with their capabilities to defend the Marine Strategic Assets directly as well as in cooperation with other entities; in our test the Users were able to assign them orders also interacting with 3D Printed Models by XR and AR Goggles
- Helicopters and Drones for Off-Shore Operations Submarines, Autonomous Systems that could be operated also to investigate possibilities to delivery and recover AUV as well as to face challenges related to Protection of Strategic Marine Assets as well as Naval Force Assets such as Rotary Wing Aircrafts and fixed wing UAV
- Submarines and UW Autonomous Systems could be operated also to investigate new solutions
- Satellites dealing with communications and services such as GPS
- Cyber elements
- Multiple potential Hostile Threats are included in terms of Aerial, Surface, Submarine and Cyber Threats
- Other external elements that could create confusion or need to be protected (e.g. general traffic, recreational craft, marine life)

CAPIAS simulates all communication over the available Heterogeneous Networks (e.g. Acoustic Modems, RF, Skynet, Link 16, etc.) as well as data package exchanges. All entities and elements have their Cyber Correspondent and are interconnected considering their Availability, Integrity and Confidentiality as well as the Cyber Ware actions to defend or compromise dynamically these characteristics

The Entities and Networks in CAPIAS are correlated and the alternative choices and actions, such as Hybrid Threats, could be evaluated in terms of impacts on different layers in critical cases such as these examples:

- Downgrading Acoustic Modem Communications avoid Awareness of Underwater Detections
- Compromising ICT System Integrity to disable C2
- GPS spoofing to transform a Cargo Ship into a torpedo against a Wind Farm
- Swarms of UAVs and AUVs coordinated attacks to Oil & Gas Rigs synchronously with Cyber Attacks
- Contamination of the Food Chain for Off-Shore Platforms
- Menaces to Data cables Generated by ROV or Divers, as well as by simple medium size boats towering modified dragnets or ad hoc threats.

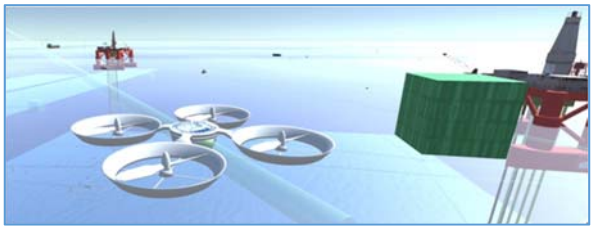


Fig. 2. Testing and demonstrating the CAPIAS Solution using XR and 3D Printed Models.

Indeed UxV and Hybrid Threats are becoming increasingly prevalent in modern conflict scenarios, presenting significant challenges for military decision-makers. To effectively counter these threats, decision-makers must be able to quickly and accurately evaluate a wide range of options and determine the best COA over multiple domains as proposed in Figure 2. In addition it is crucial to develop new solutions for maintenance in underwater environments: a very complex and challenging sector where it is necessary to evaluate a wide range of engineering and operational solutions to determine the best oes; these activities require highly trained personnel at different levels and are subject to adverse weather conditions that could benefit from M&S. In addition, off shore activities must be carefully designed, operated and executed considering also their impacts and hazards related to surrounding ecosystem. CAPIAS reproduces Off-Shore Platforms and UW Fields with their Service Requests, Failures as well as support Assets and could be affected by Natural & Human Hazards; indeed it provides users with a deep understanding of the capabilities and limitations of different systems, related maintenance strategies, allowing them to finalize convenient decisions respect clear risks, improving resilience and safety.

Fig. 3. Testing and demonstrating the CAPIAS Solution using XR and 3D Printed Models.

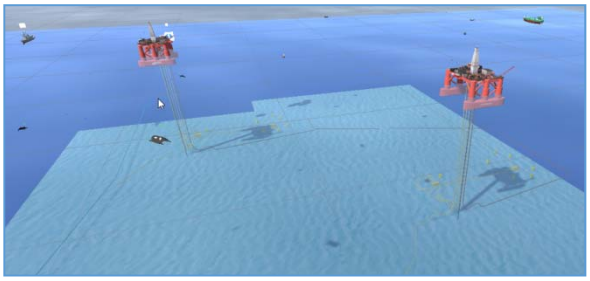


Fig. 4a. CAPIAS Scenario involving multiple assets and two Off Shore Platforms with their UW Fields.

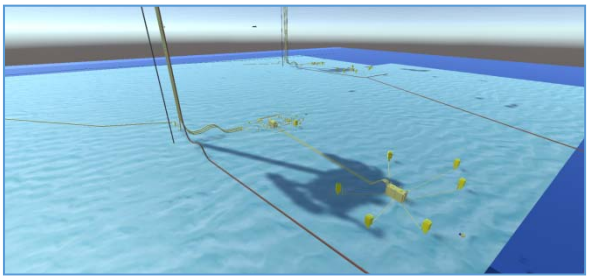


Fig. 4b. CAPIAS Scenario view of seabed extraction fields under the Off Shore Platforms

5. Conclusions

CAPIAS is a very innovative approach that make Intuitive and Interactive Multi Domain Scenarios addressing Vulnerability and Maintainability of Strategic Marine Assess.

This study demonstrates the value of combining within a Strategic Engineering approach an Innovative Simulator able to evaluate the effectiveness and efficiencies of different modalities for Surveillance and Maintenance of Off-Shore Critical Infrastructure. The system support also the analysis of alternative Configurations to prevent Vulnerabilities as well as identification of most promizing COAs for mitigation of critical situation in marine and underwater infrastructures. CAPIAS provides actually the users with a valid comprehensive understanding of the capabilities and limitations of different tactics, strategies respect processes, maintenance services, countermeasures and surveillance solutions.

Indeed the combined use of M&S, M&S and AI/IA allows to Virtually test new Solutions including new Assets, Concepts, Procedures and Strategies considering their Feasibility, Efficiency, Effectiveness, Reliability & Robustness as well as Resilience. Authors are currently working to further extend CAPIAS and to Validate, Verify and Test it over new scenarios.

References

[1]. Stefanova, B., European Strategies for Energy Security in the Natural Gas Market, *Journal of Strategic Security*, 5, 3, 2012, pp. 51-68.

[2]. Mastrangel, E., Overview of US Legislation and Regulations Affecting Offshore Natural Gas and Oil Activity. in Intellectual takeout, US Department of Energy, Energy Information Administration, *Technical Report* of *Office of Oil and Gas,* 2005, p. 20.

[3]. Bruzzone, A. G., Fontaine, J., Berni, A., Brizzolara, S., Longo, F., Dato, L., Poggi S., Dallorto M., Simulating the marine domain as an extended framework for joint collaboration and competition among autonomous systems, in *Proceedings of the DHSS*, Athens, Greece, 2013.

[4]. Santos, M. M., Jorge, P. A. S., Coimbra, J., Vale, C., Caetano, M., Bastos, L., et al., The last frontier: coupling technological developments with scientific challenges to improve hazard assessment of deep-sea

mining, *Science of the Total Environment,* V. 627, 2018, pp. 1505-1514.

[5]. Jenkins, Brian Michael, Potential Threats to Offshore Platforms, *RAND Corporation,* 1998, pp. 7406.

[6]. Prodan, T., Kasum, J., Stošić, M., & Ugrin, Ć., Security Challenges and Guideline Proposals for the Development of Underwater Security, *National Security and the Future*, Vol. 20, Issue1-2, 2019, pp. 71-84.

[7]. Bruzzone, A. G., & Massei, M., Simulation-based military training, *Guide to Simulation-Based Disciplines: Advancing Our Computational Future*, 2017, pp. 315-361.

[8] Lehto, M., & Hutchinson, B., Mini-drones swarms and their potential in conflict situations, in *Proceedings of the 15th International Conference on Cyber Warfare and Security* , March 2020, pp. 326-334.

[9]. Green, M., Brunett D., Security of International Submarine Cable Infrastructure. Time to Rethink? *Legal Challenges in Maritime Security,* 2008, pp. 557-584.

[10]. Bakir, N.O., A brief analysis of threats and vulnerabilities in the maritime domain, *NATO Science for Peace and Security Series C: Environmental Security. Managing Critical Infrastructure Risks,* 2007, pp. 17-49.

[11]. Wrathall L.R., The Vulnerability of Subsea Infrastructure to Underwater Attack: Legal Shortcomings and the way forward, *San Diego International Law Journal*, 223, 2010.

[12]. Bruzzone A.G., M. Massei, G. L. Maglione, R. Di Matteo, G. Franzitti, Simulation of manned & autonomous systems for critical infrastructure protection, in *Proceeding of the International Defence and Homeland Security Simulation Workshop,* 2016.

[13]. Bruzzone A. G., M. Massei, K. Sinelshchikov, A. Giovannetti, B. Gadupuri, Strategic Engineering Applied to Complex Systems within Marine Environment, in *Proceedings of the Annual Modeling and Simulation Conference,* 2021, pp. 1-10.

[14]. Vinas, A. C., Mansfield, T., Sobrino, P. C., & Tremori, A., Towards a Modelling & Simulation capability for training autonomous vehicles in complex maritime operations, in *Proceedings of the NATO Modelling & Simulation Group (NMSG) symposium–Towards Training and Decision Support for Complex Multi-Domain Operations (MSG-184)*, Amsterdam, 2021.

(1079)

Transition towards IEC-61499 in Factory Environment through a Learning Factory Approach

M. Foletti, P. Pedrazzoli, M. Confalonieri, A. Ferrario, F. Daniele and L. Agbomemewa

University of Applied Sciences of Southern Switzerland (SUPSI), Institute of Systems and Technologies for Sustainable Production (ISTePS), Via la Santa 1, 6962 Lugano-Viganello, Switzerland
Tel.: 41 (0)58 666 66 79
E-mail: michele.foletti@supsi.ch

Summary: The IEC 61499 standard addresses issues present in the IEC 61131-3 such as feedback connections and communication between different vendors. It provides a generic model for distributed systems with the goal of portability, reusability, interoperability, and reconfiguration of distributed applications. However, the standard has not yet been widely adopted in real factory environments due to various challenges. To address this, a stepwise integration approach through a learning factory is presented. The aim is to understand the advantages and challenges of the integration, particularly from a teaching perspective. To do this, automation students use a smart factory as a testing environment, where they implement the control of a new productive device, starting with 61131 and progressively moving towards full adoption of the 61499. Required hardware, software and skills are examined, and questionnaire results provided, where it's shown which automation topic is better approached by which standard. The results give a first look at how to structure an articulated course for an effective teaching of the subject, paving the way for the formation of new automation engineers who will promote the standard in the factories of the future.

Keywords: Programmable logic controllers, Implementation of IEC-61499 standard, Distributed automation and control systems, Smart learning factory, Smart automation, PLC, softPLC, 4DIAC, Revolution PI.

1. Introduction

The basic software architecture and programming language within programmable logic controllers (PLC) is today almost universally defined through the IEC 61131-3 [1]. Most PLCs are based on this standard, which defines a set of programming languages and associated development tools for creating control systems, including the following five languages: 1) Ladder diagram (LD): a graphical language used to create control logic diagrams that resemble electrical ladder diagrams; 2) Function block diagram (FBD): a graphical language used to create control logic diagrams that resemble block diagrams; 3) Structured text (ST): a textual language similar to the Pascal programming language, used to write control programs with a high level of structure, 4) Instruction list (IL): a low-level, machine-oriented language used to write control programs with a low level of structure, 5) Sequential function chart (SFC): a graphical language used to create control logic diagrams that resemble flowcharts. The standard also defines the rules for implementing and executing the programs written in these languages, and the method for interconnecting them. It also defines the development environment, including the editor, compiler, and debugger, and the requirements for the runtime environment, including the execution of control programs and the handling of I/O data.

However, two unresolved issues still stand: 1) Feedback connections in the application, and 2) Communication between different vendors [2]. By trying to sort these problems out, the IEC 61499 standard provides a generic model for distributed systems [3]. It aims at portability, reusability, interoperability, reconfiguration of distributed applications and its usefulness has been already proved by several research efforts.

Nevertheless, applications in real factory environment remain a rarity. The reasons are numerous and range from the upfront costs for changing technology in already existing and working productive lines, software and hardware dependences, and unclear industrial benefits due to the lack of a widely accepted engineering methodology for the IEC61499 standard adoption.

This work proposes, through a learning factory approach, a stepwise integration for the adoption of this new automation paradigm, in order to understand how and what is necessary to effectively accomplish the 61499 implementation, showing challenges and advantages for each step. The playground is the SUPSI MiniFactory [4], a laboratory-sized smart factory, where students, researchers and companies can meet, enabling knowledge transfer through practical approach.

Learning factories are a highly appreciated and proven way to teach and transfer various topics, with the possibility of using actual devices and machines, so that knowledge does not come only from theoretical lesson but is also gained through practical experiences.

Fig. 1 shows the SUPSI MiniFactory, where a lab-sized factory environment is reproduced and where actual implementation and testing can take place.

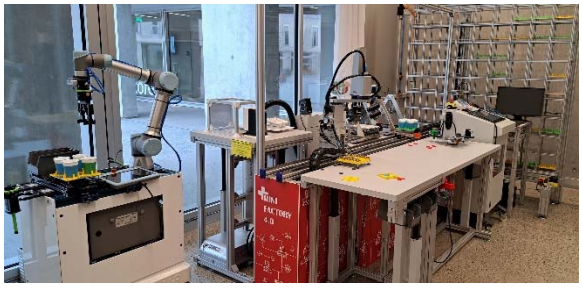


Fig. 1. The SUPSI MiniFactory, a smart learning factory where student can put into practice what learned during theoretical lectures.

2. State of the Art

Networked automation devices, as needed for Industry 4.0 or Cyber Physical Production Systems, demand for new programming languages like the one defined in the IEC 61499 standard [5]. IEC 61499 is an international standard for distributed industrial automation systems, developed by the International Electrotechnical Commission (IEC). The standard provides a generic model for distributed systems, aiming at portability, reusability, interoperability, and reconfiguration of distributed applications.
IEC 61499 is based on the function block (FB) concept, which allows for the modularization and encapsulation of functions and the creation of reusable function blocks. It also provides a standard communication model for inter-FB communication and a standard method for describing the behavior of FBs. The standard also includes a standard method for describing the behavior of FBs, which is based on state machines, and a standard method for describing the structure of FB networks.

The standard has been widely researched and has been found to be useful in various industrial automation applications such as control systems, robotics, and smart grids. Many research efforts have been made to implement and evaluate the IEC 61499 standard in real-world industrial environments [6]. For example, studies have shown that the IEC 61499 standard can improve the flexibility, scalability, and reconfigurability of industrial automation systems. Additionally, some studies have demonstrated that the standard can improve the reuse of functional components in industrial automation systems, which can reduce development costs and increase system reliability [7].

However, despite its usefulness, the adoption of the IEC 61499 standard in real-world industrial environments remains limited. This is due to various reasons such as the upfront costs for changing technology in already existing and working productive lines, software and hardware dependencies, and lack of a widely accepted engineering methodology for its adoption. In recent years, many efforts have been made to address these challenges and to promote the adoption of the IEC 61499 standard [8]. For example, some researchers have proposed the use of learning factories as a tool for education and innovation in the adoption of the standard [9]. Additionally, some researchers have proposed the use of a stepwise integration approach to adopt the standard [10].

Overall, the IEC 61499 standard has shown to be a useful tool for distributed industrial automation systems, but it still needs to be adopted widely in real-world industrial environments. Research is ongoing to address the challenges and to promote the adoption of the standard [8].

3. Methods

The proposed approach replicates an industrial case where a new machine needs to be integrated in a productive line of a factory, hence existing process parameters, communication protocols and mechanical constraints must be considered and eventually adapted. The challenge is accentuated by the high degree of automation in a context of highly flexible (mass)-production, and the heterogeneity of the machines and devices, which need to coexist and cooperate. In particular, the existing line produces and assembles customized coffee mugs, where both classical and advanced industrial devices are used (e.g., linear motor system for products movement, cartesian robot for warehousing, SCARA robot for pick&place operations, smart cameras for precise positioning and quality control, collaborative robot for human-machine assembly operations, and 3D printers and an injection moulding machine to produce components). Various PLCs from different vendors oversee machines and the sequence of operations, supported by numerous communication protocols (EIP, OPC-UA, MQTT, TCP/UDP, Modbus). The new machine, a CO2 laser marker, which need to be integrated into the line, is used for the customization of the products, and is served by a collaborative robot.

A learning factory approach is used, where students in automation, who have already familiarized with the theoretical aspects of PLCs programming, both with IEC61131 and IEC61499, can now put into practice their experience. Three 4-hours sessions are dedicated for implementing the device at various level of integration: 1) Using 61131 only (to be used as base reference); 2) A hybrid model where 61131 and 61149 coexist; 3) Full 61499 adoption. A final 2-hours slot is then dedicated for results comparison and fulfilment of a questionnaire, so that quantitative results about challenges, advantages, required resources and skills needed can be investigated.

4. Implementation

For each level of integration, hereby indicated as use case 1 to 3, a briefly description of the implementation is provided.

4.1. Use case 1

The use case is implemented through classical standard PLC programming, i.e., based on cycle scan

and through ladder code and structured text. In this case, a dedicated standard PLC from Omron with its I/O functionality is used. By using an identical PLC for the management of the rest of the Mini-Factory, the integration of such new device is quite straightforward. In fact, the two PLCs can easily communicate via Ethernet/IP or OPC-UA. Of course, there's the need to be already familiar with the software used to program the PLC. The laser marker is just able to communicate via a socket connection, but already existing function blocks managing this type of protocol are available, allowing an easy and fast integration with its dedicated PLC. In this scenario the dependence of specific hardware is very high: implementation on a different line or using different PLCs would mean setting up again everything from scratch, using different function blocks or semantic.

Fig. 2 shows an example of function block programmed via structured text used to control the laser marker. By using a case-structure is possible to easily control the process flow. However, changes still require quite a lot of effort, e.g. implementing an "error handler" or a "timeout handler" would require adaption to the code which are very error prone.

```
CASE Stage OF
    1:
       mystring:='GetStatusCode;';
       sendSktCmd:=TRUE;
       Stage:=2;
    2:
       ;//wait cmd execution
    3:
       IF laserResponse='0' THEN //ready
          Stage:=4;
       ELSIF laserResponse='-1' THEN //keyswitchopen, interlock, etc
          delay:=TRUE;//error, need operator
       ELSIF laserResponse='1' THEN //marking
          delay:=TRUE;
       ELSE
          delay:=TRUE;//in case of no responses
       END_IF;
    4:
       mystring:=CONCAT('SetJob:',JobName,';');
       sendSktCmd:=TRUE;
       Stage:=5;
```

Fig. 2. Part of the code controlling the laser, written using standard IEC61131 function block and structured text.

4.2. Use case 2

In this second scenario, 61131 and 61499 coexist. A softPLC running on a RevolutionPI is used to run code based on 61499 and programmed via 4DIAC. The softPLC controls the laser, leaving the rest of the logic still on the main PLC. Controlling the laser is quite easy as we can run C# code with any additional library (e.g., for socket connection) and the event-based programming is intuitive and easy to debug. The communication with the main PLC can again be established via OPC-UA. The main disadvantage is the need of continuous communication between the two "worlds" (e.g., between the laser and the collaborative robot), which can become quite complicated and error prone. The potential of event-based programming is only partly exploited. The advantage is that now the laser with its hardware-independent controller can be easily integrated into another factory, whereas communication protocols are re-established. Fig. 3 shows the hardware controller cabinet, where we can find PLC (Omron) and softPLC (RevolutionPI) as well as various I/O for controlling devices.

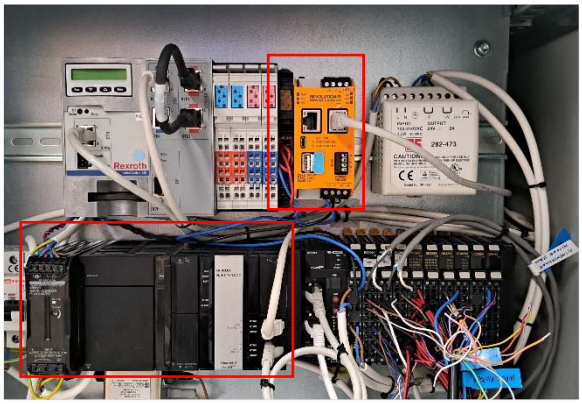


Fig. 3. PLC based on cycle-based execution (in black, bottom left) and softPLC event based (orange, top middle).

4.3. Use case 3

For the last scenario, a full 61499 implementation is performed, leaving only some basic functions still running through 61131 (e.g., ad hoc PLCs controlling specific devices). Of course, lots of resources need to be invested to convert the old logic to the new one. The implementation can happen while still running the system, but deployment and validation would require a complete stop of the factory. The advantage is that integration of future new devices, based on 61499, is straightforward and the full potential of event-based programming is used. Operations sequencing is easy to be performed as the graphical interface show the events flow in a very intuitive way.

Fig. 4 shows a portion of the function block responsible for controlling the laser. The sequence of actions is easy to be followed and changes can be quickly performed thanks to the graphical interface and event/data segregation.

5. Results

Table 1 summarize what are the main characteristics of the 3 scenarios, assigning a score between 1 (unsatisfactory) to 10 (best result) to the following criteria: 1) Time and easiness of implementation in existing 61131-based plant; 2) Possibility to distribute the control, enabling independent cyber-physical systems; 3) Readiness to Industry4.0/5.0 (Data gathering, AI, analytics, etc.); 4) Possibility to fully exploit learning-factory approach to transfer knowledge; 5) Communication protocols are easy to be integrated and to be used;

6) Availability of function blocks to communicate with devices; 7) Possibility to encapsulate algorithms using various programming language; 8) Possibility to deploy the application on different hardware effortless, 9) Easily perform changes to the operations sequence or creating a new one; 10) Modularity, reusability, adaptability, interoperability, scalability and flexibility; 11) Debugging easiness.

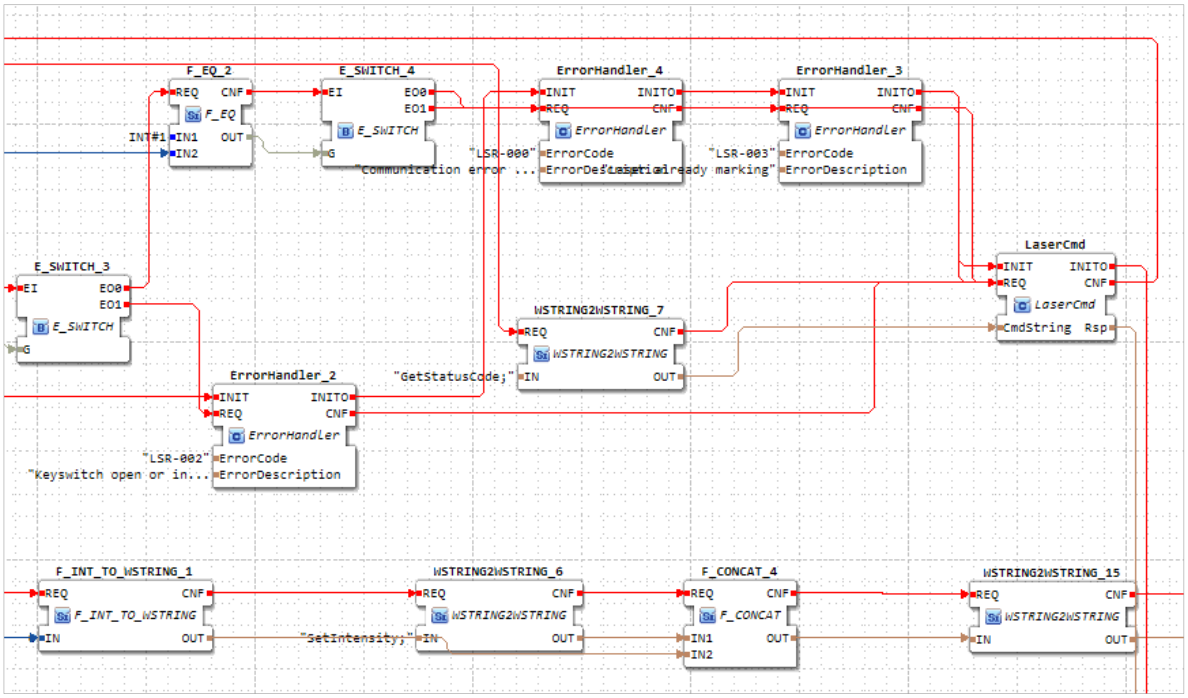


Fig. 4. Portion of the code controlling the laser, using graphical function blocks with event/data segregation.

Table 1. Use cases comparison. Numbers indicate the average score on a 1-10 scale, where 1 is the lowest and 10 the highest score.

No.	IEC 61131	Hybrid model	IEC 61499
1.	8	6	1
2.	3	5	10
3.	4	4	10
4.	2	2	8
5.	6	3	7
6.	8	2	2
7.	3	5	10
8.	1	3	10
9.	2	2	8
10.	2	2	10
11.	5	3	3

The 61131 clearly excels in the rapid integration within existing systems, under the condition to maintain the same hardware, but lacks fundamentals features offered by 61499. Moreover, adapting the sequence of operations to be performed, still require lots of work, since ladder code is a very rigid programming language. The hybrid model doesn't seem to have any advantage nor insurmountable disadvantages, but it's more a patch than a definitive solution. The full 61499 implementation requires a great initial effort and, by a merely production-based point of view, nothing get improved. But the event-based flow is an effective way to create factory process, which is also very intuitive to be used. Nevertheless, if custom function blocks or service interface FBs (which e.g. are used to run custom C# code) need to be developed from scratch, a deeper level of programming skills is required, whereas 61131 software usually already provide those function blocks.

The new standard excels in all those aspects which are of great importance for enabling the adoption of Industry4.0/5.0, namely: 1) Modularity (the use of function blocks allows for the modularization of control systems, making it easier to design, implement, and maintain complex systems); 2) Reusability

(function blocks can be easily reused in different control systems, reducing the time and effort required to develop new control systems); 3) Adaptability (the use of function blocks allows for the easy reconfiguration of control systems to meet changing system requirements); 4) Interoperability (the IEC 61499 standard specifies a common communication interface for function blocks, allowing them to interoperate with each other and with other control systems); 5) Scalability (the use of function blocks allows for the easy expansion of control systems to meet changing system requirements), and 6) Flexibility (the use of a variety of programming languages and tools, providing flexibility in the development of control systems).

It is to be emphasized how 61499 is remarkably suitable to be taught through a learning factory approach: theoretical lessons can be alternated to practical experiences, where each device with its SoftPLC can be programmed and controlled by a small group of students or professionals independently. Each of those devices represent a portion of the factory, and by combining them together a more complex system can be created, representing a full factory plant. This enables the possibility to teach the topic trough a gamification approach, a more appreciated and effective way of teaching.

6. Remarks

This experiment gave a first impression and some useful feedbacks. As part of future work, it is important to investigate a well-defined and methodological approach for effectively teaching modern PLC programming through learning factories. This includes closely examining the teaching materials, the learning factory approach, and the necessary hardware and software. This will lay the foundation for structuring an automation course, where future engineers can become familiar with modern automation techniques. Additionally, it would be beneficial to also explore the integration of industry-specific and real-world examples in the course to enhance the practical application of the learned concepts. Furthermore, incorporating project-based learning activities in the course can also be an effective way to deepen the understanding of the students and provide them with hands-on experience. This could lead to a better-prepared workforce and a more efficient use of the technology in industrial environments.

References

[1]. M. Tiegelkamp, K. H. John, IEC 61131-3: Programming Industrial Automation Systems, in Concepts and Programming Languages, Requirements for Programming Systems, Decision-Making Aids, *Springer-Verlag*, Berlin Heidelberg, 2010.

[2]. iec61499.com, [Online]. Available: https://iec61499.com/. [Accessed 01.01.2023].

[3]. R. Lewis e A. Zoitl, Modelling control systems using IEC 61499, *The IET*, 2nd Edition, 2010.

[4]. M. Foletti, F. Daniele, M. Confalonieri, P. Pedrazzoli e A. Ferrario, Implementation of a learning factory for research, education and training: the SUPSI Mini-Factory, in *Proceedings of the 11th Conference on Learning Factories, CLF2021*, 2021.

[5]. A. Zoitl, L. Prenzel e J. Provost, IEC 61499 Runtime Environments: A State of the Art Comparison, *Lecture Notes in Computer Science, EUROCAST 2019,* Vol. 12014, 2020.

[6]. P. Gsellmann, M. Melik-Merkumians, A. Zoitl, G. Schitter, A Novel Approach for Integrating IEC 61131-3 Engineering and Execution into IEC 61499, *IEEE Transactions on Industrial Informatics,* 2020.

[7]. K. Thramboulidis, IEC 61499 in Factory Automation, in Advances in Computer, Information, and Systems Sciences, and Engineering, *Springer*, 2007, pp. 115-124.

[8]. J. H. Christensen, New Developments in the IEC 61131-3 and 61499 Standards for Industry 4.0, in *Proceedings of the 10th International Conference on Model and Data Engineering,* 2021.

[9]. S. Sierla, J. Christensen, K. Koskinen e J. Peltola, Educational Approaches for the Industrial Acceptance of IEC 61499, in *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation*, 2007, pp. 482-489.

[10]. G. Lyu e R. W. Brennan, Towards IEC 61499 Based Distributed Intelligent Automation: A Literature Review, *IEEE Transactions on Industrial Informatics*, Vol. 17, Issue 4, April 2021, pp. 2295 - 2306.

(1289)

Building an Ad Hoc Network with LoRa and Transferring Field Data to the Internet

K.- P. Neitzke [1] and H. Schell [2]
[1,2] University of Applied Science Nordhausen, Weinberghof 4, 99734 Nordhausen, Germany
[2] Tel.: + 49 3631 420338, fax: + 49 3631 420818
E-mail: schell@hs-nordhausen.de

Abstract: The paper describes the structure of an ad hoc communication system, which allows to collect, transport and write field sensor data from a rural environment into the internet. The data are stored at a web server and can be read on a web page. The data are transported by LoRa transceivers which allows long distances at low data rates. The field data are collected by data collectors, sent by LoRa protocol over repeaters and written on a web server by a gateway that provides the internet access. Arduino Mega and UNO Boards and RTF95W LoRa modules were used as a base built for the data collector, the repeater and the gateway. The system is built in a modular way so any data collector witch sends messages of the certain protocol is supported. No commercial cloud provider is used; all components are self-designed and manufactured.

Keywords: Ad hoc network, LoRa, Rural environment, Education, Modular, Robust.

1. Introduction

While many systems for data collection exist for several environments, these systems are often highly dependent on urban infrastructure as access to Internet and unlimited power supply. The collection and transfer of data in a rural environment with limited means of power supply and internet access remains a challenge.

The development of an ad hoc network from the sensor to the next point of internet access over a range of several kilometres is the topic of the given paper.

To meet the requirements of working in a rough outdoor environment we developed a modular system consisting of several data collectors, repeater and a gateway to the internet. The LoRa protocol [1] was used to transport the data from data collector over the repeater to the gateway. Therefore overall distances from data collector to the gateway with internet access up to 2 km over several obstacles were reached.

Arduino Mega [2] or UNO [3] boards with RFM95W modules [4] were used as base built for data collector, repeater and gateway to the internet.

The system is built in a modular way so the access to the internet itself depends on the given opportunities. The gateway can be plugged over a RJ45 connector to any device, which provides internet access.

We tried to avoid the usual bundle between the LoRaWAN System and cloud solutions as TheTingsNetwork (TTN) [5] and others. So the data are stored on our web server where we have full access and nearly no costs. On our web page [6] all data are displayed and visualized.

In this paper we will give an insight in the structure of the system. We will discuss the bridged distances and present our test data and give some recommendations for further developments.

2. Task

The aim was to build a modular, robust, kept simple system for collecting, transmitting, storing and displaying field data. With the components one can build an ad hoc system for data measurement in a rural environment.

At our university a lot of measurement data are collected and stored. We want to invite our colleagues and students to take part and give them the opportunity to easily collect, store and display their data.

On the other hand long time environment data collection in a rural environment is still a challenge. So the system has to work with no internet access at the point of measurement.

To meet these requirements the components of the system have to be energy saving and of high range.

To build an independent and long lasting system commercial platforms as TheTingsNetwork (TTN) and other cloud solutions have to be avoided. The system has to be kept simple and flexible. For working with heterogeneous data collectors the system has to be of modular built. The data have to be stored at a web server and displayed at an according web page.

3. State of the art

The LoRa modulation is a proprietary protocol developed by Semtech and uses a chirp spread spectrum modulation [7]. The LoRa protocol provides the physical layer for the LoRaWAN network, which is an open standard developed by the LoRa Alliance. The LoRaWAN standard defines a complete network of end devices and gateways witch provide the entry to the internet. The data are usually stored on cloud servers as TheTingsNetwork.

4. Limitations of the System

The system does not provide any kind of decryption system. The messages are sent in plain text and anyone with a LoRa receiver can read and analyze them. And anyone can send messages if the message structure is provided. So no critical data must be sent.

In the test phase this is a low threshold attempt to inspire our students to take part in the system and to provide their own messages and data.

In a professional environment there needs to be added an encryption/decryption system (see 10. Recommendations).

By using the LoRa protocol there comes the disadvantage of low data rates. (see 5. Testing the LoRa connection). The LoRa system provides reglemented airtime so few data packets with short telegrams are required.

The system is not a net but a system of tunnels for the certain messages. To avoid feedback loops for messages still only one repeater per message ID is possible. To widen the range of the system by using more than one repeater a system of avoiding feedback loops has to be implemented.

5. System Topology

Fig. 1 shows the overall system topology where the sensor data are transmitted by a LoRa connection to the gateway.

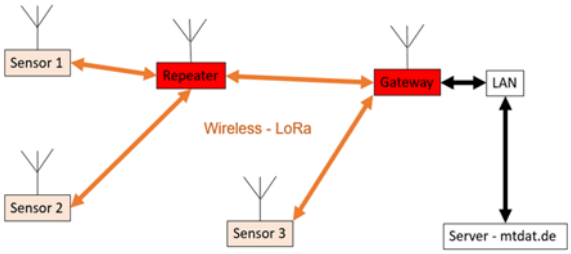


Fig. 1. System topology with sensors (data collectors), repeater, gateway and web server.

The transmission of packets is provided by defined pathways from data collector over a repeater to the gateway. The usage of a repeater is optional. The gateway provides the connection between the LoRa transmissions and the connection to the internet.

The configuration of the data collectors and the repeater is provided by messages from the gateway to the repeater or data collector. After sending a telegram the data collector listens for several seconds to a reply from the gateway. The reply can consist of the cycle time or a specific sensor configuration or simply an ok for a received telegram. The data are finally stored on a university operated web server.

The timestamp of the data is the timestamp of the gateway for now. For having the timestamps of the data collector each data collector has to have a RTC (Real Time Clock) on board (see 10. Recommendations).

The stored data on the web server are displayed on a web page.

All Messages are set in plain text. No encryption/ decryption system is provided by now. This allows everyone to read, analyze and send messages from and into the system. The message ID (MsgID) is a unique number referring to one single data collector.

The structure of the each message is: message ID and content. The content of a data telegram is defined by its message ID and may consist of several values divided by vertical bars.

There are two types of messages: data telegrams and configuration messages:

The data telegrams are surrounded by square brackets:
`[MsgId|Content]`
for example: `[50|21.5|987]`

The configuration messages are surrounded by round brackets:
`(MsgId|Content)`
for example: `(50|SET_SENDE_ZYKLUS_SEC|1800)`

By configuration messages all devices can be configurated from outside by LoRa messages. By now the cycle time of the devices can be set.

6. Testing the LoRa Connection

For sending the messages we used LoRa RTF95W modules with Arduino Mega or UNO Boards. In Europe LoRa Modules send at 868 MHz. In other countries there are other frequencies defined for LoRa (for example in the USA 915 MHz) [8].

Using the LoRa protocol provides some disadvantages. So one shares the frequency with car keys, home automation and IoT devices. Depending on the region there are regulations for using the LoRa frequency in respect of packet length, transmission power and duty cycle. For example in Europe there is a limitation to the ERP transmission power of 25 mW and to the Duty Cycle of 1% channel assignment per time [9], which means that the airtime is reglemented.

On the other hand the LoRa frequency is part of the ISM band (Industrial, Scientific and Medical *Band) [10] and therefore* there are no fees for using the LoRa frequencies. By using low transmission power, long duty cycles and short messages the power consumption of the modules is low and the range is wide.

Fig. 2 shows the test setup for the evaluation of antennas and ranges.

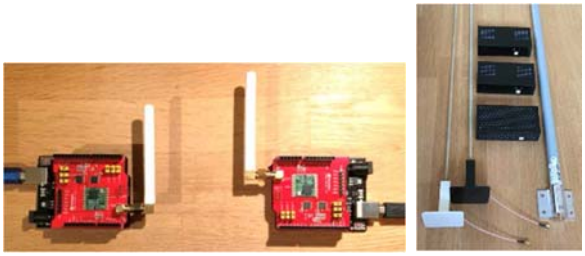


Fig. 2. Test setup for LoRa communication.

The range between the stub antennas shown in Fig. 2 was only 500 m from the inside of buildings. So bigger antennas with lager gain were tested. A range over 2 km was accomplished using a stub antenna on one side and a collinear antenna [11] on the other side.

The collinear antennas consist of several resonant parts in a GFK tube. The self-made antennas provided a slightly better range than the bought 868 Omni Antenna from MikroTik [12] (Fig. 2 on the right side) with the 6.5 dBi gain according to the data sheet [13]. Using collinear antennas on both sides we expect even wider ranges.

A lot of radio modules are available so one can choose among them. We used the *RFM95W with a maximum adjustable power of 20dBm = 100mW*.

At the given structure the distance between the furthermost data collector and the repeater are 670 m. Both antennas are collinear antennas and are located indoors. No further obstacles are given and the connection works without data losses.

The distance between the repeater and the gateway is 500 m. The gateway is also located indoors and has a collinear antenna. There are some obstacles in the way as buildings and trees. The gateway is of non-level terrain to the repeater with the repeater being of about 40 m higher altitude than the gateway. The connection works without data losses.

7. Components

7.1. The Base Build-up

All modules in the system have the same base. They consist of an Arduino Mega or UNO board and a Dragino Lora SX127x Shield [14] with RFM95W as base build-up.

They are built within a robust housing with ventilation slots and are made of 2 mm glass fiber (Fig. 3). For programming the LoRa connection we used the Radiohead Library for the RF95 module [15].

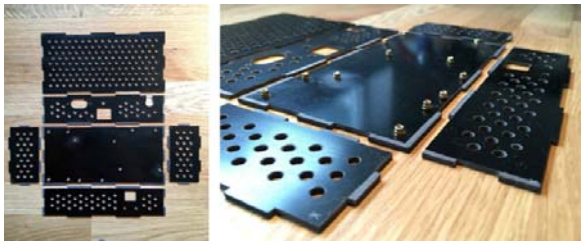


Fig. 3. Robust housing of the modules.

7.2. The Data Collector

The data collector units consist of the base build-up of an Arduino Mega or UNO board and a Dragino Lora Shield and a sensor. For data logging purposes there is also an OpenLog data logger [16] and a Real Time Clock (DS3231) [17] on board. All data are collected using a Low Pass Filter (LPF) and sent according to the configurated cycle time. The raw data are written to the on board SD card to provide further data analysis.

Because of the modular system the data collector units can be of any design provided they are able to send LoRa messages with the correct message structure. The sensors used by now are for example pressure sensors (BMP180) [18] and The Grove Loudness Sensor [19].

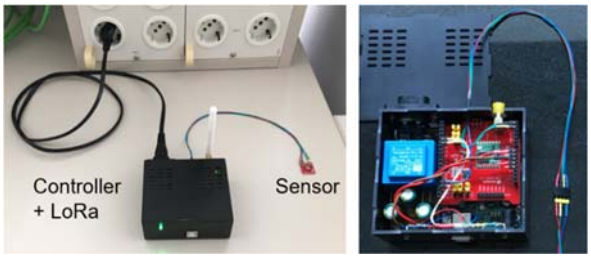


Fig. 4. A typical data collector

7.3. The Repeater

The repeater receives all LoRa messages that provide the given message structure. The messages are simply sent out again and thus amplified. By using a repeater the range of the system is widened and natural obstacles can be overcome. (see 6. Testing the LoRa connection)

The repeater consists of the same combination of Arduino Mega or UNO and Dragino LoRa shield as the base build-up. Because of the modular system the repeater can be of any design provided it is able to receive and send LoRa messages.

7.4. The Gateway

The gateway consists of a LoRa module and a connector which provides Ethernet packages over a RJ45 socket to a LAN connection (Fig. 5). The connection to the internet itself is not the topic of this paper. To be independent from a fixed network we used a LTE router so we could connect to the internet wherever the LTE net is available.

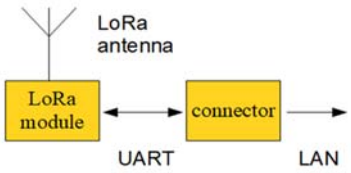


Fig. 5. Data Flow inside the gateway.

The LoRa module has the same base built as the data collectors and the repeater (Fig. 6 on the right side). It receives all incoming LoRa messages that fit the protocol. The messages are sent over an UART connection to the bridge (Fig. 6 in the middle). The bridge is an Arduino Ethernet shield [20] and sends all incoming messages over the Ethernet socket.

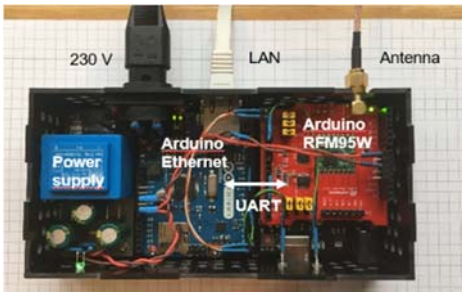


Fig. 6. The gateway.

8. Testing of the System

The first sensor we implemented was a BMP 180 pressure/temperature sensor board. In Fig. 7 the timeline of the temperature and in Fig. 8 the timeline of the pressure is shown. The data collector sends data every 10 minutes. There are two sensors in a southwards (red dots) and a northwards (green dots) bounded room.

In one of our laboratory rooms we installed a noise level sensor. The sensor determines the maximum noise level in a given time span (Fig. 9). You can clearly see the weekends, the chrismas break and the beginning of the lessons in the nearby lecture hall.

With the third sensor we measure the frequency of the power grid. Minimum, maximum and filtered frequency are given for the measurement period (Fig. 10).

We are aware that some of the measurements can be used to monitor employees. The data collector must be used according to current law. We decided to ask consent from each colleague that works in one of the monitored rooms.

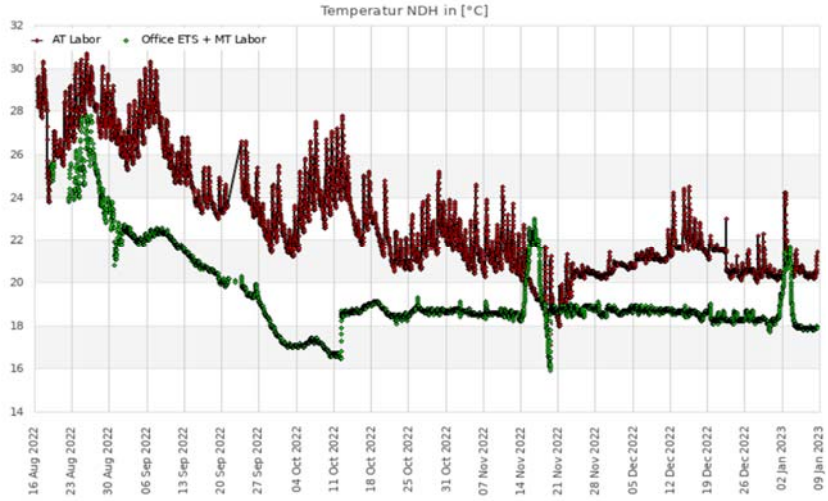


Fig. 7. Temperature over time [21].

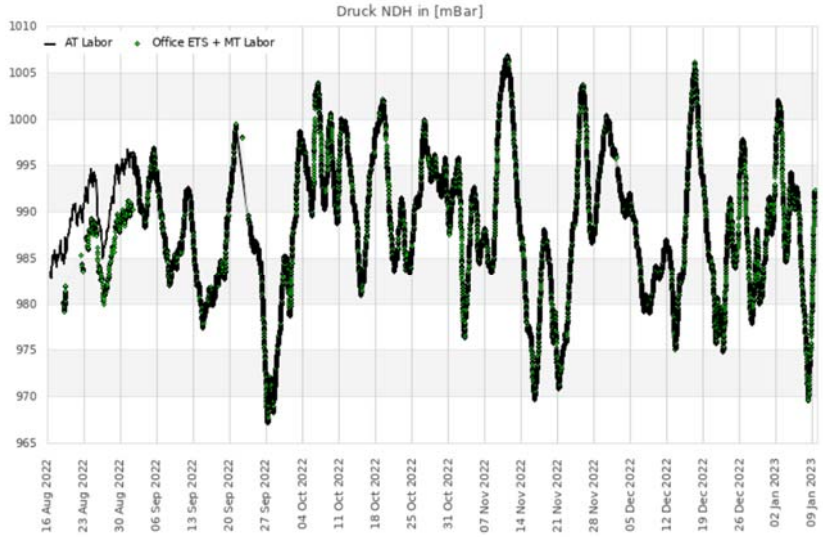


Fig. 8. Pressure over time [22].

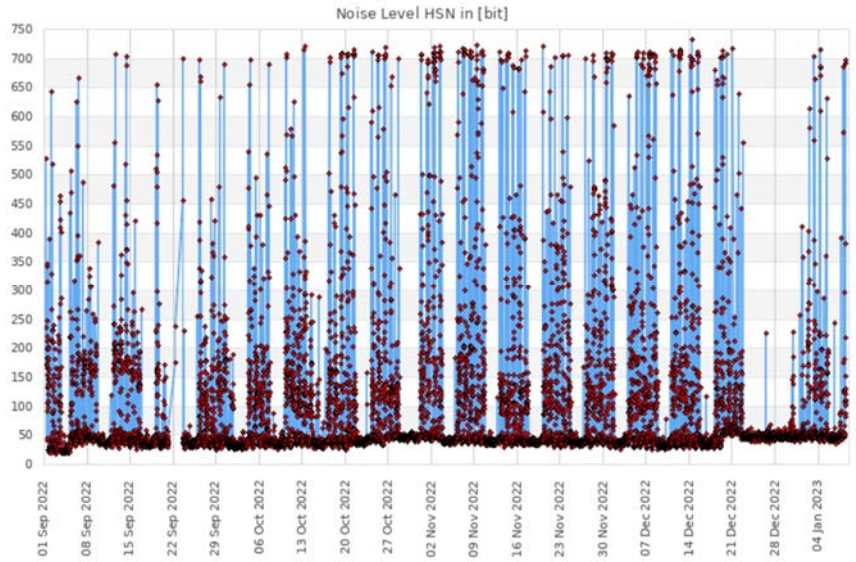


Fig. 9. Noise level sensor [23].

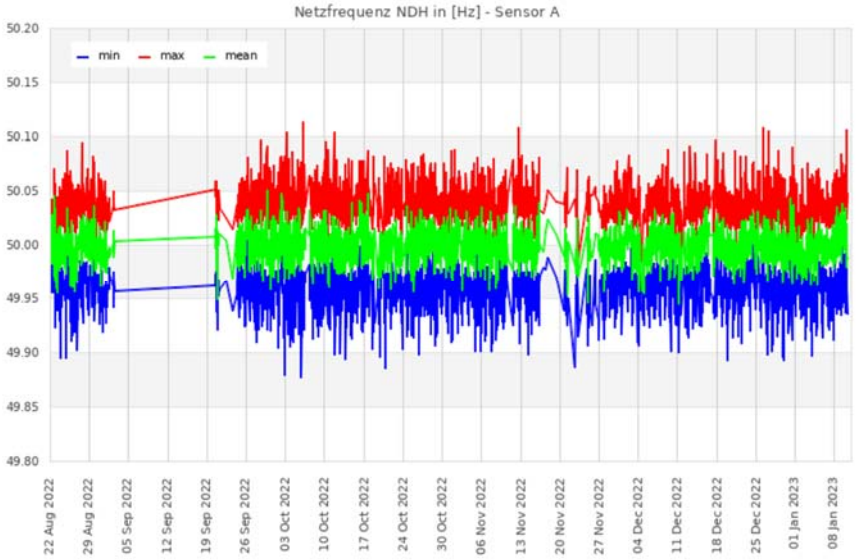


Fig. 10. Frequency of the power grid [24].

9. Conclusions

This paper presents our approach on building an ad hoc system to collect, transport, store and display field data with no internet access at the point of data collection.

By using the LoRa protocol with low data rates and amplification of the signal by a repeater we achieved ranges of 2 km with one stub antenna and one collinear antenna. With measuring cycles of 10 to 30 minutes depending on the message type and short messages we were able to meet the airtime requirements.

10 different message types are processed by now and with our configuration system it is easy to integrate new message types. The collected data are displayed on a web page.

10. Recommendations

For using the system for business data outside of an educational context an encryption/decryption system has to be included.

For offline situations a buffering of messages would be helpful. The on board SD card could be used for saving messages in an offline situation. After re-establishing the connection the data will be sent to the gateway. For a timestamp of the data a RTC has to be on board the of data collectors.

To avoid message loops only one repeater per message ID is supported by now. To reach a wider range and create longer pathways a system of loop avoiding has to be provided.

The system of configuration messages should be extended to configure the data logger and repeater from remote. Besides the cycle time there could be the turning on and off of a heartbeat of the devices and a life sign on demand.

To achieve self-sufficient data collectors a system with solar panels, a rechargeable battery and load balancer can be added.

References

[1]. LoRa Alliance Web Portal:https://lora-alliance.org/

[2]. Arduino® Mega 2560 Rev. 3 manual: https://docs.arduino.cc/static/0375a04aaba079038e05d48ea835a97d/A000067-datasheet.pdf

[3]. Arduino® UNO R3 manual: https://docs.arduino.cc/resources/datasheets/A000066-datasheet.pdf

[4]. RFM95W data sheet: https://www.hoperf.com/data/upload/portal/20190801/RFM95W-V2.0.pdf

[5]. The Things Network home page: https://www.thethingsnetwork.org/

[6]. display of data: www.mtdat.de

[7] B. Reynders and S. Pollin, Chirp spread spectrum as a modulation technique for long range communication,in *Proceedings of the Symposium on Communications and Vehicular Technologies (SCVT)*, Mons, Belgium, 2016, pp. 1-5, doi: 10.1109/SCVT.2016.7797659.

[8]. Frequency Plans for LoRa by Country: https://www.thethingsnetwork.org/docs/lorawan/frequencies-by-country/

[9]. ERC Recomendation 70-03, page 8: https://docdb.cept.org/download/2464

[10]. Bundesnetzagentur, Funkfrequenzzuweisungen: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Frequenzen/Allgemeinzuteilungen/FunkanlagenGeringerReichweite/2018_05_SRD_pdf.pdf?__blob=publicationFile&v=7

[11]. Collinear Antenna Description:https://www.electronics-notes.com/articles/antennas-propagation/phased-array-antennas/collinear-vertical-antenna.php

[12]. 868 Omni Antenna: https://www.omg.de/mikrotik/lorawan/mikrotik-lorawan-omni-antenne-824-960-mhz/a-25277/?ReferrerID=7&gclid=EAIaIQobChMIodjNlKDA_AIVu41oCR1agw6XEAQYASABEgJgXvD_BwE

[13]. 868 Omni Antenna manual: https://i.mt.lv/cdn/product_files/68_Omni_antenna_211023.pdf

[14]. Dragino LoRa shield Wiki: https://wiki1.dragino.com/index.php?title=Lora_Shield

[15] RadioHead class reference: https://www.airspayce.com/mikem/arduino/RadioHead/classRH__RF95.html (accessed on 27.01.2023)

[16]. OpenLog Wiki: https://github.com/sparkfun/OpenLog/wiki/Datasheet

[17]. DS3231 Real Time Clock data sheet: https://www.digikey.de/en/datasheets/maxim-integrated/maxim-integrated-ds3231-ds3231s

[18]. BMP180 data sheet: https://www.digikey.com/htmldatasheets/production/856385/0/0/1/bmp180-datasheet.html

[19]. Grove Loudness Sensor data sheet: https://www.mouser.com/datasheet/2/744/Seeed_101020063-1217445.pdf

[20]. Arduino® Ethernet Shield Documentation: https://docs.arduino.cc/retired/shields/arduino-ethernet-shield-without-poe-module

[21]. Screenshot taken at 09.01.2023 from:www.mtdat.de/6061t.php

[22]. Screenshot taken at 09.01.2023 from: http://www.mtdat.de/6061.php

[23]. Screenshot taken at 09.01.2023 from:www.mtdat.de/80.php

[24]. Screenshot taken at 10.01.2023 from:www.mtdat.de/50.php

(1431)

On the Application of Graphene Coatings to Prevent Local Metal Surface Overheating

U. Zaimis
Liepaja University, Institute of Science and Innovative Technologies
Liela iela 14, LV-3401, Liepaja, Latvia
Tel.: +371 29101515
E-mail: uldis.zaimis@liepu.lv

Summary: One technique for preventing local thermal breakdown of a surface is the application of a coating with a high heat transfer coefficient. Even if the base material is a metal with high thermal conductivity, such as copper, it is possible to apply a coating of graphene, which has a much higher thermal conductivity. This may be useful in engineering to prevent localized thermal damage and potentially protect a flying drone from laser weapons. The research in this case investigates the effect of a thin graphene layer on the collapse of copper foil after treatment with a pulsed laser. A mathematical model of heat propagation is developed, and heat transfer simulations are performed for a combination of materials with different heat transfer coefficients. The proposed solution has the practical advantage of maintaining the mass of the object without increasing its geometric dimensions, which is particularly important in robotics in defense, avionics and precision mechanics.

Keywords: Graphene coating, Heat transfer, Treatment by pulse laser.

1. Introduction

The development of novel materials with distinct properties offers a plethora of previously uninvestigated opportunities. These properties can include variations in electrical conductivity, heat capacity, thermal conductivity, optical transmittance, mechanical properties, among others. Graphene, in particular, possesses several of these properties, including a notably high thermal conductivity of up to 4000 W/mK (Sang *et al* 2019). In the context of the current study, the focus is on the utilization of a thin graphene layer on a metal surface for heat transfer, specifically examining the effects of short-term laser radiation on a copper foil covered with a graphene layer. The objective of this research is to utilize this heat transfer method to transfer heat to non-irradiated regions, thus preventing the collapse of mechanically resistant areas and promoting the efficient distribution of heat over a large surface area of the material.

2. Materials and Methods

2.1. Protection Possibilities of Local Thermal Influence on Metal Foil

There are several methods of protecting the metal surface from the local effects of the laser. The first and oldest is the installation of effective external thermal insulation, which does not allow thermal effects to access load-bearing structures and destroy them. A widely known example of such thermal insulation is the carbon-based thermal insulation plates of space shuttles. Such a method is effective, but it has many disadvantages - it increases the mass and dimensions of the structure and, therefore, is not acceptable for protecting small flying, floating and diving small drones (Bahadori, 2014).

The second way is the application of highly reflective surface coatings. It is possible to apply up to about 95 % reflective (in the optical range) coating, which will reflect the laser radiation and only a small part of the beam energy will be transformed into heat energy. Aluminium and silver are excellent reflectors in the visible region, their reflectance being 0.90 and 0.95 respectively. The reflectance can be calculated using the formula:

$$R(\%) = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \times 100, \qquad (1)$$

where n is the refractive index, k is the extinction coefficient and R is the reflectance in % (Prasad *et al* 2018). A list of values of n and k are obtained for ultraviolet (0.2 and 0.3 μm), visible (0.4 to 0.7 μm) and infrared wavelengths (1 to 10 μm). When a metal has k value greater than n, it appears shiny, e.g. silver. Since the wavelengths increase as we move from visible to IR, the values of k and n increase and the reflectance increases (the same).

Such a coating will be light and cheap, but will require a high quality of surface treatment, which is not always possible in mass production, in addition, various protrusions, fasteners, details on the surface will reduce the reflection effect. Dirt, dried deposits will be especially undesirable, which will significantly reduce the reflection; it will also be reduced by scratches during product use.

A third option is to apply a thin coating with a very high heat transfer coefficient, which spreads the heat over a wide area and thus reduces local heat effects.

Such can be the graphene coating, which will be discussed in this work.

2.2. Mathematical Model of Heat Transfer

The heat transfer in this study was analyzed using a modified classical heat transfer equation, as described in (Zaimis *et al* 2015). This modified equation incorporates a second-order partial derivative with respect to time, which accounts for the collapse of the desired function and accurately reflects real-world physical processes. Additionally, the equation takes into consideration the non-zero heat capacity of the materials being studied:

$$\frac{\partial u(x,t)}{\partial t} + \varepsilon \cdot \frac{\partial^2 u(x,t)}{\partial t^2} = \frac{\partial}{\partial x}\left(k(x) \cdot \frac{\partial u(x,t)}{\partial x} \right) + f(x,t) \quad (2)$$

where u is the desired function, x is the 1D spatial variable, t is the time, ε is the small smoothing factor, k is the heat transfer coefficient and f is the function of external influence – laser irradiation. By supplementing the partial differential equation with boundary and initial conditions, a mathematical initial - boundary problem was created, which was applied to numerical calculations after the convergence study and discretization step. The main focus was on the different values of the heat transfer coefficient (5 - Thermal Properties of Metals, Conductivity, Thermal Expansion, Specific Heat,), which is ~ 400 Wm-1K-1 for copper (Cu), 50-80 $Wm^{-1}K^{-1}$ for carbon steel and 14 $Wm^{-1}K^{-1}$ for stainless steel, but for graphene up to 4000 $Wm^{-1}K^{-1}$. Thus, the heat transfer in space and time was modeled at a heat transfer coefficient difference of 10 times (copper - graphene), keeping in mind that in the case of iron, steel and stainless steel the differences will be even more significant.

2.3. Laboratory Tests with Laser Treatment

After computer simulations, laboratory experiments were carried out. A copper (Cu) foil, thickness 100 nm, was irradiated with a pulsed laser - first pure copper, then coated with a layer (up to 2 nm) of graphene (Shearer *et al* 2016). Graphene deposition took place by CVD method. Laser parameters: Nd3+:YAG crystal, wavelength 532 nm, pulse power 0.15 J, pulse repetition frequency 10 Hz, pulse duration 120 microseconds. The polarization angle was not changed and is not important in this case, since the properties of all the materials under consideration are constant in the x-y plane. A single laser treatment was performed at one point; the treatment was repeated at different points and the repeatability of the obtained results was checked.

3. Results

In the simulation of laser processing of one-dimensional material $(0 \le x \le 1)$, the conditions were realized: 1) processing was performed on half $(0 \le x \le 0.5)$ of the material, so that heat transfer to the sides could be observed; 2) heat change in the time period $(0 \le T \le 1)$ was observed, which is sufficient for the function to reduce the value in one of the cases. Temperature was represented in conventional units $(0 \le T \le 25)$ in the first case (Cu, Fig. 1) and $(0 \le T \le 5)$ in the second case (graphene, Fig. 2).

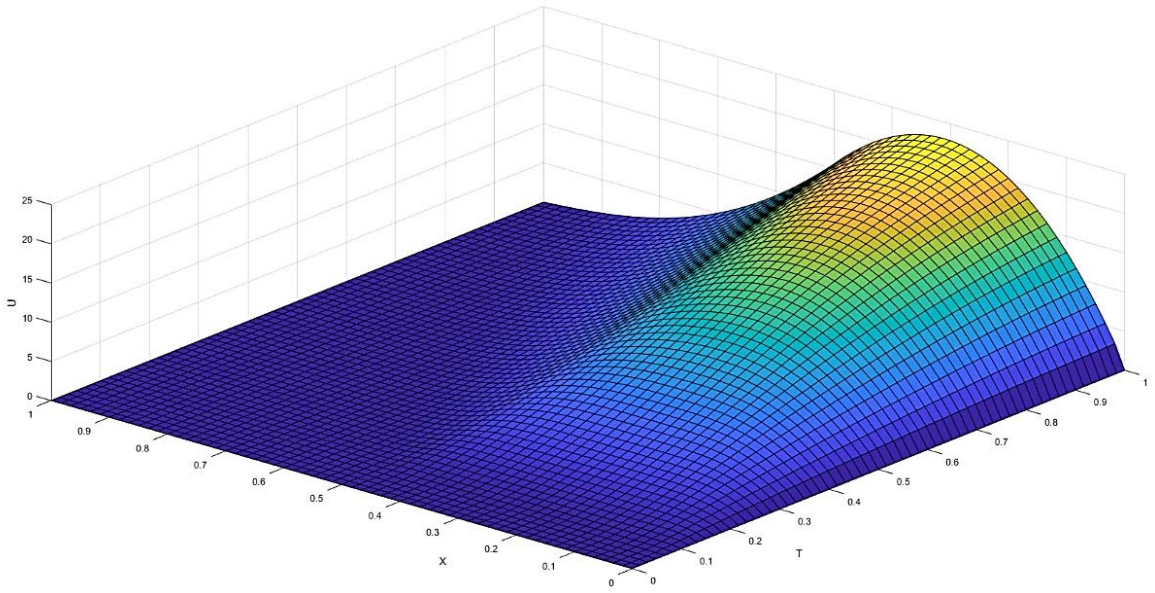


Fig. 1. Modeling of heat transfer temperature distribution with limited laser irradiation in the space and low thermal conductivity (Cu). To the left the spatial variable X, to the right the time variable T, i.e. the development of the process in time is shown; the heat effect is initially given for the first half of the considered length (max temperature 20 units).

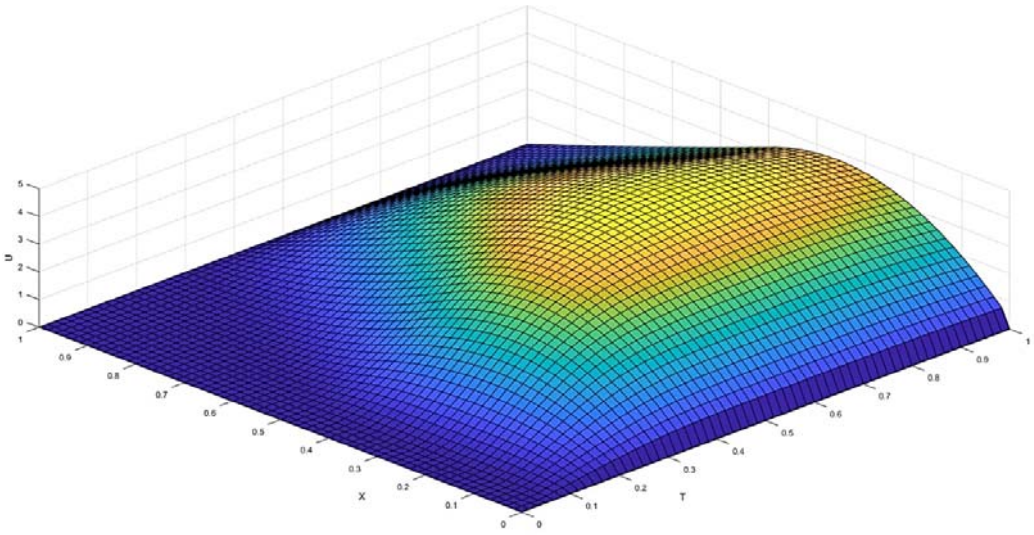


Fig. 2. Modeling of heat transfer temperature distribution with limited laser irradiation in space and high thermal conductivity (graphene, max temperature 5 units) at the same laser processing parameters as in Fig. 1.

In laboratory tests, the assumption about the sufficiency of the power of the laser pulse to destroy the material was confirmed. Fig. 3 shows a pierced Cu foil material; according to the Gaussian distribution of energy in the laser beam, the middle part of the impact plane has been overheated, melted and collapsed.

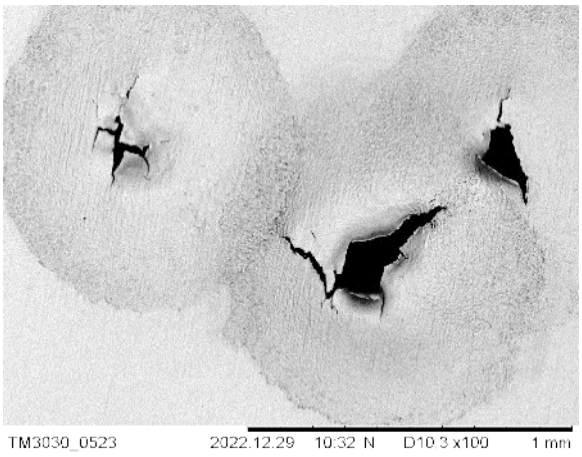


Fig. 3. Laser beam hits in pure copper foil, the material collapsed in the middle of the hit (SEM image).

Fig. 4 shows the persistence and mechanical strength retention of graphene-coated copper foil. The metal is not completely melted, the effect of heat has left traces in a wider area than the area affected by the laser beam. All processing parameters are identical to the previous one, the scales of both images are the same. A large-scale thermal effect is visible, the metal surface is cracked, and the middle part of the area affected by the laser beam is bent in the direction of the effect. In general, the exposure correlates well with the Gaussian power distribution in the laser beam.

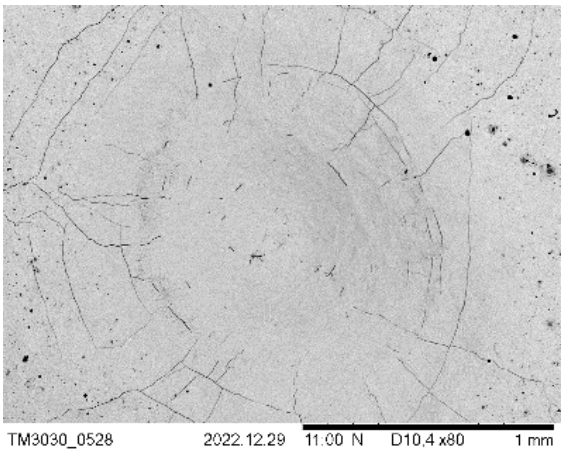


Fig. 4. A laser beam hits a copper foil covered with a thin layer of graphene, the heat is transferred to the sides, the effect of heat is visible in a wide area beyond the impact boundary (SEM image).

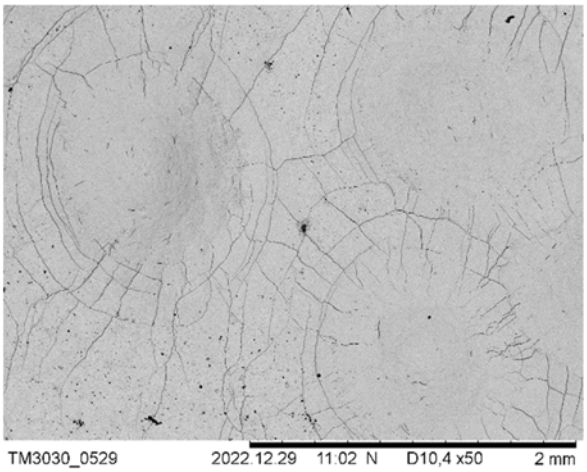


Fig. 5. Multiple laser beam hits on a copper foil covered with a thin layer of graphene, the heat effect is not accumulated (SEM image).

Another image (Fig. 5) shows the overlap of thermally affected areas, where the centers of several laser beam hits are located close to each other and the gaps have been exposed from several sides. Since the laser pulse frequency is 10 Hz, the exposure time is very short, but the heat transfer rate was high enough so that this effect did not accumulate. A different scale is used.

The question of processing graphene-coated metal in one place several times remains open. Different results have been obtained, but the conclusion is that with a high probability, the graphene coating is destroyed in the first-second hit if there is enough laser power, and the stability of the material further is lost. This has to be seen in the context of beam focusing and the technical capabilities to repeatedly hit the same point on the surface.

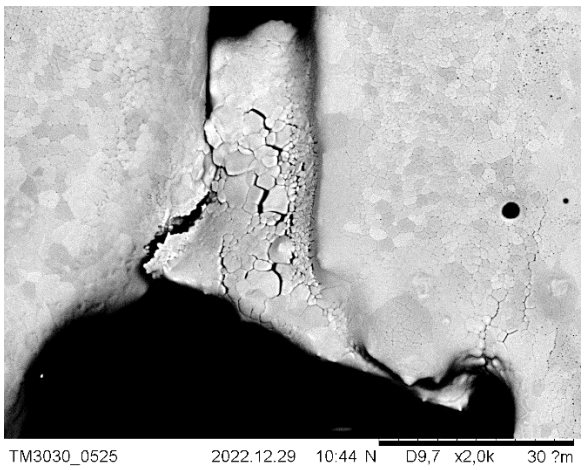


Fig. 6. Image of the collapsed region of pure copper (Cu) foil at high (2000x) magnification. Scale bar length 30 micrometers (SEM image).

Examining the collapsed areas of copper foil not coated with graphene (Fig. 6) at a high magnification (2000x) shows the breakdown of the metal in typical domains of copper, which indicates the effect of high thermal exposure. Such results are typical for copper and copper alloys close to the melting point (copper has a melting point of 1085 ^{o}C). In the center of the image, regions of fractured domains are visible, while in the unbroken region, domain splitting is observed, but the overall mechanical strength of the material has not been lost. Thin rolled metal tear edges are visible, evidence of high temperature effects.

4. Conclusions

The obtained results of mathematical modeling of heat transfer confirm the correctness of the assumption - with a higher heat transfer coefficient, a faster temperature rise is observed and, at the same time, a strong heat transfer to the non-irradiated part of the material, therefore, in the irradiated part, the laser irradiation energy is not accumulated locally in the material and as a result does not reach the critical level of material collapse.

Considering the negligible thickness of the graphene coating, the metal protection method is applicable in robotics in defense, avionics, precision mechanics and other fields.

Acknowledgements

This work was carried out within partial financial support from the European Regional Development Fund Project No. 1.1.1.5/018 "Promotion of research, innovation and international cooperation in science at Liepaja University".

References

[1]. A. Bahadori, Design and Application of Thermal Insulation, in Thermal Insulation Handbook for the Oil, Gas, and Petrochemical Industries, *Elsevier*, 2014, pp. 1-142.

[2]. K. Prasad, A. Goyal, K. Gohil, I. Jagyasi, Highly Reflective Coatings, *International Journal of Applied Engineering Research*, Vol. 13, 22, 2018, pp. 15773-15782.

[3]. M. Sang, J. Shin, K. Kim, K. Jun Yu, Electronic and Thermal Properties of Graphene and Recent Advances in Graphene Based Electronic Applications, *Nanomaterials*, Vol. 3. Issue 9, 2019, pp. 374.

[4]. U. Zaimis, Sh. Guseynov, Analytic-numerical modeling and investigation of nanostructures' dynamics on material surfaces after laser irradiation, in *Proceeding of the X Int. Conference 'Environment. Technology. Resources'*, Issue 3, 2015.

[5]. Thermal Properties of Metals, Conductivity, Thermal Expansion, Specific Heat, www.engineersedge.com

[6]. C. J. Shearer, A. D. Slattery, A. J. Stapleton, J. G. Shapter, C. T. Gibson, Accurate thickness measurement of graphene, *Nanotechnology*, Volume 27, Number 12, 2016, 125704.

(1444)

Revisiting a Matrix Characterization of Spanning Tree based on a LaSalle Invariance Principle

T. C. Lee [1] and Y. Su [2]
[1] Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, R.O.C.
E-mail: tc1120@ms19.hinet.net
[2] College of Computer and Data Science, Fuzhou University, Fuzhou 350116, China
E-mail: yfsu@fzu.edu.cn

Summary: It is well-known that a digraph has a spanning tree if and only if its graph Laplacian (matrix) has a simple zero eigenvalue and all the other nonzero eigenvalues have positive real parts. Different from the approach of matrix and graph theory, this paper proposes a novel proof based on a LaSalle invariance principle. Firstly, it is shown that the max-min function considered in present literature is a weak Lyapunov function for the error dynamic based on the coordinate transformation with any convex combination of the agents' states. Thus, the steady-state behavior can be characterized by the spanning tree condition. Secondly, the proof relates directly the steady-state studies to the graph property. Such an approach has a potential to extend similar results to more complex systems such as nonlinear systems or time-varying systems, and thus provides a new point of view in studying consensus problems of multi-agent systems. Both of them may provide a new point of view in studying consensus problems.

Keywords: Spanning tree, Laplacian matrix, LaSalle invariance principle, consensus, multi-agent systems.

1. Introduction

Cooperative control of multi-agent systems has attracted extensive studies in the past two decades [1, 6, 15, 17, 18]. In contrast with the classic centralized control, the characteristic of cooperative control lies in that the feedback information among all agents in a multi-agent system should be transmitted through a specific communication network, where the so-called digraph is one of ideal and useful descriptions. The fact that this digraph has a spanning tree has been demonstrated to be the necessary and sufficient condition for guaranteeing that the whole agents can access the information from others directly or indirectly. This condition further induces the following well-known property of graph Laplacian, which is usually adopted in stability analysis almost in every relevant reference [6, 14, 15].

Theorem 1. A digraph has a spanning tree if and only if its (any) graph Laplacian has a simple zero eigenvalue and all the other nonzero eigenvalues have positive real parts.

Theorem 1 was first proved in [16] by the eigenvalue perturbation using the Gersgorin disc theorem. It plays a significant role in many cooperative control designs. For example, with this theorem together with the famous Perron-Frobenius Theorem, one is able to show the weighted average of the agents' initial states is the steady-state of the consensus value [16], see also the discussion in Remark 1. On the other hand, one can find a strict Lyapunov function for the consensus systems based on the so-call M-matrix. A more general result can be found in [19, Theorem 4.6].

In this paper, we provide a novel proof for this theorem based on a LaSalle invariance principle. More precisely, the following extended result is proposed.

Theorem 2. For a digraph G, the following statements are equivalent where $L(G)$ is a graph Laplacian of G and n is the number of all vertices of G:
a) G has a spanning tree.
b) Every forward complete solution $x(t)$ of $\dot{x} = -L(G)x$ satisfies $\lim_{t\to\infty}[x_i(t) - x_c(t)] = 0$ for any $1 \le i \le n$, where for each i, $x_i(t)$ is the i-th component of $x(t)$ and $x_c(t) = (x_1(t) + \cdots + x_n(t))/n$.
c) Every forward complete solution $x(t)$ of $\dot{x} = -L(G)x$ reaches consensus, i.e., $\lim_{t\to\infty}[x_i(t) - x_j(t)] = 0$ for any $1 \le i \ne j \le n$.
d) $L(G)$ has a simple zero eigenvalue and all the other nonzero eigenvalues have positive real parts.
e) The rank of $L(G)$ is equal to n-1.

Obviously, Theorem 1 can be deduced by Theorem 2, which will be proven in this paper. It is worth mentioning that the novel proof directly relates the stability property of the graph Laplacian to the cooperative control system, where two distinguished features can be found.

Firstly, it is shown that the max-min function considered in present literature such as [14] is a weak Lyapunov function for the error dynamic based on the coordinate transformation with any convex combination of the agents' states (see Lemma 2 and Remark 1). Thus, the steady-state behavior can be characterized by the spanning tree condition without knowing the "only if" part of Theorem 1 in advance and the Gersgorin disc theorem (Proposition 1).

Secondly, the proof relates directly the steady-state studies to the graph property (Proposition 1). Such an approach has a potential to extend similar results to more complex systems such as nonlinear systems or time-varying systems. Indeed, the same argument can be applied to the case of jointly-connected switching

topology [12, 18] when the so-called dwell-time conditions hold (Remark 2).

To provide a self-contained statement, the following graph notations are adopted. Let $G=(V,E)$ be a digraph, where $V=\{1,...,n\}$ is the node set and $E\subseteq V\times V$ is the edge set. Let $L(G)=(l_{ij})$ be a graph Laplacian (matrix) for G, which satisfies that for any $i\neq j$, $l_{ij}<0$ whenever $(j,i)\in E$, $l_{ij}=0$ in case of $(j,i)\notin E$, and the sum of each row of $L(G)$ is 0. For any $i\neq j$, a sequence $(i,i_1),(i_1,i_2),\cdots,(i_k,j)$ of pairs in E is said a directed path (walk) from i to j. The digraph G is said to have a spanning tree when there exists $i_0\in V$ (called the root) such that for any $j\in V\setminus\{i_0\}$, there exists a directed path from i_0 to j.

Notations

1. $\Re$ denotes the set of all real numbers and $\Re^n$ is the n-dimensional Euclidean space.
2. $1_n=(1,\cdots,1)^T\in\Re^n$ and $1_n^{\perp}=\{u\in\Re^n \,|\, 1_n^T u=0\}$.
3. $\#S$ denotes the number of the elements of S.
4. For a vector space V over $\Re$ and a nonempty subset S of V, $SpanS$ denotes the smallest subspace that contains S. When S is a subspace of V, $DimS$ is the dimension of S over $\Re$.
5. For any square matrix A, a simple eigenvalue λ means that it is a root of the characteristic polynomial $p=\det(\lambda I-A)$ with $p'(\lambda)\neq 0$.
6. For any matrix A, a solution x of $\dot{x}=Ax$ defined on $[0,\infty)$ is called a forward complete solution.

2. Main Results

A LaSalle invariance principle [7, 9] is proposed. It is used to prove a consensus result that in Theorem 2, a) results in b), inducing a new proof of Theorem 2.

2.1. Revisiting a LaSalle Invariance Principle

Consider the following system:

$$\dot{x}=Ax \qquad x\in\Re^n \tag{1}$$

where A is the real-valued square matrix.

The following result is a minor extension of the so-called LaSalle invariance principle, see Appendix A for a proof [7].

Lemma 1. Consider (1). Let $X\subseteq\Re^n$ be an A-invariant subspace. Suppose there exists a continuous function $V:X\to[0,\infty)$ such that $V(x(t))$ is non-increasing for any forward complete solution x of (1) with $x(0)\in X$, and $x(0)=0$ when $V(x(t))$ is a constant function. Then, every bounded forward complete solution x of (1) with $x(0)\in X$ converges to zero. ■

2.2. Show Consensus by Weak Lyapunov Functions

In this subsection, we will show that a) implies b). i.e., the consensus can be reached.

To this end, for any solution $x(t)$ of $\dot{x}=-L(G)x$, let

$$x_+=\max_{1\le i\le n} x_i \text{ and } x_-=\min_{1\le i\le n} x_i. \tag{2}$$

Then, the following result holds with its proof provided in Appendix B.

Lemma 2. With $A=-L(G)$ and x being any forward complete solution of (1), x_+ and $-x_-$ are both non-increasing. ■

Now a weak Lyapunov candidate function can be chosen as $V=x_+-x_-$. Then, V is non-increasing along with the trajectories of (1) in view of Lemma 2 where $A=-L(G)$. Let

$$x_c=(x_1+\cdots+x_n)/n \text{ and } x_*=x-x_c 1_n. \tag{3}$$

Then, $V=(x_*)_+-(x_*)_-$ and x_* is a forward complete solution of the following equation with $x_*(0)\in 1_n^{\perp}$:

$$\dot{x}=-L_*(G)x \tag{4}$$

where

$$L_*(G)=L(G)-1_n 1_n^T L(G)/n. \tag{5}$$

Conversely, for any forward complete solution x of (4) with $x(0)\in 1_n^{\perp}$, denote the forward complete solution of (1) with initial state $x(0)$ as x^*. By the uniqueness of solutions [4], it can be seen that $x=(x^*)_*$. So the operator "$_*$" maps the all forward complete solutions of (1) onto the all forward complete solutions of (4) with their initial states lying within the set $1_n^{\perp}$.

Observe that the following inequality holds:

$$\|x_*\|/\sqrt{n}\le V=x_+-x_-\le\sqrt{2}\|x_*\| \tag{6}$$

by the definitions of *max* and *min* functions, and Cauchy inequality. Since V is nonincreasing, every forward complete solution x of (4) with $x(0)\in 1_n^{\perp}$ is bounded. Notice that $1_n^{\perp}$ is a $(-L_*(G))-$ invariant subspace. Based on Lemma 1, it remains to show that if $V(x(t))=V(x_*(t))$ is a constant function, then $x_*(0)=0$ in order to showing that $\lim_{t\to\infty} x_*(t)=0$, or equivalently, $\lim_{t\to\infty}[x_i(t)-x_c(t)]=0$ for any $1\le i\le n$. Since $V=x_+-x_-$, $V(x(t))\equiv V(x(0))$ implies that x_+ and x_- are both constant functions (Lemma 2). If $x_+(0)=x_-(0)$, $x_*(0)=0$ and the all conditions of

Lemma 1 are fulfilled. Otherwise, we will find a contradiction under the assumption that the graph has a spanning tree.

Indeed, in this case, we have $x_+(0) > x_-(0)$. Since x_+ and x_- are both constant functions, $x_+(1) = x_+(0) > x_-(0) = x_-(1)$ Suppose the graph has a spanning tree with a root denoted as i_0. Without lose of generality, we may assume that $x_{i_0}(1) < x_+(1)$. Otherwise, we may replace x by $-x$. Since i_0 is a root of a spanning tree, there exist two nodes *i* and *j* such that $x_i(1) = x_+(1) > x_j(1)$ and $l_{ij} < 0$ where

$$i \in M_+ = \{1 \le l \le n | x_l(1) = x_+(1)\},\ j \in M_- = \{1 \le l \le n\} \setminus M_+$$

based on the fact that $i_0 \in M_-$. Then,

$$\dot{x}_i(1) = -\sum_{1 \le k \ne i} l_{ik}(x_k(1) - x_i(1)) < 0.$$

where $l_{ij} < 0$ was used. Consequently, x_i is strictly decreasing near 1. So there exists 0<*t*<1 such that

$$x_+(t) \ge x_i(t) > x_i(1) = x_+(1).$$

Since x_+ is a constant function, a contradiction is reached. Therefore $x_+(0) = x_-(0)$ and $x_*(0) = 0$. Based on Lemma 1, the following result is obtained.

Proposition 1. If the digraph *G* has a spanning tree, then every forward complete solution *x*(*t*) of $\dot{x} = -L(G)x$ satisfies $\lim_{t\to\infty}[x_i(t) - x_c(t)] = 0$ for any $1 \le i \le n$, where for each *i*, $x_i(t)$ is the *i*-th component of $x(t)$, $x_c(t) = (x_1(t) + \cdots + x_n(t))/n$, $L(G)$ is a graph Laplacian and *n* is the number of the vertices. ■

Remark 1. Here the choice of x_c is just for convenience. Indeed, any convex combination of x_i can still work. As soon as Theorem 1 was proven, it can be shown that there exist left eigenvectors of $L(G)$ corresponding to the zero eigenvalue that have non-negative entries based on Perron–Frobenius theorem. Let v_l denote the special left eigenvector with the sum of all its entries being 1. Then, $x_l = v_l^T x$ is a convex combination of x_i . Since $\dot{x}_l = v_l^T \dot{x} = -v_l^T L(G)x = 0,\ x_l(t) \equiv x_l(0)$ is a constant function. Thus, $\lim_{t\to\infty}[x_i(t) - x_l(t)] = 0$ as proven in Proposition 1 implies $\lim_{t\to\infty} x_i(t) = x_l(0) = v_l^T x(0)$ for any $1 \le i \le n$. This is a classic result (see [16, Lemma 3.7]) and can be viewed as a by-product of Proposition 1. In general, it is possible to show that $\lim_{t\to\infty} x(t) = Cx(0)$ where $C = \sum_{1 \le i \le k} c_i d_i^T$ with k being the dimension of the right kernel space of $L(G)$, and for each $1 \le i \le k$, c_i and d_i being some properly chosen right eigenvector and left eigenvector of $L(G)$ corresponding to the zero eigenvalue, respectively, such that $\{d_1, \cdots, d_k\}$ is an orthogonal set [19, Corollary 6.2]. ■

2.3. Proof of Theorem 2

In this subsection, Theorem 2 is proven by showing that $a) \Rightarrow b) \Leftrightarrow c) \Rightarrow d) \Rightarrow e) \Rightarrow a)$.

It is trivial to see that $d) \Rightarrow e)$ and $b) \Leftrightarrow c)$. Moreover, $a) \Rightarrow b)$ comes from Proposition 1. The following two steps show that $b) \Rightarrow d)$ and $e) \Rightarrow a)$. This will complete the proof of Theorem 2.

Step 1: $b) \Rightarrow d)$. Let x be any forward complete solution of $\dot{x} = -L(G)x$. Based on Lemma 2, we have

$$x_-(0) \le x_-(t) \le x_i(t) \le x_+(t) \le x_+(0)\ \forall t \ge 0$$

for any $1 \le i \le n$. Particularly, x is bounded and $\dot{x} = -L(G)x$ is marginally stable [5, Table 8.1]. Hence any eigenvalue of $L(G)$ that is not on the imaginary axis must have positive real part. We also know that 1_n is an eigenvector of $L(G)$ corresponding to the zero eigenvalue. To show d), two sub-steps are considered.

Step 1-1: We claim that $u \in \mathrm{S}pan\{1_n\}$ for any $u \in \Re^n \setminus \{0\}$ with $L(G)u = 0$, i.e., *u* is an eigenvector corresponding to the zero eigenvalue. Indeed, let x be the forward complete solution of $\dot{x} = -L(G)x$ with $x(0) = u$. Then, $x(t) = u, \forall t \ge 0$. So

$$u - u_c 1_n = \lim_{t\to\infty} x_*(t) = 0 \Rightarrow u \in \mathrm{S}pan\{1_n\}.$$

Therefore zero is a simple eigenvalue of $L(G)$ (where recall that $\dot{x} = -L(G)x$ is marginally stable and hence each Jordan block related to the zero eigenvalue is one dimensional, or equivalently, the algebraic multiplicity is equal to the geometric multiplicity [5]).

Step 1-2: We claim that $L(G)$ has no nonzero eigenvalues on the imaginary axis. Otherwise, there will exist two vectors $u, v \in \Re^n \setminus \{0\}$ such that for some $\omega_0 \in (0, \infty)$, $L(G)u = -\omega_0 v$ and $L(G)v = \omega_0 u$. Let x be the forward complete solution of $\dot{x} = -L(G)x$ with $x(0) = u$. Then,

$$x(t) = \cos(\omega_0 t)u + \sin(\omega_0 t)v$$

and

$$x_c(t) = \cos(\omega_0 t)u_c + \sin(\omega_0 t)v_c.$$

So $\lim_{t\to\infty} x_*(t) = 0$ implies $u - u_c 1_n = 0$ (by choosing $t = 2m\pi/\omega_0$ for any positive integer *m*) and $v - v_c 1_n = 0$ (by choosing $t = (2m + 1/2)\pi/\omega_0$ for any positive integer *m*). Particularly, $u \in \mathrm{S}pan\{1_n\}$ and $v \in \mathrm{S}pan\{1_n\}$. Hence $v = -L(G)u/\omega_0 = 0$. A contradiction is reached. Therefore $L(G)$ has no nonzero eigenvalue on the imaginary axis.

Based on Steps 1-1 and 1-2, $L(G)$ has a simple zero

eigenvalue and all the other nonzero eigenvalues have positive real parts. The step is then done.

Step 2: $e) \Rightarrow a)$. For any node $1 \le i \le n$, let $W_{out}(i)$ consist of those nodes $1 \le j \ne i \le n$ such that there is a directed path (walk) from i to j. Similarly, $W_{in}(i)$ consist of those nodes $1 \le j \ne i \le n$ such that there is a directed path (walk) from j to i. Let i_0 be a node such that $\#W_{out}(i_0) = \max_{1 \le i \le n} \#W_{out}(i)$. We claim that $\#W_{out}(i_0) = n-1$. Then, G has a spanning tree with i_0 as its root. If $\#W_{out}(i_0) < n-1$, a contradiction is found.

By definition, $W_{in}(i_0) \subseteq W_{out}(i_0)$, otherwise, $\#W_{out}(i_0)$ is not maximal one. Let

$$W_1 = \{i_0\} \cup W_{in}(i_0),\ W_3 = W_{out}(i_0) \setminus W_{in}(i_0)$$

and W_2 is the remaining vertices that are not in $\{i_0\} \cup W_{out}(i_0)$. Since $\#W_{out}(i_0) < n-1$, W_2 is not empty. By definitions, there is no edge (j,i) such that $i \in W_1$ and $j \in W_2 \cup W_3$. Similarly, there is no edge (j,i) such that $i \in W_2$ and $j \in W_1 \cup W_3$. Thus, for any $i \in W_1$, every non-zero entry of the i-th row of $L(G)$ must appear in the j-th column of $L(G)$ for some $j \in W_1$. Since $L(G)$ is a graph Laplacian, $Dim(Span\, S_1) \le \#W_1 - 1$ where S_1 consists of the i-th rows of $L(G)$ with $i \in W_1$. Similarly, $Dim(Span\, S_2) \le \#W_2 - 1$ with S_2 denoting the set of the j-th rows of $L(G)$, $j \in W_2$. Then,

$$\begin{aligned} &\text{The rank of } L(G) = Dim(Span(S_1 \cup S_2 \cup S_3)) \\ &\le Dim(SpanS_1) + Dim(SpanS_2) + Dim(SpanS_3) \\ &\le (\#W_1 - 1) + (\#W_2 - 1) + \#W_3 \le n-2, \end{aligned}$$

where S_3 consists of the l-th rows of $L(G)$ with $l \in W_3$. This violates condition e). Therefore $\#W_{out}(i_0) = n-1$ and G has a spanning tree with i_0 as its root. That is to say that a) holds. ■

Remark 2. In the proof of Theorem 2, the main step is to prove that a) implies b), i.e., Proposition 1 holds. In this paper, it was shown based on a LaSalle invariance principle (Lemma 1). Such an approach has a potential to extend similar results to more complex systems. For example, in [1, Theorem 3], a similar argument was used to study some consensus problem for *nonlinear systems*. It is also possible to be extended to *time-varying systems* based on the concept of limiting equations proposed by Artstein [2]. In fact, the same argument can be applied to the case of switching topology where the graph G depends on the communication situation when the so-called dwell-time conditions hold, see [10, 11, 12].

3. Conclusions

A novel proof of a well-known criterion on the characterization of the existence of spanning trees for a digraph was proposed based on a LaSalle invariance principle. A similar argument could to be extended to more complex consensus problems [12]. As a future work, we are currently studying the consensus problems with agent systems being nonlinear under directed switching topology networks.

Acknowledgements

We thank the support of National Science and Technology Council of Taiwan under contract MOST-111-2221-E-110-060-MY3, and the support of National Natural Science Foundation of China under Grant No. 62173092.

Appendix A: Proof of Lemma 1

Contradiction method is used. Suppose there exists a bounded forward complete solution x of (1) with $x(0) \in X$ and $\lim_{t \to \infty} x(t) \ne 0$. Then, there exist $\varepsilon_0 > 0$ and a sequence $t_m \to \infty$ such that $\|x(t_m)\| \ge \varepsilon_0$ for any positive integer m. Since x is bounded, $\dot{x} = Ax$ is also bounded. Hence $\{x_m(\cdot) = x(\cdot + t_m)\}$ is uniformly bounded and equi-continuous. By Arzelà–Ascoli lemma [8, Th. 3.1 of Ch. 3], there is a subsequence $\{m_k\}$ of $\{m\}$ such that $\{x_{m_k}\}$ converges uniformly to a continuous function $\bar{x}: [0,\infty) \to \Re^n$ on every compact subset of $[0,\infty)$. This implies

$$\begin{aligned} &\bar{x}(t) - \bar{x}(0) = \lim_{k \to \infty} x(t + t_{m_k}) - x(t_{m_k}) \\ &= \lim_{k \to \infty} \int_0^t Ax(\tau + t_{m_k}) d\tau = \int_0^t A\bar{x}(\tau) d\tau, \forall t \ge 0. \end{aligned}$$

Hence $\bar{x}$ is also a bounded forward complete solution of (1). Moreover, $\bar{x}(0) = \lim_{k \to \infty} x(t_{m_k}) \in X$ due to $x(0) \in X$ and X being A-invariant. Now

$$\begin{aligned} &V(\bar{x}(t)) - V(\bar{x}(0)) = \lim_{k \to \infty} V(x(t + t_{m_k})) - V(x(t_{m_k})) \\ &= \lim_{s \to \infty} V(x(s)) - \lim_{s \to \infty} V(x(s)) = 0, \forall t \ge 0. \end{aligned}$$

So $V(\bar{x}(t))$ is a constant function and hence

$$\lim_{k \to \infty} x(t_{m_k}) = \bar{x}(0) = 0$$

by the assumption. A contradiction is then reached due to $\|x(t_m)\| \ge \varepsilon_0$ for any m. This shows the lemma. ■

Appendix B: Proof of Lemma 2

Since $-x_- = \max_{1 \le i \le n} -x_i = (-x)_+$, we only need to show that x_+ is non-increasing without lose of generality. Since $\max_{1 \le i \le n} u_i$ is Lipschitz continuous, $x_+(t)$ is absolutely continuous. Thus, $\dot{x}_+(t)$ exists for almost all t in $\Re$. According to [3, Theorem 2.1],

$$\dot{x}_+(t) = \max_{x_i(t)=x_+(t),1\le i\le n}\{\dot{x}_i(t)\} \le 0$$

for almost all t in $\Re$ where in case of $x_i(t) = x_+(t)$,

$$\dot{x}_i(t) = -\sum_{1\le j\ne i} l_{ij}(x_j - x_i) = -\sum_{1\le j\ne i} l_{ij}(x_j - x_+) \le 0$$

was used with $l_{ij} \le 0$ denoting the (i, j) entry of the graph Laplacian matrix $L(\mathrm{G})$. This implies that x_+ is non-increasing. The proof is then done. ■

References

[1]. V. Amelkin, F. Bullo, and A. K. Singh, Polar opinion dynamics in social networks, *IEEE Trans. Autom. Control*, Vol. 62, No. 11, Nov. 2017, pp. 5650-5665.

[2]. Z. Artstein, Uniform asymptotic stability via the limiting equations, *J. Differential Equations*, Vol. 27, 1978, pp. 172-189.

[3]. F. H. Clarke, Generalized gradients and applications, *Trans. Amer. Math*. 1273 Soc., Vol. 205, 1975, pp. 247–262.

[4]. J. K. Hale, Ordinary Differential Equations, Melbourne, FL, *Krieger*, USA, 1980.

[5]. J. P. Hespanha, Linear Systems Theory, Princeton, NJ, *Princeton Univ. Press*, USA, 2009.

[6]. A. Jadbabaie, J. Lin, and A. S. Morse, Coordination of groups of mobile agents using nearest neighbor rules, *IEEE Trans. Autom. Control*, Vol. 48, No. 6, 2003, pp. 988-1001.

[7]. H. K. Khalil, Nonlinear Systems, 3rd Edition, *Prentice-Hall*, NJ, USA, 2002.

[8]. S. Lang, *Real Analysis*, *Addison-Wesley*, MA, USA, 1983.

[9]. P. LaSalle, Some extensions of Liapunov's second method, *IRE Trans. Circs. Th.*, CT-7, 1960, pp. 520–527.

[10]. T. C. Lee and Z. P. Jiang, Uniform asymptotic stability of nonlinear switched systems with an application to mobile robots, *IEEE Trans. Autom. Control*, Vol. 53, No. 5, 2008, pp. 1235-1252.

[11]. T. C. Lee, Y. Tan, and I. Mareels, Analyzing the stability of switched systems using common zeroing-output systems, *IEEE Trans. Autom. Control*, Vol. 62, No. 10, 2017, pp. 5138–5153.

[12]. T. C. Lee, Y. Tan, Y. Su, and I. Mareels, Invariance principles and observability in switched systems with an application in consensus, *IEEE Trans. Autom. Control*, Vol. 66, No. 11, 2021, pp. 5128-5143.

[13]. Z. Lin, Coupled dynamic systems: From structure towards stability and stabilizability, Ph. D. Dissertation, *University of Toronto*, Toronto, ON, Canada, 2005.

[14]. L. Moreau, Stability of multiagent systems with time-dependent communication links, *IEEE Trans. Autom. Control*, Vol. 50, No. 2, 2005, pp. 169-182.

[15]. R. Olfati-Saber, J. A. Fax, and R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proc. IEEE*, Vol. 95, No. 1, 2007, pp. 215-223.

[16]. W. Ren and R. W. Beard, Consensus seeking in multiagent systems under dynamically changing interaction topologies, *IEEE Trans. Autom. Control*, Vol. 50, No. 5, 2005, pp. 655-661.

[17]. Y. Su and J. Huang, Stability of a class of linear switching systems with applications to two consensus problems, *IEEE Trans. Autom. Control*, Vol. 57, No. 6, 2012, pp. 1420-1430.

[18]. Y. Su and T. C. Lee, Output feedback synthesis of multi-agent systems with jointly connected switching networks: a separation principle approach, *IEEE Trans. Autom. Control*, Vol. 67, No. 2, 2022, pp. 941-948.

[19]. J. Veerman and R. Lyons, A primer on Laplacian dynamics in directed graphs, *arXiv preprint arXiv:2002.02605*, 2020.

(1565)

Impact of the Programming Language used to solve a LMI Problem on the Fault Resilience of Cyber-physical Systems

F. N. Lagunes [1], J. Bosche [2] and D. Lara-Alabazares [1,2]
[1] Instituto Tecnológico Superior de Misantla, Carretera a Loma del Cojolite, 93821 Misantla, Veracruz, México
[2] Modeling, Information and Systems, UPJV, 33 rue Saint-Leu, 80000, Amiens, France
Tel.: + 33322825915, fax: + 33322825905
E-mail: jerome.bosche@u-picardie.fr

Summary: This work deals with a robust Fault-Tolerant Control (FTC) design for a class of uncertain systems. Fault resilience is associated with a robustness bound generated by a sufficient Linear Matrix Inequality (LMI) condition for static state feedback stabilization. This design control approach is based on solving an optimization problem expressed in terms of LMI with three different programming languages which are Matlab, Scilab and Python. Numerical validations were carried out with the model of a PV energy conversion system connected to a DC-DC boost converter. Then, a robustness analysis for fault resilience associated with a control law gains, obtained using the three solvers, was realized to investigate the best performance.

Keywords: Fault, Tolerant control, Robust state feedback control, LMI, DC/DC Boost Converter, Matlab, Scilab, Python.

1. Introduction

Cyber-physical Systems (CBS) are increasingly complex and sophisticated. The challenge is to exploit new technologies and hardware to make them more efficient and autonomous [1]. Sensors are associated with CBS to get precise environment information, allowing to the system be autonomous, as much as possible. However, this poses two major problems: (a) How to make the system react effectively based on real-time sensor data? (b) How to take into account a fault sensor during operation? FTC, studied in the field of control theory, can solve this problem. It is applied as a tool in transport [2], energy [3], industry 4.0 [4], etc. In the case of a partial fault, the FTC problem can be formulated as a state feedback control (SFC) problem for an uncertain system with a norm-bounded uncertainty. Then, a controller guaranteeing a certain level of performance for the largest uncertainty domain is designed. When the fault is total, a SFC augmented with an observer or outright, an output feedback control (OFC) law, are suitable. Finally, it is possible to consider the general case of partial and/or total faults by using a robust OFC. Concerning partial fault, the approach of [5] can be exploited to design a FTC. In this work, a D-stabilization technique is used to compute a passive fault-tolerant control law that guarantees the asymptotic stability of the system in the presence of faults.

On the other hand, the pole placement of the closed-loop system in a region D of the complex plane, allows the process to react efficiently to the fault. Whatever the technique, it must lead to a control law as robust as possible, the capacity of the system to tolerate the fault depending on this robustness. In fact, most of the work relating to this issue tends to reduce the conservatism induced by the sufficient conditions, related to a LMI problem, and leading to the gain of the controller. Some LMI-based conditions consider non-quadratic Lyapunov functions [6], other works rather exploit homogeneous polynomial Lyapunov functions [7] while promising results are obtained with fuzzy Lyapunov functions [8]. Then, significant advances have been made in this area, but these new conditions, less conservative, consist in LMI systems that are increasingly complex to solve from a numerical point of view. However, it is interesting to note that the robustness of these control laws may depend on the numerical computation software used to solve the problem, and more globally, on the associated programming language.

This article compares the robustness of a control law associated with a DC-DC boost converter. The control law consists of a robust pole placement problem by static state feedback and the programming languages considered are Matlab, Scilab and Python.

2. Preliminaries

We denote by M^T, the transpose of M. $\mathbb{I}_n$ is the identity matrix of order n. $\mathbb{0}$ is a null matrix of suitable dimension. $\{\blacksquare\}^{\mathrm{T}}$ standing block (i,j) in a matrix corresponds to the transposed term of block (j,i),. $\mathcal{C}$ is the complex plane and $\mathcal{C}^-$ stands for $Re(\mathcal{C}) < 0$

2.1 Polytopic and Norm-Bounded Representation

The Polytopic and Norm-Bounded (PNB) representation is used to robust analysis and control dedicated to systems affected by both polytopic and structured uncertainties, represented in the state space by the following model:

$$\begin{cases} \dot{x}(t) = A(\theta).x(t) + B(\theta).u(t) \ + \ D.\varphi(t) & (1) \\ y(t) = C(\Delta).x(t), & (2) \end{cases}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, and $\varphi(t) \in \mathbb{R}^v$ are the state, input, output and measurable

disturbance vectors, respectively, $A(\theta) \in \mathbb{R}^{n\times n}$ and $B(\theta) \in \mathbb{R}^{n\times m}$ represent the state and control matrices, belonging to a polytopic of matrices with N vertices:

$$[A(\theta)\ B(\theta)] = \sum_{i=1}^{N} \theta_i\,[A_i\ \ B_i]\ ,\sum_{i=1}^{N} \theta_i = 1 \tag{3}$$

$C(\Delta)$ is the ooutput matrix affected by $\mathcal{JL}$-structured uncertainty whose representation is recalled here:

$$C(\Delta) = C_0 + \mathbb{E} = C_0 + J.\Delta.L \tag{4}$$

where C_0 is a nominal matrix, $J \in \mathbb{R}^{n\,x\,q}$ and $L \in \mathbb{R}^{r\,x\,n}$ are known structured matrices whereas $\Delta \in \mathbb{R}^{q\,x\,r}$ is the uncertain matrix belonging to the ball $\mathcal{B}(\rho)$ such that $\|\Delta\|_2 < \rho$. $\mathbb{E}$ is a norm-bounded uncertainty whose structure is defined par J and L.

2.2. α-stabilization

This part addresses the design of robust feedback control for a system with PNB uncertainties and particular attention is paid to the system dynamic performance, such as the time response. Also, the placement of all spectrum of the state-matrix $\Gamma(A_C = A(\theta) + B(\theta).K)$ in $Re(\mathcal{C})$, is treated, defined by $x < \alpha$, and included in $\mathcal{C}^-$, is denoted as $\mathcal{C}^\alpha$, $x \in Re(\mathcal{C}^-)$. Then, when $\Gamma(A_C) \in \mathcal{C}^\alpha$, A_C is said α-stable.

A sufficient condition of robust α-stabilization by static state feedback deduced from that proposed by [5] and is recalled here. Therefore, it does not constitute an original contribution in the context of this work but rather a robust control law technique which will be processed by different solvers and analyzed.

Theorem 1. There is a matrix K associated with the robust state control law $u(t) = K.x(t)$ α-stabilizing system (1)-(2), if there is a solution to the following convex optimization problem

$$\begin{matrix} min \\ X,R,\zeta \end{matrix} \qquad \gamma = \rho^{-2} \tag{5}$$

where ζ is a positive scalar fixed a priori, $\mathrm{X} = X^T > 0$, $\mathrm{X} \in R^{n\times n}$, $\mathrm{R} \in R^{m\times n}$ and γ are variables verifying the LMI constraints

$$H_i(A_i, B_i, J_i, L_i, D) \ < \ 0\ ,\ \ \forall i \in \{1,..,N\} \tag{6}$$

with

$$H_i = \begin{bmatrix} \{-\alpha\mathrm{X} + A_i\mathrm{X} + B_i\mathrm{R}\} + \{\blacksquare\}^{\mathrm{T}} & \{\blacksquare\}^{\mathrm{T}} & \{\blacksquare\}^{\mathrm{T}} & \{\blacksquare\}^{\mathrm{T}} & \{\blacksquare\}^{\mathrm{T}} \\ L_i\mathrm{X} & -\mathbb{I}_r & 0 & \mathbb{O} & \mathbb{O} \\ L_i\mathrm{X} & \mathbb{O} & -\mathbb{I}_n & \mathbb{O} & \mathbb{O} \\ J_i^T & \mathbb{O} & \mathbb{O} & -\gamma\mathbb{I}_q & \mathbb{O} \\ D^T & \mathbb{O} & \mathbb{O} & \mathbb{O} & -\zeta\mathbb{I}_v \end{bmatrix}$$

A static state feedback Is then associated with:

$$K = R.X^{-1} \tag{7}$$

and the uncertainty domain for which the α-stability is guaranteed is given by $\|\Delta\|_2 \le \rho^* = \gamma^{-\frac{1}{2}}$, the optimal value of ρ.

2.3. Control Objective

Our control objective is to exploit the α-stabilization condition to develop a proportional-integral regulator, whose architecture is shown in Fig. 1.

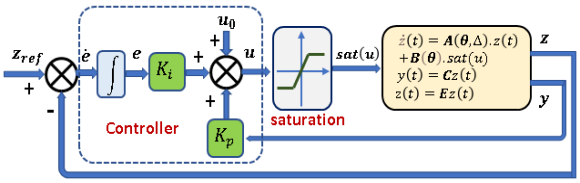


Fig. 1. PI controller architecture.

$z(t) \in \mathbb{R}^z$ is the controlled output vector and the problem consists in regulating $z(t)$ to a reference value $z_{ref}(t)$. The control law is then defined by:

$$u(t) = u_0 + K_p y(t) + k_i \int_0^t (z_{ref}(\tau) - z(\tau)).d\tau \tag{8}$$

The input command is saturated, $u(t) \in [\underline{u}; \bar{u}]$, with $\underline{\boldsymbol{u}} = u_0 - \delta_u(t)$ and $\bar{\boldsymbol{u}} = u_0 + \delta_u(t)$, where u_0 is the nominal value of u(t). In order to take into account the integral action inherent in this control architecture, the state-space representation (1)-(2) is extended to an augmented model with a state vector resulting from the concatenation of $x(t)$ and $e(t)$, the integral of the error between the reference signal $z_{ref}(t)$ and $z(t)$, the signals to be regulated. Without difficulty, we obtain the corresponding representation.

$$\begin{cases} \dot{X(t)} = \mathbb{A}(\theta).X(t) + \mathbb{B}(\theta).u(t) + \mathbb{D}.\varphi(t) + \mathbb{F}.z_{ref}(t) & (9) \\ y(t) = \mathbb{C}(\Delta).X(t) & (10) \end{cases}$$

$$\text{with} \begin{cases} X(t) = \begin{bmatrix} x(t) \\ e(t) \end{bmatrix},\ \mathbb{A}(\theta) = \begin{bmatrix} A(\theta) & \mathbb{O} \\ -E & \mathbb{O} \end{bmatrix},\ \mathbb{B}(\theta) = \begin{bmatrix} B(\theta) \\ \mathbb{O} \end{bmatrix} \\ \mathbb{D} = \begin{bmatrix} D \\ \mathbb{O} \end{bmatrix},\ \mathbb{F} = \begin{bmatrix} \mathbb{O} & \mathbb{O} \\ \mathbb{O} & \mathbb{I}_z \end{bmatrix},\ \mathbb{C}(\Delta) = \begin{bmatrix} C(\Delta) & \mathbb{O} \\ \mathbb{O} & \mathbb{I}_z \end{bmatrix} \end{cases} \tag{11}$$

3. FTC for DC-DC Boost Converter

3.1. Control Objective of a PV Conversion System

The DC-DC boost converter system, associated with a photovoltaic module, connected to a load as shown in Fig. 2, fits into ours problem and will be used to support a comparison of the fault resilience resulting from the resolutions of control optimization problems.

The converter consists of: a) a power transistor Q and a rectifier diode D (threshold voltage v_D) constituting the switching-mode power device, b) a smoothing inductance L with internal resistance r_L,

and c) a filtering voltage capacity C_1 and a bus capacity C_2. The control objective is to generate the duty cycle of signal $u(t)$ for the gate of Q to regulate the voltage $v_{PV}(t)$ to the reference photovoltaic voltage $v_{PVref}(t)$. An MPPT algorithm generates the reference PV voltage from data (weather, current, voltage) acquired on an experimental set-up whose parameters are given in Fig. 2. The corresponding $v_{PVref}(t)$ is computed over a 12 hour period starting at 7:00 am, showed at left side of Fig. 3, with the PV current $i_{PV}(t)$ generated by the panel shown at right side.

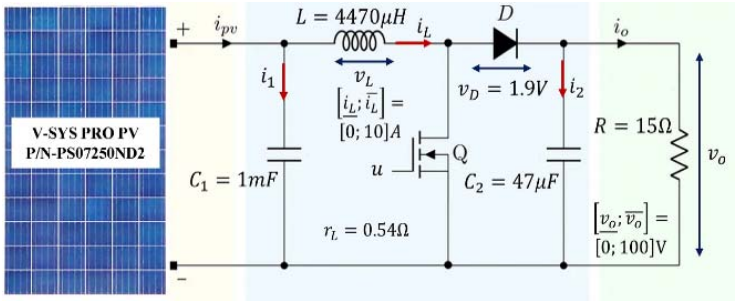


Fig. 2. DC-DC boost converter for a PV module.

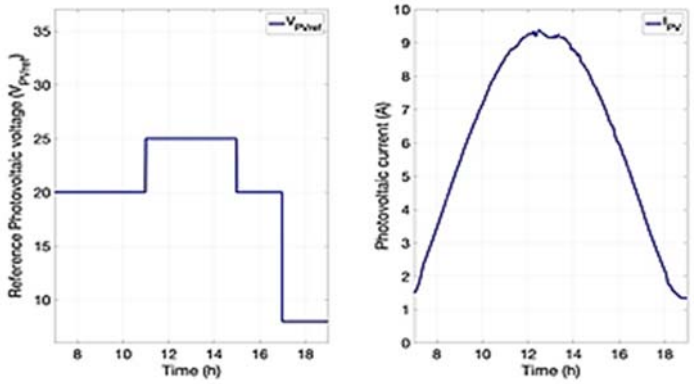


Fig. 3. Photovoltaic reference voltage and current.

3.2. PNB Representation of a Boost Converter

The averaged model of this system considers two sub-models: one is associated with an "off state" of the thyristor (u(t)=0), the other with an "on state" (u(t)=1). It can be written in the form (1) with $x(t) = [i_L(t) \;\; v_0(t) \;\; v_{PV}(t)]^T$, $\varphi(t) = [v_D \;\; i_{PV}(t)]^T$ where $i_L(t)$ is the inductor current, $v_0(t)$is the output voltage, and $v_{PV(t)}$ the PV voltage generated by the module. Basically, the dynamic equations of the boost converter can be expressed by the Linear Parameter-Varying (LPV) state-space equation (1) with:

$$\begin{cases} A = \begin{bmatrix} -\frac{r_L}{L} & \frac{(u_0-1)}{L} & \frac{1}{L} \\ \frac{(1-u_0)}{C_2} & \frac{-1}{RC_2} & 0 \\ \frac{-1}{C_1} & 0 & 0 \end{bmatrix} \\ B(\theta(t)) = \begin{bmatrix} \frac{(v_0(t)+v_D)}{L} \\ \frac{-i_L(t)}{C_2} \\ 0 \end{bmatrix} \; and \; D = \begin{bmatrix} \frac{(u_0-1)}{L} & 0 \\ 0 & 0 \\ 0 & \frac{1}{C_1} \end{bmatrix} \end{cases} \quad (12)$$

Note that the control matrix $B(\theta(t))$ depends on time-varying states $i_L(t)$ and $v_0(t)$, leading to a nonlinear the model. Nevertheless, the range of signal variation is bounded, as shown in Fig. 2, so $B(i_L(t), v_0(t))$ can be considered as a polytope of matrices as in (3) with N=4, $B_1 = B(\underline{i_L}, \underline{v_0})$, $B_2 = B(\underline{i_L}, \overline{v_0})$, $B_3 = B(\overline{i_L}, \underline{v_0})$ and $B_4 = B(\overline{i_L}, \overline{v_0})$.

On the other hand, under normal operating conditions, the states are measured by current and voltage sensors, respectively noted S_{I_L} , S_{v_0} and $S_{v_{PV}}$. Then, it is assumed that these sensors can be affected by a fault. Depending on the value of $C(\Delta)$ in (4), three cases are considered associated with $C_0 = \mathbb{I}_3$ and the corresponding values for *J* and *L*, as shown Table 1.

Table 1. Three fault cases for the DC-DC boost converter

Case	J	L	S_{I_L}	S_{v_0}	$S_{v_{PV}}$
1	$B(\theta)$	[1,0,0]	partial	none	none
2	$B(\theta)$	[0,1,0]	none	partial	none
3	$B(\theta)$	[0,0,1]	none	none	partial

According to part 2.3, it is clear that $z_{ref}(t) = v_{PVref}(t)$ and $z(t) = v_{PV}(t)$. It leads to $E = [0 \quad 0 \quad 1]$ and then, the augmented DC-DC boost converter model is easily deduced from (9), (10), (11) with (12). A fault-tolerant control law can then be computed if there is a solution to the convex optimization problem (5) verifying LMI conditions:

$$H_i(\mathbb{A}_i, \mathbb{B}_i, \mathbb{J}_i, \mathbb{L}_i, \mathbb{D}) \quad < \quad 0\,, \quad \forall i \in \{1,..,4\} \tag{13}$$

$$\begin{bmatrix} \mu_1 \mathbb{I}_n & R^T \\ R & \mathbb{I}_m \end{bmatrix} > 0 \tag{14}$$

$$\begin{bmatrix} \mu_2 \mathbb{I}_n & \mathbb{I}_n \\ \mathbb{I}_n & X \end{bmatrix} > 0 \tag{15}$$

where $\mathbb{J}_i = [J_i \quad \mathbb{O}]^T$ and $\mathbb{L}_i = [L_i \quad \mathbb{O}]$ and LMIs (14) and (15) respectively limit the 2-norm of R and X^{-1}, and consequently the controller gain values.

4. Numerical Simulations

The Fault-Tolerant Control technique presented in the section 2.3 has been tested on the DC-DC boost converter of section 3.1 for $\alpha = [-1 \; -10 \; -100]$ (transient performance), $\zeta = 1$ (disturbance rejection) and $\mu_1 = \mu_2 = 1000$ (limitation of gain values). Then, the optimization problem (5) considering LMI conditions (13), (14) and (15) has been solved using Matlab, Scilab and Python programming language for the three cases presented by Table 1, each time leading to a robustness bound ρ^*. Table 2 show the robustness bound values obtained for each case, respectively for α. Each robustness bound represents the fault tolerance of the remaining sensor affected, S_{I_L}, S_{v_0} or $S_{v_{PV}}$. Colored cells correspond to the highest bound.

Table 2. Robustness bounds ρ^*.

Case	α	Language		
		Matlab	**Scilab**	**Python**
1	-1	1.0836	0.9971	0.9474
	-10	0.4582	0.3824	0.4113
	-100	0.720	0.1346	0.0995
2	-1	0.8935	0.8396	0.75783
	-10	0.7364	0.6713	0.7003
	-100	--	--	--
3	-1	0.4578	0.4198	0.3901
	-10	0.6348	0.5768	0.6860
	-100	0.0848	0.0490	0.0320

Generally, Matlab seems to give significantly better results for $\alpha = -1$ and $\alpha = -10$ except for case 3 where the best fault tolerance is obtained with Python. It is logical to obtain less robust control laws for $\alpha = -100$, however it should be noted that Scilab is the most efficient in this case. In order to compare the quality of these FTC control laws, we carried out a simulation with Simulink for case 3 (fault on $S_{v_{PV}}$ sensor) and $\alpha = -10$. The magnitude of the considered fault is the one corresponding to the largest robustness bound, $\rho^{*Py} = 0.6860$, obtained with Python. For information, we give the gains of the controllers obtained by the three solvers and associated with the control law given in Table 3.

$$u(t) = 0.5 + k_i e(t) + k_{p1} i_L(t) + k_{p2} v_o(t) + k_{p3} v_{PV}(t) \tag{16}$$

Table 3. Gains obtained with the three solvers.

	Gain			
Solver	k_i	k_{p1}	k_{p2}	k_{p3}
Matlab	-28.3943	-0.1921	0.0136	0.1831
Scilab	-60.5522	-3.2052	0.1301	3.0070
Python	-1644.892	-5.1347	0.4208	4.3108

For simulation purposes, the fault is associated with an additive uncertainty on the output matrix, such as:

$$C(\Delta) = \mathbb{I}_3 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} . \rho^{*Py} . sin(2.\pi.t) . [0 \quad 0 \quad 1] \tag{17}$$

Photovoltaic voltage $v_{PV}^*(t)$ respectively obtained with Matlab, Python and Scilab are compared with $v_{PVref}(t)$ in Fig. 4. The three control laws are rather fault resilient but it is clear that the voltage regulation is better with the state feedback matrix K generated by Python. This is highlighted by the Fig. 5 which presents a zoom of Fig. 4. Finally, Fig. 6 shows the shape of the control inputs associated with the three previous control laws. Note that the signal generated by the controller respects $0 \leq u(t) \leq 1$.

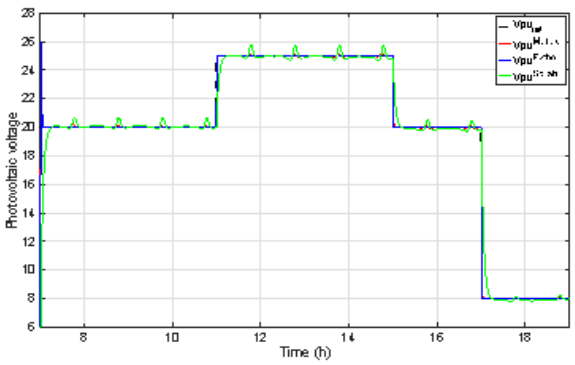


Fig. 4. Regulated photovoltaic voltages $v_{PV}^*(t)$.

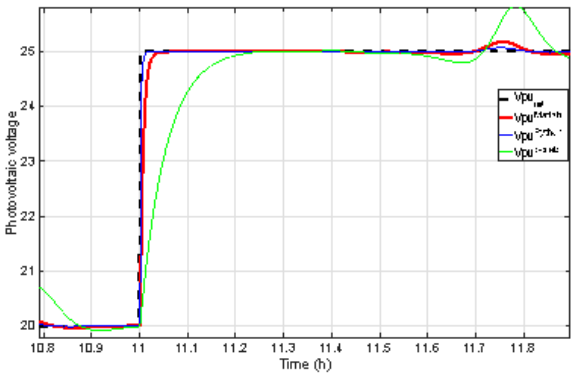


Fig. 5. Zoom of regulated photovoltaic voltages $v_{PV}^*(t)$.

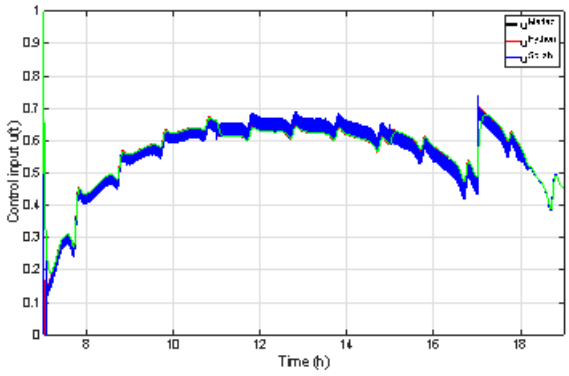


Fig. 6. Control signal $u(t)$.

4. Conclusions

This paper deals with fault resilience of DC-DC boost converter sensors associated with a robustness bound generated by a state feedback control technique. The resulting problem was solved using three programming languages. The contribution of this study is not to the development of a new fault-tolerant control technique but rather to the robustness analysis (and therefore the fault resilience) associated with a control law, depending on the programming language. The simulations carried out in this work tend to show that the most robust control laws are quite regularly obtained with the Matlab software. However, a much more in-depth study, with more case studies, would be needed to identify a real trend. The only possible statement following this work is that there is a certain disparity concerning the generated control laws, in terms of fault resilience, depending on the solver.

References

[1]. C.V. Lozano and K.K. Vijayan, Literature review on Cyber Physical Systems design, *Procedia Manufacturing*, Vol. 45, 2020, pp. 295-300.

[2]. A. Baldini et al., Fault Tolerant Control for remotely operated vehicles with thruster faults using non linear disturbance observers, *IFAC PapersOnLine*, Vol. 55, Issue 31, 2022, pp. 275-280.

[3]. Y. Zhang et al., Fault Diagnosis and Fault Tolerant Control of energy maximization for wave energy converters, *IEEE Transactions on Sustainable Energy*, Vol. 13, Issue 3, 2022, pp. 1771-1778.

[4]. A.B. Masood et al., Control over blockchain for Data-Driven Fault Tolerant Control in Industry 4.0, *IEEE Mediterranean Communication and Computer Networking Conference (MedComNet)*, Pafos, Cyprus, 01-02 June 2022.

[5]. M.A. Ragaieg et al., A robust output feedback input constrained design for a class of LPV system, in *Proceedings of the Mediterranean Conference on Control and Automation (MED'2020)*, Saint-Raphaël, France, September 2020, pp. 284 - 291.

[6]. A. Cherifi et al., Global non-quadratic D-stabilization of Takagi-Sugeno systems with piecewise continuous membership functions, *Applied Mathematics and Computation*, Vol. 351, Issue 1, 2019, 10.1016/j.amc.2019.01.031.

[7]. G. Chesi, LMI conditions for time-varying uncertain systems can be non-conservative, *Automatica*, Vol. 47, Issue 3, 2011, pp. 621-624.

[8]. X. Fan et al., A fuzzy Lyapunov function method to stability analysis of fractional-order T-S fuzzy systems, *IEEE Transactions on Fuzzy Systems*, Vol. 30, Issue 7, 2022, pp. 2769-277.

(1609)

Artificial Intelligence Assisted Ultrasonic Technology for Efficient Seaweed Material Recovery

S. Ozolina, A. Kukuskins and U. Zaimis
Liepaja University, Faculty of Science and Engineering, Institute of Science and Innovative Technologies,
Liela iela 14, LV-3401, Liepaja, Latvia.
Tel.: +371 29101515
E-mail: sintija.ozolina@liepu.lv, liepu.andrejs.kukuskins@gmail.com, uldis.zaimis@liepu.lv

Summary: This paper examines the potential of using ultrasound as a method to clean Furcellaria lumbricalis, a type of macroalgae found in the Baltic Sea. Ultrasound is becoming increasingly popular as a cleaning method due to its non-invasive and efficient nature. This paper investigates the effects of different solution densities and durations on the cleaning success of F. lumbricalis and AI application to the process. A series of experiments are conducted on samples of F. lumbricalis collected from the Baltic Sea and results are analyzed. This study provides a better understanding of the effectiveness of ultrasound for cleaning F. lumbricalis and provides further evidence of its potential as a non-invasive and efficient cleaning method and AI usage in automation process.

Keywords: Ultrasound cleaning, Furcellaria lumbricalis biomass, AI application, Automation.

1. Introduction

Furcellaria lumbricalis is a type of macroalgae that can be cleaned using ultrasound devices. Ultrasound devices produce sound waves that vibrate the water, causing the algae to be removed from sand particles.

Sand particles and other impurities [1] in macroalgae Furcellaria lumbricalis can cause damage to further processing equipment. These impurities can clog machinery and interfere with normal operation, leading to slower processing times and reduced efficiency. They can also cause wear and tear on the equipment, leading to higher maintenance costs and a shorter lifespan. Sand particles and other impurities can also contaminate the product, leading to a reduced shelf life and lower quality. It is important to remove these impurities before further processing to ensure safety, quality, and efficiency.

This paper discusses the process of cleaning macroalgae Furcellaria lumbricalis using ultrasound device, as well as the advantages and disadvantages of this technology and incorporation AI to the macroalgae cleaning process.

2. Materials and Methods

The best ultrasound frequency for removing sand impurities from macroalgae would depend on the size of the particles that need to be removed. Generally, frequencies between 20 kHz and 50 kHz are used for removing particles in the range of 0.1 to 1 mm [2].

The size of sand particles in the Baltic Sea can vary greatly, ranging from very fine particles (less than 0.063 mm in diameter) to coarse particles (up to 1 mm in diameter).

50 MHz ultrasound device for research applications needed.

The device consists of an ultrasound bath and transducers.

In this study, a primary experiment was conducted to examine the cleaning effectiveness of ultrasonic cleaning using various density salt solutions on Furcellaria lumbricalis. For each sample, 100 g of untreated Furcellaria lumbricalis was used and 200 ml of different density salt solutions were applied. Four ultrasonic cleaning experiments were performed using a 50 Hz ultrasonic transducer. The results of these experiments were analyzed to determine the influence of solution density on the cleaning effectiveness of ultrasonic cleaning.

The density of the solution in an ultrasound bath has been shown to impact the cleaning performance of the bath. The density of the solution can affect the cleaning power. If the solution is too dilute, it may not effectively remove contaminants from the items being cleaned [3]. Conversely, if the solution is too concentrated, it may be too harsh and potentially cause damage to the items being cleaned. It is crucial to find best density of the solution in the ultrasound bath to ensure optimal cleaning performance.

In order to optimize the cleaning process, the addition of dissolved carrageenan to the cleaning solution may be considered as a thickening agent. The use of excess production carrageenan in the cleaning process may be facilitated through the use of carrageenan as solution for ultrasound cleaning.

3. AI Usage to Optimize the Cleaning Process

Artificial intelligence (AI) can be used to automate the process of visual inspection for detecting sand in algae through the use of computer vision. This field of AI involves using computers to interpret and analyze images and video.

To utilize computer vision for this purpose, images of the algae and sand mixture must be captured using a microscope or other imaging device and analyzed using image processing algorithms. These algorithms can be trained on a large dataset of images with and without sand to recognize the presence of sand in the images and classify the images as "clean" or "contaminated" based on the presence or absence of sand. Cleaning process algorithm shown in Fig. 1.

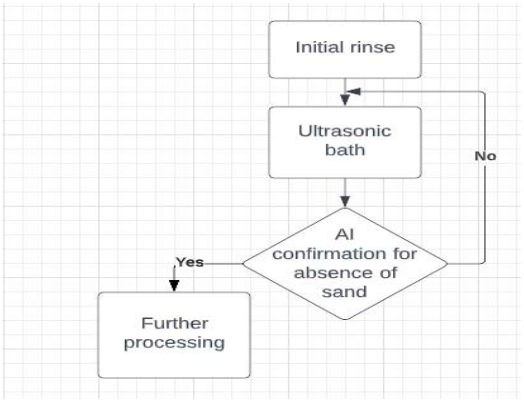


Fig. 1. Cleaning process algorithm.

The accuracy of the system depends on the quality of the images, the training data, and the complexity of the image processing algorithms used.

There are various libraries and frameworks for computer vision, including scikit-image, SimpleCV, MATLAB Image Processing Toolbox, and OpenCV-Python. The latter, OpenCV (Open Computer Vision), is a popular library of computer vision algorithms written in C++ that can be used to analyze and interpret images and video, including functions for image processing, computer vision, and machine learning [4].

A tank with a see-through bottom can potentially be used to train an AI system to detect sand separation and accumulation in algae by capturing images or video of the mixture from below. However, using a single perspective may not provide a complete view of the mixture and may make it difficult for the AI system to accurately detect sand. Additional perspectives, such as a view from the top, can provide a more comprehensive view of the mixture and improve the accuracy of the AI system.

In general, higher image quality, including factors such as resolution, contrast, and lighting, is more useful for accurately detecting the presence of sand in the algae. Higher resolutions and pixel densities can also be helpful in accurately detecting small particles of sand in the mixture. As a guideline, images with a resolution of at least 300-600 pixels per inch (ppi) and a pixel density of at least 300-500 pixels per square millimeter (ppmm) may be sufficient for distinguishing small particles of sand in the algae mixture, depending on the size and distinctness of the algae cells and sand particles.

3. Conclusions

In conclusion, this study demonstrated the potential of using ultrasound as a method to clean Furcellaria lumbricalis, a type of macroalgae found in the Baltic Sea. The results showed that different solution densities and durations had an impact on the cleaning success of F. lumbricalis, and that the addition of carrageenan to the cleaning solution may be useful for optimizing the cleaning process.

The incorporation of artificial intelligence in the cleaning process considered to be beneficial for optimizing the cleaning efficiency and automation of the process.

These findings provide further evidence of the effectiveness of ultrasound as a non-invasive and efficient cleaning method for F. lumbricalis, and the potential for AI to improve the automation of the cleaning process.

Acknowledgements

This work was carried out within partial financial support from the European Regional Development Fund Project No. 1.1.1.5/018 "Promotion of research, innovation and international cooperation in science at Liepaja University".

References

[1]. Drinkwater, Bruce W., Dynamic-field devices for the ultrasonic manipulation of microparticles, *Lab on a Chip,* 16, 13, 2016, pp. 2360-2375.

[2]. Rapouch, K., and M. Mrázek, Use of ultrasound for cleaning of components of historical vehicles in Technical Museum in Brno, *Koroze a ochrana material*, 64, 3, 2020, pp. 79-86.

[3]. Mason, Timothy J., Ultrasonic cleaning: An historical perspective, *Ultrasonic Zoochemistry*, 29, 2016, pp. 519-523.

[4]. OpenCV. 2021. About – OpenCV. [online] Available at: <https://opencv.org/about/> [Accessed 19 December 2022]

(1985)

Deep Learning Based Anomaly Detection for Predictive Maintenance

B. Hilliger [1], Y. Rebahi [1], B. A. Lepadatu [2] and A. Cardoso [3]
[1] Fraunhofer Fokus, Kaiserin Augusta Allee 31, 10589 Berlin, Germany
[2] Continental Automotive Romania S.R.L, Siemens Str 1, 300704 Timisoara, Romania
[3] University of Coimbra, Centre for Informatics and Systems of the University of Coimbra (CISUC), Coimbra, Portugal
E-mail: yacine.rebahi@fokus.fraunhofer.de

Summary: In this paper, two algorithms for deep learning based anomaly detection in predictive maintenance are compared. The two algorithms are a Long-Short-Term-Memory (LSTM) Autoencoder and a LSTM Predictor. As a variation both algorithms are also combined with Principal Component Analysis (PCA). The underlying anomaly detection works by taking the output of the neural networks and using thresholds and a moving-average to define what is labeld as an anomaly. To compare the models, cross validation is done to get the best shape of the deep learning model, and then a scoring system is developed to compare the anomaly prediction performance of each model. The comparison showed that the models have their own strengths as the LSTM Autoencoders have a lower prediction error and are more robust with respect to some parameters, while the LSTM Predictor if configured correctly can achieve a better anomaly detection performance. The data that was used to evaluate the algorithms came from a real production environment.

Keywords: Deep learning, Predictive maintenance, Industry 4.0, LSTM, Autoencoder, PCA, KYKLOS 4.0.

1. Introduction

Traditionally, industry maintenance is done in one of two ways, corrective, or preventive. Corrective maintenance aims to repair a given part once it is broken. Preventive maintenance aims to replace a part before it is broken, for example by defining a fixed lifetime for each part. The problem with corrective maintenance is that, one might have long downtimes. Preventive maintenance, on the other hand, might be cost inefficient because most of the time, the replaced parts are still in a decent condition and could run for longer [1]. To find a better maintenance strategy, predictive maintenance (PdM) was introduced. The goal of PdM is to detect machine failures, degraded performance, or a downtrend in product quality before one of these occur [2]. Due to its importance, the German Federal Ministry of Affairs and Energy has published "The Standardization Roadmap of Predictive Maintenance for Sino-German Industry 4.0/ Intelligent Manufacturing" [3], while the European Federation of National Maintenance Societies (EFNMS) has published the maintenance Body of Knowledge (BoK) [4]. Predictive maintenance can be implemented in different forms. One form is a data driven way. This means, based on machine learning techniques, a model can learn the machine behaviour by looking at past data from the machine. Another way is to use expert knowledge to build a predictive maintenance model. This has the disadvantage that one needs a lot of expert knowledge to build a model that would not be easily adapted if the machine changes behaviour [5]. With all the advantages that artificial intelligence and machine learning can bring to the PdM field, it seems that data driven solutions in the context of industry 4.0 are still in their early stage deployment [6]. The study from 2019 simply reveals that many companies are either not recording the data or have no way to use it to improve their services. This paper is organized as follows. Section 1 introduces the topic being investigated. Section 2 discusses the related state of the art. Section 3 highlights the algorithms being implemented, section 4 presents the experimental results, and section 5 concludes the paper.

2. Related Work

The detection of anomalies is at the core of PdM, with the main focus early detection of anomalies in equipment and its components, as well as the generation of alerts that allow the proper scheduling of maintenance activities. Being a relevant topic for industry 4.0 and circular manufacturing, there have been several activities related to the detection of anomalies with predictive maintenance, using different data-driven approaches and applications. For example, Kamat and Sugandhi present in [7] a survey where the challenges of traditional anomaly detection strategies are described and where a deep learning technique is proposed to early predict anomalies. Tercan and Meisen in [8] conduct a systematic review on predictive quality in manufacturing, Carvalho et al in [9] present a systematic literature review of machine learning methods applied to predictive maintenance, and Sharma et al in [10] present a review on condition-based maintenance using machine learning. In particular, deep learning algorithms used for predictive maintenance are discussed by Soahaib et al in [11].

In terms of application examples, the work by Davari et al in [12] propose a data-driven PdM framework for a system in the railway industry using a deep learning based approach to detect abnormal data and reduce the false alarm rate. A PdM approach using

deep learning is developed by Neto et al, in [13] for predicting the current health status of rolling bearing components. In this article, two algorithms for detection of anomalies based on deep learning in predictive maintenance of metal press machines are implemented and compared.

3. Anomaly Detection based Predictive Maintenance: The KYKLOS 4.0 Approach

Preventive maintenance is often used in industry to avoid machines beakdowns and production interruptions regardness the costs that can lead to. Such approaches can make the use of data driven techniques difficult. For instance, supervised learning is a method requiring the knowledge of data labels that reflect some information related to previous occurred failures. Unfortunately, with frequent preventive maintenance operations, it would be very difficult to obtain such labels. For this reason, unsupervised learning seems to be a reasonable approach to tackle data driven predictive maintenance. To achieve this, we will be monitoring the past behaviour of a machine over a certain period of time. The longer the time interval is, the more accurate the prediction algorithms would be. LSTM is a type of recurrent neural network architectures that has shown efficiency in various areas including time series based prediction [14]. In this paper, anomaly detection will be realised through two optimizations of LSTM: LSTM autoencoder (or reconstruction model) and LSTM predictor (prediction model). The two models will be compared to each other in terms of accuracy and performance. The comparison will be done on data collected from a real production environment in the Continental premises in Romania. For the used dataset, a scaled version and a PCA dimensionality reduced version will be used as input for the neural networks. The anomaly detection algorithm is based on [15] and can be split into three parts. The first part is the pre-processing of the data. Mainly meaning feature engineering and scaling or dimensionality reduction with principal component analysis. Then the pre-processed data is fed through a previously trained reconstruction or prediction model. The predicted values are then fed into the evaluation algorithm. The evaluation is used for transforming the predicted values to an anomaly score.

3.1. The Used Dataset

The Continental Dataset was recorded over a period of 7 hours on a Printed Circuit Board (PCB) manufacturing machine at the company Continental. This machine has overall 50 sensors and additionally a state indicator called "Step Number". Of these 50 sensors, 46 are binary sensors and the other four are continuous variables. The sensors were measured each second in the 7-hour period, producing 25695 rows. Since there is only a 7-hour period available, it is unlikely that there are any anomalies in the dataset that point to the machine needing maintenance. But there are 2 main reconfiguration events happening in this 7-hour period which can be detected by the anomaly detection algorithms. The information that those are reconfiguration events came from Continental, but the partners from Continental were not able to reconstruct the exact timing those events occurred. The machine follows a certain cycle for producing PCB Boards. These cycles are not fixed but differ always a little from the previous cycles, even if it is only the timing between each step.

3.2. Data Pre-processing

The pre-processing for this dataset is as follows: First of all, 9 columns are dropped because they are constant, so they do not provide any information. As mentioned before, there are 25695 rows in the dataset but many of them are repeating because the machine does not change its state each second, for instance, when waiting for the operator to take out the produced peace or when a production step takes longer than one second. Therefore, the repeated rows were taken out, and a duration column was added to the data storing the number of seconds the machine was in that state. Because the time it took for the operator to take out the produced piece varies a lot and can get quite long (up to 83 Minutes), the duration for these steps was set to 1 so the networks do not have to try to predict the time for operator to do this. This is because the time the operator takes, points to human error and not to machine degradation. After the rows are dropped this leaves a dataset with 4690 rows.

3.3. Models Building

With the pre-processed data, a model is trained. For training, it is useful to only include data where it is known the machine is in a "healthy" state, hence it is a semi supervised predictive maintenance technique. To get a better understanding of the models, it is useful to look at the Input and Output shapes of the model. As described, the data is transformed into N sequences of length $seqlen$. These sequences are fed into the model recursively. Because the model mainly consists of LSTM cells, this sequential data is fed into the model step by step. The output of each of the two model types is different. The output of a LSTM-Autoencoder is a sequence with the same shape as the input sequence. The output of a LSTM-Predictor, on the other hand, will be a vector of the values the network believes to be in the next timestamp. So, in other words, it is a sequence of length one. When training neural networks, it is always important to have a technique to avoid overfitting. For this, early stopping is applied when training the network. This means, after each epoch of training, the prediction error will be measured on a small randomly selected validation set. If this error does not improve for a certain number of rounds, the training will be stopped [16]. After the training is

done, the training error is measured because the training error is used for the evaluation part.

3.4. Evaluation

The evaluation algorithm works as follows:

1. Measure Error of predicted values. There are different options to measure the error. In this case, the squared error was chosen because it penalizes a large difference between predicted and true values. The squared error is defined by the following formula:

$$error_i = (y_{true_i} - y_{predicted_i})^2 \quad (1)$$

2. Define a threshold from training error. With N being the number of training samples, the parameter σM can control the likelihood of sample to be classified as anomalous and $\sigma(trainerror)$ being the standard deviation of the train error computed with Equation (1).

$$threshold = mean(trainerror) + \sigma M * \sigma(trainerror) \quad (2)$$

$$threshold = \mu + \sigma M * \sqrt{\frac{1}{N}\sum_{j=1}^{N}(trainerror_j - \mu)^2}, \quad with\ \mu\ \frac{1}{N}\sum_{j=1}^{N} trainerror_j \quad (3)$$

3. Test the error of each sample against this threshold:

$$low\_level\_anomalies = \begin{cases} 1 & \text{if } error_i > threshold \\ 0 & else \end{cases} \quad (4)$$

4. Because LSTM Autoencoder outputs a sequence of predicted values for each vector in the original dataset, there will be $seqlen$ vectors to test against the threshold. To reduce this down to one majority voting is applied. This means that if more than half of the errors corresponding to a certain vector is higher than the threshold, the corresponding vector will be classified as an anomaly. This idea was also discussed in [17].
5. Look at how many samples surpassed the threshold in last WS timesteps ($WS = window\ size$):

$$anomalyscore_i = \frac{1}{WS}\sum_{j=i-WS}^{i} low\_level\_anomalies_i \quad (5)$$

The resulting $anomalyscore_i$ can be seen as the likelihood of the of the i-th sample to be anomalous.

This evaluation algorithm is for a univariate dataset. If it is applied on a multivariate time series, the steps 1-5 are the same, but they need to be applied for each variable in the dataset. Before this algorithm is deployed, it is useful to do two more steps.

6. For each column in a subcomponent in the machine, we take the mean of each anomaly score, thus deriving with an $anomalyscore$ for the whole subcomponent.

$$high\_level\_anomalies_i = \begin{cases} 1 & \text{if } anomalyscore_i > AST \\ 0 & else \end{cases} \quad (6)$$

7. Define an anomaly-score-threshold (AST) for each subcomponent above which the subcomponent will be defined as "anomalous" or "broken".

If $high\_level\_anomalies_i$ is equal to one, maintenance personnel should check out the machine or a certain component of the machine.

A subcomponent in this case can be for example a motor with different sensors, like torque, energy consumption and drawn current. Of course, it is also useful to have a subcomponent which includes all columns of the machine which could be described as the overall state.

3.5. Comparison Method

As already mentioned, the goal of this paper is to compare and evaluate reconstruction and prediction models. This comparison and evaluation will be done in the following way:

1. Take a portion of the dataset where no anomalies are assumed to be in as the training set.
2. Use k-fold cross validation to find a good number of layers and determine a good size of layers. The number of parameters per model will be kept roughly in the range of the rows in the dataset.
3. Evaluate the prediction performance of each model by taking the mean of the errors for each model size.
4. Use the best model parameters to train a new model on the whole training set.
5. Use this model to calculate the anomaly score for the test set.
6. Compare anomaly scores of the different models

To get a baseline for step 6. A PCA based model is also used. This model does not use deep learning to detect anomalies but only relies on PCA and will be called PCA Model in the following sections.

4. Experiments and Results

4.1. Model Selection

The Model selection was done using 5-fold cross validation. The goal of the cross validation was to find a good model but not necessarily to find the optimal model, and to test as many different configurations as possible, so a general idea of which models perform best can be derived. For the training, a model had up to 5000 epochs to finish training but usually with early stopping the training stopped between 500 and 1500 epochs of training. The training was done in mini batches of size 256, meaning the update of the weights was done after the error was measured for 256 samples. Before starting the cross validation, the number of layers and the size of each layer had to be chosen. This was achieved in a way to include as much variation as possible, for example a network with very few nodes and only one layer, and a network with many nodes

and many layers. The only constraint to this, was that the number of parameters of each network was to be kept below 10 000. This number was chosen because it seemed like a decent value to have enough possible variation in the number of layers and size of each layer, while also not having too many parameters compared to the relatively small training sets. To find this threshold, larger networks were also tested but there turned out to be no significant improvement, so they were not included. Another important parameter to choose for the autoencoders, was the size of the latent space. The smaller the latent space, the more the network must learn how the features relate to each other, because the network must compress the values much more. But if the latent space is too small, the model will only be able to reconstruct the main features and loose some smaller nuances of the data. But the smaller nuances could be important because they might be anomaly indicators. The tested latent spaces were 1, 5, 10. The architecture of each network is displayed with each datapoint, so for example [5,1,5] means the network had an input layer, which depends on the number of input variables, three LSTM layers with 5, 1, and 5 nodes and last an output layer which is the same size as the input layer. For the scaled dataset (not the one transformed with PCA), the Autoencoders perform better than the LSTM Predictors. The results are depicted in Fig. 1. Interestingly even an autoencoder with a latent space of one is able to perform better than the LSTM predictors. This means that it is quite easy for a model to reconstruct the data but quite difficult to predict what will happen next. This indicates that Autoencoders might be more useful for this use case. Similar results can be observed in the PCA transformed model selection in Fig. 2 except that the Predictors are performing as good as an autoencoder with latent space 5 and better than one with latent space of 1.

Fig. 1. Model selection results for scaled continental dataset.

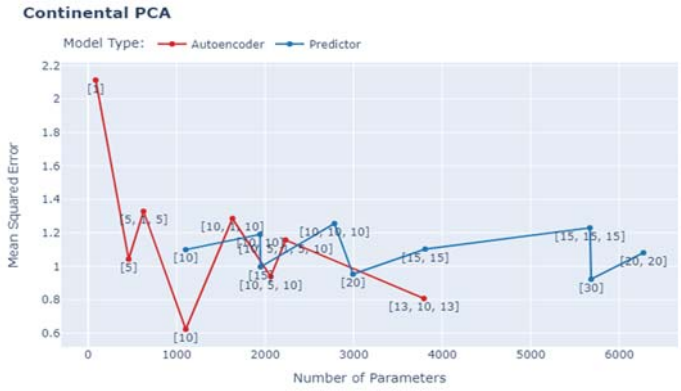


Fig. 2. Model selection results for the PCA transformed continental dataset.

4.2. Anomaly Detection Performance

To measure the anomaly detection performance, the best prediction model from model selection and the best two autoencoders with different latent spaces are compared. The evaluation of the Anomaly Score on this dataset is complex because no information about when exactly which part of the machine is anomalous is available. Continental confirmed that the two anomalies found in the test set, are reconfiguration events, but they could not reconstruct the exact timing as to when these reconfiguration events happened. The problem with the anomalies being reconfiguration events instead of failures, is that reconfiguration events do not mean there happened degradation in the data before. Because of this, the goal of this evaluation will be to detect the reconfiguration events as clear as possible, as this could be an indicator that other anomalies might be recognized more clearly as well. This evaluation will be done by comparing different

metrics of the anomaly score. The first two metrics will be the mean and standard deviation of the anomaly score of the training set. On the test set, the maximum value will be taken as the third metric. A good performing model is expected to have a low mean and a low standard deviation on the training set but a high maximum value in the test set because this points to a truly clear detection of the anomalies. To make this clearer, a simple scoring algorithm was derived, and it can be described by the following formula:

$$score = \frac{max_{test} - mean_{train}}{std_{train}}$$

In other words, the formula is saying how often the standard deviation would have to be added to the training mean to get to the test max. The higher the score, the better because this means the anomaly is clearer. Before looking at these metrics, it is good to get a feeling of the anomaly score by looking at the anomaly score for each model. This is shown in Fig. 3. It shows that the anomaly score for scaled models has much clearer anomalies, than the deep learning models that used PCA as an Input.

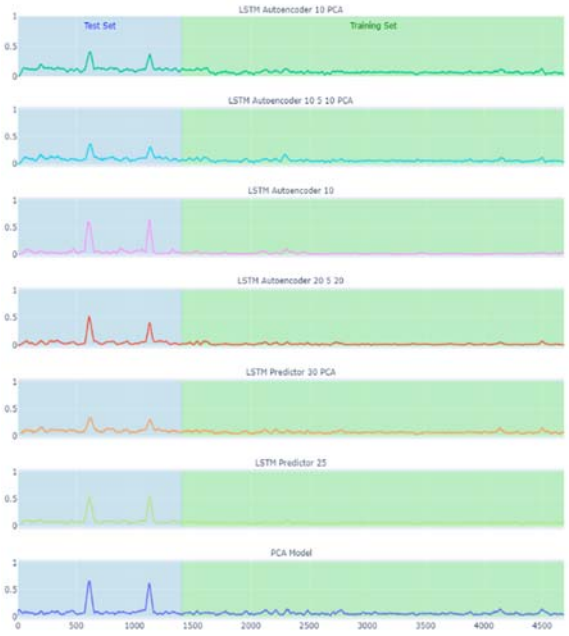


Fig. 3. Anomaly scores for the three models that were trained on the Continental scaled dataset.

All this is also reflected when looking at the metrics. The results from the metrics are shown in Table 1. There, it is confirmed what was already derived by the qualitative analysis from the graphs: The LSTM Predictor 25 and LSTM Autoencoder 10 perform best, while the models where the data was transformed by PCA before it was fed into the model perform worst. This is quite surprising because the PCA Model is performing quite average but the models where the PCA transformation was also used were not able to use this advantage.

Table 1. Results of the Continental Anomaly Detection with retransforming to original number of features.

$\sigma M = 1, WS = 30$	Train mean	Train std	Test max	Score
Predictor 25	0.063	0.011	0.55	44.52
Autoencoder 10	0.014	0.014	0.649	44.13
Autoencoder 20 5 20	0.022	0.015	0.518	33.94
PCA Model	0.06	0.021	0.653	28.69
Autoencoder 10 PCA	0.071	0.02	0.415	17.02
Autoencoder 10 5 10 PCA	0.061	0.02	0.362	15.16
Predictor 30 PCA	0.066	0.018	0.333	14.9

In Table 1, the PCA Autoencoder and PCA Predictors were retransformed before evaluation. If this step is left out, the results for those are much better. The values for this are shown in Table 2. The Predictor 25 and the Autoencoder 10 are still the best performing models, but then two of the models that used PCA pre-processing, the Predictor 30 PCA and the Autoencoder 10 PCA perform almost the same. The Autoencoder 20 5 20 and the Autoencoder 10 5 10 PCA have similar scores as well. Last in the scoring is the PCA Model. What is also noticeable is that the values in the Test max column are around 0.8 for the LSTM PCA models in Table 2, while they were between 0.33 and 0.42 in Table 1.

Table 2. Results of Continental anomaly detection without retransforming the LSTM PCA models.

$\sigma M = 1, WS = 30$	Train mean	Train std	Test max	Score
Predictor 25	0.063	0.011	0.55	44.52
Autoencoder 10	0.014	0.014	0.649	44.13
Predictor 30 PCA	0.057	0.017	0.797	43.71
Autoencoder 10 PCA	0.04	0.019	0.818	40.63
Autoencoder 20 5 20	0.022	0.015	0.518	33.94
Autoencoder 10 5 10 PCA	0.056	0.023	0.805	33.21
PCA Model	0.06	0.021	0.653	28.69

With this, it can be concluded that it is not recommendable to retransform the output after the deep learning model predicted the values of the PCA transformed dataset. This poses the problem that all subcomponents have to be known before training and cannot be defined after the training. But knowing the subcomponents beforehand might be good anyway because it might be good to separate certain similar subcomponents and train only a single more general model for them. For this, it is important that the setup of the machine is well known by the data scientist that creates the model. There are also a couple of other conclusions from this analysis. The first one is that the

LSTM Predictor and LSTM Autoencoder perform almost equally well. The second conclusion is that the autoencoder with a latent space of 10 performs better than an autoencoder with a latent space of 5. So a larger latent space seems to be beneficial in this domain.

4.3. Parameter σM

In this section, a closer look at one of the three parameters of the evaluation algorithm will be taken because the results of this paper are highly dependable on σM.

The multiplier σM mainly controls how easily a sample is classified as an anomaly. A higher σM means fewer samples will be classified as an anomaly.

In Fig. 4, the Score of the seven different models for different σM is shown. For small and large σM, the LSTM Predictor 25 performs best, but there are also times where it performs worst. Nonetheless, the scaled models perform better than the PCA transformed across most tested σM.

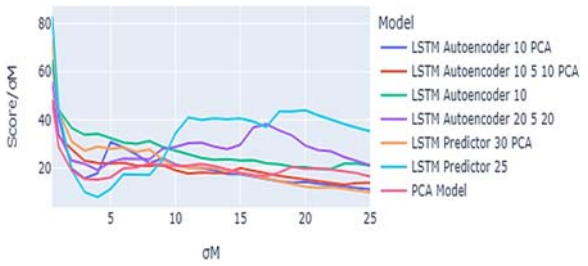


Fig. 4. Scores based on different σM for the Continental dataset.

Even though the models are highly dependent on σM when comparing them to each other in a real production environment, with only one or two models deployed, this would not matter too much. Instead, trial and error can be a strategy to find a good σM. First, a rather low anomaly threshold could be taken. If the threshold is passed very often and leads to many false positives, it could be increased. This depends on if it is more important for a company to catch all anomalies but have some false alarms or if the company rather does not catch all anomalies but it is sure there are no false alarms. Probably, more often the first scenario will be the case. If the second scenario is the case, it is good to start with a high σM and then decrease it using trial and error. σM can also be used to find the intensity of an anomaly. Usually, an anomaly with an anomaly score describes the likelihood of an anomaly, but if the anomalies are tested with higher σM and still show up, this means the intensity of the anomaly was higher as well. Because if there was a small error between predicted and true values, the anomaly score will drop very rapidly with an increased σM. But if the error was large, increasing σM will barely change the anomaly score and the anomaly will remain.

5. Conclusions

In this work, deep learning approaches were considered for anomaly detection in predictive maintenance, using a LSTM Autoencoder and a LSTM Predictor. The two algorithms were also combined with PCA for comparison purposes. The results showed that each model has its own strength. LSTM Autoencoders have a lower prediction error, but are more robust with respect to some parameters. Properly configured LSTM Predictor can achieve better performance in anomaly detection. The data used to evaluate the algorithms came from a metal press machine integrated in a real production environment. The obtained results showed the relevance of using machine learning approaches as PdM tools to early detect anomalies, so contributing to improve equipment availability and support optimized maintenance management.

Acknowledgements

This work has been carried out within the H2020 KYKLOS 4.0 project (Grant Agreement Number 872570). This project, funded by the European Commission, aims at developing an innovative Circular Manufacturing ecosystem based on novel Circular Production Systems and Artificial Intelligence technologies, targeting personalized products with extended life cycle and promoting energy efficient and low material consumption intra-factory production processes, resulting in reduced greenhouse gas emissions and air pollutants.

References

[1]. R. K. Mobley, Introduction to predictive maintenance, 2nd ed., Butterworth-Heinemann, Boston, 2002.

[2]. E. Lughofer and M. Sayed-Mouchaweh, Predictive Maintenance in Dynamic Systems: Advanced Methods, Decision Support Tools and Real-World Applications, 1st ed., *Springer International Publishing,* Cham. 2019.

[3]. Standardization Council Industry 4.0: The Standardization Roadmap of Predictive Maintenance for Sino-German Industry 4.0/ Intelligent Manufacturing, Version 2.0, *Federal Ministry of Economic Affairs and Energy,* 2019.

[4]. EFNMS: Body of Knowledge (BoK), 2019, Available online at: https://www.efnms.eu/about-us/bok/

[5]. Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, A Survey of Predictive Maintenance: Systems, Purposes and Approaches, December 2019. Accessed: Aug. 1 2021. [Online]. Available: https://arxiv.org/pdf/1912.07383.pdf

[6]. C. G. Machado, M. Winroth, D. Carlsson, P. Almström, V. Centerholt, and M. Hallin, Industry 4.0 readiness in manufacturing companies: challenges and enablers towards increased digitalization, *Procedia CIRP*, Vol. 81, 2019, pp. 1113–1118, doi: 10.1016/j.procir.2019.03.262.

[7]. P. Kamat and R. Sugandhi, Anomaly Detection for Predictive Maintenance in Industry 4.0 - A survey, in *Proceedings of the 6th Int. Conf. on Energy, City and Future (EVF'19)* , 2020, pp. 195-202, doi: 10.1051/c3sconf/202017002007.

[8]. H. Tercan and T. Meisen, Machine learning and deep learning based predictive quality in manufacturing: a systematic review, *Journal of Intelligent Manufacturing,* 33, 2022, pp. 1879–1905, doi: 10.1007/s10845-022-01963-8.

[9]. T. P. Carvalho, F. A. Soares, R. Vita, R.d.P. Francisco, J. P. Basto and S. G. Alcalá, A systematic literature review of machine learning methods applied to predictive maintenance, *Computers and Industrial Engineering*, 137, 2019, 106024, doi: 10.1016/j.cie.2019.106024.

[10]. J. Sharma, M.L. Mittal and G. Soni, Condition-based maintenance using machine learning and role of interpretability: a review, Int. J. Syst. Assur. Eng. Manag., 2022, 20: 1–16, doi: 10.1007/s13198-022-01843-7.

[11]. M. Sohaib, S. Mushtaq and J. Uddin, Deep Learning for Data-Driven Predictive Maintenance, Intelligent Systems Reference Library: Vision, *Sensing and Analytics: Integrative Approaches*, 2021, pp, 71–95, doi: 10.1007/978-3-030-75490-7_3.

[12]. N. Davari, B. Veloso, R. P. Ribeiro, P. M. Pereira and J. Gama, Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry, in *Proceedings of the 8th IEEE Int. Conf. on Data Science and Advanced Analytics (DSAA),* 2021, pp. 1-10, doi: 10.1109/DSAA53316.2021.9564181.

[13]. D. Neto, J. Henriques, P. Gil, C. Teixeira and A. Cardoso, A Deep Learning Approach for Data-Driven Predictive Maintenance of Rolling Bearings, In: Brito Palma, L., Neves-Silva, R., Gomes, L. (eds.) *CONTROLO 2022. Lecture Notes in Electrical Engineering*, Vol. 930, *Springer*, Cham., doi: 10.1007/978-3-031-10047-5_52

[14]. M. Klapper-Rybicka, N. N. Schraudolph, J. Schmidhuber, Unsupervised Learning in LSTM Recurrent Neural Networks, in *Proceedings of the International Conference on Artificial Neural Networks, ICANN 2001*, 2001, pp. 684-691.

[15]. R. Silipo, P. Winters, and I. Adae, Anomaly Detection in Predictive Maintenance: Anomaly Detection with Time Series Analysis, 2014. [Online]. Available: https://www-cdn.knime.com/sites/default/files/inline-images/Anomaly_Detection_Time_Series_final.pdf

[16]. I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. Cambridge, Massachusetts, London, England: *The MIT Press*, 2016.

[17]. Hsieh Ruei-Jie, Chou Jerry, and Ho Chih-Hsiang, Unsupervised Online Anomaly Detection on Multivariate Sensing Time Series Data for Smart Manufacturing, in *Proceedings of the 12th IEEE Conference on Service-Oriented Computing and Applications (SOCA)*, 2019, pp. 90–97.

(2252)

Fast Terminal Sliding Mode Control of the Active Suspension System

B. El Kassmi, A. Hmidani and E. M. Mellouli
University of Sidi Mohamed Ben Abdellah,
National School of Applied Sciences of Fez Lisa Laboratory, Fez, Morocco
Tel.: + 212 772-064611
E-mail: badre.elkassmi@usmba.ac.ma

Summary: To ensure the vehicle's stability and smoothness while moving even on a road with bumps, nowdays manufacturers use the active suspension system instead of the conventional passive system. This study focuses on the control, simulation and evaluation of the active suspension system using three different controllers based on the sliding mode control (SMC) strategy. A simplified quarter-car model of the active suspension system is employed to control the hydraulic force of the electo-valve to absorb the disturbances due to the irregularity of the road by delivering the control current calculated with the control law. The closed loop control stability is proven by Lyapunov stability theory. The convergence is observed and the performances of the studied controllers are compared in simulation results.

Keywords: Suspension system, Sliding-mode control, Vibration control, Feedback control, chattering.

1. Introduction

Active suspension system affects the handling and braking quality of cars. It also serves to protect the vehicle itself and any cargo or luggage from damage and wear. Searchers and engineers propose some control strategies of the active suspension system to improve the passengers comfort and the road-holding ability of the car [1, 2].

In this work, we study the active suspension system based on a hydraulic actuator, which is controlled with the use of hydraulic servomechanism that receives its input from the controller and delivers the hydraulic pressure to monitor body movement and vehicle ride level. The most important features expected to be improved is disturbance absorbing.

Significant and insightful conclusions for vehicle suspensions can be observed from a simple quarter car model with an undamped tire and a sinusoidal road disturbance.

2. Control of the Active Suspension System with FTSMC Method

The proposed system model is controlled by the sliding mode strategy with three different sliding surfaces: conventional SMC, terminal SMC and Fast terminal SMC. The control law for each surface is determined and the stability is proven with the Lyapunov stability theory [3-7].

The simulation is done with a MATLAB code, and the results are given in the simulation section.

2.1. System Description and Dynamic Model of the Active Suspension

The model used of the active suspension system is a simplified quarter-car model [1] shown in Fig. 1.

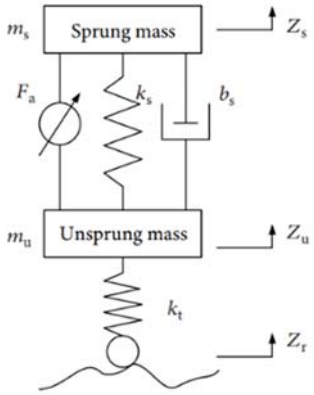


Fig. 1. Simplified model of a quarter-car active suspension system.

The dynamic characteristics between the command u and hydraulic force of valve F_a[1] with k_1, k_2, and k_3 are positive constants determined by the servo-valve:

$$\dot{F_a} = -k_1 F_a + k_2(\dot{z_s} - \dot{z_u}) + k_3 u \qquad (1)$$

2.2. State-space Form of the System

Consider the state variables $x_1, x_2; x_3, x_4$ *and* x_5 knowing z_s the vertical displacement of the sprung mass m_s, z_u the vertical displacement of the unsprang mass m_u, z_r the vertical road profile. k_t, k_s and b_s are the equivalent stiffness coefficient of the tire, the spring and the dumper, respectively:

$$x_1 = z_s, x_2 = \dot{z_s}, x_3 = z_u, x_4 = \dot{z_u}, x_5 = F_a \qquad (2)$$

According to Newton's second law ,(1) and (2), the system dynamics are rewritten as the following state-space form model:

$$\begin{aligned}
\dot{x}_1 &= x_2,\\
\dot{x}_2 &= -\frac{b_s}{m_s}x_2 + \frac{b_s}{m_s}x_4 - \frac{k_s}{m_s}x_1 + \frac{k_s}{m_s}x_3 + \frac{1}{m_s}x_5,\\
\dot{x}_3 &= x_4,\\
\dot{x}_4 &= -\frac{b_s}{m_u}x_4 + \frac{b_s}{m_u}x_2 - \frac{k_s}{m_u}x_3 + \frac{k_s}{m_u}x_1 - \frac{k_t}{m_u}x_3 + \frac{k_t}{m_u}z_r - \frac{1}{m_u}x_5,\\
\dot{x}_5 &= -k_1x_5 + k_2(x_2 - x_4) + k_3u,
\end{aligned} \tag{3}$$

The suspension displacement y is :

$$y = z_s - z_u = x_1 - x_3 \tag{4}$$

The tracking error is defined as:

$$e = y_d - y = y_d - x_1 + x_3 \tag{5}$$

The control objective of the active suspension system is to make the actual suspension displacement y fast and accurately track its set point y_d.

2.3. Sliding-Mode Control of the System

The sliding-mode surfaces are written as:

$$S = \ddot{e} + \alpha_1\dot{e} + \beta_1|\dot{e}|^{\varphi_1}sign(\dot{e}) + \alpha_2 e + \beta_2|e|^{\varphi_2}sign(e) \tag{6}$$

The surface parameters take different values depending on the sliding mode method used:

- SMC: $\alpha_1, \alpha_2 > 0$, $\beta_1 = \beta_2 = 0$, $\varphi_1 = \varphi_2 = 0$
- TSMC: $\alpha_1 = \alpha_2 = 0, \beta_1, \beta_2 > 0,\ 0 < \varphi_1 < 1$ and $1 < \varphi_2 < 2$
- FTSMC: α_1, α_2 , $\beta_1, \beta_2 > 0$, $0 < \varphi_1, \varphi_2 < 1$

The first appearance of u is in the 3rd derivative of the error and exactly in $\dddot{x}_5$:

$$\dddot{e} = \dddot{e}_x(x_1, x_2, x_3, x_4) - \left(\frac{1}{m_u} + \frac{1}{m_s}\right)\dddot{x}_5$$

with $\dddot{e}_x$ doesn't depend on x_5.

The sliding mode control signal, in which S_x doesn't depend on u is:

$$u = \frac{1}{\rho}\dot{S}_x + K.atan(S) \tag{7}$$

with $\rho = k_3\left(\frac{1}{m_u} + \frac{1}{m_s}\right)$ and $\dot{S}_x = \dot{S} + \rho u$

Using the Lyapunov function candidate $v = \frac{1}{2}S^2$ we demonstrate the stability of the closed loop system choosing K large enough:

$$\dot{v} = S\dot{S} = S(\dot{S}_x - \rho u + \frac{k_t}{m_u}\dot{z}_r) \tag{8}$$

Substituting the controller (7) into (8), we have:

$$\dot{v} = -K|S| + S\frac{k_t}{m_u}\dot{z}_r < 0 \qquad \text{with} \quad K \gg \frac{k_t}{m_u}$$

2. Simulation Results

We consider the desired output: $y_d = 0.2\sin(0.4t)$.

The disturbance signal is:

$$z_r(\mathrm{t}) = 0.06\sin(2t) + 0.03\sin(4t) + 0.02\sin(7t) \tag{9}$$

The vertical road profil z_r (disturbances) is shown in the Fig. 2. And in the same figure we plot the three output controllers. It has been observed that all the outputs converge and reach the desired trajectory.

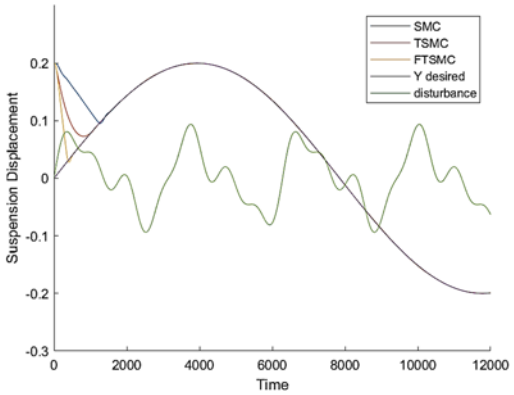


Fig. 2. Suspension displacement of the active control with different method SMC, TSMC and FTSMC.

It is observed that the FTSMC is faster than TSMC and faster than SMC.

We can notice in Fig. 3 that the problem of chattering in the control signal $u(t)$ is suppressed by using the atan(.) function.

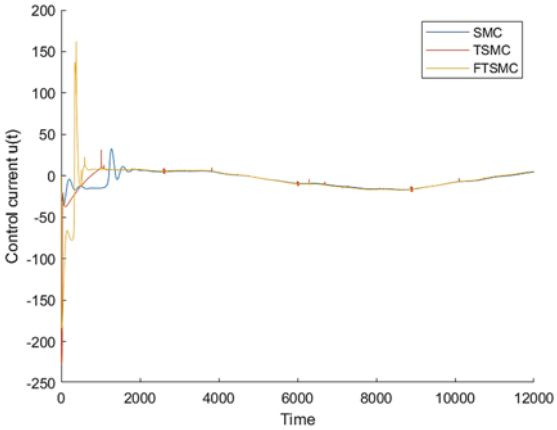


Fig. 3. The servo-valve control current *u(t)* in the hydraulic servo-system for the three controllers.

4. Conclusions

The hydraulic active suspension system has been controlled by the sliding mode control strategy. The simulation proves the convergence and stability of the system. The FTSMC has the best performances of time convergence and stability compared with the TSMC and SMC.

To improve this work, it is recommended to use optimization methods such as PSO or MPC to calculate the sliding surfaces gains and to use an observer to estimate accelerations.

References

[1]. Rui Bai, Dong Guo, Sliding-Mode Control of the Active Suspension System with the Dynamics of a Hydraulic Actuator, *Complexity,* Vol. 2018, Article ID 5907208.

[2]. W. Y. Wang, M. C. Chen, and S. F. Su, Hierarchical fuzzy neural control of anti-lock braking system and active suspension in a vehicle, *Automatica*, Vol. 48, No. 8, 2012, pp. 1698–1706.

[3]. YAO, Meibao, XIAO, Xueming, TIAN, Yang, et al., A fast terminal sliding mode control scheme with time-varying sliding mode surfaces, *Journal of the Franklin Institute,* Vol. 358, No 10, 2021, p. 5386-5407.

[4]. R. Naoual, E. M. Mellouli and I. Boumhidi, Adaptive fuzzy sliding mode control for the two-link robot, in *Proceedings of the 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14),* Rabat, Morocco, 2014, pp. 1-6.

[5]. R. Naoual, E. M. Mellouli, S. Sefriti, I Boumhidi, Fuzzy sliding mode control for the two-link robot, *International Journal of Systems, Control and Communications,* 6, 1, 2014, pp. 84-96.

[6]. E. M. Mellouli, M. Alfidi, I. Boumhidi, Fuzzy sliding mode control for three-tank system based on linear matrix inequality, *International Journal of Automation and Control,* 12, 2, 2018, pp. 237-250.

[7]. Mellouli, El Mehdi et al., Combined fuzzy logic and sliding mode approach's for modelling and control of the two link robot, in *Proceedings of the IEEE International Conference on Complex Systems (ICCS' 2012),* 2012, pp. 1-6.

(2320)

Development of Laser Cleaning Technology and its Potential for Process Automation

R. Kozera, B. Przybyszewski, M. Gloc, D. Kuczyńska-Zemła, J. Pura and H. Garbacz
Warsaw University of Technology, Faculty of Materials Science and Engineering,
ul. Woloska 141, 02-507 Warsaw, Poland
E-mail: Rafal.Kozera@pw.edu.pl

Summary: This study presents the potential of laser cleaning technology as an alternative for applicable in the presence in industry method for chemical and mechanical coating removal from painted surfaces used in aviation, automotive, etc. These methods are costly, time consuming, ergonomically unsafe, and produce large amounts of hazardous waste and emissions. Thus, laser technology can be an interesting solution with all their advantages like: 50% less processing time, 90% reduction in labour, low operating costs, minimum maintenance, applicable for both metal and composite structures and for all colours of paint and coating. In order to increase laser cleaning efficiency, process is dedicated for automation by combining laser device with arm of robot. In present study a SPI20W laser with wavelength of 1060 nm was applied in order to verify efficiency of laser cleaning process and potential for to remove a painting system from aluminum alloy sheets covered with a pure aluminium layer called CLAD. The results showed that the number of scans directly influences the efficiency of coating removal and surface structure. The efficiency of laser cleaning in aircraft applications was evaluated by microscopic observation of the surface. Investigation also confirms the high potential of laser cleaning technology for upscaling and process automation by means of a robot arm.

Keywords: Laser cleaning, Aviation coatings, Aluminium alloy, Fiber laser, Automation.

1. Introduction

In nowadays, many specialized maintenance and servicing companies remove paint and coatings from e.g. aircraft by using methods such as aggressive chemicals, abrasive sanding, or the newest methodology like plastic media blasting. These methods are costly, time consuming, ergonomically unsafe, and produce large amounts of hazardous waste and emissions. Composite structures in new aircrafts no longer allow the use of chemicals and the application of complex primers. Coating systems require more accurate and controlled techniques that allow for selective removal of layers of paint without damaging the underlying layers [1].

Hazardous waste are commonly generated from these activities are usually require high disposal costs under environmental regulations. Large amounts of wastes are associated with coatings removal include the disposal of liquid paint removers and contaminated rinse water from chemical stripping technique and media waste from a variety of blasting processes. Chemical paint removers are the only one process currently authorized for removing paint from metal bodies.

The body area of a medium-sized aircraft is approximately 1600 m2. From the point of view process automation, airplane be completely stripped of paint in about three days, which means the process speed should exceed 22m2 /h. When the task is performed, for example, by three laser units, each of them should be capable of removing more than 7m2 of paint in one hour. The planes are covered with a 100 µm thick layer of paint, which requires a pulse energy of approximately 90 J to accumulate until it is completely removed. For a cleaning speed of 8 m2 /h, an average laser power of 2 kW is required. For coatings with a thickness of the order of 100 µm, from an economic point of view, pulsed CO2 laser radiation is the best. In this case, approximately 15 pulses are required and the energy density to remove 100 µm of paint is approximately 6 J / cm2. Ideally, the pulse energy exceeds 6 J, making it possible to work with a beam spot greater than 1 cm2, which allows reducing the necessary accuracy of the handling system when machining larger surfaces. One of the main problems that can arise when removing paint from metallic surfaces is melting or damaging them. Consequently, the primary focus is on enabling greater process control (complete and / or selective paint stripping) and improving paint stripping efficiency. It can be noticed that it varies widely, depending on the laser parameters (laser wavelength, operating mode, pulse energy, pulse duration, average power) and the type of paint [2-4].

1.1. Methodology

Painting process

The coatings for the laser cleaning process were deposited using spraying technology, the same as in the case of industrial painting process. For coating, standard AkzoNobel polyurethan aviation paint was utilized.

Material Characterization

The surface observations before and after laser cleaning were analysed using a Zeiss Axio Observer light microscope.

2. Results

In Figs. 1-3 can be observed surface of samples after different number of laser beam repetitions. It is clearly visible that one time repetition give only partly removal of coating leaving stripes of material (Fig. 1a-b). After double repetition (Fig. 2a-b), no mentioned above phenomena were found, while after three repetitions parallel grooves are visible. This suggests texturization of aluminium substrate.

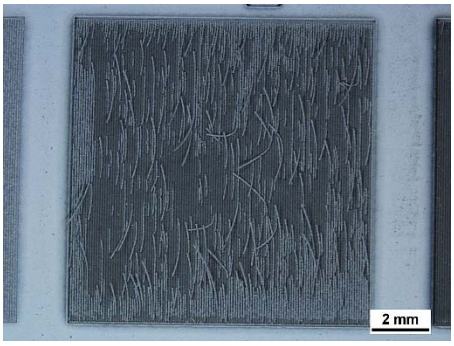


Fig. 1a. Surface of sample after one repetition.

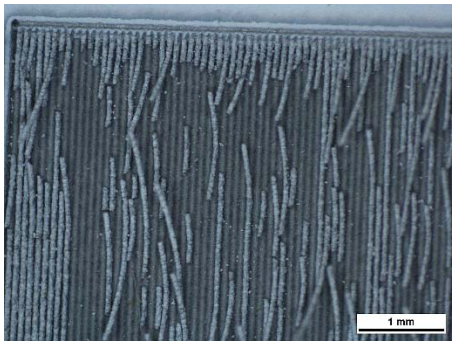


Fig. 1b. Surface of sample after one repetition.

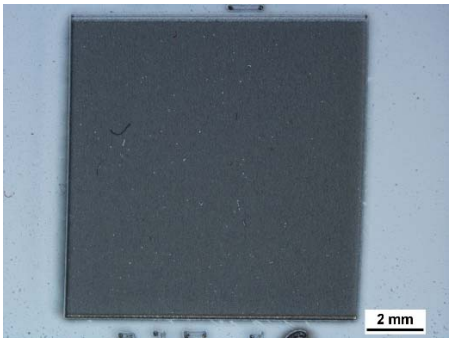


Fig. 2a. Surface of the sample after two repetitions.

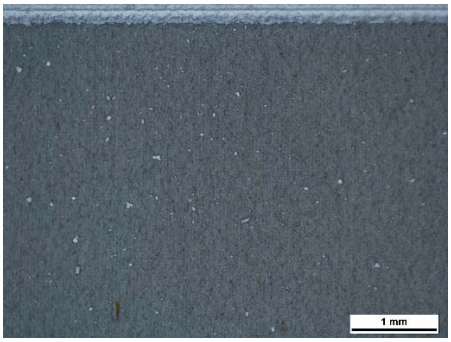


Fig. 2b. Surface of the sample after two repetitions.

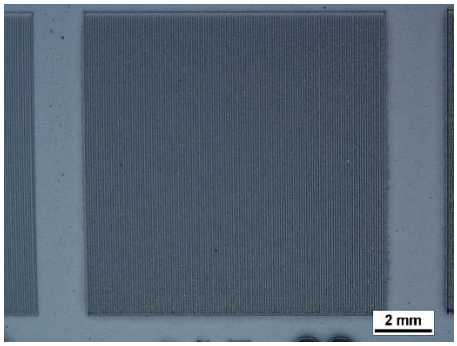


Fig. 3a. Surface of the sample after three repetitions.

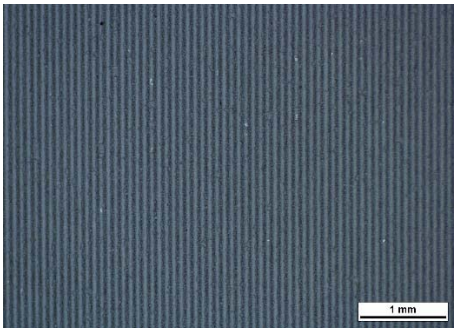


Fig. 3b. Surface of the sample after three repetitions.

3. Conclusions

The efficiency of laser cleaning in aircraft applications was evaluated. The results showed that the number of scans directly influences the efficiency of coating removal and surface structure. After one repetition, not complete removal was observed while after three texturization of substrate material was indicated. Good control and optimization of parameters suggests possibility of upscaling of the process and its automation.

Acknowledgements

This research was financially supported by the National Centre of Research and Development under Grant MAZOWSZE/0211/19

References

[1]. ESTCP Cost and Performance Report, Environmental Security Technology Certification Program (ESTCP), *U.S. Department of Defense*, August 2008.

[2]. Y. K. Madhukar, S. Mullick, D. K. Shukla, S. Kumar, A. K. Nath, Effect of laser operating mode in paint removal with a fiber laser, *Applied Surface Science,* 264, 2013, pp. 892–901.

[3]. M. Barletta, A. Gisario, V. Tagliaferri. Advance in paint stripping from aluminium substrate, *Journal of Materials Processing Technology* 173, 2006, pp. 232–239.

[4]. M. J. J. Schmidt, L. Li, J.T. Spencer, An investigation into the feasibility and characteristics of using a 2.5 kW high power diode laser for paint stripping, *Journal of Materials Processing Technology,* 138, 2003, pp. 109–115.

(2671)

Security Solutions for Industrial Radio Systems

A. Weinand [1], M. Karrenbauer [1] and H. D. Schotten [1,2]
[1] RPTU Kaiserslautern-Landau, Institute for Wireless Communication and Navigation,
Paul-Ehrlich-Straße 11, 67663 Kaiserslautern, Germany
[2] German Research Center for Artificial Intelligence (DFKI), Intelligent Networks Research Group (IN), Trippstadter Str. 122, 67663 Kaiserslautern, Germany
E-mail: {andreas.weinand, michael.karrenbauer}@rptu.de, hans_dieter.schotten@dfki.de

Abstract: Industrial Radio Systems (IRS) have rised high attention within the mobile network research and development communities within recent years. They differ from conventional Mobile Broadband (MBB) systems with respect to Quality of Service (QoS) requirements such as data rate, transmission reliability and latency. Further, security aspects have been put into a higher focus since the transmitted payloads in IRS are more sensitive and prone to cyber attacks. This is especially the case when services and physical processes are depending on the transmitted data (e.g. closed loop control). Therefore, we propose solutions in order to prevent cyber attacks within IRS by the introduction of novel security functionalities to the field. The most important security goal in the context of IRS is device authentication since further security goals depend on this (e.g. message integrity and confidentiality). A root of trust is necessary in order to enable that goal and has therefore to be included within the system security architecture. Further security functionalities such as device authentication, certificate issuing, validation and revocation, as well as session key management and secure message transmission schemes can be derived from that.

Keywords: Industrial radio, Machine type communication, Security, Authentication, Certificate authority, Plug&Trust.

1. Introduction

The concept of Industrial Radio (IR) can be understood as local wireless networks or campus networks which have a limited coverage and allow only access for predefined non-public users in contrast to conventional cellular public networks. Industrial Radio Systems (IRSs) have received wide attention during the upcoming of the fourth industrial revolution. Within IRSs the application scope is shifted towards applications within industrial environments (e.g. compared to consumer applications in cellular networks). The main difference between these two application areas are the existing QoS and security requirements due to different traffic types. While MBB traffic occurring from consumer applications is rather triggered by user interactions, Machine Type Communication (MTC) as in IRSs is triggered by physical machines or processes. Current solutions within the IT (information technology) area can therefore be only partly exploited or have to be completely redesigned for the application within industrial scenarios.

New solutions for IRSs have e.g. been proposed in [1, 2] where applications such as closed loop control processes were considered. MTC is typically divided into two main categories which are massive Machine Type Communication (mMTC) and Ultra Reliable Low Latency Communication URLLC. While mMTC includes applications such as smart metering, condition monitoring or GPS tracking, within URLLC applications such as factory and process automation or vehicle control (e.g. UAV, ground conveyor) are considered. The main advantage of the utilization of wireless transmission systems within such scenarios is their lower cost compared to wired solutions (e.g. no cables and physical interfaces required) and higher flexibility. The number of connected devices can e.g. be increased easily (no cable installation required). One of the drawbacks of the IR concept however is, that there are high requirements with respect to security. Compared to wire based systems, a risk of cyber attacks due to the open nature of the wireless channel exists. Network security has therefore to be considered during the design process of IRSs. Within this work, security solutions are presented in order to meet the given IR security requirements.

The remainder of this work is organized as follows. Within Section 2 possible cyber attacks within IRS are presented. Further, security requirements are derived and a security analysis of currently existing IR solutions is conducted in order to identify gaps towards the requirements. Security solutions for IRS are presented in Section 3 in order to overcome these gaps and discussed in Section 4. Further, an outlook to future work is given and Section 5 finally concludes the paper.

2. Industrial Radio Security Analysis

This section presents the potential cyber attacks on IRSs. Security requirements are then derived based on that and the state of the art IRSs security is analysed subsequently. Limitations of the current IRS solutions with respect to the requirements are outlined finally. However, only security functionalities for securing the wireless interface I_W (see Fig 1) of IRSs are considered here.

IR networks provide a substantial target for cyber attacks since sensitive information such as machine control data are transmitted over a wireless channel. Initially, every user within the coverage of the network

is able to interact with the transmitted information and consequently able to interact with the server applications as well due to the open nature of wireless channels. Dedicated security measures are therefore required in order to protect such sensitive information and guarantee the safe operation of the controlled machines and processes within IR applications. Further, intellectual property (e.g. machine configurations, production volume) may lead to attackers.

2.1. Potential Cyber Attacks within Industrial Radio Networks

Cyber attacks can be roughly divided into two categories. Attackers can either actively interfere with IRSs or passively. Possible active attacks are replay attacks, Man-in-the-Middle (MITM) attacks or spoofing attacks, which prohibit the secure transfer of data on wireless interface. Passive attacks such as eavesdropping are less harmful for the operational safety of IRS applications but can however lead to further attacks such as blackmailing.

Fig. 1 shows how an attacking node (Attacker) is able to interfere with an IRS. There are two possible ways for an attacker in order to tap into the IRS. He can either act as the IR gateway (IRGW) or as one or more of the IR devices (IRDs). In the first case, the attacker can gain device control by transmitting wrong control data to an IRD consuming data (e.g. actuator). He can then aim to generate physical damage to the respective IR application if no safety measures exist. Further, the attacker can collect feedback data of the IRDs and modify or leak it. Direct interactions with the IR application are possible if an attacker takes over the role of an IRD. He can then further transmit wrong feedback data (e.g. wrong sensor status) to the IRGW and backend in order to manipulate the IR application indirectly.

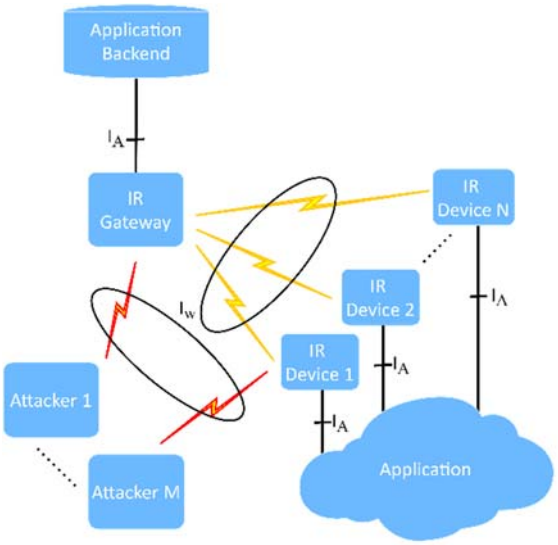


Fig 1. Cyber attacks within IRSs.

Further, the attacker is assumed to use advanced radio equipment such as directed antennas and high sensitivity receivers in order to maximize the communication range to his benefit (e.g. when physically located outside of a factory). It is additionally assumed, that he has full knowledge of the respective IR protocols and due to that able to interact with the IRS over I_W. Cyber attacks occurring by e.g. attacking IRDs or IRGW physically (hardware attacks) or by intrusion into the IRS remotely from the backend side of the network are out of scope of this work and therefore not considered. Security measures on the interface I_A can be additionally applied independent of the underlying network (IR or fixed network). They can however not prohibit wireless attacks on I_W.

2.2. Industrial Radio Security Requirements

Several security requirements need to be fulfilled in order to establish the secure execution of IR applications due to the aforementioned potential cyber-attacks. Beside the common security requirements of the CIA triad (confidentiality, integrity and availability), authenticity is additionally another important requirement within IRSs. There has to be an initial trust anchor in order to manage the access of devices and gateways within IRSs. Authentication of IRDs and IRGWs is therefore required. The transmission of user data has further to be secured by ensuring message integrity as well as message authenticity in order to prevent active cyber-attacks. Confidentiality is further required in order to prevent passive cyber attacks. Resulting from that, the secure exchange of cryptographic keying material is required as an initial step.

Further, IRSs should be flexible in such a sense, that it is easily possible to add new devices to an existing IR network or replace network devices without huge configuration efforts. This has an indirect influence on the security requirements, since changing the network configuration in such a way requires to (re)run user authentication routines. Further, the installation of new IR networks should be possible with a high degree of usability for non-experts (no manual configuration required by e.g. security engineers) at a low (monetary) cost. This requires automated routines for user authentication and revocation. Further, cost in terms of resource and energy efficiency should be considered as well and access rights should be granted independent of the dedicated IR technology. In summary, the basic requirements for IR network security, which are followed throughout this work are

- R1: Limitation of network access to desired user devices independent of IR technology
- R2: Insurance of confidentiality, integrity and authenticity of messages transmitted on the wireless interfaces of the system
- R3: Secure exchange of cryptographic material (e.g. session keys)

• R4: Insurance of life cycle security of network devices

• R5: Insurance of a high degree of usability and a low degree of maintainability simultaneously at low cost.

The focus of this work is to present solutions in order to meet the requirements R1, R4 and R5.

2.3 Security gap Analysis of IR Technologies

Many wireless solutions for the application in IRSs are existing currently. However, only a subset of the security requirements as introduced in Section 2.2 can typically be fulfilled by a single technology.

Beside, many IR technologies provide already a wide range of security features such as device authentication, key management and secure message transmission on the interface I_W. Known security vulnerabilities have been eliminated within new technology generations respectively. Some drawbacks within current IRS technologies are however still existing. Most state of the art technologies have e.g. simplified mechanisms for the management of user devices, such as passphrase or PIN based authentication schemes (e.g. WPA2-PSK within IEEE 802.11, Bluetooth) due to their primarily application area within consumer scenarios. When it comes to IRSs, systems are however higher scaled compared to consumer systems and additionally more prone to cyber attacks due to more sensitive data payloads. Some technologies already allow the application of device management schemes, e.g. IEEE 802.1X or LoRaWAN Join Server. The deployment of such architectural components is however left to the owner or operator of the network and therefore requires advanced knowledge and skilled personnel or third party support. This is in contrast to the IR security requirement R5 (high usability, low maintainability and low cost). Only 3GPP based systems provide a built-in device management and security architecture. In Section 3 more automated authentication and device management routines (Plug&Trust) including life-cycle security are therefore proposed for IRSs.

3. Security Enhancements for Industrial Radio Systems

The IRDs and IRGW need to set up trust initially towards each other in order to guarantee secure communication on the wireless interface I_W. This is enabled by Plug&Trust (P&T) routines and happens pairwise between the IRDs/IRGW and the network side (e.g. Security Backend). P&T protocols are automated setup routines for provision and validation of security credentials and can be used in order to limit the IRS access to predefined user groups. However, further security functionalities are required in order to enable these such as a Public Key Infrastructure (PKI), Secure Elements (SE) and an Authenticator device. The high level IRS security architecture is depicted in Fig. 2.

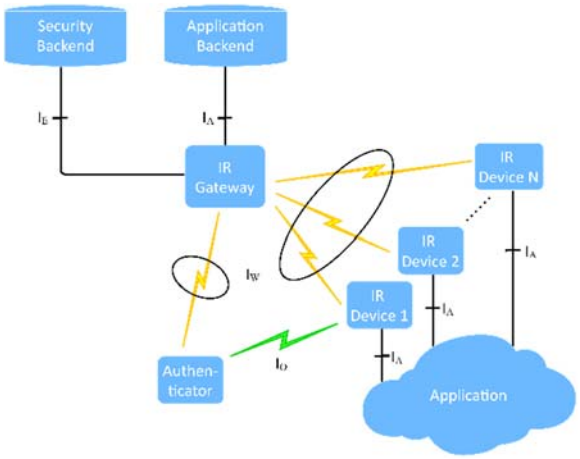


Fig. 2. IRS security architecture.

3.1. Prerequisites

There are some prerequisites required in order to enable P&T protocols. The IRDs need to be equipped with cryptographic credentials initially in order to verify and exchange P&T credentials respectively. Beside a PKI, which deals as a cryptographic baseline for subsequent security operations, SEs are required in order to store cryptographic materials such as keys and execute respective cryptographic operations. In order to cope with the special demands of resource constrained MTC devices which are typically insecure during the network setup phase, the concept of an authenticator device is further introduced.

3.1.1. Public Key Infrastructure

A PKI as depicted in Fig. 3 is required. Our proposal is a two level certificate authority (CA), where the root CA trust is enabled legally (e.g. contracting) between the stakeholders (S1 to SN) involved within the IRS value chain (e.g. chip and device manufacturers, network operators and users). Every stakeholder can then derive CAs individually from the root CA in order to issue own local certificates and validate global certificates. The advantage of such an approach is, that any stakeholder can then verify certificates issued by any other stakeholder within the IRS value chain and life cycle security (IR security requirement R4) can be achieved. External parties are however not able to issue or verify certificates respectively [3].

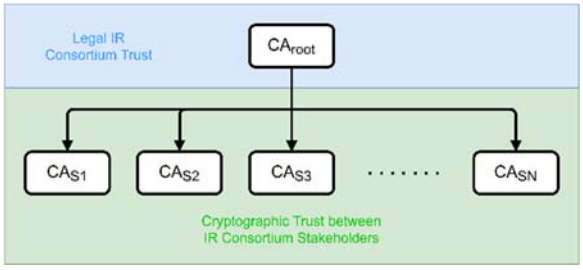


Fig. 3. PKI certificate environment.

3.1.2. Authenticator

The interface I_W cannot be used for the transfer of initial cryptographic credentials, since it is insecure during network setup. Therefore, the concept of an authenticator device is proposed, which has already a secure connection established towards the Security Backend, in order to load credentials into the IRDs and IRGW. The communication link with the device side is then realized using Out-of-Band (OOB) protocols such as Near Field Communication (NFC) on the interface I_O. Towards the network side (e.g. Security Backend) the interface I_E encapsulated within secured I_W is used. Further, resource constrained IRDs (e.g. in mMTC) can benefit from such an approach if computational expensive cryptographic operations have to be executed. The main drawback of that approach is the additional cost for the interface I_O on IRDs and IRGW.

3.1.3. Secure Elements

Though physical attacks on devices are not considered during the operational phase, cryptographic credentials should be stored in a secure and tamper proof way. This can be achieved by Secure Elements (SEs). Further, SEs such as Hardware Security Modules (HSMs) provide additional security functionalities for the execution of cryptographic operations such as generation and verification of cryptographic material (e.g. Elliptic Curve Digital Signature Algorithm (ECDSA), Elliptic Curve Diffie-Hellmann (ECDH)). Symmetric cryptographic functionalities such as AES cipher suite, cryptographic hashing functions such as SHA-256 and random number generator can further be provided. Compared to other solutions such as Trusted Platform Modules (TPMs), which aim avoid software related attacks, HSMs are in contrast to that higher in cost but provide additional protection against hardware based attacks (e.g. side channel attacks) [3].

3.2. Plug&Trust Protocols

After the PKI based credential provisioning, the IR devices can execute Plug&Trust routines in order to authenticate towards the Security Backend. The Plug&Trust routines are divided into two steps. In a first step, both I_W ends (IRD/IRGW and Security Backend) need to verify their trustworthiness mutually and exchange initial cryptographic credentials.

This has to be done only once for each time a network device joins the IRS and is described within Section 3.2.1. In a second step, which is executed every time a new communication session (e.g. MTC service) is started, both devices mutually authenticate each other and prove their access rights, e.g. based on handshake techniques. During this step, the devices are additionally generating common keying material which is used in order to secure the messages exchanged within the execution of a MTC service respectively. Further details of this step are introduced within section 3.2.2. Fig. 4 shows an overview of the respective procedures involved for the setup and execution of secure MTC services. Finally, considerations on the cipher operation mode within MTC are outlined.

Fig. 4. Plug&Trust routines overview.

3.2.1. Initial Plug&Trust Phase

Initially, both ends on the interface I_W (IRD or IRGW) need to verify their trustworthiness mutually and exchange initial cryptographic credentials. This step has to be executed for each time an IR network device joins the IRS. Fig. 5 shows the procedure of the initial connection setup within a link on the I_W interface.

The CAs as introduced in section 3.1.1 can be used for that purpose. This offers the possibility to identify devices by global certificates (e.g. specific for the manufacturer of the network equipment or of the SE) as well as local certificates (e.g. operator or user specific certificates). While the global certificates enable verification of the respective manufacturer of a network device over the whole device life cycle, local certificates enable the verification of the access rights of a network device to specific networks of an operator or user (e.g. within a factory). After setting up the physical connection, the IRDs and IR network side (Security Backend) will check their global certificates mutually in order to make sure that no third party devices are involved (which could e.g. be compromised due to product piracy). This step must be possible at any time for IRS network devices, even when it is not deployed within a network, in order to ensure lifecycle security. In case of a successful verification of the global certificates, the local certificates are checked which provide access rights, e.g. for a specific operator and factory. This means, that network devices which are equipped with these certificates could by default not be operated by a third party (e.g. an attacker) and vice versa (third party devices will not get access to the respective IR network).

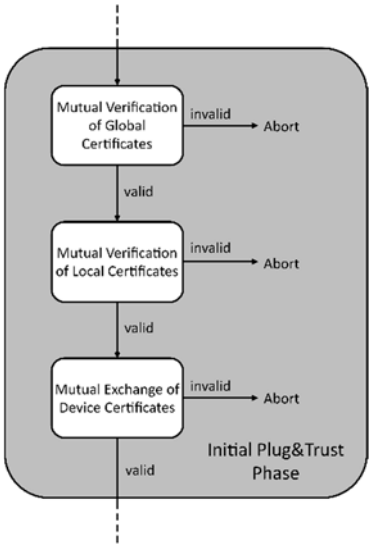


Fig. 5. Initial Plug&Trust Phase.

Both sides need to exchange their own device specific certificates in a final step, in order to make use of cryptographic techniques introduced within the next section. The authentication process is aborted, if in any steps the outcome of the certificate verification yields invalid results.

3.2.2. MTC Session Handshake

In a second step, which is executed every time a new communication session (e.g. MTC service) is started, both devices (IRD and IRGW) authenticate each other mutually and prove their access rights, e.g. based on handshake techniques (see Fig. 6). During this step, the devices are additionally generating common keying material which is used in order to secure the messages exchanged within the execution of a MTC service on I_W.

The IRDs and IRGW need to prove the possession of the private key matching to the public key which was exchanged within the device specific certificate in the previous step. A simple mutual challenge-response scheme including random nonce values can e.g. be used (such as within a TLS 1.3 handshake [4]). After a successful authentication at both sides, they can agree on initial keying material. Mutual authentication and negotiation of keying material can be achieved by making use of Public Key Cryptography (PKC) techniques such as signing/verifying and DHKE. If the mutual authentication step yields an invalid outcome on any side, the handshake is aborted respectively. Since PKC operations are resource intensive, the authenticator device can be used for the MTC session handshake as well. Symmetric keypairs used within a Secure MTC service need to be updated periodically in order to ensure a high degree of security such as e.g. forward secrecy. Conventional key management solutions used within the initial key exchange such as e.g. DHKE come with the cost of an additional control signalling overhead which rises with the number of connected devices.

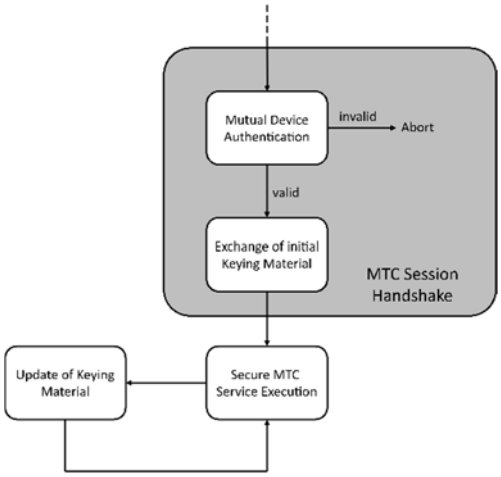


Fig. 6. MTC Session Handshake.

3.2.3. MTC Cipher Operation Mode

For confidentiality protection conventional cryptographic schemes such as the AES-128 cipher suite are used. However, the cipher operation mode has a high impact on the IRS efficiency and Key Performance Indicators (KPIs), since modes such as Cipher Block Chaining (CBC) can introduce delays in case of transmission errors. An appropriate mode for MTC traffic with high latency requirements is therefore the counter (CTR) mode where subsequent encryption blocks are independent.

4. Discussion & Outlook to Future Work

Within the previous Section, solutions in order to achieve the IRS security requirements R1, R4 and R5 were presented. Beside the insurances of the hard requirement R1, special focus was simultaneously put on the requirements R4 and R5 which are rather soft requirements (life cycle security, usability, maintainability). Latter are however very important within the IRS context since security solutions in existing state of the art technologies are often switched off due to high configurations efforts. Especially small and medium sized IRS users can therefore benefit from the presented solutions which do not require highly skilled personnel.

However, some drawbacks are still existing within conventional security solutions as outlined here. The secure transmission of MTC messages, as well as the update of keying material is lacking in resource efficiency due to transmission overheads and power consumption for cryptographic operations. Physical Layer Security (PLS) based techniques [5] can therefore be applied during the MTC service execution. The resource efficiency can be increased by that compared to the application of conventional security techniques for secure message transmission and keying material generation such as e.g. Message Authentication Codes (MACs) and Elliptic-Curve Diffie–Hellman (ECDH) key exchange. Such schemes exploit the wireless transmission channel and security

functionalities can be derived respectively. For secure message transmission in terms of message authenticity and integrity, Physical Layer Authentication (PLA) [6, 7] can be applied, whereas Secret Key Generation (SKG) [8, 9] can be applied as session key management scheme, e.g. within resource constrained IoT and mMTC devices.

5. Conclusions

Security solutions for IRS were introduced in this work with a focus on setting a root of trust. Beside many technological enablers for that purpose, such as PKI and SEs, legal trust between involved parties is required additionally. This can e.g. be achieved in terms of an ecosystem consisting of IR device manufacturers, network operators and users. Considering the value and volume of worldwide IR economic potentials, this seems however to be a promising approach in order to prevent many cyber attacks.

Acknowledgements

This work was funded in by the Federal Ministry of Education and Research of Germany within the projects "Industrial Radio Lab Germany" (under contract 16KIS1011) and "6G-Campus" (under contract 16KISK198).

References

[1]. C. Bockelmann, A. Dekorsy, A. Gnad, L. Rauchhaupt, A. Neumann, D. Block, U. Meier, J. Rust, S. Paul, F. Mackenthun, A. Weinand, H. D. Schotten, J. Siemons, T. Neugebauer and M. Ehlich, HiFlecs: Innovative Technologies for Low-Latency Wireless Closed- Loop Industrial Automation Systems, in *Proceedings of the Mobilkommunikation – Technologien und Anwendungen - 22. ITG-Fachtagung Conference*, 2017.

[2]. R. Kraemer, M. Methfessel, R. Kays, L. Underberg and A. C. Wolf, ParSec: A PSSS approach to industrial radio with very low and very flexible cycle timing, in *Proceedings of the 24th European Signal Processing Conference (EUSIPCO' 2016),* 2016, pp. 1222 - 1226.

[3]. N. Pohlmann, Cyber-Sicherheit, *Springer*, 2019.

[4]. IETF, The Transport Layer Security (TLS) Protocol Version 1.3, 2018.

[5]. E. Jorswieck, S. Tomasin and A. Sezgin, Broadcasting Into the Uncertainty: Authentication and Confidentiality by Physical-Layer Processing, *Proceedings of the IEEE*, Vol. 103, October 2015, pp. 1702-1724.

[6]. A. Weinand, R. Sattiraju, M. Karrenbauer and H. D. Schotten, Supervised Learning for Physical Layer Based Message Authentication in URLLC Scenarios, in *Proceedings of the 90th IEEE Vehicular Technology Conference (VTC2019-Fall),* 2019, pp. 1-7.

[7]. H. Forssell, R. Thobaben and J. Gross, Performance Analysis of Distributed SIMO Physical Layer Authentication, in *Proceedings of the IEEE International Conference on Communications (ICC' 2019),* 2019, pp. 1-6.

[8]. A. Ambekar and H. D. Schotten, Enhancing Channel Reciprocity for Effective Key Management in Wireless Ad-Hoc Networks, in *Proceedings of the IEEE Vehicular Technology Conference (VTC Spring),* 2014, pp. 1-5.

[9]. C. Lipps, S. B. Mallikarjun, M. Strufe, C. Heinz, C. Grimm and H. D. Schotten, Keep Private Networks Private: Secure Channel-PUFs, and Physical Layer Security by Linear Regression Enhanced Channel Profiles, in *Proceedings of the 3rd International Conference on Data Intelligence and Security (ICDIS' 2020),* South Padre Island, TX, USA, 2020, pp. 93-100.

(2683)

Design and Mechatronics of a Neurosurgical Robot for Tumor Ablation under MRI

Qingpeng Ding [1], Wanquan Yan [1], Jianghua Chen [1], Kim Yan [1], Chun Ping Lam [1] and Shing Shin Cheng [1,2]
[1] Dept. of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Sha Tin, HK
[2] Multi-Scale Medical Robotics Center, Hong Kong Science Park, N.T., HK
E-mail: qpding@mae.cuhk.edu.hk, sscheng@cuhk.edu.hk

Summary: Straight rigid instruments are still dominant tools used in most existing surgical cases in the human brain. The lack of distal dexterity and compatibility with the magnetic resonance imaging (MRI) limit the usage of these instruments in image-guided minimally invasive surgeries. Continuum robots built with MRI compatibility can potentially lead to wide adoption of surgical robots in intraoperative procedures. This paper introduces a 2-DoF 3D-printed continuum end effector and modular shape memory alloy (SMA) actuators for real-time tip motion in 3D workspace. The robot takes advantage of modular design to decouple linkage from the actuators to the end effector. This modular architecture facilitates establishment of a sterile interface and maintenance of the entire robotic system. Design of the robot and the control system are described, and the performance under MRI is evaluated experimentally.

Keywords: Medical robotics, Robotic surgery, Neurosurgery, Shape memory alloy, Magnetic resonance imaging.

1. Introduction

Brain tumours, including glioma and other metastatic tumours, are among the list of cancers with significant toughness to treat. Patients suffering from brain tumours have a 5-year survival rate of less than 35% [1]. In most cases, radiotherapy conducted in a stereotactic or whole brain manner remain the standard care of brain metastases alongside surgical resection. To improve the clinical outcome of treatment for brain tumor, various ablative techniques have been proposed to reduce patient suffering and shorten hospital time. Laser ablation stands out among the existing treatment approaches for its higher ablation efficiency and sharper boundary of the ablated zone.

Compared to traditional rigid surgical instruments, continuum robots have enhanced distal dexterity and can serve as a compliant interaction interface with human tissue [2]. Continuum robots are typically composed of an elastic backbone, including super elastic rod and elastomeric tube, and are essentially hyper redundant [3]. Many of the current designs utilize ferromagnetic materials, which are not compatible with MRI and lead to imaging artefacts [4]. Therefore, plastic or silicone alongside additive manufacturing are receiving more attention for building continuum robots and ensuring MRI compatibility.

Driven by the vision that a MRI compatible continuum robot will significantly enhance the surgical outcome of tumor removal with laser ablation, we develop a surgical robot proven to be compatible with the restrictive MR environment. The contributions of this work include: the development of a 2-DoF continuum robot, a modular SMA actuator with real-time performance and the mechatronics system of the robotic system.

2. Components and Design

2.1. SMA-Driven Steerable Robot

Fig. 1(a) shows the components of the SMA-driven steerable robot, namely the end effector module and the actuation module. The end effector module is composed of a cable-driven flexible tip, sensors, pulley mechanisms and parts for cable tensioning. The rotary motion from the SMA actuators is measured by the two MRI compatible optical encoders. To ensure a higher clinical application potential of the robot, the outer diameter of the end effector is designed to be less than 4 mm, and a tubular central lumen with a diameter of 2 mm runs through its center as a channel for the laser fiber. The end effector has a length of approximately 110 mm, which can accommodate the size of human brain and the distance of a long path from the burr hole to a deeply seated brain tumor.

The actuation module houses two modular MRI conditional SMA actuators used for 2-DoF actuation of the steerable tip. The working principle of the SMA actuator can be inspected in Fig. 1(b). The two antagonistic SMA springs are placed inside the shells and connected with each other via a cable. The cable is then routed through a pulley for rotary motion in both directions. The spring pre-displacement is critical for defining the performance of the actuator, thus an adjustment mechanism is placed on the proximal end. When one SMA spring is heated and recovers to its memorized shape, the opposite SMA spring gets pulled to generate rotary motion. The SMA material is thermally activated for motion output. Nichrome wires are routed and glued on the surface of the SMA springs for a high heating rate. A stream of cold air is guided to the housing chamber to cool down the deactivated SMA spring and improve actuation speed of the

actuator. Two independent chambers are designed for each spring to lower the possible thermal interference.

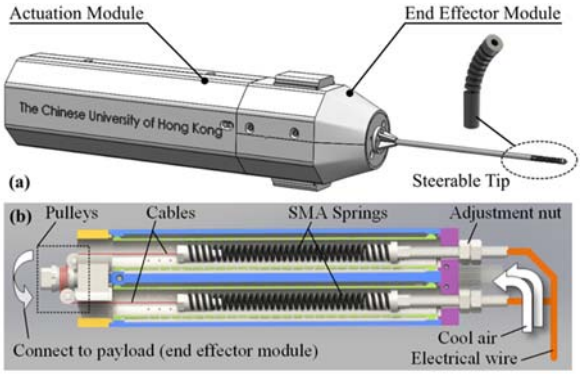


Fig. 1. (a) Components of the neurosurgical robot, and (b) cross-section view of the SMA actuator.

2.2. Control System

The controller box, as shown in Fig. 2, is capable of signal processing, sensor data acquisition and closed loop control of the neurosurgical robot. The control program is coded in Simulink and runs on a desktop PC in real-time to perform control tasks. An I/O interface board (Sensoray 826) bridges the data between the host PC and variable sensors equipped on the robot. The PWM signals from the control board are sent to an H-bridge driver and control the heating current to the SMA actuators. The cold air streams are controlled by pneumatic valves, which are driven by a switch circuit with on/off signals.

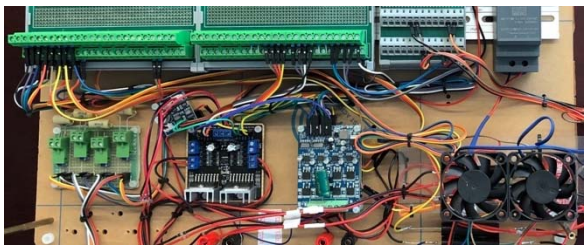


Fig. 2. I/O, sensor interface and amplifier of the controller.

A PI controller with model-based feedforward compensation was implemented to improve the tracking capability of the robot [5]. In trajectory tracking tests, model predictive control (MPC) was also implemented to compensate the tracking error induced by inaccuracy in kinematic modeling. The input signal of the MPC controller is taken from the reference signal X^r and the optimized trajectory is generated with $X^{rN} = X^r + \Delta X^r$ by minimizing the following cost function L

$$\sum_{i=1}^{N} Q_{k+i}(X_{k+i}^r - x_{k+i})^2 + \sum_{i=1}^{N} P_{k+i}{\Delta X_{k+i}^r}^2, \quad (1)$$

where ΔX^r is the derivative of the reference, $N = 15$ is the length of the prediction horizon, the subscripts k and i indicate time instants, x_{k+i} is the predicted tip position computed with calibration results of forward kinematics, P_{k+i} is a weighting factor relative to ΔX^r, and Q_{k+i} is a normalized weighting factor.

3. Results

The robotic system is set up in the MRI room to evaluate its safety and performance for delivering laser energy under image guidance. The controller and host PC are placed in the control room, while the SMA-driven robot is placed on the patient bed near the head coil inside the scanner room. The steerable end effector is inserted into an agar phantom, which mimics the human brain under surgery. The robot is controlled to follow a step signal and steers from one direction to another, when the MR machine conducts a dynamic scan and visualizes the tip bending motion. The sensor signals from the rotary encoders are converted to tip bending angle and shown as Fig. 3.

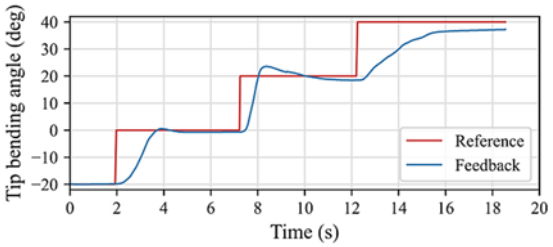


Fig. 3. Bending angle of the steerable end effector under semi-closed loop control in the MRI environment.

4. Conclusions

This paper presents a MRI-compatible robot driven by SMA actuators for neurosurgery. The design of the robot and the mechatronics architecture of the control system are introduced. The SMA actuators are proven safe to drive the steerable end effector in MRI scanner. Performance of the robotic system is also evaluated.

References

[1]. S. Lapointe, A. Perry, and N. A. Butowski. Primary brain tumours in adults, *The Lancet*, Vol. 392, Issue 10145, Aug. 2018, pp. 432–446.

[2]. J. Burgner-Kahrs, D. C. Rucker, and H. Choset. Continuum Robots for Medical Applications: A Survey, *IEEE Transactions on Robotics*, Vol. 31, Issue. 6, Dec. 2015, pp. 1261–1280.

[3]. F. Jelínek et al., Classification of Joints Used in Steerable Instruments for Minimally Invasive Surgery-A Review of the State of the Art, *Journal of Medical Devices*, Vol. 9, Issue. 1, Mar. 2015, pp. 010801.

[4]. Q. Ding, et al., A High-Performance Modular SMA Actuator with Fast Heating and Active Cooling for Medical Robotics, *IEEE/ASME Transactions on Mechatronics*, Vol. 27, Issue. 6, Dec. 2022, pp. 5902-5913.

[5]. S. Shao et al., Design, Modeling, and Control of a Compact SMA-Actuated MR-Conditional Steerable Neurosurgical Robot, *IEEE Robotics and Automation Letters*, Vol. 5, Issue 2, Apr. 2020, pp. 1381–1388.

(2880)

Robotic Bioprinting: a 7 Degree of Freedom Robotic Arm towards Volumetric Printing for Tissue Engineering

V. C. David, M. Kalogeropoulou, P. Fucile and L. Moroni
MERLN Institute for Technology Inspired Regenerative Medicine, Complex Tissue Regeneration department, Maastricht University, Maastricht, the Netherlands
E-mail: v.david@maastrichtuniversity.nl

Summary: Recent advances in technology have shown the limitations of the typical fabrication approaches for scaffolds for Tissue Engineering (TE). In order to better mimic the architecture of native tissue towards their regeneration, new technologies and approaches are needed. Several attempts of integration between Additive Manufacturing (AM) and robotics have been made in large scale manufacturing. This study aims to show initial results on the development of a novel AM apparatus based on a 7 Degree Of Freedom (DOF) robotic arm, which will allow for volumetric bioprinting. A precise control on the extrusion and the robot motion is achieved through the integration of custom-made codes, both for robotic control in Robotic Operating System (ROS) environment and for AM implementation. Also, custom-made slicing procedures, which are crucial to achieve such printing "in air", were implemented. First results of successful non-planar and volumetric printing will be presented.

Keywords: Additive manufacturing, Robotic bioprinting, Tissue engineering, Scaffolds.

1. Introduction

In the field of Tissue Engineering (TE), a typical approach consists in the fabrication of porous 3-Dimensional scaffolds that enhance the regeneration of a damaged tissue. Scaffolds are fabricated through Additive Manufacturing (AM) methodologies, in a layer-by-layer fashion of fabrication. They must mimic the structural and chemical composition of the native tissues, thus to promote biologic activity and regeneration process.

However, recent technological advances show limitations in such approaches, especially for extrusion-based systems, when complex architectures need to be reproduced (i.e., anisotropic structures and specific fibers orientation).

A shift from the typical x-y layer-by-layer based AM is needed in order to achieve devices that more closely mimic the native tissue architecture. Applications in the industrial and construction field (i.e., large scale manufacturing) have already implemented systems with a higher number of degrees of freedom (DOF) [1-3], which have allowed for the fabrication of complex structures in a "less planar" fashion.

Implementation of robotic arms in the AM chain is promising, though full control on the process for small scale manufacturing is not achieved yet. First attempts have been performed with soft materials and simpler architectures [4, 5]. Furthermore, the extrusion of typical TE medical-grade biomaterials towards a full volumetric printing remains very challenging.

This study aims to show initial results on the development of a new extrusion-based AM system based on a 7 DOF robotic arm towards the fabrication of scaffolds for TE.

2. Materials and Methods

A novel custom-made "robotic bioprinting" system was developed integrating a traditional custom-made Fused Deposition Modelling (FDM) extruder with a 7 DOF robotic arm (Sawyer robotic arm, Rethink Robotics), as shown in Fig. 1.

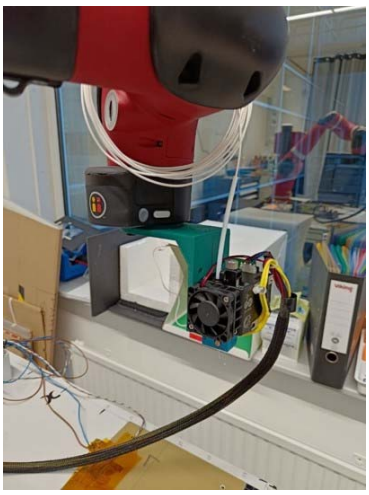


Fig. 1. The robotic bioprinting apparatus based on a 7 DOF robotic arm.

The additional joint and thus DOF can significantly extend the capabilities of the AM process by allowing changes of direction of the material deposition during the fabrication process, as well as the deposition of complex nonplanar layers. However, the advantages in terms of final results come with more complexity from a problem-solution point of view, since the Inverse Kinematics (IK) resolution for such a system is more complex and requires high computational resources.

To fully exploit the potential offered by a 7 DOF robotic arm an ad hoc slicing platform has been developed, which allows for the generation of custom-made gcode for non-planar and volumetric printing.

In particular, the Grasshopper plugin within the environment of the Computer-Aided Design (CAD) software Rhinoceros was used to develop such software, starting from the 3D model of the desired object. The output of this process (i.e., the gcode) was used to extract the cartesian coordinates of every waypoint, which were fed to the robot in order to achieve a precise control on motion and trajectory generation.

The integration between robot motion and material extrusion was achieved through the development of a custom code based on the open-source linux-based Robotic Operating System (ROS), which allowed to send synchronous signals to the extruder (via serial port control) and the robot. The motion planning process and accuracy of motion were optimized by using different open source planning algorithms and changing some relevant parameters, such as joints acceleration and motion speed.

First results were achieved through the extrusion of a Polylactic Acid (PLA) filament 200° C through a 0.4 mm nozzle attached to the robot. The temperature was set to 200° C. Many processing parameters, such as printing speed, joints acceleration, and printing flow were varied through the robot control code according to the architectural features and requirements.

3. Results and Discussion

The successful integration between robot manipulation and extrusion control allowed for the fabrication of nonplanar paths (Fig. 2), paving the way to even more intricate scaffolds towards a better mimicking of the native tissue architectures.

Fig. 2. Non-planar trajectories achieved with the robotic bioprinting apparatus.

Custom-made gcode were successfully generated and tested, and resulted in an improvement in terms of parameters control and optimized trajectory generation if compared to commercial slicer software.

First attempts of xyz based printing (i.e., "printing in air") were performed starting from such codes. These results are promising towards *in vivo* bioprinting and regenerative medicine (RM) in general.

4. Conclusions

The developed new AM robotic system is promising towards biofabrication applications. It allows for a more precise control on the orientation of the extruder during motion, thus varying the extruding direction and acting on the fibers alignment. This system will overcome the limitations of the current layer by layer systems used in RM. It will allow for the fabrication of TE scaffolds that better mimic native tissue and for the extrusion of complex structures (i.e., specific fibers alignment). This could help improving the regenerative capacity of the scaffold itself.

Further optimizations will allow for smoother movements in air and implementation of collision avoidance algorithms, which will be crucial for advanced manufacturing, especially in situ. In conclusion, this new platform opens up to huge opportunities in small scale manufacturing, especially for TE.

Acknowledgements

This publication is part of the project 3D-MENTOR (with project number 18647) of the VICI research programme, which is financed by the Dutch Research Council (NWO).

References

[1]. L. D. Evjemo, G. Langelandsvik, J. T. Gravdahl. Wire arc additive manufacturing by robot manipulator: Towards creating complex geometries, *IFAC-PapersOnLine,* Vol. 52, Issue 11, 2019, pp. 103-109.

[2]. L. D. Evjemo, S. Moe, J. T. Gravdahl. Robotised wire arc additive manufacturing using set-based control: experimental results, *IFAC-PapersOnLine,* Vol. 53, Issue 2, 2020, pp. 10044-10051.

[3]. L. D. Evjemo, S. Moe, J. T. Gravdahl, O. Roulet-Dubonnet, L. T. Gellein, Additive manufacturing by robot manipulator: An overview of the state-of-the-art and proof-of-concept results, in *Proceedings of the 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA' 2017)*, 2017, pp. 1-8.

[4]. X. Li, Q. Lian, D. Li, H. Xin, S. Jia. Development of a robotic arm based hydrogel additive manufacturing system for in-situ printing, *Applied Sciences,* Vol. 7, Issue 1, 2017, pp. 73.

[5]. Z. Zhang, C. Wu, C. Dai, Q. Shi, G. Fang, D. Xie, X. Zhao, Y.-J. Liu, C. C. Wang, X.-J. Wang, A multi-axis robot-based bioprinting system supporting natural cell function preservation and cardiac tissue fabrication, *Bioactive Materials,* Vol. 18, 2022, pp. 138-150.

(2922)

Vehicle Localization by Optimally Weighted Use of MEMS Sensor Data

Takayoshi Yokota [1] and Taiga Yamagiwa [2]

[1] Faculty of Business and Informatics, Tsukuba Gakuin University, 3-1 Azuma, Tsukuba, Ibaraki, Japan
[2] Graduate School of Sustainability Science, Tottori University, 4-101 Koyamacho-Minami, Tottori, Japan
Tel.: +81 298584811
Email: yokota@tsukuba-g.ac.jp

Summary: The Global Navigation Satellite System (GNSS) has been widely used as a method to accurately determine the position of a vehicle, but at present, the accuracy may be degraded depending on the radio wave reception conditions from the satellite. There have also been growing concerns about cyber-attacks on GNSS. We have been developing a method for estimating the position of a traveling vehicle using MEMS sensor data derived from air pressure, acceleration, gyro, and geomagnetic sensors. In the current work, we set up an experimental environment to collect data at 50 Hz and implemented a running test on an actual road. By optimizing the weighting of each item of sensor information, we were able to capture characteristics of roads that could not be captured in the past. Our findings confirmed that the position estimation accuracy was improved on the order of one meter.

Keywords: Localization, Atmospheric pressure, Acceleration, Geomagnetic, MEMS sensor, Optimum weight.

1. Introduction

The Global Navigation Satellite System (GNSS) is widely used, but its accuracy may be degraded depending on the reception status of the radio waves from the satellite. Urban areas with high-rise buildings have long been a problem in this regard, but in recent years, although rare, radio interference caused by solar flares and radio interference caused by intentional jamming have occurred in some areas. This has created the need for a new technology that can compensate for the inherent disadvantages of the vehicle position estimation technology, which is highly dependent on GNSS and requires huge maintenance costs. In this study, we propose a localization algorithm for a traveling vehicle based on the characteristics of the road obtained from MEMS sensors including acceleration, angular velocity, and magnetization for x, y, and z axes relying only minimally on GNSS.

2. Vehicle Localization Algorithm by MEMS Sensor

Terrain-based localization is an interesting research field that has emerged in recent years [1-12]. We have been working on a vehicle localization algorithm based on MEMS sensors for the past several years [1-5]. In our algorithm, accurate location information obtained by high-precision GNSS including RTK-GNSS (u-blox ZED-F9P) as well as acceleration, angular velocity, and geomagnetism information (Invensense 9250) are collected in advance on an actual road for use as reference data. A test vehicle for evaluating the performance of the localization algorithm is then run to acquire sensor data, which are compared with the above reference data to obtain an accurate location. The outline of the proposed algorithm is shown in Fig. 1. Pattern matching is performed on the basis of the normalized mutual correlation function defined as

$$(i,j) = \frac{\sum_{k=0}^{N-1}\left(f_{i+j+k}-\bar{f}_{i+j}\right)\left(g_{i+k}-\bar{g}_i\right)}{\sqrt{\sum_{k=0}^{N-1}\left(f_{i+j+k}-\bar{f}_{i+j}\right)^2}\sqrt{\sum_{k=0}^{N-1}\left(g_{i+k}-\bar{g}_i\right)^2}}, \quad (1)$$

$$\bar{f}_{i+j} = \frac{1}{N}\sum_{k=0}^{N-1} f_{i+j+k}, \quad (2)$$

$$\bar{g}_i = \frac{1}{N}\sum_{k=0}^{N-1} g_{i+k}, \quad (3)$$

where i is the index of evaluation for the vehicle's travel time and j is the index of the time lag between two vehicles (reference vehicle and evaluation vehicle). If we can determine this time lag accurately, the location of the evaluation vehicle can be substituted by the location (latitude and longitude) of the reference vehicle at the matched time. The pattern matching can then be achieved by finding the maximum value of similarity of the sensor data with a mutual correlation function calculated over a specific time window (e.g., one second).

3. Experiments

3.1. Experimental Setup

We evaluated the proposed algorithm on a flat road section (Fig. 2) with very little change in the road attributes, where the localization error had been as large as 100 m or more when we tested an earlier version of our method [1, 2]. Moreover, even when we

had used an accelerometer and gyro sensor in addition to the barometric sensor, there was no improvement of the position accuracy at the processing interval of 0.1 s. This was due to the influence of aliasing noise caused by the sampling interval being too large.

Therefore, our aim here was to reduce the influence of aliasing noise by performing normalized correlation at 0.02 s (50 Hz). The experimental setup is summarized in Table 1. The MEMS sensors (Invensense Mpu9250) and Bosch BME280s were installed in the rear seat of the vehicle as shown in Fig. 3. We processed the resultant data with the resolution of 50 Hz for the Mpu9250 data. The atmospheric sensor data was used only to identify the flatness of the test course. Fig. 4 shows the altitude along the test course.

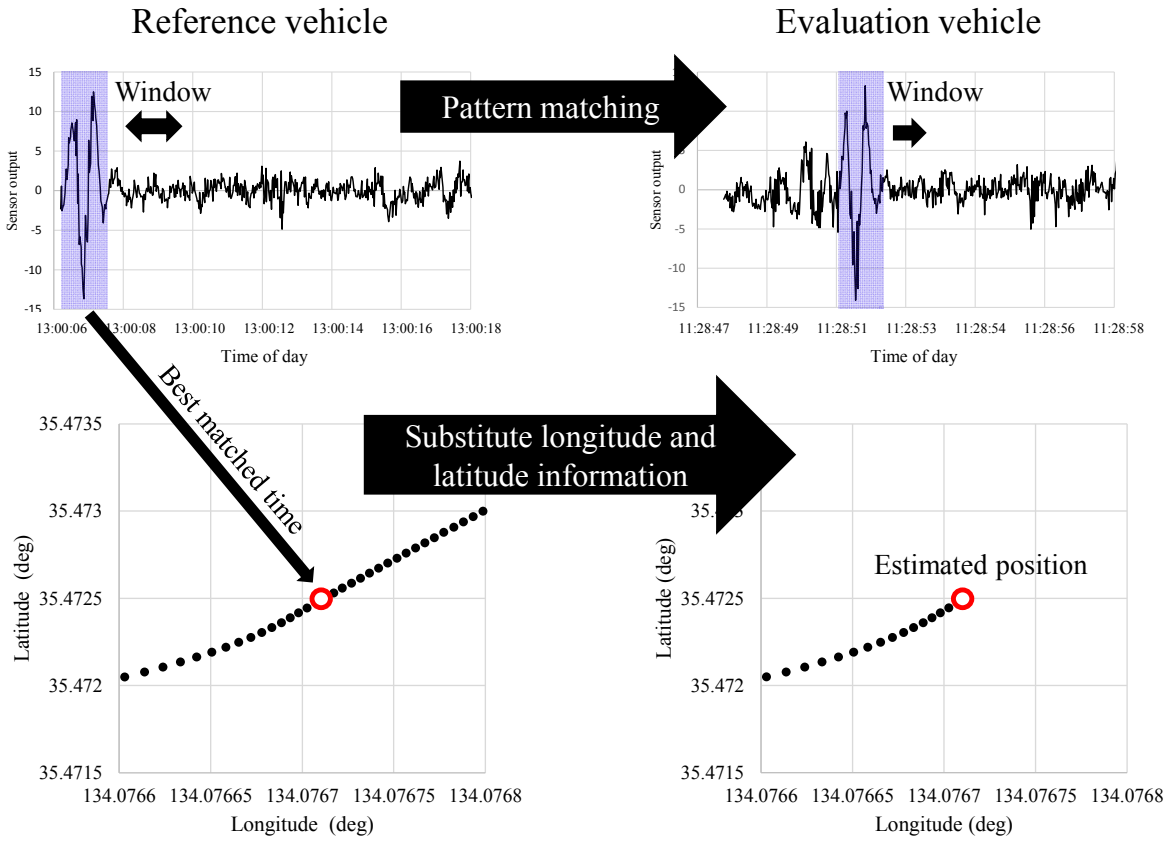


Fig. 1. Outline of proposed algorithm. The reference vehicle stores sensor data together with RTK-GNSS position data for later reference.

Fig. 2. Flat and straight 1.36-km test course along a Grade 2 river.

Table 1. Experimental setup.

Date	Dec. 5 (Sun), 2021, Mar. 6 (Sun), 2022
Vehicle	Daihatsu Mira Geno (Small car)
MEMS sensor	1. Invensense Mpu-9250, 50-Hz data acquisition, acceleration, angular velocity, geomagnetism 2. Bosche BME280 20-Hz data acquisition atmospheric pressure
RTK-GNSS	u-blox ZED-F9P,5Hz data acquisition
Test course	1.36 km, Shikano-cho, Tottori-shi, Japan

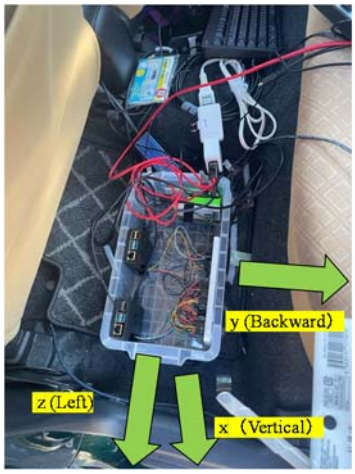


Fig. 3. Sensor axes of Mpu-9250 are indicated in arrows. Three Raspberry Pi 4Bs are used for data acquisition: one for an Mpu-9250 and two for 16 modules of the BME280s. 16 outputs of BME280s are averaged to reduce random noise.

3.2. Pattern Matching by Normalized Mutual Correlation Function

Nine normalized mutual correlation functions were obtained for each sensor data (acceleration, angular velocity, and magnetization for x, y, and z axes), as shown in Figs. 4, 5, and 6. Each sensor data of the evaluation vehicle on Mar. 6, 2022 was correlated with the reference vehicle data of Dec.5, 2021.

3.2.1. Normalized Mutual Correlation of Acceleration Sensor Data

Fig. 4 shows the results of the normalized mutual correlation of the acceleration data. Among them, a time lag profile can be seen slightly for ax, (acceleration for vertical direction).

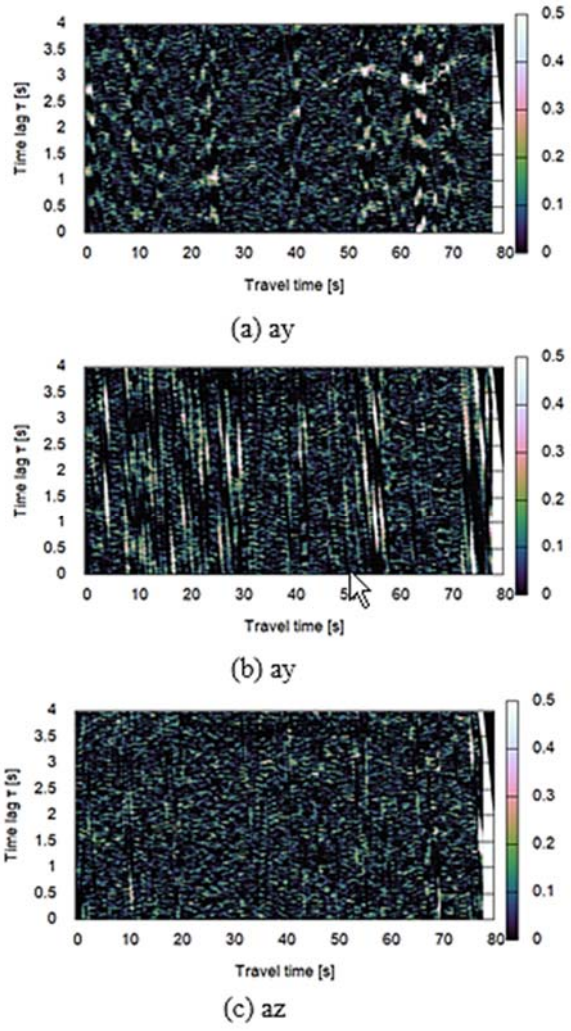


Fig. 4. Normalized mutual correlation of acceleration sensor data. Acceleration data toward x ,y, and z axes are denoted as axial and az, respectively.

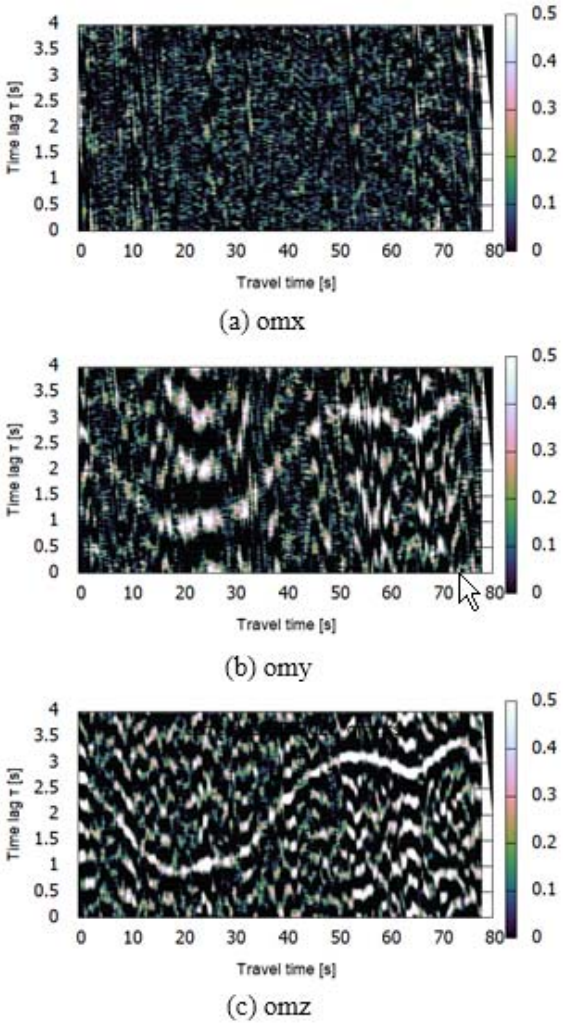


Fig. 5. Normalized mutual correlation of gyroscope data. Gyroscope data around x ,y, and z axes are denoted as omx,omy and omz, respectively.

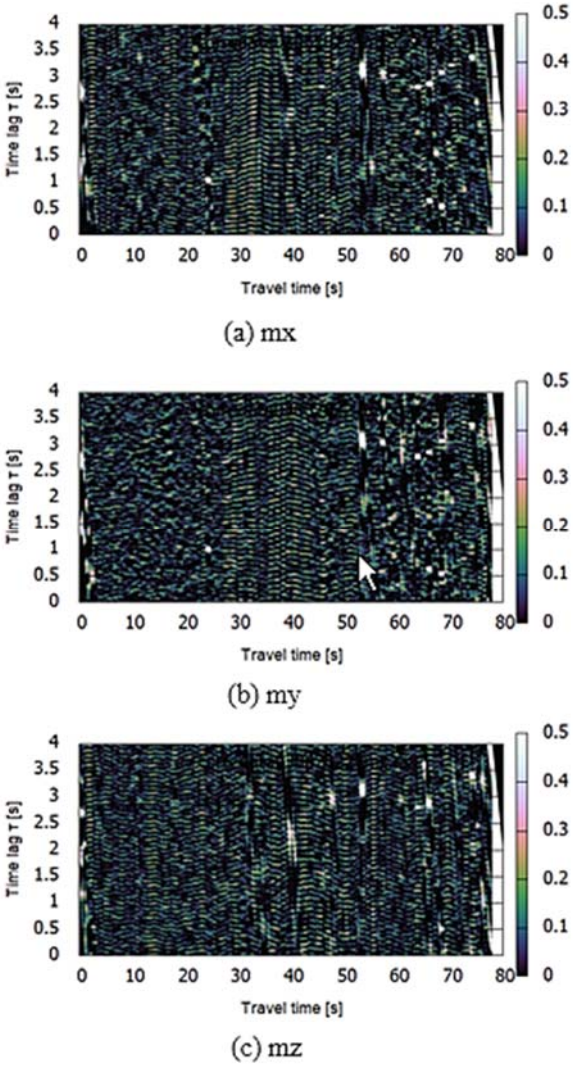


Fig. 6. Normalized mutual correlation of geomagnetism data. Geomagnetism data along x, y, and z axes are denoted mx,my and mz, respectively.

3.2.2. Normalized Mutual Correlation of a Gyroscope Data

Fig. 5 shows the results of the normalized mutual correlation of the gyroscope data. Among them, omz clearly shows a time lag profile. Omy also shows a time lag profile, although its resolution is lower. Gyroscope data around the z axis represents pitch rate and the data around the y axis represents roll rate.

Yaw rate, that is, the rotation rate around the x axis, does not effectively captures the feature of roads and the time lag profile cannot be seen. This is because the test road is an almost completely straight road and thus very little change exists in yaw rate.

3.2.3. Normalized Mutual Correlation of Geomagnetism Data

Fig. 6 shows the results of the normalized mutual correlation of the geomagnetism data. Some portions of the time lag profile are visible here, especially for the latter part from 40 to 80 s of the evaluation vehicle travel time. The resolution is rather high.

3.3. Optimum Weight of Normalized Mutual Correlation Functions

In the previous sections we were able to obtain nine normalized mutual correlation functions, some of which clearly show the time lag profile and others do not. In this section, we explain how to find an optimally weighted average for them to form an optimum normalized mutual correlation function (Eq. (4)).

The optimum weights were calculated by a simulated annealing method so that the measure function F (defined in Eq. (7)) was maximized. F measures the signal-to-noise ratio (SNR) of the optimum normalized mutual correlation function of Eq. (4).

$$R_{opt}(i,j) = \sum_{n=0}^{M-1} w_n R_n(i,j), \tag{4}$$

$$w_n \geq 0, n = 0,1,\dots,8 \tag{5}$$

$$\sum_{n=0}^{8} w_n = 1 \tag{6}$$

$$F = \frac{1}{M}\sum_{i=0}^{M-1} \frac{max_j R_{opt}(i,j)}{\left|min_j R_{opt}(i,j)\right|} \tag{7}$$

Fig. 7 shows the optimum weights we derived. As we can see, the pitch rate, vertical magnetism, roll rate, and backward magnetism play significant roles (74.7 %) in vehicle localization in the proposed algorithm. Fig. 8 shows the optimally weighted mutual correlation function, where the time lag profile is clearly recognizable. Fig. 9 show the estimation results of the time lag between the two vehicles. The ground truth of the time lag was calculated by comparing the RTK-GNSS trace of each vehicle. RTK-GNSS was operating at 5 Hz and the traces were over-sampled at 50 Hz by introducing a linear interpolation of 5-Hz position data. It is reasonable that a vehicles velocity cannot be changed so rapidly within 0.2 seconds. In estimating the optimum time lag, we introduced a particle filter so as to reduce the influence from the noise that still exists in the optimum averaged mutual correlation function (Fig. 8). As we can see in Fig. 9, the RMS error of the time lag estimation was 0.091 seconds, which is equivalent to about 1.5 m in position error (assumed vehicle speed was 16.7 m/s) and is far more precise than the results with our earlier method [1, 2]. At the travel time of 33 s, the time lag estimate reached the maximum error of about 0.3 s, which is almost equivalent to 5 m in position error (again, vehicle speed was assumed to be 16.7 m/s)

Next, we estimated the position error precisely by the proposed algorithm following the data flow in Fig. 10. The result is shown in Fig. 11. The rough estimate from the time lag estimate result of Fig. 9 is rather overestimated assuming the vehicle speed is 60 km/h. The details of the position error results along the evaluation travel time are provided in Fig. 11.

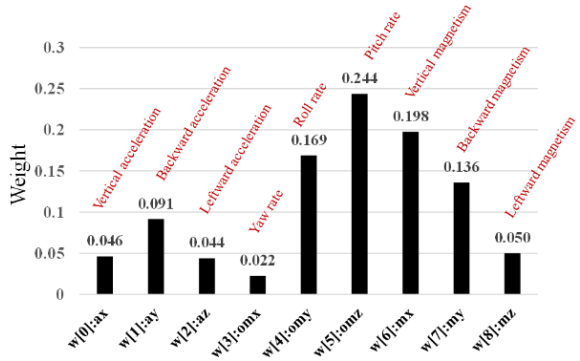


Fig. 7. Optimal weight for nine mutual correlation functions.

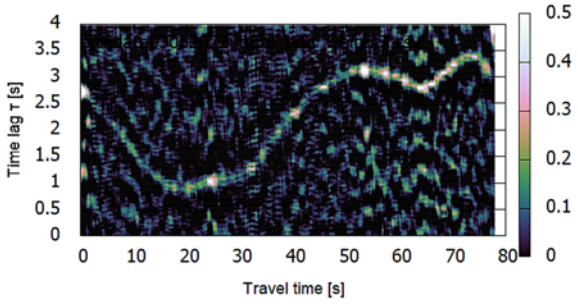


Fig. 8. Optimally weighted normalized mutual correlation function.

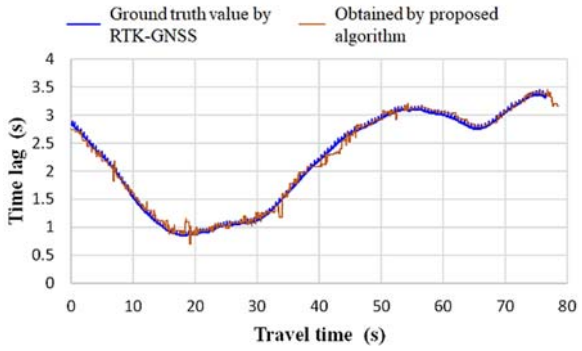


Fig. 9. Estimation results of time lag between the two vehicles.

We can see that even without RTK-GNSS, the position of the evaluation vehicle was successfully estimated with an error that was far lower than the results of our previous studies. Specifically, the RMS error was 0.38 m and the maximum error was 1.74 m. The maximum error occurred at the travel time of 33 s, which is also where the time lag estimate error is apparent in Fig. 9.

Throughout this work all software were written in C++ and Python for both Windows PC and Raspberry Pi 4B.

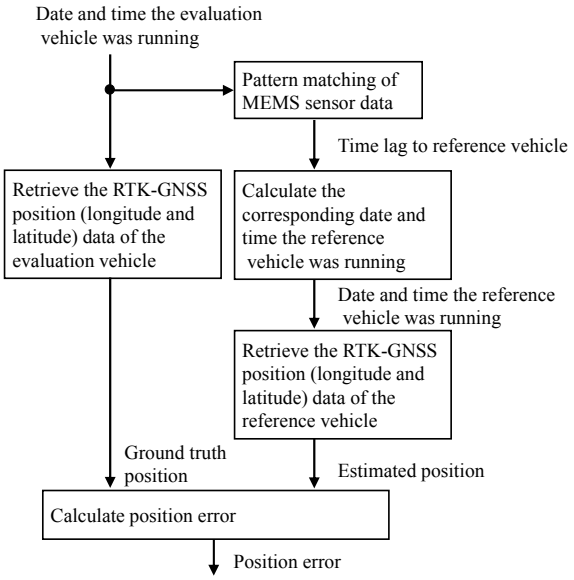


Fig. 10. Data flow diagram for precise position error evaluation.

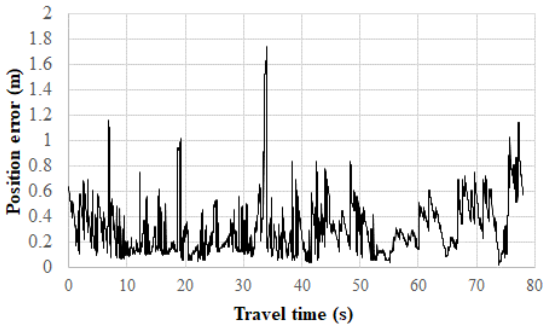


Fig. 11. Detailed position error along the evaluation travel time. The RMS error was 0.38 m and the maximum error was 1.74 m (at the travel time of about 33 s.

4. Conclusions

We presented a vehicle localization algorithm based on MEMS sensor data and showed through evaluation tests on a real road that it achieved highly precise position estimation. Our main contributions are summarized below:

1. We set up MEMS sensor data acquisition equipment operating at 50 Hz to calculate mutual correlation functions with a sufficient time resolution.
2. Among the nine mutual correlation functions we examined, the pitch rate, vertical magnetism, roll rate, and backward magnetism played significant roles (74.7%) in the vehicle localization.
3. After performing optimal weighted average of the nine mutual correlation functions using a particle

filter, the time lag estimation worked quite well and its estimation RMS error was 0.091 seconds, which is equivalent to about 1.5 m in localization error (assumed vehicle speed was 16.7 m/s)

4. The detailed position error analysis showed that the position RMS error was 0.38m and its maximum error was 1.74m, both of which are far lower than our previous results.

Future extensions will involve combining a dead reckoning algorithm to stabilize the estimation and extending the algorithm to more general road networks and different vehicles and drivers.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20K04723.

References

[1]. T. Yamagiwa and T. Yokota, Vehicle Localization Method Based on MEMS Sensor Data Comprising Pressure, Acceleration and Angular Velocity, in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA' 2022)*.

[2]. T. Yokota, Vehicle Localization by Altitude Data Matching in Spatial Domain and its Fusion with Dead *Automation*, Vol. 8, No. 4, January 2021, pp. 208-216.

[3]. T. Yokota,Vehicle localization by dynamic programming from altitude and yaw rate time series acquired by MEMS sensor, *SICE Journal of Control, Measurement, and System Integration*, 14, 1, April 2021, pp. 78-88.

[4] T. Yokota, Vehicle Localization Based on MEMS Sensor Data, in *Proceedings of the SICE Annual Conference*, September 2021.

[5]. T. Yokota, Localization Algorithm Based on Altitude Time Series in GNSS-Denied Environment, in *Proceedings of the SICE Annual Conference*, September 23-26, 2020, pp. 952-957.

[6]. T. Yokota, M. Okude, T. Sakamoto and R. Kitahara, Fast and robust map-matching algorithm based on a global measure and dynamic programming for sparse probe data, *IET Intell. Transp. Syst.*, Vol. 13 Issue 11, 2019, pp. 1613-1623.

[7]. J. Tsurushiro and T. Nagaosa, Vehicle localization using its vibration caused by road surface roughness, in *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety*, Yokohama, Japan, November, 2015, pp.164-169.

[8]. A. J. Dean, R. D. Martini, and S. N. Brennan, Terrain-based road vehicle localisation using particle filters, *International Journal of Vehicle Mechanics and Mobility*, Vol. 49, Issue 8, 2011, pp.1209-1223.

[9]. E. Laftchiev, C. Lagoa, and S. Brennanvehicle, localization from real-time data using dynamical models, ACC, Maui, HI, USA, 2012.

[10]. J. Gim and C. Ahn, Ground feature-based vehicle positioning, in *Proceedings of the SICE Annual Conference*, 2020, pp. 983-984.

[11]. J. Gim and C. Ahn, IMU-based virtual road profile sensor for vehicle localization, *Sensors*, Vol. 18, No. 10, 2018, pp. 33-44.

[12]. X. Qu, B. Soheilian, and N. Paparoditis, Landmark based localization in urban environment, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 140, June 2019, pp. 90-103.

[13]. https://www.boschsensortec.com/products/environmental-sensors/humidity-sensors-bme280/(accessed on January 27, 2023)

[14]. https://invensense.tdk.com/products/motion-tracking/9-axis/%20mpu-9250 (accessed on Jan.27.2023).

[15]. https://www.u-blox.com/en/product/zed-f9p-module (accessed on January, 27, 2023)

(2969)

Test Strip for Surface-enhanced Raman Spectroscopy, Method for its Preparation and use Thereof

Z. Chaloupková [1] and V. Ranc [2]
[1] Catrin, Regional Center of Advanced Technologies and Materials, Palacky University Olomouc, Olomouc, Czech Republic
[2] Laboratory of Experimental Medicine, Faculty of Medicine and Dentistry, Institute of Molecular and Translational Medicine, Palacky University Olomouc, Olomouc, Czech Republic
Tel.: +420585634390
E-mail: Zuzana.Chaloupková@upol.cz

Summary: New methods for the most sensitive direct detection of trace amounts of chemical compounds in complex mixture are constantly being sought. The test strip serves as a comprehensive analytical tool that allows the separation of chemical compounds based on their various physicochemical interactions of the sample components with the polymer and the part containing plasmonic silver and/or gold nanoparticles in its structure, which are subsequently detected by surface-enhanced Raman spectroscopy. These silver and/or gold nanoparticles, are covalently immobilized on the nitrocellulose membrane, allow their further functionalization by suitable antibodies. The test strip allows direct detection of low molecular weight substances and proteins, such as IgG, which is an important indicator of immune system function. The detection is based on a combination of modular polymer strips capable of separating the analyte from the biological matrix based on various physicochemical interactions of the sample components with the polymer and the part containing silver and/or gold nanoparticles in its structure.

Keywords: Raman spectroscopy, Silver nanoparticles, Gold nanoparticles, Strips, SERS.

1. Introduction

Structural analysis of organic compounds is one of the most difficult tasks in analytical chemistry [1]. New technologies and the continuous development of instrumental techniques provide new possibilities, but even so, the analysis of trace amounts of a chemical compound present in a complex mixture is still considered a difficult task [2]. Therefore, new methods for the most sensitive direct detection are constantly being sought. This work describes a test strip for surface-enhanced Raman spectroscopy, which consists of a plastic substrate on which is attached a nitrocellulose membrane containing covalently anchored silver and/or gold nanoparticles. The test strip is used for the direct detection of low molecular weight substances and proteins, where detection is based on a combination of modular polymer strips capable of separating the analyte from the biological matrix based on various physicochemical interactions of the sample components with the polymer and the part containing silver and/or gold nanoparticles in its structure.

Surface-enhanced Raman scattering (SERS) is used here to significantly reduce the detection limit [3]. The analysis of substances using surface-enhanced Raman spectroscopy has so far been based on substrates containing mainly metal colloids (silver or gold). The disadvantages of colloidal systems are the temporal and chemical instability of particle size or morphology and the inhomogeneity caused by the concentration of residual ions resulting from the chemical reduction of the metal from the respective salt. Solid substrates in the form of metal layers or glass substrates coated with gold or silver nanoparticles can also be advantageously used as substrates for surface-enhanced Raman spectroscopy. However, these substrates are very fragile and their preparation is time and cost consuming.

Another option is to spray the prepared silver nanoparticles onto a nitrocellulose membrane. The disadvantage of spraying is inhomogeneity and instability, which negatively affects the reproducibility of results and the high detection limits of SERS measurements. However, the nanoparticles are electrostatically deposited on the strip, i.e. unsystematically and inhomogeneously, and in principle cannot provide a uniform analytical signal. This method of detecting a single analyte in a simple solution does not, in principle, allow multiplex analysis of multiple analytes in complex samples.

Structural analysis of organic compounds is one of the most difficult tasks in analytical chemistry. Therefore, new methods are constantly being sought for the most sensitive direct detection (i.e. detection of the analyte under investigation directly, rather than detection of another substance with which the analyte of interest interacts). This work describes a system for the direct detection of trace amounts of chemical compounds in complex matrices.

The test strip contains silver or gold nanoparticles covalently bound to the nitrocellulose membrane via nitrogen-containing functional groups present in the nitrocellulose structure [4]. The covalent binding of the nanoparticles in the test strip ensures their homogeneity. This leads to very low detection limits at the ng/l concentration level and stability of the analytical signal, which is not possible with methods based on, for example, spraying nitrocellulose with colloidal silver or gold.

2. Results & Discussion

2.1. Description of the Test Strip

The present invention [WO 2022/179647 A1] provides a test strip for surface-enhanced Raman spectroscopy, which comprises a substrate that prevents the flow of a sample through a porous nitrocellulose membrane on which a nitrocellulose membrane having a pore size of 0.30 µm is fixed. In this nitrocellulose membrane, silver and/or gold nanoparticles are covalently anchored. Nanoparticles are defined as particles with a size in the range of 1 nm to 1 000 nm. The test strip comprises a sample coating section comprising a nitrocellulose membrane having a pore size of 0.30 µm, with an adhesion treatment consisting of a non-wetting surface (commercial product, high retention volumes > 50 µl/cm2), and an absorption section for absorbing the sample passing through the test strip, wherein in the absorption section the nitrocellulose membrane has a pore size of 0.45 µm to 0.50 µm for easy absorption of the sample. The sampling and absorption portions are located at opposite ends of the test strip in contact with the nitrocellulose membrane and are arranged so that the liquid passes freely from its application to the sampling portion through the nitrocellulose membrane, the silver and/or gold nanoparticles, the nitrocellulose membrane, to its absorption in the absorption portion of the test strip.

2.2. Test Strip for Adenine Detection

The test strip for SERS contains silver nanoparticles covalently bound to a nitrocellulose membrane according to the protocol described by S. Agnihotri [5] using NaBH4 as the primary reducing agent and trisodium citrate (TSC) as the secondary reducing agent as well as stabilizer.

The test strip was further tested for the detection of Adenine at a concentration of 10^{-7} mol/L. 50 µL of Adenine solution at a concentration of 10^{-7} M was applied to the test strip, and after the solution passed through the entire test strip spontaneously and dried, Raman microscope measurements were performed. A He-Ne laser with an excitation wavelength of 633 nm and a laser power per sample of 3 mW was used for excitation.

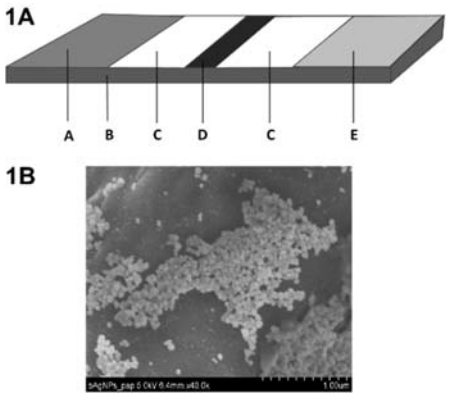


Fig. 1. A) Schematic drawing of a test strip containing silver and/or gold nanoparticles serving as a substrate for surface-enhanced Raman spectroscopy, where A is a sample plain part used for sample deposition, B is a substrate (e.g., a plastic sheet), C is a nitrocellulose membrane attached to the plastic substrate, D is a nitrocellulose membrane to which silver and/or gold nanoparticles are bonded and covalently anchored in the nitrocellulose membrane, E is a porous absorber portion for trapping the sample passed through the nitrocellulose membrane. B) Electron microscope image showing silver nanoparticles (bright spherical objects) covalently immobilized on nitrocellulose membrane C (dark background).

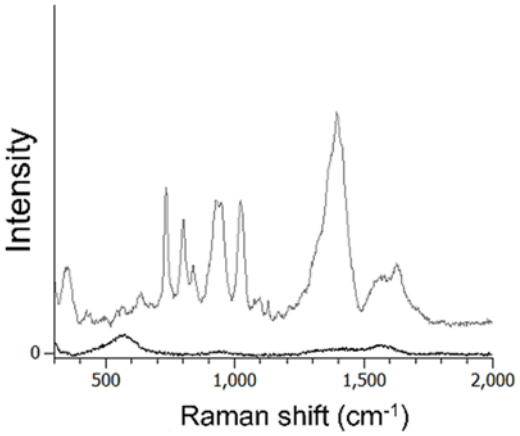


Fig. 2. Raman spectra of Adenine at a concentration of 10^{-7} M immobilized on covalently bound silver nanoparticles compared to a blank (signal originating from the substrate - silver nanoparticles bound to a nitrocellulose membrane (black spectrum))

3. Conclusion

The test strip shows the possibility of direct detection of low molecular weight substances and proteins, such as Adenine, based on a combination of modular polymer strips capable of separating the analyte from the biological matrix. This is based on the different physicochemical interactions of the sample components with the polymer and the part containing silver and/or gold plasmonic nanoparticles in its structure, which are subsequently detected by surface-enhanced Raman spectroscopy.

Acknowledgements

This work was supported by the ERDF/ESF project "Nano4Future" (grand No. CZ.02.1.01/0.0/0.0/16_019/0000754).

References

[1]. E. C. Minor et al. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci.: Processes Impacts* 16, 2014, pp. 2064-2079.

[2]. B. J. Bolann et al. Evaluation of methods for trace-element determination with emphasis on their usability in the clinical routine laboratory, *Scand J Clin Lab Invest,* 67, 4, 2007, pp. 353-366.

[3]. Y. Tzeng and B-Y. Lin, Silver SERS Adenine Sensors with a Very Low Detection Limit, *Biosensors,* 10, 5, 53, 2020, pp. 1-13.

[4]. L. Wei et al., Silver nanoparticles: synthesis, properties, and therapeutic applications, *Drug Discovery Today,* 20, 5, 2015, pp. 595-601.

[5]. S. Agnihotri and S. Mukherji, Size-controlled silver nanoparticles synthesized over the range 5–100 nm using the same protocol and their antibacterial efficacy, *Royal Society of Chemistry*, 2014, pp. 3974-3983.

(3177)

Exploring Applications of Industry 4.0 and Lean Manufacturing Techniques within a Product Label Manufacturer: A Simulation Case Study

Stefan Harrison [1*] and Boppana Chowdary [2]
[1, 2] Faculty of Engineering, The University of the West Indies, Trinidad
E-mail: stefan.harrison@my.uwi.edu, Boppana.Chowdary@sta.uwi.edu

Abstract: Industry 4.0 is a manufacturing design philosophy and industry standard for the integration of Information Communication Technology (ICT) and Cyber-Physical Systems (CPS) into production processes. Industry 4.0 aims to improve productivity by creating a waste averse smart manufacturing system using novel applications of current and emerging technologies. It is based on the concepts of data-driven decision support services, Horizontal and Vertical Information Technology-Operations Technology (IT-OT) value-chain integration and decentralized control and production. Lean Manufacturing is a value-to-customer-focused manufacturing philosophy which applies procedures designed to ingrain waste reduction and efficient, competency-building practices into workplace culture. The aim of this paper is to examine the synergistic benefits of these manufacturing philosophies on a local firm using Discrete Event Simulation (DES). The key performance indicators (KPI) of Flow Time (FT), Waiting Time (WT) and Work in Process (WIP) were used to determine the efficacy of models investigated. The study results indicated a 46%, 57% and 61% decrease in each KPI respectively.

Keywords: Industry 4.0, Lean manufacturing, Discrete event simulation, Case study.

1. Introduction

Since the advent of modern globalised markets, there has existed competition amongst firms from developed and developing nations for increased market share. This is particularly true of low-skilled industries [1] from areas including the Caribbean. This competition motivated the development of innovative thinking and technologies in service of a trend toward more efficient, less labour and resource intensive industrial processes [2]. One philosophy based on this trend of innovative thinking is Lean Manufacturing (LM), a value-to-customer-focused application of procedures designed to ingrain waste reduction and competency-building practices into workplace culture [3].

Amongst the novel technologies developed were increasingly cost-effective CPS and ICT devices. These technologies, when combined with advances in artificial intelligence and high-volume data analytics, then made the concept of Industry 4.0 (I4.0) possible. I4.0 is the philosophy of integrating efficiency and value-creating technologies into all aspects of a product's value-chain. In reductive terms, I4.0 uses new applications of technology to simplify work and offer access to previously unfeasible solutions.

2. Literature Review

Current research indicates that some applications of LM practice actually serve to enhance the ability of I4.0 technologies to further improve productivity [4-6]. While mature LM practice is not a requirement for introducing basic I4.0 technological solutions, there is some evidence to suggest a correlation between the advanced application of I4.0 technologies and a mature understanding of LM [7]. This is particularly evident when considering the stresses that enacting the digital transformation required of I4.0 places on a firm's organizational structure and workplace culture. Organisations must adopt continuous improvement and flexible thinking in order to keep pace with the new capabilities offered by I4.0. Meanwhile workers must build upon their technical skills and core competencies if they wish to take advantage of novel technologies and opportunities [8]. LM provides excellent procedural tools and approaches in both of these cases.

However, there exists a scarcity of academic research into the impact of the core technologies of I4.0 on LM with regard to application within developing nations [4, 9]. Factors including a firm's socioeconomic environment, supply chain positioning and access to technology and training have been proposed as influences that moderate the interaction between I4.0 and LM [5, 10]. This paper therefore aims to examine the application of I4.0 and LM as improvement measures within a Trinidad manufacturing firm using a Discrete Event Simulation (DES) case study.

3. Case Study

3.1. System Definition

The subject of the case study was a printed label and packaging supply firm servicing the Caribbean region. The firm had identified time and quantity shortfalls in production as areas in need of improvement. A lean approach to production was seen as a way to improve operations. Value Stream Mapping (VSM) was used define the current state of

the system. Based on this understanding of existing processes (see Fig. 1), a simulation model was formulated to represent the existing system. The Key Performance Indicators (KPI) of Flow Time (FT), Waiting Time (WT) and Work in Process (WIP) were compared to those produced by a second model which incorporated a proposed strategy utilising I4.0 and LM practices. Statistical testing via SPSS was used to determine the significance of any differences observed between the two models.

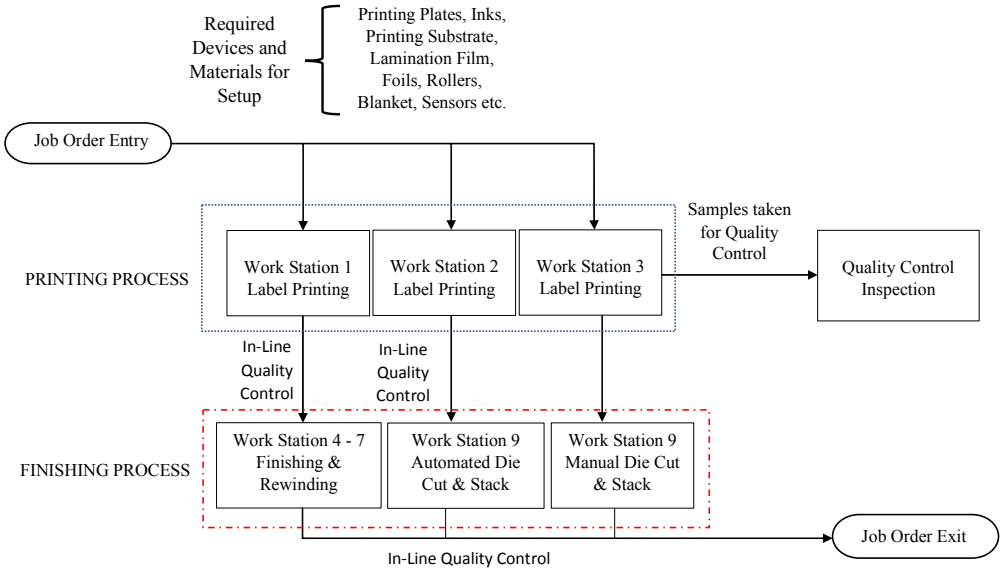


Fig. 1. Label Printing Process Map. A Physical Job Ticket and e-Kanban Ticket follow the Job Order from Entry to Exit.

3.2. Model Formulation

Rockwell Automation's Arena® simulation software was used to develop models for the existing and proposed systems (see Fig. 2). These models were designed as non-terminating simulations reflecting the existing system where machine tools were infrequently idle during working hours. Truncated replications were used to compensate for initiation bias. This required that a warm up period, the time needed the operation to achieve a steady-state conditions, be determined.
A graph of Flow Time versus Simulation Time was generated using Arena Output Analyzer® data. Examination of this graph allowed for a conservative estimation of 30 hours as the duration of the warm up period [11]. Based on the available resources the simulation run length was set to 10 days, 16 hours a day for 150 replications. These parameters provided data output for the KPI with a precision of ± 1 hour for FT and WT, and ± 1 label for WIP. The required number of replications to achieve this precision, n_e, was derived using Equation (1).

$$n_e = n_0 \frac{h_0^2}{h^2} \quad (1)$$

where n_0 is the number of replications required to derive a precision of h_0 and h is the desired precision of the examined KPI [11].

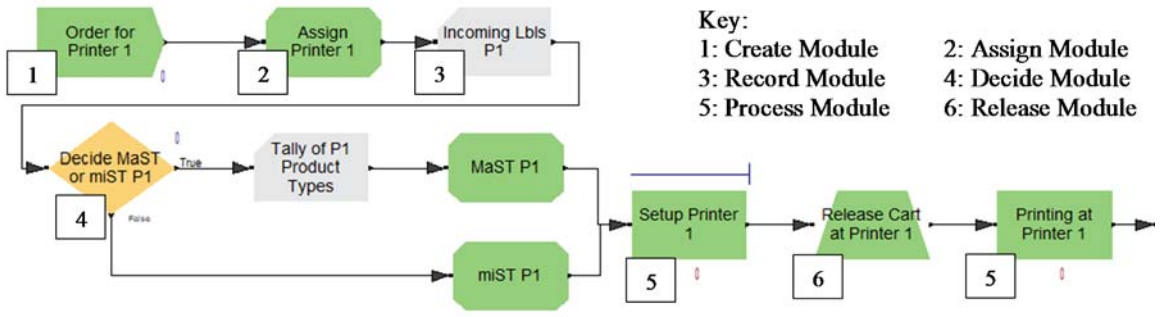


Fig. 2. Extract from Existing Model: Product Label Printing Process at Printer 1. Simulated Job Orders Enter at the Ctreate Module, are assigned user variables at the Assign Modules, recorded metrics are logged by the Record Module while the job orders are processed at the Decision Module based on predefined conditions, the Process Modules simulate the use of system resources. Resources not used by following processes can be freed by the Release Module.

4. Results and Discussion

Findings from literature have indicated that I4.0 results in suboptimal productivity improvements when employed as a wholly technological solution [5, 12, 13]. In contrast, a beneficial, synergistic effect has been noted when I4.0 is applied in along with LM practice [14-16]. This is perhaps due the technology focused, value-creation approach of I4.0 complementing the customer focused, worker competency-building approach of LM. In essence, I4.0 provides novel applications of technology to simplify complex work while LM provides simple procedures that encourages workers to accept and adapt to using new technology efficiently [13].

In light of these findings, the proposed strategy utilized a combination of LM practices and I4.0 technologies to address wastes observed during the VSM exercise. First, job orders entered the existing system as a physical work order which was then compared to a matching e-kanban ticket provided to the workstation. This process added unneeded complexity to the process. In notable instances, delays were observed when the physical job order did not match the information provided by the electronic ticket.

Experienced operators tended to ignore the discrepancy between the two documents instead of pausing work to request clarification of the issue. When questioned, they noted that it was common for the physical document to fail to keep pace with changes to the production plan. As a result, when instructions were in doubt, the e-kanban ticket was preferentially referenced. Removing the requirement of a physical job order simplified the process and decreased the time spent on setup activities.

Second, the processes for retrieving the storage location of printing plates and logging completed job orders were automated. This addressed commonly indicated sources of wasted time during setup activities. These changes were represented in the proposed model by utilizing the firm's best practice standard for setup times in the relevant processes as determined by prior time and motion studies.

The KPI for the existing and proposed models were presented in Table 1 and showed improvements to FT, WT and WIP when examined. Comparing the two models, production FT decreased by 46% from an average of (50 ± 1) hours to (27 ± 1) hours. The average WT similarly decreased by 57% from (42 ± 1) hours in the existing model to (21 ± 1) hours in the proposed model. Finally, a 61% decrease was observed in the WIP, (1,182,099 ± 1) labels to (459,318 ± 1) labels, when the proposed strategy was implemented. Independent Samples t-tests (see Table 2) determined that the differences observed were significant at the 95% Confidence Interval. The reported test statistics were FT: $t(58) = 10.73$, $p < 0.001$; WT: $t(58) = 9.96$, $p < 0.001$ and WIP: $t(58) = 6.35$, $p < 0.001$.

Table 1. Key Performance Indicators of Existing and Proposed Models: Flow Time (FT), Waiting Time (WT) and Work in Process (WIP).

KPI	Existing Model	Proposed Model	Precision
FT (hour)	50	27	± 1 hour
WT (hour)	42	21	± 1 hour
WIP (label)	1,182,099	459,318	± 1 label

Table 2. Analysis of Key Performance Indicators - Levene's Test and Independent Samples t-Test: Flow Time (FT), Waiting Time (WT) and Work in Process (WIP).

KPI	Levenes's Test		Independent Samples t-Tests		
	F - value	p value	t - value	Degrees of Freedom	p value
FT (hour)	1.702	0.197	10.73	58	<0.001
WT (hour)	2.238	0.140	9.96	58	<0.001
WIP (label)	0.434	0.513	6.35	58	<0.001

5. Conclusions

The proposed strategy was formulated with consideration to the issues revealed by the VSM exercise, resources available to the firm and the interviews of factory floor personnel. The proposed changes leveraged existing but underutilized capabilities of the firm's ERP system. Through application of LM practices and basic I4.0 techniques for automating processes and closer integration of the Printing and Production Scheduling Departments, the proposed model indicated the possible improvement to the KPI of FT, WT and WIP by 46 %, 57 % and 61 % respectively. This result supports the postulation that there is a beneficial synergy when efficiency-promoting I4.0 technologies are combined with the competency-building procedures of LM practice.

References

[1]. Donno, Daniela, and Nita Rudra. David and Goliath? Small Developing Countries, Large Emerging Markets, and South-South Preferential Trade Agreements, *International Studies Quarterly*, Vol. 63, No. 3, 2019, pp. 574-588. DOI: 10.1093/isq/sqz041.

[2]. Jian Qin, Ying Liu, Roger Grosvenor. A Categorical Framework of Manufacturing for Industry 4.0 and Beyond, *Procedia CIRP*, Vol. 52, 2016, pp. 173-178.

[3]. Hozak, Kurt, and Eric O. Olsen. Lean habits, *International Journal of Business Excellence*, Vol. 19, No. 3, 2019, pp. 305-322. DOI: 10.1016/j.procir. 2016.08.005.

[4]. Pagliosa, Marcos, Guilherme Tortorella, and Joao Carlos Espindola Ferreira. Industry 4.0 and Lean

Manufacturing, *Journal of Manufacturing Technology Management*, Vol. 32, No. 3, 2021, pp. 543-569. DOI: 10.1108/jmtm-12-2018-0446.

[5]. Tortorella, Guilherme Luz, Ricardo Giglio, and Desirée H. van Dun. Industry 4.0 adoption as a moderator of the impact of lean production practices on operational performance improvement, *International Journal of Operations & Production Management*, Vol. 39, Nos. 6/7/8, 2019, pp. 860-886. DOI: 10.1108/ijopm-01-2019-0005.

[6]. Sanders, Adam, Chola Elangeswaran, and Jens Wulfsberg. Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing, *Journal of Industrial Engineering and Management*, Vol. 9, No. 3, 2016, pp. 811-833. DOI: 10.3926/jiem.1940.

[7]. Moura, Luciano Raizer and Holger Kohl, Maturity Assessment in Industry 4.0 – A Comparative Analysis of Brazilian and German Companies, *Emerging Science Journal,* Vol. 4, no. 5, 2020, pp. 365-375. DOI: 10.28991/esj-2020-01237.

[8]. Santos, Reginaldo Carreiro, and José Luís Martinho. An Industry 4.0 maturity model proposal, *Journal of Manufacturing Technology Management*, Vol. 31, No. 5, 2019, pp. 1023-1043. DOI: 10.1108/jmtm-09-2018-0284.

[9]. Aboal, Diego, and Ezequiel Tacsir, Innovation and productivity in services and manufacturing: the role of ICT, *Industrial and Corporate Change*, Vol. 27, No. 2, 2018, pp. 221-241. DOI: 10.1093/icc/dtx030.

[10]. Shahin, Mohammad, F. Frank Chen, Hamed Bouzary, and Krishnan Krishnaiyer. Integration of Lean practices and Industry 4.0 technologies: smart manufacturing for next-generation enterprises, *The International Journal of Advanced Manufacturing Technology*, Vol. 107, No. 5-6, 2020, pp. 2927-2936. DOI: 10.1007/s00170-020-05124-0.

[11]. Kelton, W. D., R. P. Sadowski and N. B. Zupick, Simulation with Arena, 6th ed., *McGraw Hill*, 2015.

[12] Saabye, Henrik, Thomas Borup Kristensen, and Brian Vejrum Wæhrens. Real-Time Data Utilization Barriers to Improving Production Performance: An In-depth Case Study Linking Lean Management and Industry 4.0 from a Learning Organization Perspective, *Sustainability,* Vol. 12, No. 21, 2020, pp. 1-21. DOI: 10.3390/su12218757.

[13]. Vereycken, Yennef, Monique Ramioul, Sam Desiere, and Michiel Bal. Human resource practices accompanying industry 4.0 in European manufacturing industry, *Journal of Manufacturing Technology Management,* Vol. 32, No. 5 2021, pp. 1016-1036. DOI: 10.1108/jmtm-08-2020-0331.

[14]. Buer, Sven-Vegard, Jan Ola Strandhagen, and Felix T. S. Chan. The link between Industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda, *International Journal of Production Research,* Vol. 58, No. 8, 2018, pp. 2924-2940. DOI: 10.1080/00207543.2018.1442945.

[15]. Abdulnour, S., C. Baril, G. Abdulnour, and S. Gamache. Implementation of Industry 4.0 Principles and Tools: Simulation and Case Study in a Manufacturing SME, *Sustainability*, Vol. 14, No. 10, 2022, pp. 1-20. DOI: 10.3390/su14106336.

[16]. Pereira, Clinton, and H. K. Sachidananda. Impact of industry 4.0 technologies on lean manufacturing and organizational performance in an organization, *International Journal on Interactive Design and Manufacturing,* Vol. 16, No. 1, 2022, pp. 25-36. DOI: 10.1007/s12008-021-00797-7.

(3232)

Semi-active Damper Suspension Road Estimation and Control based on Neural Networks

Diana Hernandez-Alcantara [1], Luis Amezquita-Brooks [2] and Luis Rivera-Perez [1]
[1] Universidad de Monterrey, Av. Ignacio Morones Prieto 4500-Pte, 66238, Nuevo Leon, Mexico
[2] Universidad Autonoma de Nuevo Leon, Av. Universidad SN, 66455, Nuevo Leon, Mexico
E-mail: diana.hernandeza@udem.edu

Summary: The use of semi-active damper systems in commercial automotive vehicle applications has become widespread in recent years. In contrast with typical passive shock absorbers, semi-active dampers allow modifying the damping coefficient online, introducing the requirement of appropriate control algorithms. In this article a novel approach for semi-active damper manipulation in automotive vehicles is presented. The proposed scheme is based on a Neural Network (NN) road-profile reconstruction together with a NN-based predictive controller. The main objective of the system is to improve passenger comfort; however, the algorithm can also be tuned for road holding or a combination of both. The resulting suspension control system was validated trough simulations using the CarSim™ software, which includes a comprehensive representation of the vehicle, and a well-known non-linear damper model. The results show that significative improvements can be achieved with a relatively simple NN topology, which is well posed for real-time implementation due to its low computational complexity.

Keywords: Road estimation, Neural network, Semi-active suspension.

1. Introduction

In recent years vehicular drive-train and driving-aids technologies have evolved because of automation and the use of electric energy. However, another crucial element for these vehicles is the suspension. This subsystem is responsible of a major part of the vehicle handling characteristics, most notably passenger comfort and road holding [1, 2].

Traditional suspensions use passive dampers which have a fixed damping profile. However, recently semi-active dampers have become widely available in commercial vehicles. These devices are based in technologies such as magnetorheological, electrorheological, or electrohydraulic systems, and are able to modify their damping coefficients during the operation of the vehicle [3].

The simplest approach for semi-active damper operation is an open loop scheme, where a fixed damper setting is used depending on the pilot configuration or the road-type. In this setup the road-type can be obtained with an online classifier. On the other hand, fully online approaches modify the damper setting constantly using live vehicle measurements [1, 4, 5].

2. Road Estimation

The main variables which are involved in the suspension dynamics are the sprung mass position z_s, the unsprung mass position z_{us}, the damper deflection z_{def} and the road profile z_r, as well as their first and second time-derivatives.

The sprung mass acceleration $\ddot{z}_s$ can be used to measure passenger comfort. In particular, it is desired for $\|\ddot{z}_s\|$ to be as low as possible for the current road profile z_r. In this setup the z_r effectively acts as an external perturbation. Thus, knowledge regarding this variable can be very valuable for the suspension management system, [6]. Nonetheless, in a typical commercial setup, the only relevant variables measured directly are $\ddot{z}_{us}$ and $\ddot{z}_s$ trough the aid of accelerometers, since the addition of road profile measuring sensors is currently very costly.

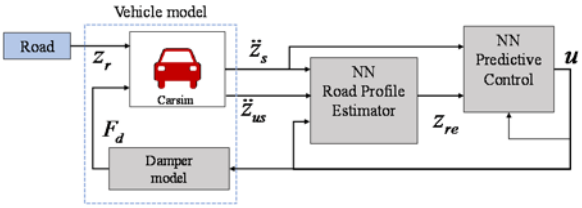


Fig. 1. Block diagram of the proposed system.

In the proposed approach, measurements of $\ddot{z}_{us}$ and $\ddot{z}_s$ are feed to a Neural Network (NN) to reconstruct z_r, as shown in Fig. 1. This NN was trained using simulated vehicle suspension data obtained from the CarSim™ software using the randomly-generated and standardized road profiles of Table 1. Since semi-active dampers have particular non-linear behaviors, currently not included in CarSim™, the simulation was complemented with the well-known Guo model [7].

Table 1. ISO 8601 Road profile PSD parameters.

Type of road	Class	Gain	Slope
Smooth runway	A	3.2×10^{-7}	3.8
Smooth highway	B	1.2×10^{-6}	3.5
Highway with gravel	C	4.8×10^{-6}	2.1
Rough runway	D	8.1×10^{-6}	2.0

Several NN topologies and training algorithms were tested; however, a simple feedforward time-series NN topology with Bayesian-regularization training was preferred due to its low overall computational cost and acceptable performance. Table 2 shows a summary of the performance of several simple NN topologies, with the 1-hidden layer 8-neurons topology being selected as the final structure for the road profile estimator.

Table 2. Performance of road profile estimator NNs.

Hidden layers	1			
Neurons	5	6	7	8
SME ($\times 10^{-5}$)	6.03	5.95	5.99	1.09
Hidden layers	2			
Neurons	5	6	7	8
SME ($\times 10^{-5}$)	6.11	6.01	5.95	5.92

In this context it is important to note that several experimental and theoretical studies have determined that estimating the road profile from measurable suspension variables is an ill posed problem [6]. For instance, if z_r is treated as a perturbation input, then a linear disturbance observer can be designed for this variable using a linear vehicle suspension model. In this setup it was found that the observability matrix tends to be close to singular. This indicates that it is, in principle, difficult to estimate the road profile. Nonetheless, the NN estimator accuracy was overall acceptable as indicated by the low levels of Square Mean Error (SME) obtained in Table 2. In addition, Fig. 2 shows some typical examples of the performance of the resulting road profile estimator in the time domain.

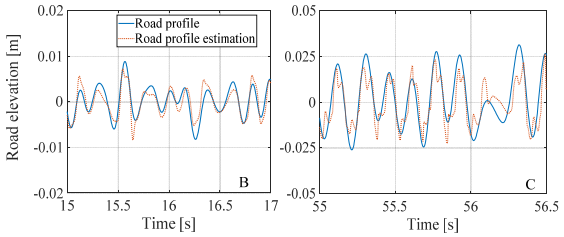


Fig. 2. Time-domain responses of the NN road profile estimator.

3. Predictive Semi-active Damper Control

Although the accuracy of the road profile estimator is important, the main objective of the suspension control system is to improve the vehicle dynamics. In this article the goal is to improve the passenger comfort, which is achieved by reducing $\|\ddot{z}_s\|$.

For the manipulation of the semi-active damper a recurrent time-series NN is proposed with the sprung mass acceleration $\ddot{z}_s$ and the reconstructed road profile z_{re} as inputs and the semi-active damper manipulation u as output, as shown in Fig. 1. In a similar manner to the road estimation, the training and validation datasets were obtained using the CarSim™ software plus the modified Guo non-linear damper model.

The training of this NN is configured so that a predictive control configuration is obtained by setting up the desired control input u as the NN output in a supervised training scheme. In particular, during the training the predicted decay for $\ddot{z}_s$ is specified as a smoothed first order dynamic with a stable α decay-rate. For stability it is required that $\alpha \in (0,1)$. In addition, this parameter is important for improving the overall system performance. An α value closer to 1 will tend to a smother $\ddot{z}_s$ decay while a value closer to 0 implies a more sudden decay. In practice, values close to 1 tend to produce the better results because they are less sensitive to sensor noise and other perturbations. As an example, Table 3 shows the Root Mean Squared (RMS) of the sprung mass acceleration with several values for α considering a mixture of road profiles from Table 1.

Table 3. Performance with respect to α decay-rate.

α	$\ddot{z}_s$ **RMS**
0.90	0.9818
0.95	0.9695
0.99	0.9602

On the other hand, although the road estimator provides data regarding the current and past road profiles, the predictive controller requires z_r predictions, both during the training phase and the NN implementation. In this case the predicted road profile is setup as a repetition of the previous road profile time-series for the prediction horizon H. Although other road-profile prediction schemes were also tested, this presented the most consistent results.

As a summary of the previous discussion, Fig. 3 shows an example of the predicted $\|\ddot{z}_s\|$ decay for several α values as well as the predicted road profile used for the training of the NN with a prediction horizon of 4. These time-series datapoints together with the previous damper input u values (i.e. the recurrent input of the NN) are used as inputs for the predictive control NN during the training stage.

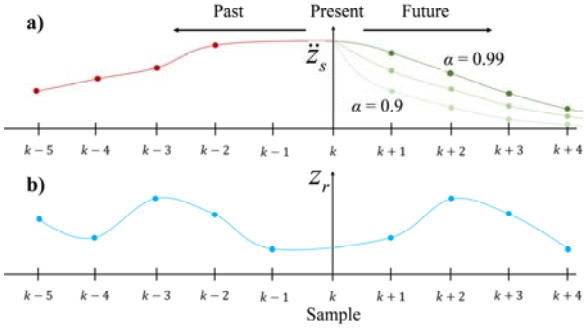


Fig. 3. a) Predicted $\ddot{z}_s$ decay; b) Predicted road profile.

Several NN topologies and training algorithms were tested for the predictive control subsystem; however, a simple 2-layers 7-neurons per layer with Bayesian-regularization as training algorithm yielded acceptable results and a stable recurrent NN, which is an important property for this subsystem.

4. Results

The proposed system was tested using CarSim™ and a set of randomized and standardized road profiles. The results, presented in Table 4, show that the proposed algorithm was able to reduce the RMS of the sprung mass acceleration in all cases. The fact that improvement is more significant in road classes A and B may be explained due to the selection of a smoother decay rate, since road classes C and D would require a more aggressive setting.

Table 4. Performance of the suspension system.

Road Class	Nominal RMS	NN Control RMS	Improvement
A	0.0198	0.0178	10.1%
B	1.8195	1.6031	11.89%
C	3.7619	3.6493	2.99%
D	5.5272	5.4538	1.328%

Finally, a series of pseudo-Bode diagrams for the vehicle suspension was obtained by introducing a chirp-like road profile in the CarSim™ simulator. Figure 4 shows the results for several fixed damper settings as well as the proposed NN approach. From this figure it is clear that there is not a single fixed setting which is better than the NN controller for all the frequency range.

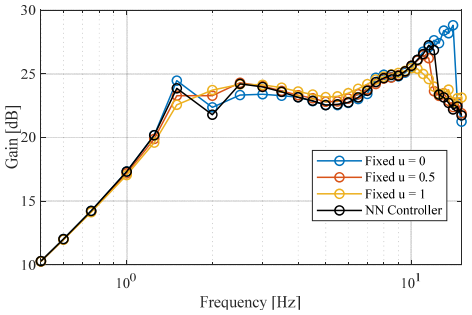


Fig. 4. Pseudo-Bodes of $\|\ddot{z}_s\|$ using several fixed damper values and the closed loop controller.

5. Conclusions

A NN road-estimator working in conjunction with a NN predictive-controller were able to improve the comfort of a semi-active suspension system with a relatively simple topology and without the need of additional sensors, which is a salient characteristic of the proposed system.

The results show that the separation of the road estimation problem from the semi active damper control allows tuning these two subsystems as separate NN, each with particular requirements and characteristics.

Although the results presented here are focused in passenger comfort, they can be easily modified for additional goals, such as road-holding. Future work may focus in setting the NN training to improve particular frequency domain goals, measured by the vehicle suspension pseudo-Bode responses.

References

[1]. J. Park, K. Min, H. Kim, W. Lee, G. Cho, and K. Huh, Road Surface Classification Using a Deep Ensemble Network with Sensor Feature Selection, *Sensors*, Vol. 18, No. 12, 2018.

[2]. Y. Qin, C. Xiang, Z. Wang, and M. Dong, Road excitation classification for semi-active suspension system based on system response, *Journal of Vibration and Control*, Vol. 24, No. 13, Feb. 2017, pp. 2732–2748.

[3]. M. M. Morato, M. Q. Nguyen, O. Sename, and L. Dugard, Design of a fast real-time LPV model predictive control system for semi-active suspension control of a full vehicle, *Journal of the Franklin Institute*, Vol. 356, No. 3, 2019, pp. 1196–1224.

[4]. W. Liu, R. Wang, R. Ding, X. Meng, and L. Yang, On-line estimation of road profile in semi-active suspension based on unsprung mass acceleration, *Mechanical Systems and Signal Processing*, Vol. 135, 2020, p. 106370.

[5]. Y. Qin, Z. Wang, C. Xiang, E. Hashemi, A. Khajepour, and Y. Huang, Speed independent road classification strategy based on vehicle response: Theory and experimental validation, *Mechanical Systems and Signal Processing*, Vol. 117, 2019, pp. 653–666.

[6]. M. Yousefzadeh, S. Azadi, and A. Soltani, Road profile estimation using neural network algorithm, *Journal of Mechanical Science and Technology*, Vol. 24, No. 3, 2010, pp. 743–754.

[7]. S. Guo, S. Yang, and C. Pan, Dynamic Modeling of Magnetorheological Damper Behaviors, *Journal of Intelligent Material Systems and Structures*, Vol. 17, No. 1, Jan. 2006, pp. 3–14.

(3798)

Complex Multi Robot Hybrid Platform for Augmented Virtual Skiing Immersion

T. Houda [1], A. Beghdadi [2], L. Beji [2] and A. Amouri [2]
[1] ESIEE-IT, 8 Rue Pierre de Coubertin, 95300 Pontoise, France
[2] IBISC EA 4526 Laboratory, University of Paris-Saclay, 91080 Evry-Courcouronnes, France
E-mail: thouda@esiee-it.fr, aymanaymar.beghdadi@univ-evry.fr lotfi.beji@univ-evry.fr, ali.amouri@univ-evry.fr

Summary: In recent years, the integration of virtual reality in conjunction with mechatronic systems such as motion simulators has become very important in industry in general. The different types of simulators are increasingly being integrated into the various application areas in aviation, at sea and on land. An extension of the simulator in the field of sports and especially functional rehabilitation has gained great importance in the public and private sector. In this context, the goal of our team in the laboratory is to study the design and produce a simulator mastering all low-level elements in order to explore the various possibilities of full control of the integration of a virtual environment into a complex robotic system. In this paper, we aim to validate the precision and shape tracking of motion perception in skiing for simulator users with the Nao robot.

Keywords: Handiski, Motion simulator, Motion Cueing Algorithm, Humanoid Robot, Disabled people, Washout filter.

1. Introduction

Ski simulators are becoming increasingly popular in the sliding sports industry due to their ability to replicate real-world sliding conditions and provide users with an immersive experience. Ski simulators are simulation devices that replicate the movements and sensations of sliding sports such as skiing, snowboarding and skating. They use advanced technologies to replicate the sliding motions, gravity, and vibrations of snow, such as the Motion Cueing Algorithm (MCA) [1]. These simulators can be used for sports training, injury recovery, mental preparation, and functional rehabilitation [2]. Ski simulators can also be coupled with virtual reality to increase immersion and human sensation [3].

MCAs are computer-based techniques used to simulate the effects of motion in skiers. These algorithms are designed to mimic the physical sensation of motion, such as acceleration and deceleration, by adjusting the visual cues that the skier receives. These cues can include changes in the skier's body position, speed, and slope angle, as well as changes in the skier's body position and motion itself. The MCA produces a trajectory that maximizes human sensation while respecting the physical limitations of the simulator.

Motion simulators are a multidisciplinary engineering system that can be traced back to the early 20th century. One of the first such devices was the Gough-Stewart platform [4], which was primarily used for military training and research [5, 6]. This platform has a six-degree-of-freedom range of motion. The upper platform can be mounted in a cockpit and uses a combination of rotating and linear actuators to simulate motion. Over time, motion simulators have evolved significantly thanks to advances in technology and engineering. The hybrid platform was introduced by coupling the Gough-Stewart platform with the X-rail or XY table to create a more realistic and immersive experience. For example, the Renault Driving Simulator, which adds an X-rail at Technocentre Renault, France [7] and the National Advanced Driving Simulator (NADS) at IOWA, USA [8] and the XY-6 DOF from IBISC Laboratory [9], which adds an XY table. The aim of the project is to contribute to technological progress in the field of rehabilitation and to provide a unique approach to physiotherapy for people with disabilities. For that, a low-cost simulator called the XY-6 DoF was designed and built from scratch in the IBISC laboratory [10]. The XY-6 DoF simulator is characterised by being completely open and allowing low-level access to the platform, which enables the development and adaptation of new algorithms.

In this paper, we aim to validate the precision and shape tracking of motion perception in skiing for simulator users and evaluate the effectiveness of the simulator as a rehabilitation tool. The humanoid robot Nao has a skeletal structure that mimics a human, which enables it to emulate human-like motion patterns. It is also equipped with appropriate sensors to perceive and interact with its environment. The Nao humanoid robot has been used to interact with humans for many years and has proven its efficiency. Nowadays, it is more and more used in robot-robot interaction, replacing humans, such as in the paper [11], which uses the Nao robot for gesture-based interaction between humans and multi-robot swarms. For all these reasons and for human safety, Nao is a good choice to replace humans with a good similarity.

This paper is organized as follows: Section 2 describes the technical and system architecture of the XY-6 DoF simulator. Section 3 discusses the robot coupling and interaction between the XY-6 DoF and the humanoid robot Nao. Finally, Section 4 presents simulations and discussions.

2. Technological Presentation

The ski simulator system integrates complex multi-robot interaction through the XY-6 DoF platform and the Nao humanoid robot. The XY-6DoF simulator presented in Fig.1 and Fig.2, was designed and developed from scratch by our research team in the IBISC laboratory [11].

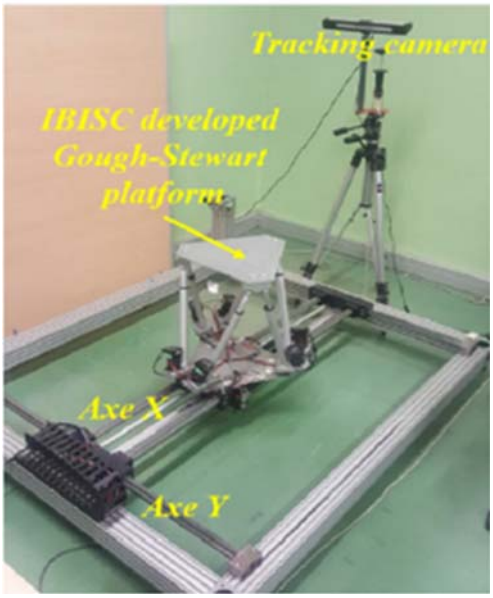


Fig. 1. The XY-6 DoF physical ski simulator.

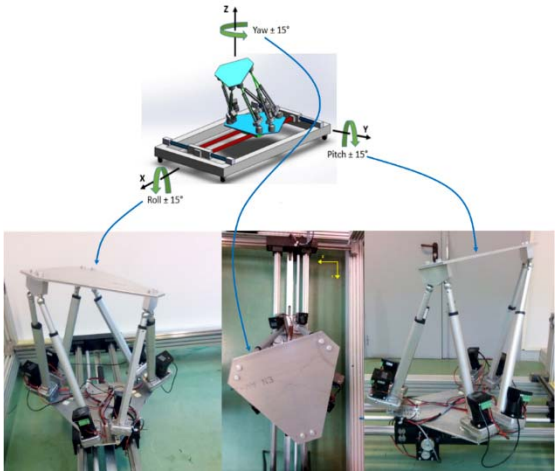


Fig. 2. The pseudo-rotation of the XY-6 DoF simulator.

The main advantage of the XY-6 DoF Simulator is that it is a completely open low-level access platform that makes it very easy to create, develop, and evaluate new algorithms. Our hybrid XY-6 DoF simulator takes advantage of the XY Table and the parallel Gough-Stewart platform characteristics.

Linear XY motions from the bottom of the platform and rotational motions from the upper part are generated by 8 PD4- CB Nanotec motors, guaranteeing the system a total of 6 degrees of freedom. Cylinders actuated by the six motors of the upper part allow rotational movements with an angular amplitude of +/- 15°.

The following table shows the characteristics of the XY-6 degrees of freedom in each direction.

Table 1. XY-6 DoF simulator characteristics.

Axis	Acceleration	Trajectory
X	1.25 (m/s²)	115 (cm)
Y	0.71 (m/s²)	115 (cm)
Z	0.52 (m/s²)	30 (cm)
Roll	0.17 (rad/s²)	+/- 15 (°)
Pitch	0.17 (rad/s²)	+/- 15 (°)
Yaw	0.17 (rad/s²)	+/- 15 (°)

Communication with the Gough-Stewart platform is carried out using a National Instruments card PCI -8513 CAN, a CAN FD interface (Controller Area Network/Flexible Data-Rate). This device allows communication with the PD4-CB brushless DC motors at 1 Mbps and their control via a special Matlab toolbox. The 8 joints are driven by the Nanotec PD4-CB servo motor, where the rotational movement of the leg is converted into a translational movement by the high-quality ball screw system.

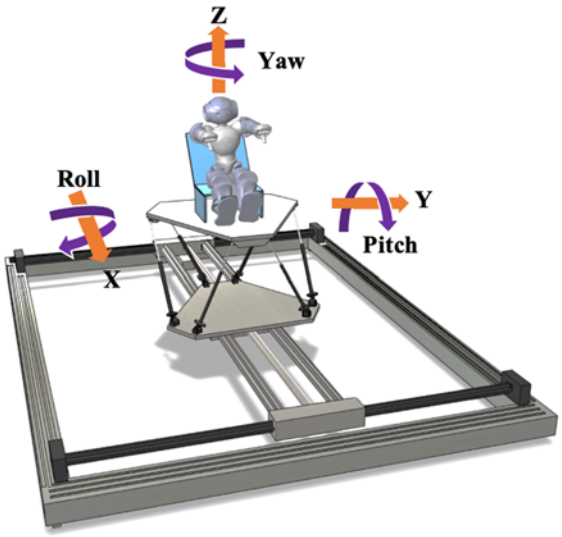


Fig. 3. The XY-6 DoF ski simulator coupled with the Nao robot.

Fig. 3 shows the integration of the robot Nao into our simulator XY-6 DoF in an immersive environment. Thanks to the advantages of its pelvic kinematics design, the well-known Nao robot [12,13] was chosen to be integrated into the system, respecting the mechatronic constraints related to handy-skiing context (inactive lower limbs).

The connections between mechatronic and software layers presented in Fig. 4 highlight the existing multi-robot interaction and the technical methods that allow its observation.

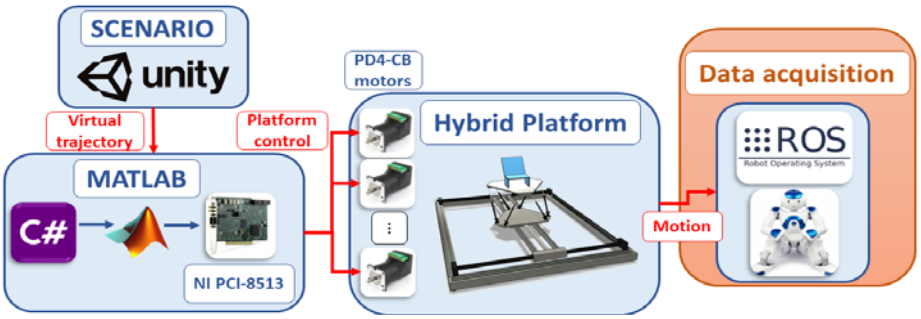


Fig. 4. Technological system architecture.

3. Multi-robot Interaction

The XY-6 DoF platform tracks trajectory information from the virtual slide scenario designed using Unity software. The dynamic modeling behavior and sliding mode control of the XY-6 DoF robot simulator were developed and validated in a previous research work [3, 10, 14], where inertial and friction parameters were optimized and identified. The dynamics of the XY-6 DoF simulator can be represented as follows:

$$\Gamma = {}^{b}J_p^T\left({}^{b}F_e + {}^{b}F_p\right) + \sum_{i=1}^{6}\left(\frac{\partial \dot{q}_i}{\partial \dot{q}_a}\right) H_i(q,\dot{q},\ddot{q}), \quad (1)$$

where Γ is the vector of the actuator torques, ${}^{b}J_p^T$ is the Jacobian matrix, ${}^{b}F_e$ is the external wrench exerted on the platform, $\dot{q}_i$is the passive joints speeds, $\dot{q}_a$is the active joints speed, $H_i(q,\dot{q},\ddot{q})$ is the inverse dynamic model of leg i considering as serial robot, ${}^{b}F_p$ is the total forces, is calculated by the following equation:

$${}^{b}F_p = {}^{b}\Pi_p\ {}^{b}\dot{v}_p + \begin{bmatrix} {}^{b}w_p \times ({}^{b}w_p \times {}^{0}MS_p) \\ {}^{b}w_p \times ({}^{b}I_p\ {}^{b}w_p) \end{bmatrix} - \begin{bmatrix} M_p I_3 \\ {}^{b}M\hat{S}_p \end{bmatrix} {}^{b}g \quad (2)$$

where M_p, ${}^{b}\mathrm{I}_p$, ${}^{b}\Pi_p$, ${}^{0}MS_p$ are the mass, inertia tensor, spatial interia matrix moment of inertia vector of the platform, respectively.

Parallel and serial robots are a known challenge which is the large number of dynamic parameters, which requires a lot of processing time, whether in dynamic modeling, identification or control. Therefore, the dynamic parameters of the XY-6 DoF hybrid structure simulator have been optimized and identified in previous works [14]. As a result, the standard 210 inertial parameters are reduced to only 90 basic inertial parameters.

The motion of the XY-6 DoF simulator motion is validated by the acceleration data from the NAO IMU and the robot model, as shown in Fig. 5.

Many works [15, 16] highlight interactions in multi-modal robotic systems, such as human-robot and multi-robot interactions, to improve the collaboration and/or the impact between systems. First, the integrated system applies a validation process to the platform motion management algorithms to improve the reliability of virtual trajectory recovery. It provides a reliable solution to assess the motion control of the XY-Gought-Stewart platform and multi-robot interaction.

The interaction evaluation will be done through the inertial and articular information coming from NAO. Afterward, a human motion perception model will be integrated into the system to improve the emulation of human sensory capabilities within NAO.

Fig. 6 shows the homothety between the present integration of multi-robot interaction and the objective integration of robot-robot human interaction (outlined in red). This future integration will provide a tool for measuring the quality of immersion for simulators by estimating through sensor data the impact of the movements of the virtual scenario and the platform on a human.

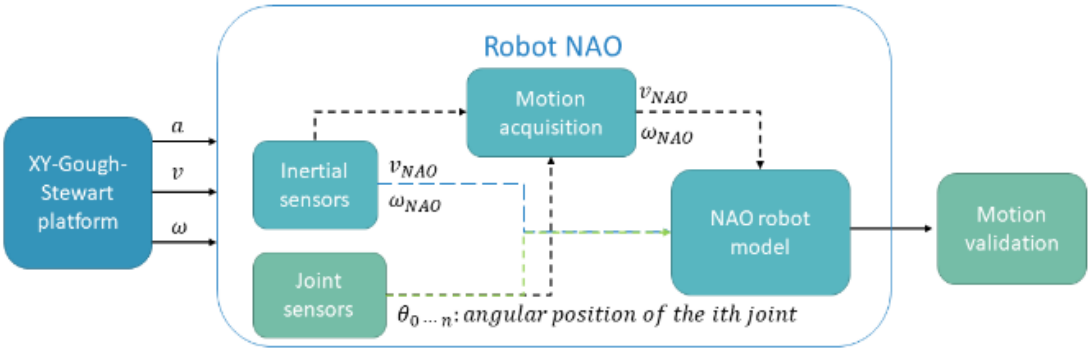


Fig. 5. XY-6 DoF simulator/NAO robot interaction.

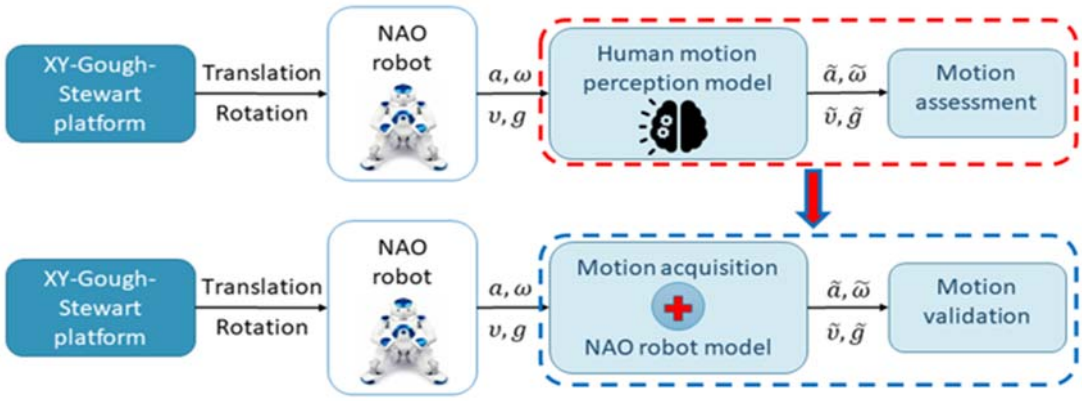


Fig. 6. Multi-robot and robotic human interaction.

4. Experimental Results

The validation of the platform controlled motion is illustrated in the following figures through the study of multi-robot interaction. In this work, we have neglected the Z-axis and yaw rotation as they have a minor effect on the ski motion sensation comparing to the X and Y axes and roll and pitch rotation. Translation motions on axis X and Y in Figures 7 and 8 are oscillatory movements of one meter in amplitude. Similarly, the rotational motions illustrated in Figures 9 and 10 are oscillatory angular movements from -10 to 0 degrees. The platform behaviors are obtained by extrapolating the platform position from XY motors through the kinematic model.

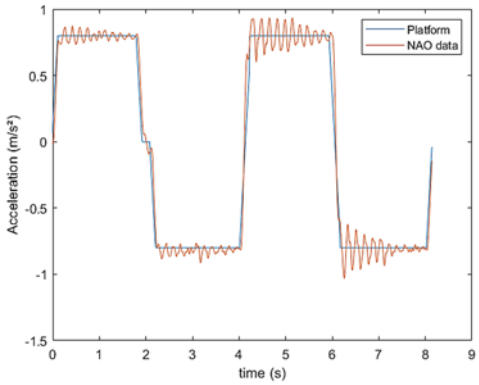


Fig. 7. Motion validation for X-axis movement.

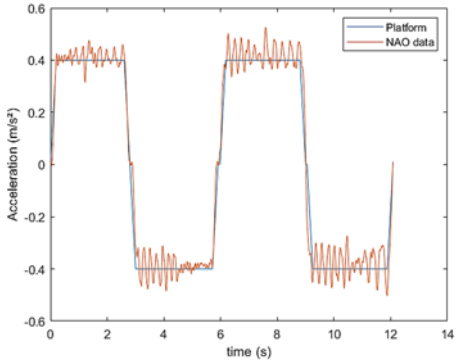


Fig. 8. Motion validation for Y-axis movement.

Observing the NAO robot's dynamic behavior in Fig. 7 and 8 validates the platform's accurate behavior tracking for an axis movement. However, some oscillations are observed for high dynamics. They result from slight platform vibrations due to axis friction and IMU acquisition noise.

Rotation motions do not suffer from axis friction. Thus, the angular speed behavior tracking is also precise but also less noisy.

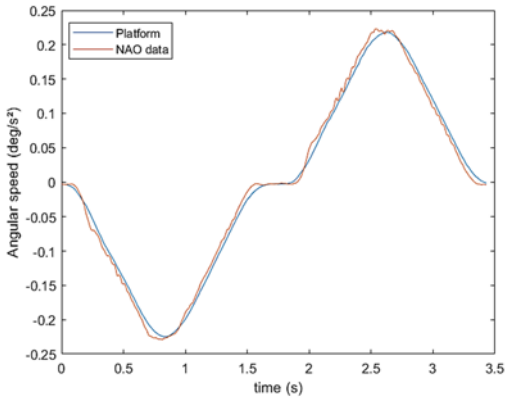


Fig. 9. Motion validation for Pitch movement.

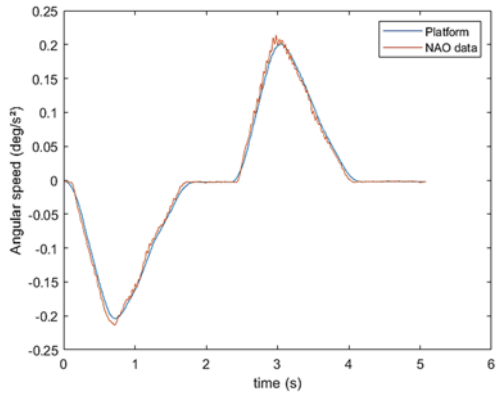


Fig. 10. Motion validation for Roll movement.

This study validates the ability of NAO to perceive the multi-robot interaction and the interest in integrating it into our system for future work.

5. Conclusions

In this paper, we have presented an original robotic platform system. The final goal is an operating system for a better immersive sliding simulator.

As a first step, we validated the response with a device like the Nao humanoid robot, which is a real human homothetic system. The embedded sensors give us all the information we need to properly assess the robot-robot interaction and system behavior.

Experimenting with all the systems gives us realistic feedback. According to our research plan, the global behavior is in accordance with our goals.

The obtained result is realistic and follows the acceleration shape. As we can see in Fig.7 to Fig.10, the response is very close to the real motion.

The direction of our future work is to manage all different parts of the system to achieve human perception integration and realize a better immersive way.

References

[1]. Casas, Sergio, Ricardo Olanda and Nilanjan Dey, Motion Cueing Algorithms: A Review, Algorithms, Evaluation and Tuning, *International Journal of Virtual and Augmented Reality (IJVAR),* 1, 1, 2017.

[2]. Petrović, Đorđe, Radomir Mijailović and Dalibor Pešić, How to Improve the Inclusion of Drivers with Disabilities: Measures to Enhance Accessibility, Mobility, and Road Safety, *Journal of Transportation Engineering, Part A: Systems*, 148, 10, 2022.

[3]. Houda, Taha, Lotfi Beji, Ali Amouri and Malik Mallem. Dynamic Behavior of an Interactive XY - 6 DoF Simulator for People with Reduced Mobility, in *Proceedings of the IEEE Conference on Control Technology and Applications* (CCTA' 2020), 2020, pp. 522-527.

[4]. Nahon, Meyer A. and Lloyd D. Reid, Simulator motion-drive algorithms - A designer's perspective. *Journal of Guidance Control and Dynamics,* 13, 1990, pp. 356-362.

[5]. Pan, C. T., Pei-Yuan Sun, Hao Li, Cheng-Hsuan Hsieh, Zheng-Yu Hoe and Yow-Ling Shiue. Development of Multi-Axis Crank Linkage Motion System for Synchronized Flight Simulation with VR Immersion. *Applied Sciences,* 11, 2021, p. 3596.

[6]. Wei, Ming-Yen, Design and Implementation of Inverse Kinematics and Motion Monitoring System for 6DoF Platform, *Applied Sciences,* 11, 19, 2021, 9330.

[7]. Reymond, Gilles and Andras Kemeny. Motion Cueing in the Renault Driving Simulator, *Vehicle System Dynamics,* 34, 2000, pp. 249 - 259.

[8]. Salaani, Mohamed Kamel, Dennis Guenther and Gary Heydinger. Vehicle dynamics modeling for the national advanced driving simulator of a 1997 jeep Cherokee, *SAE Transactions,* 108, 1999, pp. 443-465.

[9]. Miunske, Tobias, Christian Holzapfel, Edwin Baumgartner and Hans-Christian Reuss. A new Approach for an Adaptive Linear Quadratic Regulated Motion Cueing Algorithm for an 8 DoF Full Motion Driving Simulator, in *Proceedings of the International Conference on Robotics and Automation (ICRA' 2019),* 2019, pp. 497-503.

[10]. Houda, Taha, Lotfi Beji, Ali Amouri and Malik Mallem. Handiski simulator performance under PSO-based washout and control parameters optimization. *Nonlinear Dynamics,* 110, 2022, pp. 649 - 667.

[11]. Amouri, Ali and Fakhreddine Ababsa. Sliding Movement Platform for Mixed Reality Application, *IFAC-PapersOnLine*, 49, 2016, pp. 662-667.

[12]. Gouaillier, David, Vincent Hugel, Pierre Blazevic, Chris Kilner, Jérôme Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre and Bruno Maisonnier. Mechatronic design of NAO humanoid, in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2009, pp. 769-774.

[13]. Kofinas, Nikolaos, Emmanouil Orfanoudakis and Michail G. Lagoudakis. Complete Analytical Forward and Inverse Kinematics for the NAO Humanoid Robot, *Journal of Intelligent & Robotic Systems*, 77, 2015, pp. 251-264.

[14]. Houda, Taha, Ali Amouri, Lotfi Beji and Malik Mallem. Dynamic Parameters Optimization and Identification of a Parallel Robot, *Multibody Dynamics 2019*, 53, 2019, pp.367-374.

[15]. Darmanin, Rachael N. and Marvin K. Bugeja. A review on multi-robot systems categorised by application domain, in *Proceedings of the 25th Mediterranean Conference on Control and Automation (MED' 2017)*, 2017, pp. 701-706.

[16]. Sharkawy, Abdel-Nasser. Human-Robot Interaction: Applications, in *Proceedings of the 1st IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0,* February 2021, pp. 98-103.

(4005)

Software in the Loop Simulation of Robotic Systems

Dimitrios Chasapis [1], G. Kalogiannis [2], S.Sirianidou [2] and G. Hassapis [2]
[1] Barcelona Supercomputing Centre, Group SONAR, Plaça Eusebi Guell 1-3,
08034, Barcelona, Spain
[2] Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki,
54124 Thessaloniki, Greece
Tel.: +30 2310996324
E-mail: ghass@ece.auth.gr

Summary: One of the development phases of modern automation systems is the evaluation and testing of the software that implements all their functions. To this end, it is required to create a platform which will provide the necessary environment for writing the system software and evaluating the correctness of this software. In this paper such a platform is proposed consisting of a robotic system simulator linked with an open-source software engineering environment in which one can write control software for a robotic system according to the IEC 61499 standard, when the runtime of the robotic system is a real time operating system. The case study of the control of a rehabilitation device is used to demonstrate the use of this platform and assess its feasibility and efficiency.

Keywords: Software engineering, Robotics, Automation system simulation, IEC61499 standard.

1. Introduction

Software engineering tools for automation systems are under continuous development over the last 15 years. Mainly they address the programming of computer systems build with real time operating systems. Software engineering tools were initially provided by the manufacturer of the real time computer system, but the publication of the IEC 61131-3 [1] and the IEC 61499 [2] standards allowed third parties to develop very powerful workbenches allowing software development for automation in graphical languages for any computer make and making in this way modular and transferable software from any type of computer to another. Programming languages of the IEC61131-3 and IEC 61499 standards are based on Function Bock Diagrams (FBD) and Sequential Function Charts, concepts very close to those used by a control engineer rather than a computer scientist. Especially, the type of Function Blocks proposed in the IEC 61499 standard can be easily used to program the required functions for Industry 4.0 components [3]. However, there is always the need of testing and verifying the correctness of the software. The best way to do this is by connecting the actual real time computer system with the physical system, which may be a robotic arm, a chemical process plant, or an entire manufacturing facility and run the appropriate tests. This is not always possible as issues related with plant operation safety are raised in case that the software under development is malfunctioning. The next closest approach to the real plant operation under the control of the developed software is to connect the real time computer running the developed automation software with a validated simulator of the plant and run all the tests almost in real time. Usually, this kind of simulation is called software in the loop simulation. Representative work on software in the loop simulators for robotic systems is reported in [4, 5] and [6]. The approach taken in the first reference is to develop software in typical programming languages such as C^+ and perform a simulation of the specific robotic system under the control of the developed software also in a typical programming language, whereas in the second reference the MATLAB suite is used to model the physical robotic system and link it with the control software also developed in a typical programming language. The third reference utilizes the ROS operating system [7] and the GAZEBO [8] robotic simulator

The proposed in this article work extends the programming and simulation environments of the above mentioned work by forming a platform in which the programming of the control software is carried out in the high level graphical languages recommended by the IEC 61499 standard which are appropriate for designing block diagrams of control systems. Then, the developed programs are loaded and run under the control of the same real time operating system used in robotic systems, which is the ROS operating system. The computer running the written control software under the control of ROS is linked with the GAZEBO simulator of robotic systems which executes the robotic system functions that are defined by the written control software. This approach is closest to the real world situation.

To make this platform operational the necessary interfaces between a workbench for writing programs in IEC 61499 languages with the ROS operating system were constructed and the computer running the robotic system control software under the ROS operating system was linked with the GAZEBO robotic system simulator, In specific the 4DIAC workbench [9] for writing programs in IECE61499 languages was interfaced with the ROS operating system.

The platform was used to demonstrate how one can write and evaluate the correct functioning of software written to control a rehabilitation device, namely the ARTROMOT- E2 [10] device for exercising the muscles of a human hand.

2. The Platform Components

2.1. The Robotic Simulator

The GAZEBO robotic simulator provides a modular tool for modeling joint rotations and lateral movements of multiple robotic arms in a defined space where gravity, floor characteristics and other parameters are considered. Shapes and graphical library components can be used to describe the robotic arms and their joints along with their dynamic characteristics. Control inputs to the drives of the joints allow the user to perform rotational and lateral actions on any defined mechanical model of a robot system by applying static or time algorithms and monitor the robot responses.

2.2. The IEC 61499 Standard and the 4DIAC Tool

Any robotic system with a real time operating system will be controlled by one or more computer systems which will run the control software. Instead of writing routines for the real time operating system in low level C language, a more efficient way and in line with current practice of control engineering is to write the software according to the recommendations of the IEC 61499 standard. A tool which allows writing programs according to the recommendations of IEC 61499 is the 4DIAC workbench. The basic programming element of 4DIAC is the Function Block Diagram, defined as an event driven programming component that contains one or more control algorithms, event and data inputs and outputs and event-based algorithm execution control. In Fig.1 the typical form of the graphical programming symbol is depicted.

Data are entered from the left side of the graph whereas their reading by the encapsulated algorithm depends on the occurrence of events on the respective event input, as shown in Fig. 1.

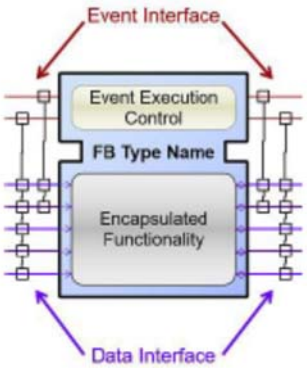


Fig. 1. Typical form of a Function Block.

A program can be a network of standard and custom-made function blocks. Also, service function blocks are offered for transferring data over a computer network.

3. The ROS Operating System

Tools for implementing rotation and lateral movement algorithms are provided by the ROS robotic operating system. It can accept software programs in C^{++}, library components and data sets organized as nodes of a network of software entities. A robotic system might contain many nodes, i.e., one node for the control of a sensor, a second node for the control of a joint motor and a third node to make the trajectory plan.

4. The Software in the Loop Simulation Platform

The proposed platform for assessing the correct operation of the control software, written according to the IEC 61499 recommendations, is shown graphically in Fig. 2.

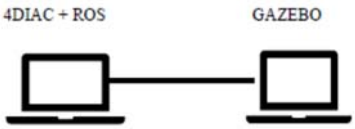


Fig. 2. Software in the loop simulation platform

It consists of two networked computers, one running the 4DIAC workbench and the ROS operating system and the other the GAZEBO robotic simulator. Interfaces have been built between the 4DIAC and the ROS systems and between the ROS system and GAZEBO.

5. The Software in the Loop Simulation of a Rehabilitation Device

A case study is presented in this section which demonstrates how to access the correct functioning of software written to control an exoskeleton arm, namely the ARTROMOT E2 device. Fig. 3 is a picture of the ARTROMOT device.

Fig. 3. The ARTROMOT E2 exoskeleton device for exercising hand muscles.

As Fig. 3 shows the device consists of a chair and an arm on which the hand of the patient is tied the way shown in Fig. 4.

Fig. 4. ARTROBOT E2 trajectory of hand movements.

The exoskeleton arm is used for exercising the muscles of the human arm by rotating at different speeds the wrist and moving up and down the arm around the elbow. By using the platform one can verify that the written control software rotates the hand at angles defined by the trainer and the repetition of the movements are in compliance with the set values

The model of the moving parts of the ARTROMOT E2, shown in Fig. 5, were created in the GAZEBO simulator. The parts were graphically drawn as AUTOCAD suites and were converted to .STL files, a format supported by GAZEBO.

Fig. 5. Basic moving parts of the ARTROBOT E2.

In order to simulate the combined dynamic movement of these parts, three connectors, according to the terminology used in GAZEBO, have been created, one between the two moving parts and one between a moving part and a steady base, which may be the handle of the chair. The third connector stabilizes the model within its operating environment. Fig. 6 shows the three connectors and their rotational axis.

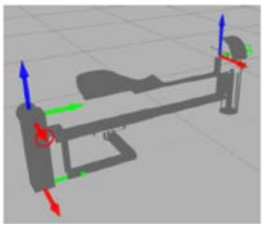


Fig. 6. The GAZEBO model of the exoskeleton arm.

In the formed file, information has been inserted for the package concerning the gazebo-ros control used to move the arm as well as for the virtual actuators which transmit the movement to the connectors. The connector movements are controlled by PID algorithms offered as plugins of the ROS operating system. The exchange of data between the controllers and the connectors are programmed in the DIAC environment and the PUBLISH/SUBSCRIEB data exchange protocol is used to input to the controllers the angle set points of the arm gazebo model and transmit the computed outputs of the controllers to the connectors of the gazebo model which adjust the angles of the arm.

6. The Simulation Results

The following simulation runs have been conducted for different values of human hand mass, speed of movement and angles of elbow and wrist rotation. .

Table 1 shows the provided by the simulator angle of the elbow movement against the requested by the trainer angle value for different hand mass. The observed angle deviation (error) from the requested value is also shown in Table 2.

Table 2 shows the provided by the simulator angle of wrist rotation again for the same hand masses given in Table 1. Table 3 shows the provided by the simulator angles for the elbow and wrist rotation when the speed of movement changes.

Table 1. Movement of connector 1: elbow angle.

Hand Mass (kg)	Desired angle (rad}	Simulation angle (rad)	Error (rad)
2	1.0	0.942228498198	0.057771501802
	1.5	1.49493339784	0.0050666021613
	2.5	2.44346002106	-0.05653997894
	1.0	0.942228498272	0.0577715017281
3.1	1.0	0.924448692776	0.0755513072236
	1.5	1.49333220467	0.00666779533258
	2.5	2.44346006656	0.05653993344
	1.0	0.924448692709	0.0755513072913
4.3	1.0	0.904022403439	0.095977596561
	1.5	1.49146367202	0.0085363279817
	2.5	2.44346015244	0.00000015244
	1.0	0.904022403948	0.0959775960519

Table 2. Movement of connector 2: wrist angle.

Hand Mass (kg)	Desired angle (rad}	Simulation angle (rad)	Error (rad)
2	1.0	0.999999999724	0.000000000276
	2.0	2.00000000031	-0.00000000031
	3.14	3.14000000019	-0.00000000019
	1.0	0.999999999588	0.000000000412
3.1	1.0	0.999999999351	0.000000000649
	2.0	1.99999999935	0.00000000065
	3.14	3.13999997374	0.00000002626
	1.0	0.924448692709	0.0755513072913
4.3	1.0	0.9999999985030.	0.000000001497
	2.0	1.9999999985	0.0000000015
	3.14	3.13999996788	0.00000003212
	1.0	1.0000000015	-0.0000000015

Table 3. Elbow and wrist angles for different speed of movement

Connection	Speed (rad/sec)	Desired angle (rad)	Simulation angle (rad)	Error (Rad)
1(elbow)	100	1.0	0.999999999724	0.0755513073802
	50	1.0	0.92444869267	0.0755513073797
	10	1.0	0.924448692718	0.0755513072823
2(wrist)	100	1.0	1.00000000065	-0.00000000065
	50	1.0	1.00000000065	-0.00000000065
	10	1.0	1.00000000046	-0.00000000046

7. Conclusions

In this work a platform for writing software for the control of robotic systems and the testing of the functioning of this software according to desired specifications has been created. This platform allows the writing of the software according to the IEC61499 standard by using the 4DIAC tool running it under the control of the ROS real time operating system and evaluating its operation by linking the ROS-DIAC modules with the GAZEBO robotic simulator. By using a validated robotic system model one can monitor the robotic variables related with the joint rotation and arm displacements. When the simulation results comply with the desired values one can conclude that the control software is functioning correctly.

A case study concerning the control of a device used to exercise the muscles of a human hand has been studied demonstrating in this way he feasibility of constructing such a platform.

References

[1]. Sysl-Heinz John and Michael Tiegelkamp, IEC 6113: Programming Industrial Automation Systems, *Springer-Verlag*, 1995.

[2]. Aloi Zoitl &Robert Lewis, Modeling Control of System Using IEC61499, *The Institution of Engineering and Technology,* London 2014.

[3]. F. Zesulka, P. Macron, I. Vesely, O. Sajdl, Industry 4.0 - An Introduction to the Phenomenon, *IFAC-PapersOnLine,* Vol. 49, Issue 25, 2016, pp. 8-12.

[4] Mossaad Ben Ayed, Lilia Zunari, Mohamed Abid, Software In the Loop Simulation for Robot Manipulators, *Engineering Technology and Applied Science Research*, Vol. 7, No 5, 2017, pp. 2017-2021.

[5]. Giuseppe Silano and Luigi Ianneli, CrazyS: A software -in-the-Loop Simulation Platform for the Crazyflie 2.0 Nano - Quadcopter, *Studies in Computational Intelligence book series (SCI),* Vol. 831, 29 June, 2019.

[6]. M. F. Syed Ahamed, G. Tewolde, and J. Kwon, Software - in-the-Loop Modeling and Framework for Autonomous Vehicles, in *Proceedings of the International Conference on Electro/Information Technology (EIT),* Rochester, MI, USA, 2018, pp. 0305-0310.

[7] ROS Web Portal: http://www.ros.org/Eclipse 4
DIAC Web Portal: http://www.eclipse.org/4diac/i

[8]. Control plugin
http://gazebosim.org/tutorials?tut=guided_i5

[9] Eclipse 4 DIAC Web Portal:
http://www.eclipse.org/4diac/index.plp

[10]. ARTROMOT-E2 compact user manual, ORMED GmbH & Co. KG.

(4123)

IOTA Data Preservation Implementation for Industrial Automation and Control Systems

I. C. Lin [1], P. C. Tseng [2] Y. S. Chang [1] and T. C. Weng [2]
[1] National Chung Hsing University, 145 Xingda Road, Taichung, Taiwan
[2] Ph.D. Program of Business, Feng Chia University, Taichung, Taiwan
Tel.: +886 422840864, fax: +886 422857173
E-mail: iclin@nchu.edu.tw

Summary: The blockchain 3.0 currently under development has a wider range of applications, including identity authentication, logistics, medical care, industry 4.0 / 5.0, etc. The combination of blockchain and the Industrial Automation and Control Systems is also an important application at this stage. With the popularity of the industrial automation and control systems and 5G networks, the Internet of Things devices can be regarded as a node in the blockchain network, and nodes can communicate and verify each other through the blockchain network, thereby achieving a completely decentralized network. The paper will show the implementation and execution results of data preservation from industrial automation and control systems to IOTA.

Keywords: IOTA, Industry 4.0 / 5.0, Blockchain, Data preservation, Industrial automation and control systems.

1. Introduction

The application of the Industry 4.0 / 5.0 is growing exponentially in various fields. When the Internet of Things devices are built, a large amount of data will be generated. These data will be concentrated on the client server or the cloud server. Such a centralized infrastructure will produce a single point of failure and a lack of trust between devices [1]. A distributed system based on blockchain is one of the solutions. The application of blockchain in the Industry 4.0 / 5.0 has been very diverse [2]. Datta et al. [3] established a security scheme that allows information and energy transactions between vehicles to be stored in encrypted records; Hang et al .[4] used Hyperledger Fabric to build a fish farm platform to ensure data integrity; Guan et al [5] designed a two-tier distributed energy transaction to protect transaction information through the blockchain consensus mechanism; Grecuccio et al. [6] used IoT devices to record the process of the food supply chain, and established an Ethereum node as a gateway to broadcast the data to the smart contract of the blockchain network for storage.

However, there are still many problems in the application of blockchain in the Industry 4.0 / 5.0. For example, sensor data cannot be monitored for a long time. Due to handling fees and scalability issues, it will cost a lot of costs to import the blockchain into the Internet of Things. Although there have been many results using smart contracts in the supply chain, the history data of the Internet of Things cannot be recorded in the blockchain, and the data transmission process from the perception layer of the Internet of Things to the network layer is not open and transparent. In order to solve the problems of scalability, efficiency, and handling fees, Popov proposes a distributed ledger Tangle based on DAG (Directed Acyclic Graph, DAG) [7]. Different from the general blockchain ledger which is packaged into a chain-like structure, Tangle uses the DAG structure as the basis of the ledger, which can increase the speed of transactions and can completely record the history of the Industry 4.0 / 5.0.

According to Ericsson's report [8], by 2021, about 28 billion smart devices will be connected globally. In addition, more than 15 billion devices have adopted machine-to-machine (M2M) communication [8]. With the establishment of the Internet of Things in life, WSNs (Wireless Sensor Networks) with low energy consumption, small size and convenient deployment have become the key to the application of the Industry 4.0 / 5.0. In general wireless sensing network system architecture, a large number of sensors are scattered in the area to be sensed to collect various environmental data, and then the data is transmitted back to the manager or user through the data collector. Since the sensors do not know the existence of themselves and other sensors, they can only be managed through a centralized server or cluster head. How to manage, identify, and encrypt connections is a problem worthy of discussion. Ensure the confidentiality and integrity of the data from the collection end to the user end.

2. Related Works

2.1. IOTA

IOTA was officially launched around 2018, which is an open source decentralized ledger technology, and runs on a peer-to-peer network. An IOTA account is a way to prove the ownership of transactions in the Tangle, like a bank account, but the difference is that instead of creating an account through a name and password, it uses a seed to create it. Unlike banks, Seed will not be accessed by anyone other than itself. Seed represents an identity on the Internet. This approach

can decentralize IOTA and maintain anonymity. IOTA uses WOTS (Winternitz One Time Signature) to generate addresses. Seed is the master key of the account. Many private keys can be generated through Seed. Each private key can generate an address, but this address can only be used once. Each address can contain any amount of IOTA currency, and the total amount of the account is the sum of the currencies of all addresses in the account.

The seed is the only master key that can prove the ownership of any IOTA currency in the message or address. The seed is a string of 81 Trytes composed of 26 uppercase English letters and the number 9. The number of seeds is very huge, there can be a total of 8.7×10^115 seeds. Therefore, the probability of two identical seeds is very low. A seed can generate different private keys according to different index and security levels, and each private key can generate an address. There are three security levels, which correspond to different private key lengths. The higher the security level, the longer the private key length and the less likely to be stolen. Security level 1 is relatively simple, suitable for storing low-value IoT device data. Security level 2 is used for wallet transactions and high-value IoT devices. Security level 3 is used for transactions that require high security such as exchanges.

There are five main steps in the transaction sending process, which are described in detail as follows:

1. Generate transaction information:

The first step is to create a single transaction. You must specify Address and Value for each transaction. You can also define a Tag to classify different transactions. Message is the message content of the transaction. In zero-value transactions, Message is used to trace the root address, and Timestamp is automatically generated by IOTA. To send a valuable transaction, at least two IOTA transaction messages are required, one with a positive value to allow the receiver to obtain encrypted currency, and one with a negative value to allow the sender to deduct the transferred amount.

2. Packaged into Bundle:

Bundle is a collection of transactions. After the transaction information is generated, all transactions will be packaged into Bundles. The sum of the Value of all transactions in the Bundle must be zero, and the Bundle also has a unique address for querying transactions. After the packaging is complete, use the Seed to generate the private key, and use the private key to sign the Bundle.

3. Choose two Tips:

The POW calculation must be performed before the transaction is attached to the ledger. The object of performing POW is in the Tangle, using Markov chain Monte Carlo (MCMC) to randomly select two tips, called Branch and Trunk, and then add the Hash of the two Tips to the Bundle.

4. Do proof of work:

Perform POW (proof of work) operation on the selected two Tips, and then append the calculated Nonce value to the Bundle.

5. Broadcast to the Tangle network:

After verifying other transactions, the Bundle is broadcast to the Tangle for storage, and the new transaction is Tips, waiting to be verified by others.

2.2. Signature and Verification

As shown in Fig. 1, the signature is generated when the Bundle is generated, the Bundle is signed to verify the ownership, and finally the signature is appended to the Signature Fragment in the Bundle. In order to ensure the security and validity of the signature, first obtain the Hash value of the Bundle and obtain the private key. The private key is generated from Seed, generated using WOTS. There are different lengths according to the security. In the case of the lowest security level, the length of the private key is 27 x 81 Trytes. The private key is formed into a segment for every 81 Trytes, and hash operations are performed on them with unequal times. The number of times comes from the Hash value of the Bundle. After converting the Bundle Hash to the last decimal, an integer between -13 and 13 can be obtained. Then subtract this integer from 13, and then hash the obtained values to the segments, and the integration is the signature. It can also be known that the higher the security level, the longer the private key and the longer the signature. As shown in Fig. 5, the first and second Tryte of Signed Data (Normalized Bundle Hash) are L and W, and the converted decimal is 12 and -4. Then the values subtracted by 13 are 1 and 17, so Segments 1 and 2 perform 1 and 17 sponge function calculations respectively. After all the values are combined, it is the signature.

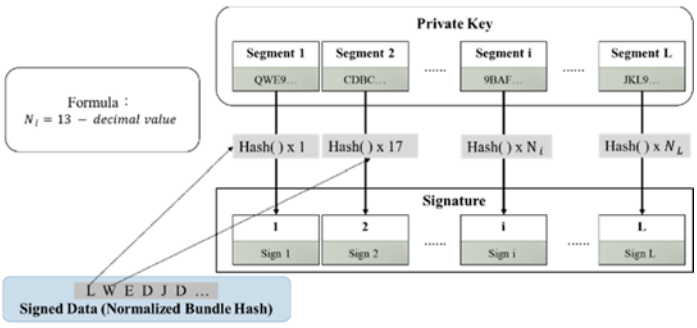


Fig. 1. Transaction signature.

The node verifies the signature in the transaction through the Bundle Hash and address. The method is similar to the signature. First get the value of Bundle Hash, and then convert it to decimal. Then add 13 to each Tryte value, and then each signature fragment in the signature corresponds to the result of the Bundle Hash operation, and the number of hash operations is different. Finally, combine the hash values of the signature fragments and perform two hash operations to get the transaction address. Verify that the address of the Bundle is consistent with this address to know whether the signature is correct. Because the address generation method is also WOTS, in the step of the private key, the segment is subjected to 26 sponge function operations, and the 13-d (decimal value) operation is performed when signing, and 13 + d (decimal value) is performed again during verification. Value), the result of 13 - d + 13 + d is 26 sponge operations. This method can achieve the purpose of signature verification without leaking the private key and seed. As shown in Fig. 2, the decimal places of the first two Trytes of Signed Data are 12 and -4, and the values after adding 13 are 25 and 9. After 25 and 9 operations, the private values of Segment 1'and Segment 2'are obtained. Key hash value. After all the fragments are merged and hashed, the address can be obtained. If the addresses match, the transaction is valid.

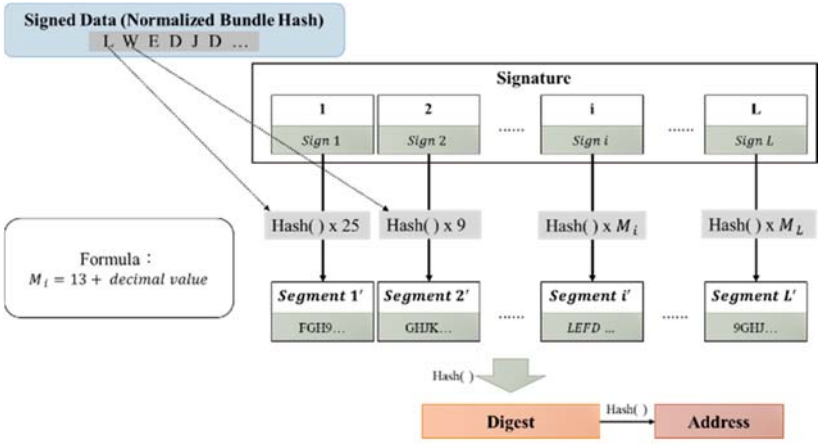


Fig. 2. Signature verification.

2.3. Proof of Work

Proof of work is a cryptographic certificate used to prove that the problem has been solved and the right to write into the ledger can be obtained. These problems are difficult to solve, but they can be easily verified for correctness and can be used to protect the network from spam attacks, because the attacks cost a certain amount of cost. IOTA's POW is not executed by miners, but is transferred to nodes for execution by the person who wants to trade. Any node must execute POW first in order to write into the ledger. As shown in Fig. 3, there is 81 Trytes in the transaction message that is the Nonce value. To execute POW, you must first fill in the Nonce value randomly, and then convert the transaction message including the Nonce into the Trits format and perform the hash operation. The Minimum Weight Magnitude (MWM) after the last mantissa must all be zero. The difficulty of MWM varies according to the network. In the main network, the MWM is 14, which means that the next 14 bits after the POW is executed must be zero, and the MWM in the test network is 9. When the node that initiated the transaction performs the POW and broadcasts it to other nodes for storage, the other nodes will throw the calculated Nonce value into the sponge function for calculation. If the calculation result meets the MWM, the transaction can be written into the Tangle.

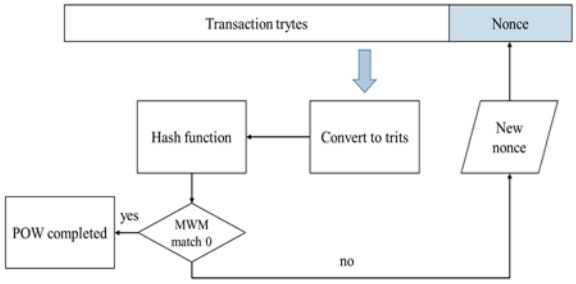


Fig. 3. Proof of work.

2.4. Channel

MAM's Channel is like streaming media. Publishers can publish messages regularly, viewers can receive messages through subscription, and only the owner can publish messages. In IOTA, this owner is the seed owner. If the Seed is stolen, the attacker can publish messages at will. There are three modes in Channel to control the message flow, namely Public, Private, Restricted. In the public mode, everyone can directly obtain the message content, while in the private message content, only the owner of the Seed can unlock the encrypted content. In Restricted mode, encrypted content can be unlocked by using a key called Sidekey. By sharing Sidekey, confidential information can be easily opened to specific people for

viewing. The address generation methods of the three modes are not the same. Although they all use the Merkle-tree Signature Scheme, they are different in the last Root. The detailed address is generated as follows:

1. Public: $Address = Root$.
2. Private: $Address = hash(Root)$, hidden messages are decrypted using Root.
3. Restricted: $Address = hash(Root)$, hidden messages are decrypted using sidekey.

2.5. Sponge Function

The sponge function is a cryptographic algorithm. It uses a limited state to receive input bit streams of any length. After the data is "absorbed" into the sponge, the desired result is "squeezed out", and then it can satisfy any Length of the output. Sponge function can be divided into two stages, namely Absorbing and Squeezing. The information is first input and compressed repeatedly, and then the result is repeatedly extruded. As shown in Fig. 4, the values M0~M3 need to be input, after inputting M0, go through the calculation of XOR and function, then input the value of M1, and repeat the calculation until all the values are counted. Then when the value is taken out, it will go through the function calculation again to get the value of Z0. If the length is not enough, continue the calculation until the required length is obtained.

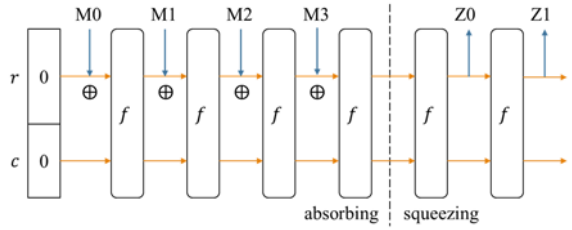


Fig. 4. Sponge function.

3. Implementation

The system designed in this paper is divided into three main members: Base Station, Cluster Head, and Sensor Node. The hardware configuration of the system is as follows.

3.1. Base Station

The base station is divided into two parts, namely Gateway and IOTA nodes. The Base Station runs on the Windows 10 operating system and is equipped with Intel's i7 processor and 16G of memory. The Proxy Server run by Gateway is written in node.js language, while IOTA nodes are set up using Hornet, and only the 15600 Port is open for external node synchronization. In the data storage part, use Chronicle to store the data in Scylla's NoSQL database.

Because MAM data upload must choose a website with SSL/TLS standards, the Base Station must apply for a Domain Name and generate a certificate. To set up a node, you must first download the public account snapshot from the IOTA website, then select a fixed node, and wait for the node to complete synchronization before uploading data to the node.

3.2. Cluster Head

Cluster Head uses a Raspberry Pi 4/8G device with a network card, as shown in Fig. 5. The operating system uses a Linux system, and uses the version of ubuntu-20.04.2-preinstalled-server-arm64+raspi.
Cluster Head also uses node.js to write, uses mosca's suite to run MQTT Server, and uses IOTA's node.js Client Library to encrypt and sign data.

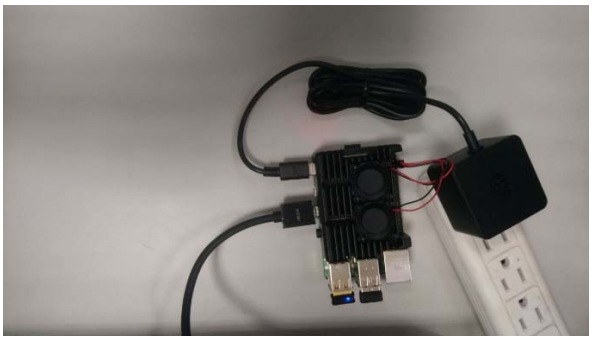


Fig. 5. Raspberry pi 4.

3.3. Sensor Node

Sensor Node uses the development version of Arduino Nano 33 IoT, with 256 KB of CPU Flash Memory and 32 KB of SRAM. And the part of the sensor uses RFID Reader of RC522, as shown in Fig. 6, uses C/C++ language to write the program.

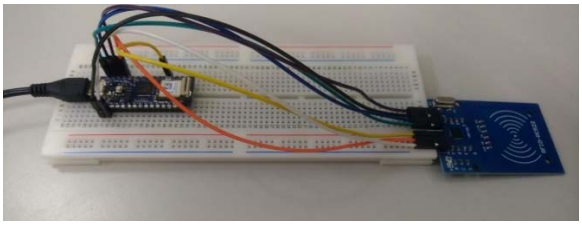


Fig. 6. Arduino nano 33 IoT.

3.4. Gateway

The Gateway in the Base Station is responsible for configuring new identity data for the new device, as shown in Algorithm 1 in Fig. 7. When a new device wants to join the network, it will first indicate whether it is Cluster Head or Sensor Node, and then obtain the latest data from Channel 3, and determine whether the new device is in the blacklist. If everything is correct, Gateway will configure a new AES key and ID for the new device, and obtain the whitelist of the network from Channel 2. Then append the ID of the new device to the whitelist and upload it to Channel 2.

Algorithm 1: New Device

Input: field, ch, information
Output: device_id, encryption key
$count = 2$
tag = SECONDCHANNEL
$mac \leftarrow get\ address\ from\ information$
$blacklist \leftarrow fetch\ mam\ data\ from\ channel\ 3$
If *mac in the blacklist* **then**
    $return\ failed$
else
    $fetchdata \leftarrow fetch\ mam\ data\ from\ channel\ 2$
    $ID \leftarrow generate\ random\ number\ for\ new\ device$
    If $field\ is$ cluster **then**
        $generate\ cluster\ key\ by\ hash(BSkey\ ||ID)$
    else if $field\ is\ sensor$ **then**
    $generate\ sensor\ key\ by\ hash(BSkey\ ||ch\ ||ID)$
    end if
    $channelState \leftarrow read\ channel\ detail\ from\ channel2.json$
    $channelState.count = count$
    $newdata \leftarrow pack\ ID\ to\ fetchdata$
    message $\leftarrow createmessage(channelState, newdata)$
    $mamAttach(message, 3, 14, tag)$
    $Strore\ the\ mam\ root\ to\ channel2.json$
end if
$return\ ID, key$

Fig. 7. New device pseudocode.

Assuming that the wireless sensor network has a detection tool for malicious devices, when a malicious device is detected, the Blacklist program can be called, as shown in Algorithm 2 in Fig. 8. Pass in the ID and MAC address of the malicious device, then get the latest blacklist from Channel 3, and get the latest network identity list from Channel 2. Then record the ID, MAC and time of the malicious device on the blacklist and upload it to Channel 3. Finally, the malicious device is removed from the list of Channel 2, and then re-uploaded to Channel 2, and the latest root is stored in channel2.json.

Algorithm 2: Blacklist

Input: id, mac
Output: root
$count = 3$
tag = THIRDCHANNEL
$blacklist \leftarrow fetch\ mam\ data\ from\ channel\ 3$
$fetchdata \leftarrow fetch\ mam\ data\ from\ channel\ 2$
$newlist \leftarrow pack\ id\ and\ mac\ to\ blacklist$
$channelState \leftarrow read\ channel\ detail\ from\ channel3.json$
$channelState.count = count$
$message \leftarrow createmessage(channelState, newlist)$
$mamAttach(message, 3, 14, tag)$
$Strore\ the\ mam\ root\ to\ channel3.json$
If $id\ is\ in\ fetchdata$ **then**
    $newdata \leftarrow remove\ id\ from\ fetchdata$
    $channel2 \leftarrow read\ channel\ detail\ from\ channel2.json$
    $channel2.count = 2$
    $msg \leftarrow createmessage(channelState, newdata)$
    $mamAttach(msg, 3, 14, "SECONDCHANNEL")$
    $Strore\ the\ mam\ root\ to\ channel2.json$
end if
$return\ mam\ root$

Fig. 8. Blacklist pseudocode.

As shown in Algorithm 3 in Fig. 9, when the Cluster Head wants to register the connection identity data, it will send the Cluster Head ID and encrypted data to the Gateway. Then Gateway will use the Cluster Head ID and its own key to perform MD5 operations, and use the first 16 Digest as the decryption key, and the last 16 Digest as the IV of the AES CBC. Then use the Key and IV to unlock the encrypted content to obtain the time stamp and identity ID. After the Gateway verifies that the Cluster Head identity is correct and the Channel 2 list exists, a temporary certificate and the validity of the certificate will be generated. Then use the key and IV of the Cluster Head to encrypt the certificate data and send it back to the Cluster Head.

Algorithm 3: Register

Input: CHID, encryptdata
Output: BSID, return_data
$return\_data$ = ""
$concat\_id \leftarrow BSkey\ concatenate\ with$ CHID
$SKey \leftarrow MD5\ hash(concat\_id\ )$
$key \leftarrow get\ first\ 16\ digest\ from\ SKey$
$iv \leftarrow get\ last\ 16\ digest\ from\ SKey$
$content \leftarrow AES\ CBC\ decrypt$(encryptdata,$key, iv$)
$TS1 \leftarrow content\ XOR\ CHID$
If $TS1\ is\ timeout$ **then**
    $return\ failed$
else
    $fetchdata \leftarrow fetch\ mam\ data\ from\ channel\ 2$
    If $CHID\ is\ not\ in\ fetchdata$ **then**
        $return\ failed$
    else
        $TE \leftarrow compute\ the$ overdue time
        $Pi \leftarrow MD5\ hash\ the\ value\ of\ (CHID||BSID||TE)$
        $TC \leftarrow MD5\ hash\ the\ value\ of\ (Pi||$BSkey)
        $phrace \leftarrow package\ TC, Pi, TE, and\ the\ timestamp\ TS2$
        $return\_data \leftarrow AES\ encrypt(phrase, key, iv)$
    end if
end if
$return\ return\_data$

Fig. 9. Registration pseudocode.

The logged-in Pseudocode is shown in Fig. 10. The Cluster Head will send the certificate-related data to the Gateway for identity verification, and at the same time send the first root data of the MAM and the decrypted sidekey to the Gateway. When Gateway receives the message, it will verify whether the certificate has expired and whether the Cluster Head has the correct certificate data. Then Gateway will generate the AES CBC Key and IV, and decrypt the encrypted content, and at the same time verify whether the timestamp is within a reasonable transmission time. Finally, the MAM information of the Cluster Head is recorded in Channel 1, which is used to unlock the MAM data stream uploaded by the Cluster Head. Then Gateway stores the latest root of Channel 1 in channel1.json, and records the IP, $sidekey,\ TE$ of the Cluster Head in whitelist.json.

In the part of secure communication, because the Cluster Head has been authenticated during the previous registration and login, the secure communication does not verify too much information. Only to confirm whether the IP of the Cluster Head is in the whitelist and whether the validity of the certificate of the Cluster Head has expired. If it is correct, it will be transferred to the IOTA node of the Base Station, as shown in Algorithm 5 in Fig. 11.

Algorithm 4: Login

Input: CHID, encryptdata, C, TE, R, P
Output: BSID, confirm
IP ← get ip address from TCP/IP
If *TE is timeout* **then**
    return failed
end if
P ← concatenate BSID, CHID, TE*
If *P ≠ P** **then**
    return failed
end if
Compute TC by MD5 hash(P ||BSkey)
C ← MD5 hash(CHID ||TC ||R)*
If *C ≠ C** **then**
    return failed
else
    concat_id ← BSkey concatenate with CHID
    SKey ← MD5 hash(concat_id)
    key ← get first 16 digest from SKey
    iv ← get last 16 digest from SKey
    content ← AES CBC decrypt(encryptdata,key, iv)
    TS3 ← get timestamp from content
    If *TS3 is timeout then*
        return failed
    else
        sidekey ← get sidekey from content
        root ← get first root from content
        *Store IP 、sidekey 、*TE *to whitelist.json*
        rootKey ← MD5 hash(CHID ||sidekey)
        mam publish(root, sidekey)
        Store mam publish root to channel1.json
    end if
end if
return confirm ← true

Fig. 10. Login pseudocode.

Algorithm 5: MAM publish

Input: mam transaction data
Output: forward transaction to hornet node
IP ← get ip address from TCP/IP
If *IP is not in the whitelist.json* **then**
    return failed
else
    TE ← get TE from the whitelist.json
    If *the TE is timeout* **then**
        return failed
    else
        Forward to 14265 port where the hornet node is located
    end if
end if

Fig. 11. Publish MAM pseudocode.

3.5. Raspberry pi

As shown in Fig. 12, Algorithm 6 is the Pseudocode of Cluster Head. The Cluster Head will receive the data passed by the Sensor Node, and use the ID of the Sensor Node and its own Key to perform the MD5 hash function calculation. Then the first 16 Digest of the generated value is used as the Key, and the last 16 Digest is used as the IV. Then the Key and IV are calculated by AES CBC to obtain the encrypted content of the Sensor Node. When it is confirmed that the encrypted content of the Sensor Node can be correctly unlocked. And the timestamp is in line with the upload delay, the Cluster Head will package the sensor data and send it to the 3000 Port of the Base Station, so that the Gateway will forward the packet to the IOTA node for upload. Finally, the Cluster Head will store the latest Root data in channel.json.

Algorithm 6: Cluster Head

Input: SID, mqtt_data
seed ← get mam seed from config
sidekey ← get mam key from config
TC ← get temporary confirm from storage
TE ← get time expired from storage
If *not register* **then**
    call base station port 3001 *to register*
end if
If *not login* **then**
    create new channel by using seed, sidekey
    create mamMessage by channelState, TC, TE
    call base station port 3002 *to login*
else
    If *TE is timeout* **then**
        login again
    else
        concat_id ← CHkey concatenate with SID
        SKey ← MD5 hash(concat_id)
        key ← get first 16 digest from SKey
        iv ← get last 16 digest from SKey
        content ← AES CBC decrypt(mqtt_data,key, iv)
        channel ← read channel detail from channel.json
        msg ← createmessage(channelState, content)
        mamAttach(msg, 3,14,*"CLUSTERHEAD")*
        Strore the mam root to channel.json
    end if
end if

Fig. 12. Cluster head pseudocode.

3.6 Arduino Nano

As shown in Fig. 13, Algorithm 7 is the upload process of the Sensor Node. The Sensor Node can collect data on a regular basis, but because the experimental device is an RFID Reader, the Sensor Node will passively obtain the data. When the Sensor Node obtains the data, it will obtain the Key and IV from the configuration file, and convert the sensor data to hexadecimal, and then encrypt it with AES CBC. Then it is sent to the Broker of Cluster Head through the MQTT protocol. After the Subscriber of the Cluster Head obtains the Sensor Node data from the Broker, the data will be packaged and uploaded to the Tangle ledger.

Algorithm 7: Sensor Node

Input: sensor data
Output: SID, mqtt_data
iv ← get AES iv from config
key ← get AES encryption key from config
If *WiFi is not connected* **then**
    reconnect()
end if
If *MQTT broker is not connected* **then**
    reconnect()
end if
While *read the RFID tag from reader* **do**
    hexdata ← change data to hex string
    TS4 ← get timestamp of sensor node
    encdata ← *AES CBC encrypt(TS4, hexdata, iv, key)*
    publish(SID, encdata) to cluster head
end while

Fig. 13. Sensor node pseudocode.

4. Execution Result

Fig. 14 shows the verification screen of Gateway's registration and login of the Cluster Head, and the successfully logged-in IP of the Cluster Head is displayed on the screen, and then the MAM message is forwarded to the IOTA node.

Cluster head execution and login screen shown In Fig. 15, the encrypted message generated when the Cluster Head is registered and logged in and the screen of the first MAM message stream sent after the login is successful.

The cluster head receives the message from the sensor node shown in Fig. 16, when the Cluster Head receives the sensor node message, it repackages the message and uploads it to the Tangle.

Sensor node reads sensor data shown in Fig. 17, after Arduino Nano receives the sensor data, it performs encryption and forwards the packet to Raspberry Pi via MQTT.

After the sensing data is successfully uploaded to the Tangle, you can use the IOTA Tangle Explorer to query the data recorded in the Tangle. Because the functions of the Base Station and the Cluster Head are different, the information recorded in the IOTA MAM is also different.

Fig. 18 shows the First Root, certificate information, certificate validity and hashed $sidekey$ of the Cluster Head in WSNs.

Fig. 19 shows the identity ID and sensor information of all devices in WSNs, as well as the relationship between Cluster Head and Sensor Node.

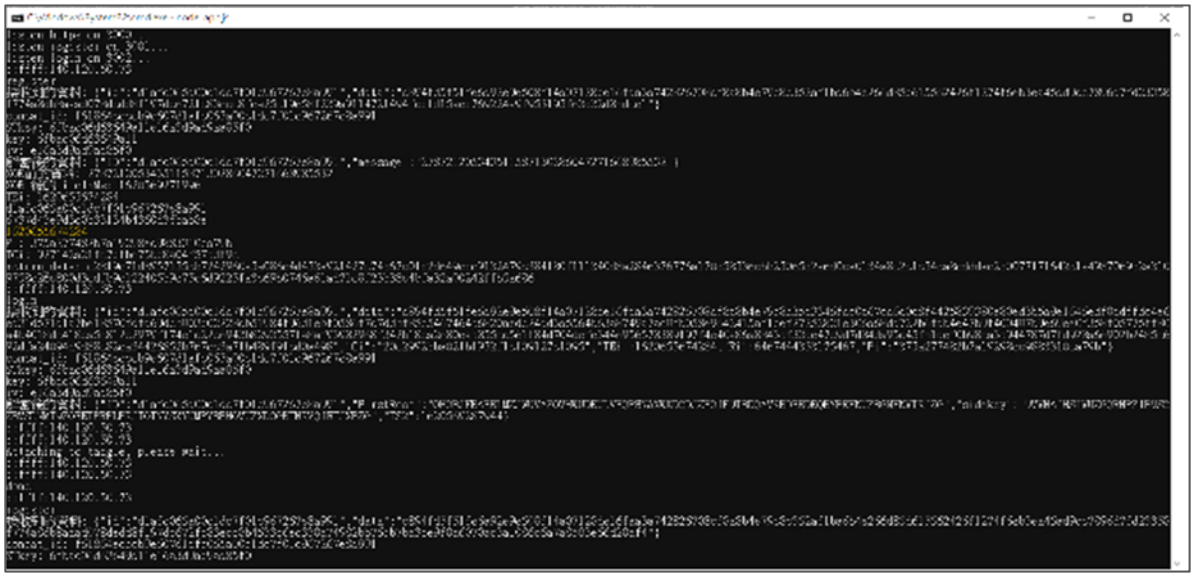


Fig. 14. Base station operation screen.

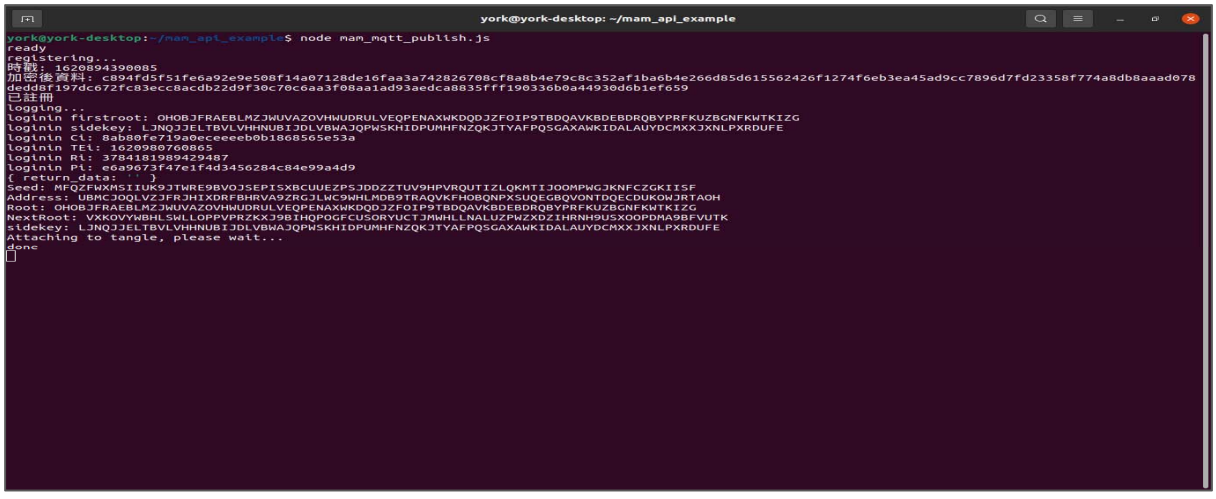


Fig. 15. Cluster head login and registration screen.

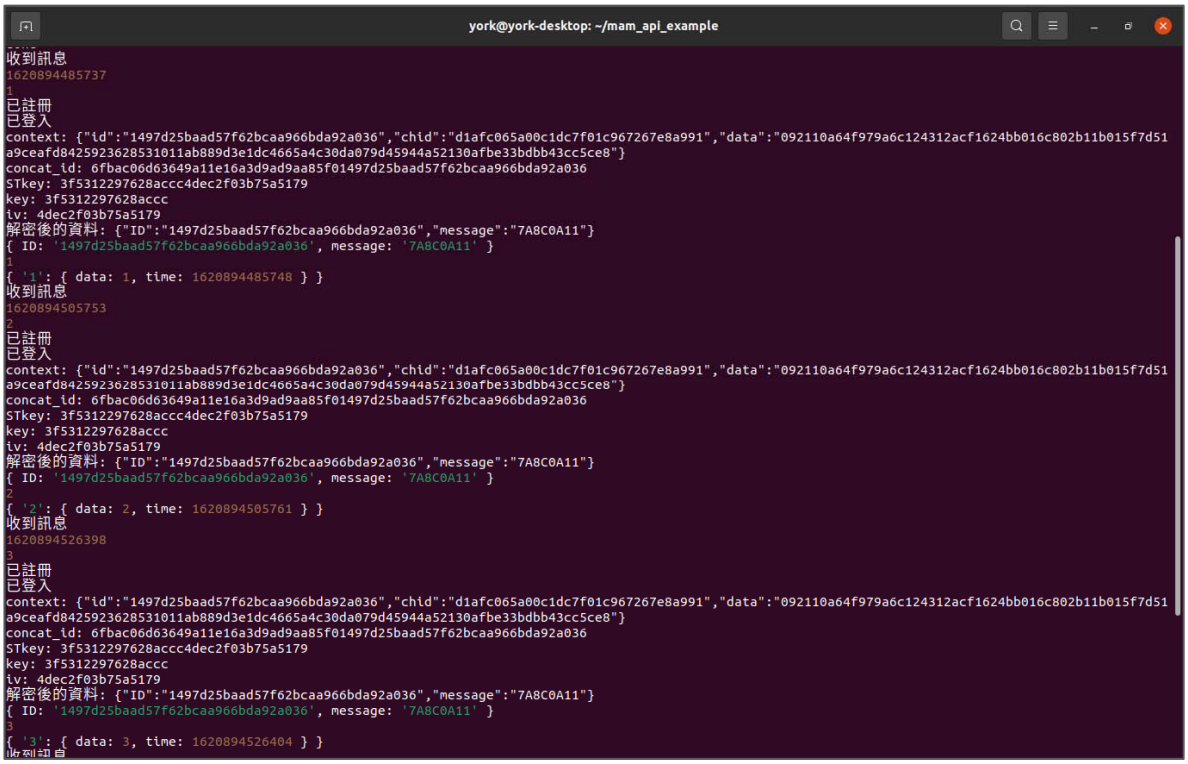


Fig. 16. The cluster head gets the packet screen.

Fig. 17. Sensor node upload package screen.

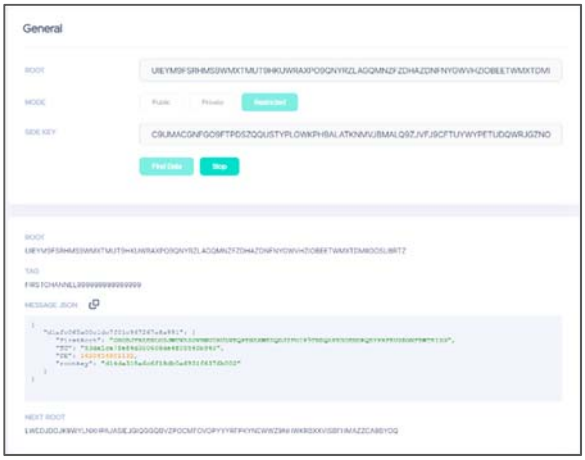


Fig. 18. The content of the message stored in the base station channel1.

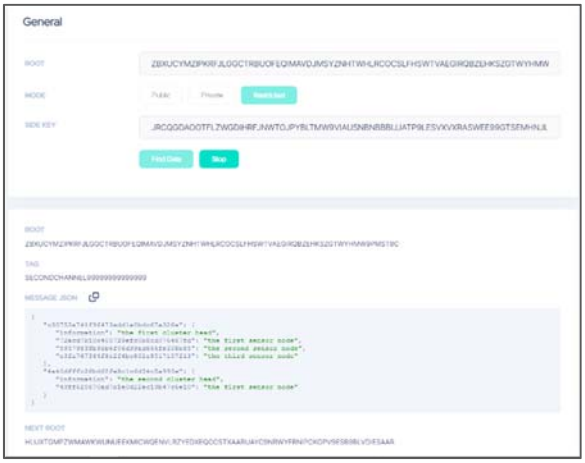


Fig. 19. Message content stored in the base station channel2.

Fig. 20 shows the blacklist in WSNs, recording the ID and MAC Address of the device.

Fig. 21 shows the data uploaded by the Cluster Head to the Tangle, and the RFID Tag obtained from the Sensor Node is recorded in the Channel.

Fig. 22 shows the Dashboard of the node in the Base Station, you can view the current running status, and the Synced in the node shows that it has been running synchronously with the IOTA network.

Fig. 20. The content of the message stored in the base station channel3.

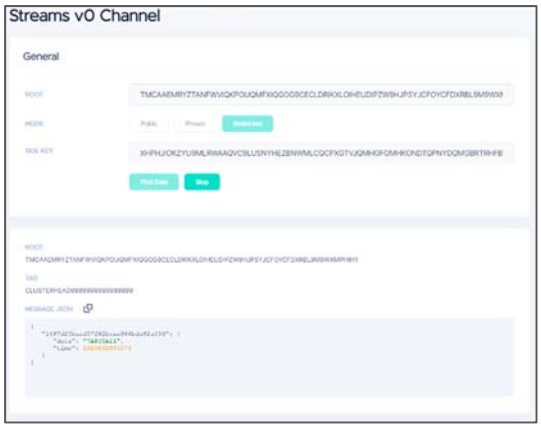


Fig. 21. Message content stored in the cluster head channel.

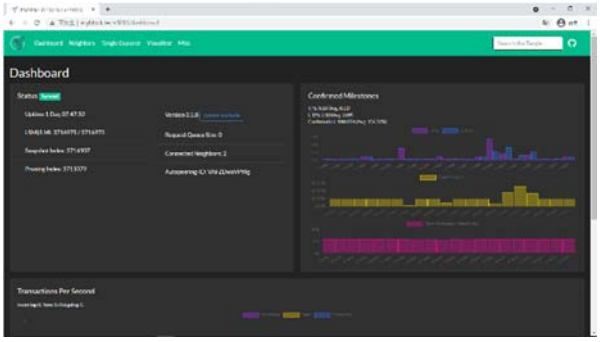


Fig. 22. Node dashboard screen.

5. Conclusions

There have been many achievements in the application of IOTA in industrial automation and control systems, and many papers have also proposed solutions for IOTA. Although these papers all put forward good solutions and application areas for how to connect IOTA and the industrial automation and control systems, they do not consider that in wireless sensor networks, many sensor nodes do not have high computing power and storage space. Especially in industrial control networks, sensing data must pass through many network layers before being uploaded to the distributed network, and the security and privacy in the middle must be protected. For IoT devices, the IOTA Foundation also proposes solutions. IOTA designs many hardware devices and firmware, such as CryptoCore, STM X-Cube-IOTA1, etc. However, the lack of hardware support hinders the development of the industrial automation and control systems, and these devices still require high computing power. In addition, IOTA also has improved lightweight Bee nodes that can perform data upload actions, but still not all hardware can perform them. Therefore, this paper will conduct identity authentication for resource-constrained sensing devices, and upload secure sensing node data to the Tangle for storage. In this paper, we have shown that how to implement IOTA in industrial automation and control systems, the execution result can be acceptable.

Acknowledgements

This research was funded by Ministry of Science and Technology, grant number 111-2218-E-005 -010.

References

[1]. R. Agrawal et al., Continuous security in IoT using blockchain, in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2018, pp. 6423-6427.

[2] Q. Wang, X. Zhu, Y. Ni, L. Gu, and H. J. I. o. T. Zhu, Blockchain for the IoT and industrial IoT: A review, *Internet of Things,* Vol. 10, 2020, p. 100081.

[3] S. K. Datta, J. Haerri, C. Bonnet, and R. F. Da Costa, Vehicles as connected resources: Opportunities and challenges for the future, *IEEE Vehicular Technology Magazine*, Vol. 12, No. 2, 2017, pp. 26-35.

[4] L. Hang, I. Ullah, D.-H. J. C. Kim, and E. I. Agriculture, A secure fish farm platform based on blockchain for agriculture data integrity, *Computers Electronics in Agriculture,* Vol. 170, 2020, p. 105251.

[5] Z. Guan, X. Lu, N. Wang, J. Wu, X. Du, and M. J. F. G. C. S. Guizani, Towards secure and efficient energy trading in IIoT-enabled energy internet: A blockchain approach, *Future Generation Computer Systems*, Vol. 110, 2020, pp. 686-695.

[6] J. Grecuccio, E. Giusto, F. Fiori, and M. J. E. Rebaudengo, Combining Blockchain and IoT: Food-Chain Traceability and Beyond, *Energies,* Vol. 13, No. 15, 2020, p. 3820.

[7] S. J. W. P. Popov, The tangle, *White paper*, Vol. 1, 2018, p. 3.

[8] A. J. N. J. Ericsson, Cellular networks for massive IoT-enabling low power wide area applications, January 2016, pp. 1-13.

(4246)

Industrial 5G to Boost Data-Driven IIoT Applications – Opportunities and Barriers

J. Collin [1], J. Pellikka [2] and J. T. J. Penttinen [3]
[1] Aalto University, Otakaari 1B, 02150 Espoo, Finland
E-mail: jari.collin@aalto.fi
[2] Nokia, Espoo, Finland
E-mail: jarkko.pellikka@nokia.com
[3] Syniverse Technologies Atlanta, GA, USA
E-mail: jyrki.penttinen@outlook.com

Summary: The digitalization of products and services is increasingly disrupting competition. Traditional, physical product-based business models are challenged by data-driven digital ecosystems that pursue new ways to increase end-customer value. Modern digital platforms and the use of real-time data enable a substantial leap in value – as already witnessed in numerous IoT (Internet of Things) applications for consumers. In the enterprise markets, a lack of common business rules for sharing data between enterprises slows down the similar development in industry IoT (IIoT) applications. This paper describes the main opportunities and barriers in utilizing 5G mobile technology to boost data-driven IIoT applications. The viewpoint is an industrial enterprise that seeks new business opportunities to leverage 5G technology. Findings presented are based on a multiple case-study covering the mining, forest, energy, construction, and telecom industries. The paper focuses on sharing key learnings and best practices between the industry verticals, as well as understanding which elements are common and which are industry specific.

Keywords: Industrial 5G, industry IoT, Private networks, Digital services.

1. Introduction

The fourth industrial revolution, Industry 4.0, will involve the technical integration of Cyper Physical Systems (CPS) into the use of the Internet of Things (IoT) in industrial processes, having implications for value creation, business models, downstream services and work organisation [1]. Industry 4.0 will create new economic value through radical transformation of multiple industries, triggered by the convergence of information technology (IT), operational technology (OT), and communication technology (CT) [2].

New IoT-enabled services extend a value offering from delivering products to using them in the most optimal way. Increased customer value is created with online services at the very moment when a consumer uses a product in that unique situation and environment [3]. As Peter Drucker [4] aptly pointed out already decades ago: "what the customer considers value is not a product itself but utility, that is, what a product does for him/her".

Compared to consumer customers, enterprises are more cautions on using data-driven online services in industrial processes. In addition, business risks on data security and privacy combined with mission critical operations can easily become barriers to utilize cross-company data. The 5th generation mobile technology (5G) can provide industrial ecosystems with such an open and trusted platform [5].

Compared to previous wireless technologies, 5G provides users with tens of times higher speed levels, connection reliability with latency of few milliseconds, hundreds of times more capacity for IoT devices per square kilometre, and several new features that are important especially for industrial use, like private 5G networks [6]. Today, there is a growing interest on private 5G networks among industrial enterprises and governments – e.g. in Germany the local regulator has already granted 265 licenses for private 5G network [7].

This paper describes the main opportunities and barriers in utilizing 5G mobile technology to boost data-driven industrial IoT solutions and related use cases in different industry sectors. The viewpoint is an industrial enterprise that seeks new business opportunities to leverage 5G technology in its digital transformation journey. Findings presented are based on a multiple case-study covering the mining, forest, energy, construction, and telecom industries. The research is to identify and describe contemporary industrial 5G use cases in driving the customer value, productivity, and sustainability in a selected industry ecosystem. The paper focuses on sharing key learnings and best practices between the industry verticals, as well as understanding which elements are common and which are industry specific.

2. Opportunities from Industrial 5G

2.1. Underlying Concepts

The concept of IoT has its roots back to 1991 when Mark Weiser introduced his vision of *The Computer for the 21st Century* including a basic idea on ubiquitous computing with wireless sensor networks [9]. A term "Internet of Things" started its life as a title of Kevin Ashton's presentation in 1999 [10]. More broadly, the term IoT started to appear in research

innovations, and commercial contexts in the beginning of this millennium when new cloud computing and wireless sensor technologies emerged in the consumer markets. Over the last 10-15 years, IoT technology and its central role in enabling digital transformation have received enormous attention in public discussions across industries [1, 3, 11, 12]. The emergence of the IoT concept has resulted in numerous definitions, with earlier references primarily focused on the technological aspects [13]. This paper rather considers IoT from a value driven service viewpoint focusing on new functionalities of a system to improve the overall process. Data, information, and services are the fundamental parts of creating value through the IoT [13]. The prime value is created when data is turned into meaningful information and digital services in real-time.

Industry IoT, also referred as Industrial Internet, undertakes an enterprise viewpoint to the IoT. Simply, the term refers to IoT technology used in industrial settings. However, it is not limited to any specific industry sectors but covers all segments having industrial operations – from dairy, slaughterhouse, and bakery to heavy industries such as steel mill, paper mill, machine shop, mine, or power plant [3]. Industry IoT consortium, established in 2014 as a global not-for-profit partnership of industry, government, and academia, defines the term as follows: "*Industrial Internet is an internet of things, machines, computers, and people enabling intelligent industrial operations using advanced data analytics for transformational business outcomes*" [14]. An Industrial Internet solution contains four key characteristics: real-time data processing, transaction predictability, mobility of operations, and increased automation [3].

Industrial 5G refers to an Industrial Internet solution that in real-time combines a massive amount of data from multiple networked sensors with a secure connectivity and computing platform enabling ultra-reliable, low-latency, and high-bandwidth communication in designated locations. It is a secured, wireless data platform dedicated for an industrial enterprise to integrate different Industrial IoT (IIoT) applications between its ecosystem players. 5G technology provides industrial ecosystems with such an open and trusted platform [5]. The new connectivity technology provides users with tens of times higher speed levels (enhanced mobile broadband, eMBB), connection reliability with latency of few milliseconds (ultra-reliable and low latency, uRLL), hundreds of times more capacity for IoT devices per square kilometre (massive machine type communication, mMTC), and several new features that are important especially for industrial use, like network slicing, mobile edge computing, and non-public private 5G networks [15, 16].

2.2. Industry Case Studies – High-level Summary

Our research represents a multiple case-study that aims at creating new theory from contemporary use cases. Focus is on selected industry verticals and sharing learnings with best-practices between different industries. Industrial 5G opens new kinds of opportunities to drive customer value, productivity, and sustainability in the selected industries.

The list of industry case studies shown in Table 1 represent high-level business problems related to a respective industry transformation. These areas demonstrate potential value from the new 5G technology.

Table 1. Summary of industry case studies.

Industry	Case Study Description
Mining	Wireless, remote control of autonomous vehicles is already a reality in certain processes – compared to existing wireless technologies, what benefits, and new business opportunities 5G-technology enables in building, operating, and maintaining wireless solutions in mining area – both underground and on-ground?
Forest	How to improve the efficiency of pulp mill operations by utilizing 5G-technology in production, maintenance, and safety?
Construction	During a building construction project, how to flexibly provide the building ecosystem players with a common, real-time 5G data platform of construction site related information?
Telecom	How can real-time data of system's components be applied to improve energy efficiency of 5G network, edge computing, and customer industrial processes?
Energy	Compared to the private LTE (4G), how 5G enabled video analytics can improve the reliability and efficiency of sustainable production, maintenance, and safety operations in a large refinery area?

To truly understand how operational value is created, a deep dive for each industry use case is needed on how the Industrial 5G solution can be built and how it works. Sharing learnings and best practices between different industry sectors is an essential part of this analysis.

2.3. The Components of Industrial 5G Solution

Any Industrial Internet solution requires an integrated technology infrastructure that works seamlessly together in a secure way [3]. Our study demonstrates how operational value is created in different parts of the infrastructure. Industrial 5G infrastructure consists of following generic parts: sensors, connectivity, data storage, analytics, application, and digital service. Operational value is dependent on how these elements, as shown in the Fig. 1, work together in a secure way.

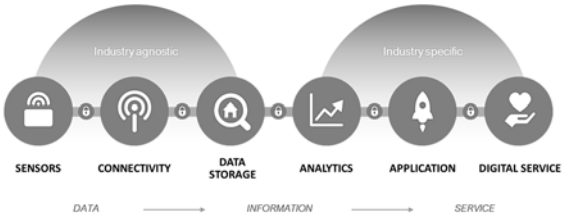


Fig. 1. Technology Infrastructure for Industrial 5G.

In general, technology is no longer a bottleneck, but the challenge is that the parts need to be integrated together. Our learnings from cross-industry analysis show that sensors, connectivity, and data storage are highly standardized parts. These are common elements regardless the industry sector – data is collected, transferred, and stored as bits (0/1). Whereas analytics, application, and digital services are elements where data is turned into valuable information that users should access with a well-functioning user interface. These parts are industry specific – and they include processes where the biggest value is created.

A full scale Industrial 5G solution is, thus, not limited to the connectivity part only but need to cover other elements, too. A short explanation of all these infrastructure elements, especially from 5G technology angle, is needed to understand full potential of opportunities.

Sensor is a device that produces an output signal for the purpose of sensing a physical phenomenon [3], like temperature, pressure, luminosity, sound, humidity, location, turbidity, or height of the liquid surface. In industrial settings, it typically collects data from machines, vehicles, toolsets, buildings, and assets to which it is installed, or from its surroundings. Sensor data is used to monitor industrial processes, events, and human behaviors, too. 5G opens new opportunities to expand the sensor coverage and capacity, decrease their life-cycle costs, and build new kinds of process monitoring, controlling, and optimization capabilities. For instance, 5G supported video cameras and drones are being actively tested to automate industrial processes in large industry areas.

Connectivity part transfers sensor data from multiple sources into a common data storage in real-time, as the 5G network latency drops onto the level of few milliseconds. Still today, many industry connections are wired legacy technologies to comply with high reliability needs those old wireless solutions, like WiFi and 2G/3G/4G, do not meet. Naturally, 5G technology revolutionizes this area by bringing ultra-reliable, high-capacity, and low-latency wireless connections easily available in a whole industrial area. The connectivity part can be handled either through a private or public 5G network. 5G private networks, also referred to as non-public networks (NPNs) by 3GPP standards, are physical or virtual 5G cellular systems deployed for private use [5].

Data storage refers to a secure data centre, where sensitive sensor data is passed and stored for further processing. Today, this often is a physical on-site server, as public cloud solutions do not provide safe and fast enough service levels. 5G technology allows mobile edge cloud that can provide such a solution for industrial customers.

Analytics turns the real-time data into a meaningful information that is essential for decision-making and forecasting analysis. Deep analysis of data patterns with various solutions based on machine learning (ML) and artificial intelligence (AI) are utilized in this part. Typically, this a part where OT meets IT: the sensors data of automation systems is enriched for the use of management systems. 5G data platform can play a central role in this crossroads and enrich analytics with communications network data.

Application part utilizes the enriched information based on data analytics and provide users with mobile apps with a well-functioning user interface. This IIoT application part should be linked to the company's enterprise architecture and API strategy. 5G enabled devices and networks enable the use of data-driven applications that work in real-time. For instance, wireless virtual reality (VR) and augmented reality (AR) applications can drive automation and intelligence into industrial maintenance operations in a massive scale.

Digital service acts as a productized "sellable" service for the industry ecosystem that consists of customers, partners, subcontractors, technology vendors, as well as own employees. The digital service part ties together the whole Industrial 5G infrastructure allowing new online business models based on data-driven IIoT applications.

Our research demonstrates that opportunities lie in all parts of the infrastructure – not only in wireless connectivity. In addition, many commonalities and synergies across industries do exist in sensor, connectivity, and data storage parts. Whereas, synergies are not easily reachable in data analytics, application, and digital service parts – they seem to be highly industry specific components.

However, all infrastructure parts must seamlessly and securely work together, as witnesses in the cases studies. It is, therefore, essential to manage the whole value creation process from turning real-time data into meaningful information and valuable services for the different players of industry ecosystem.

3. Barriers to Implement Industrial 5G

The implementation of Industrial 5G involves many barriers and is everything but a simple straight-forward initiative. It enforces the integration of processes and systems between OT and IT. Traditionally, industries have seen and handled OT and IT as two different specific domains, keeping separate technology stacks, protocols, standards, management, and organizational units [8]. The modernization of legacy processes and tools may lead to overwhelming complexity and cost increases. Cyber security risks and insufficient competences can also become a bottleneck for the implementation.

In each industry case study, implementation barriers were identified and elaborated in many discussions with the stakeholders on management and operational levels. Main barriers per industry sector is summarized in the Table 2.

Table 2. Main barriers for Industrial 5G.

Identified barriers for Industrial 5G	Industry Sectors				
	1	2	3	4	5
Technology	x		x		x
Total Cost of Ownership	x	x			
Competencies	x	x	x	x	x
Change Complexity		x	x		x
Data Security and Privacy	x	x		x	x

As our research is to learn and share best practices between the industries, we used these findings in our common workshop and prioritized them together. The prioritization was conducted in a simple manner by utizing the known *100-Point Method* that is globally used in agile software development to prioritize product backlogs in a scrum team. The method is basically a voting scheme of the type that is used in brainstorming exercises [17].

In our settings, the method was utilized by asking from each industry sector representative (5 persons in total): what are the main barriers in creating data-driven services from industrial 5G (rank the following areas)? The areas are similar to "Identified barriers for Industrial 5G" as presented in Table 2. Each person simultaneously prioritized these five barriers from 1st (as the biggest barrier) to 5th (as the smallest barrier) without seeing answers from the others. So, each barrier had to be chosen and prioritized by everybody. Mathematically, the 1st biggest barrier was scored as 33,33 points, 2nd as 26,67 points, 3rd as 20,00 points, 4th as 13,33 points, and 5th as 6,67 points – so that the total score per person equals 100 points. The points of all answers per the barrier were then summed together.

The results of the common prioritization are presented in Fig. 2.

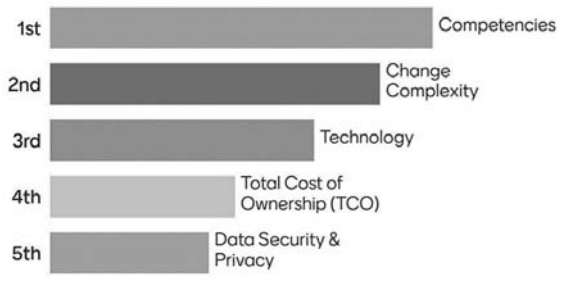


Fig. 2. Technology Infrastructure for Industrial 5G.

The biggest barrier after the voting is *Competencies*. The second biggest one is *Change Complexity*. The third place goes to *Technology*, which is followed by *Total Cost of Ownership* as the fourth place. Surprisingly, the smallest barrier is *Data Security & Privacy*. After the voting each barrier and its ranked position was briefly elaborated more together.

3.1. Barrier #1: Competencies

All industry sectors consider the main barrier is a lack of competent people who can design and build use cases to maximize the business value from Industrial 5G. The biggest competence challenge is how to merge business and technology perspectives together to discover operational innovations that turn real-time data into meaningful information and valuable digital services.

3.2. Barrier #2: Change Complexity

Industrial 5G is seen a major transformative initiative impacting processes, systems, and data models. The implementation should be linked to the strategy, enterprise architecture, and digitalization journey of the company. The complexity of changes depends on the change approach – evolutionary approach seems to be the only feasible option. All industries prefer to start with a piloting approach that first tests Industrial 5G in a selected industry location with a restricted scope. After piloting, the solution can be "productized" for a wider implementation.

3.3. Barrier #3: Technology

Technology is considered a potential barrier for the implementation especially in legacy industrial settings. 5G technology is standardized, but IoT technology is not. Therefore, an all-encompassing and one-size-fits-all technical package is not available, rather enterprises need to integrate the different parts together. Furthermore, combining different "worldviews" of IT and OT easily becomes a barrier. All industry sectors, however, consider technology must not become a showstopper, as it should constantly renew itself.

3.4. Barrier #4: Total Cost of Ownership (TCO)

The total costs of Industrial 5G, covering the whole life-cycle of the system, is seen a potential barrier in a large scale implementation. The value of having a dedicated connectivity and computing platform differs significantly per industry sector. Therefore, it depends highly on industry ecosystem whether the total cost of ownership calculations pay off. In the evolutionary implementation approach, the cost is not considered a significant challenge for any sector.

3.5. Barrier #5: Data Security & Privacy

Data security and privacy is identified as a major risk area in all industry sectors. The risk is not seen dependent on Industrial 5G alone, as these risks need to be managed anyhow. In fact, 5G can mitigate the risk level, as enterprises can build solutions based on private networks and edge computing.

4. Conclusions

The prime value of Industrial 5G is created when data is turned into meaningful information and digital services in real-time. Improvement opportunities from 5G technology lie in all infrastructure parts, although the core new capabilities are focused around connectivity piece. The generic parts that are common across industries are sensors, connectivity, and data storage. The components are fairly standardized already today. However, the biggest customer value is generated from data analytics, applications, and digital services that are highly industry specific and customized. For these parts there is no standard, common solution available. Therefore, the implementation may often require extra system integration efforts.

The implementation of Industrial 5G involves obstacles and challenges that need to be proactively managed. Main barriers in a prioritized order are competencies, change complexity, technology, total cost of ownership, and data security & privacy. Managing these barriers proactively is essential for a successful implementation.

References

[1]. H. Kagermann, J. Helbig, A. Hellinger, and W. Wahlster. 2013. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry, *Forschungsunion, acatech*, 2013.

[2]. A. Mahmood et al., Industrial IoT in 5G-and-Beyond Networks: Vision, Architecture, and Design Trends, *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 6, June 2022, pp. 4122-4137.

[3]. J. Collin and A. Saarelainen, Teollinen internet, *Talentum Pro*, 2016.

[4]. P. Drucker, Management: Tasks, Responsibilities, Practices, *Harper & Row*, New York, 1974.

[5]. Attaran, M., The impact of 5G on the evolution of intelligent automation and industry digitization. *Journal of Ambient Intelligence and Humanized Computing*, 2021, pp. 1-17.

[6]. 3GPP (https://www.3gpp.org/)

[7]. Übersicht der Zuteilungsinhaber für Frequenzzuteilungen für lokale Frequenznutzungen im Frequenzbereich 3.700-3.800 MHz, 2022-xx-xx_Zuteilungsinhaber, 15.11.2022.

[8]. C. Giannelli and M. Picone, Editorial 'Industrial IoT as IT and OT Convergence: Challenges and Opportunities', *IoT*, Vol. 3, 1, 2022, pp. 259-261.

[9]. M. Weiser, The Computer for the 21st Century, *Scientific America*, September 1991, pp. 94-104.

[10]. K Ashton, That "Internet of Things" Thing, *RFID Journal,* Vol. 22, Juli 2009.

[11]. P. Varga, et. al, 5G Support for Industrial IoT Applications— Challenges, Solutions, and Research Gaps Sensors; *Basel*, Vol. 20, Issue 3, 2020.

[12]. C. Giannelli and M. Picone, Editorial "Industrial IoT as IT and OT Convergence: Challenges and Opportunities", *IoT*, Vol. 3, 1, 2022, pp. 259-261.

[13]. K Sorri, N Mustafee, M. Seppänen, Revisiting IoT definitions: A framework towards comprehensive use, *Technological Forecasting and Social Change*, Vol. 179, June 2022, 121623.

[14]. Home - Industry IoT Consortium (iiconsortium.org)

[15]. System Architecture for the 5G System (Release 17), Mar. 2021. https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/16.06.00_60/ts_123501v160600p.pdf

[16]. A. Ghosh, R. Ratasuk, S. Redana, P. Rost, 5G-Enabled Industrial IoT Networks, *Artech House*, 2022.

[17]. D. Leffingwell and D. Widrig, Managing Software Requirements, First Edition, *Addison Wesley,* October 22, 1999, 528 pages.

(4516)

Reinforcement Learning-Based Odometry Prediction Network Trained Without Ground-Truth

Yeongmin Ko [1], Junbom Pyo [2], Yeonsu Seol [2] and Moongu Jeon [1]
[1] Gwangju Institute of Science and Technology, Gwangju, South Korea
[2] Korea Research Institute for Culture Technology
Tel.: +82 062-715-3254
E-mail: koyeongmin@gist.com, xenoluv@gist.ac.kr, comm8123@gist.ac.kr, mgjeon@gist.ac.kr

Summary: This paper presents a deep reinforcement learning-based odometry prediction method for Simultaneous Localization And Mapping (SLAM) trained without ground truth. Odometry prediction methods commonly predict the vehicle's odometry from consecutive sensor data. However, most existing methods based on deep learning are trained in a supervised learning manner. These methods require ground-truth about the odometry; collecting ground-truth consumes numerous resources because additional precise sensors are essential. To solve this problem, we propose a new method equipped with a reinforcement learning manner. In the objective function, the given lidar data are transformed to generate a partial map using the predicted odometry, and virtual sensor data is collected from the partial map. Because the objective function is non-differentiable, we consider the objective function a reward function and apply reinforcement learning to train the model. The proposed method is evaluated on the simple simulation environment.

Keywords: Odometry, Deep learning, Reinforcement learning, SLAM, LiDAR.

1. Introduction

Simultaneous Localization And Mapping (SLAM) is a fundamental technology for autonomous driving systems. Some state-of-the-art SLAM methods generate a High Definition (HD) map and predict the location and pose. The vehicle's location and pose are essential information to determine the optimal path or control values. This paper focuses on odometry prediction, one of the important modules for SLAM.

Although deep learning methods show outperforming performance in various areas, many state-of-the-art odometry prediction methods do not utilize deep learning methods [1, 2]. The difficulty of collecting data is one reason. Most deep learning-based odometry prediction methods are usually trained in a supervised learning-based manner [3]; they require the vehicle's exact odometry or pose as ground-truth. To collect ground-truth, precise sensors like GPS are essential. However, these sensors consume enormous resources and budgets, which is more critical for mobile robot laboratories or businesses that do not use many resources.

We introduce a method to train a deep odometry prediction network without ground-truth to address this problem. We focus on the 2D LiDAR, the most popular mobile robot sensor in this study. First, we define a new objective function that does not depend on ground-truth and correspondence detection using a virtual sensor; the virtual sensor collects data in the partial map. Our proposed objective function calculates a score from the difference between real and virtual sensor data. Next, because this objective function is non-differentiable, we apply a reinforcement learning method to train the network.

2. Method

Our proposed method includes two networks, actor- and critic-network. This actor-critic architecture is trained by a modified Deep Deterministic Policy Gradient (DDPG) algorithm [4] due to continuous action space. Both networks are simply composed of fully connected layers. In the remaining part, we will introduce our training process.

First, we combine two consecutive sensor data to generate the partial map, as shown in Fig. 1. We consider points closer than the threshold of the same neighborhood, and the same neighborhood points are connected to construct walls.

Next, we collect data using a virtual sensor in the partial map as Fig. 1. The virtual sensor is implemented to have the same feature as the real sensor, such as the number of samples, maximum length, minimum length, and the angle between samples. We set the virtual sensor at the position following the predicted odometry. We implement virtual 2D LiDAR using Bresenham's line algorithm and measure distances between the virtual LiDAR and generated walls.

Third, we calculate a score based on the difference between real LiDAR data and virtual LiDAR data. We simply count the number of points that are matched within the small threshold and return a reward value according to the counting result. In Fig. 1, if correct odometry is fed to the objective function, real LiDAR data (blue) and virtual LiDAR data (green) are almost matched and get a higher reward; the case of the wrong odometry is not matched and gets a lower reward. This reward function is non-differentiable, but we can train the network using the reinforcement learning method.

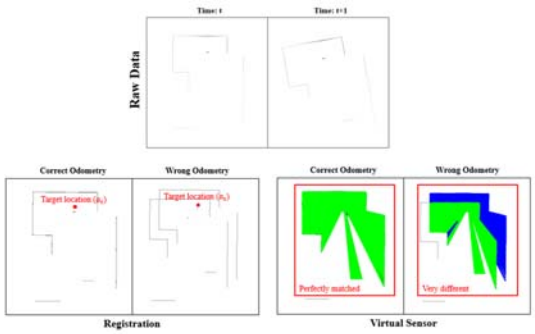


Fig 1. Virtual sensor example.

Finally, we apply the reinforcement learning method to train our odometry prediction network using the proposed reward function. The following equation is the original DDPG loss for critic-network [4]:

$$y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'\left(s_{i+1}\middle|\theta^{\mu'}\right)\middle|\theta^{Q'}\right)$$
$$Loss = \frac{1}{N}\sum(y_i - Q(s_i, a_i|\theta^Q)^2\,, \quad (1)$$

where r_i denotes a reward of i-th step, Q and μ denote the critic-network and actor-network, Q' and μ' denote each target network, and θ denotes parameters of network.

However, if we use the original equation, we cannot get the target Q-value because s_{i+1} is not derived by the current action, a_i. To solve this problem, we express the target Q-value by the following equation:

$$y_i = R(\{l_{i-1}, l_i, l_{i+1}\}, \{a_i, a_{i+1}\}, t = 1) + R(\{l_{i-1}, l_i, l_{i+1}\}, \{a_i, a_{i+1}\}, t = 2), \quad (2)$$

Since the odometry prediction problem has a unique solution, even a slightly greedy training strategy works well.

3. Result

The total training step is about 200k, and the learning rate is initially set by 0.0001; we multiply 0.1 by the learning rate for every 100k train step. LiDAR data is normalized by the maximum range of LiDAR. We make a simple simulation environment using Gazebo and Turtlebot3 model because the simulation environment is comfortable for collecting ground-truth of the vehicle's exact location. In this paper, we use simulated Hokuyo 2D LiDAR; the sensor measures the surroundings of the vehicle at 1 degree intervals and generates 360 samples.

We construct a baseline that has the same architecture as our actor-network. The baseline is trained by a supervised learning manner, and other conditions are the same as our proposed method. The performance is measured by the root mean square error between the estimated location and ground-truth. Our proposed method is trained without ground-truth, it shows a similar performance, as shown in Fig. 2. Blue lines are ground-truth, and green lines are prediction.

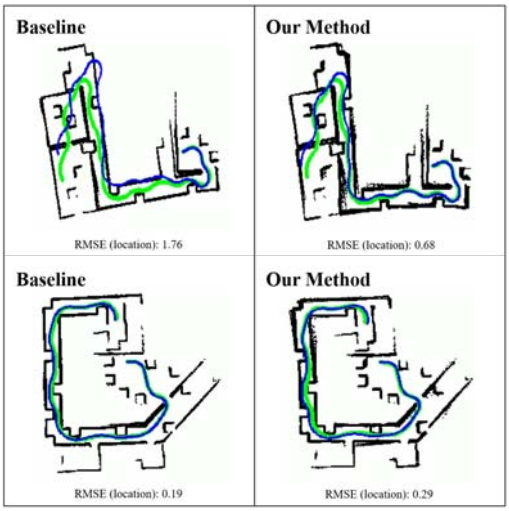


Fig 2. Experimental result.

4. Conclusion

Our proposed method utilizes reinforcement learning to train the network without ground-truth. Learning without ground-truth shows many positive points in many ways. This method can save costs and resources to collect ground-truth. It also allows researchers to develop their system using more data, increasing its scalability.

Acknowledgements

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea ggovernment (SIT) (No.2014-3-00077, Development of Global Multi-target Tracking and Event Prediction Techniques Based on Real-time Large-Scale Video Analysis) and Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism (No. R2022060001, Development of service robot and contents supporting children's reading activities based on artificial intelligence)

References

[1]. P. Biber and W. Strasser, The normal distributions transform: a new approach to laser scan matching, in *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Vol. 3, 2003, pp. 2743–2748.

[2]. J. Zhang and S. Singh, LOAM: Lidar odometry and mapping in real time, in *Proceedings of the Robotics: Science and Systems Conference,* 2014.

[3]. A. Valada, N. Radwan, and W. Burgard, Deep auxiliary learning for visual localization and odometry, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA*, 2018*),* pp. 6939–6946.

[4]. T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, Continuous control with deep reinforcement learning, in *Proceedings of the International Conference on Learning Representations (ICLR' 2016),* 2016.

(4530)

Advanced Methods for Anomaly Detection and Event Recognition by IoT Sensors Immersed in Heterogeneous and Evolving Environments

R. Ait Ouammi, A. Staron[1], H. Snoussi[2] and A. Bittar[2]
[1]ARTIFEEL, 16 rue Washington 75008 Paris FRANCE
[2] University of Technology of Troyes, LIST3N Laboratory, 12 Rue Marie Curie, 10300 Troyes FRANCE
Tel.: + 33636146116
E-mail: redwane@artifeel.com

Summary: This paper presents a research topic that addresses the use of advanced techniques such as artificial intelligence on very complex topics. This will be developed in the context of a real use case: the security of infrastructures by intelligent and autonomous sensors. The aim is to make possible the surveillance of any fenced area by communicating fully autonomous electronic sensors, with the assurance of a very low or zero false alarm rate. This document highlights the constraints that limit the application of current artificial intelligence technologies and presents the methods and advances that are being put in place to address this need.

Keywords: Artificial intelligence, IoT, Security, Sensors.

1. Introduction

The aim of this research is to improve the current surveillance systems, which have some limitations, namely: the complexity of installation to operate properly, but mainly the high rate of false alarms despite the imposed requirements. This generates very high costs related to communications and interventions. The overall objective of this study is to provide surveillance of infrastructures, using a single box, equipped with sensors, which will be fixed to the openings (access gates) protecting these infrastructures.

2. State of the Art

Whether it is an alarm to be installed and controlled by oneself or a remote monitoring system, the installation remains a complex process and sometimes requires specialists. Indeed, current systems need several pieces of equipment to work (the control panel, the sensors, the hotline, the WiFi connection, etc.) but also the power supply and the local Internet connection. As a result, many locations are left unattended. Moreover, these systems often require user intervention (activation/deactivation), which naturally increases the rate of false alarms. A change of environment in these systems always results in false alarms or the need for professional operators to parameterize the monitoring system. This setting is based on algorithms that are based on a binary output, normal or abnormal, which means that there is no flexibility in exploiting the acquired data. This is also the case for camera surveillance systems, where there are false alarms related to moving objects (e.g. animals) passing through the field of view of these cameras.

3. Objective

The box contains battery, transmission module, memory, microprocessor and vibration sensors, which makes it compliant with the GDPR (General Data Protection Regulation) requirements. Fig. 1 illustrates the device architecture.

Fig. 1. The device.

The main focus is to use the vibration data to detect anomalous usage, related to intrusion attempts, but also to identify the different manipulations through the different vibration profiles, so that scenarios can be built up, and normal and unusual activities can be accurately recognized. Fig. 2 reports an example of measured vibration signal used to detect and classify the user action.

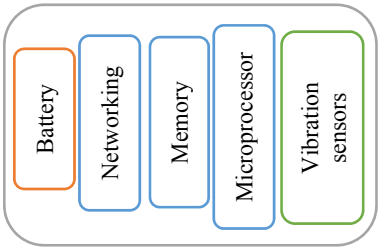


Fig. 2. Device architecture.

4. Development

Recognizing intrusion attempts in the field of surveillance requires prevention, which means that the anomaly must be detected before it is completely produced. This presents a significant challenge, as we will have to classify vibrations before we have the complete signal profile. Incomplete vibrations can have many similarities, so it is difficult to detect the anomaly and even worse when it comes to categorizing the elementary events that compose the global vibration.

To avoid false alarms, it is important that the system be designed, is intelligent and flexible according to the context and habits of the users. The context or environment may experience changes or variation in data sets related to normal use, the system is developed in order to detect these changes and to fit the context to avoid false alarms.

This project involves developing an intelligent learning system that will allow devices to learn continuously from acquired data, and will also enable sharing of knowledge between the different devices that monitor the same type of equipment, in order to help them detect anomalies at the earliest possible stage. A tradeoff between a global AI model learned from all devices data streams and a local AI model adapted to specific users environments should be found in order to achieve high detection performances.

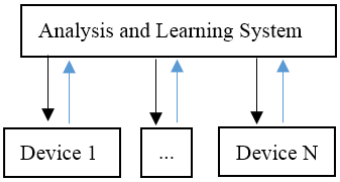


Fig. 4. Lifelong learning architecture.

The learning system uses the data from all the sensors to develop a general model of anomaly detection and classification of different forms of vibration. The system plans to specialize each sensor for its own context, and for those devices that cannot accommodate enough data (low solicitations), the global learning system will be able to feed it with enough knowledge.

For this purpose, the system interacts with the sensors, in the way that the sensor provides vibration data and the system sends back knowledge as necessary. Fig. 3 shows the principle of this bi-directional mechanism aimed at updating the general model and local models.

The analysis and learning system is based on raw data transformation and machine learning techniques, in order to classify the data and establish specialised models for each device according to its context as early as possible. Vibrations are transformed by calculating features such as the maximum and minimum energy in specific time ranges, the percentage change of values between two consecutive samples, the presence or not of sudden peaks, etc. Such information is logged and used for consumption by a K means learning algorithm, specifically the Mini Batch K-Means variant, which uses mini-batches to reduce the computation time, while still attempting to optimise the same objective function. Mini-batches are subsets of the input data, randomly sampled in each training iteration. These mini-batches drastically reduce the amount of computation required to converge to a local solution, which is in accordance with the purpose of the system to incorporate an algorithm based on early classification.

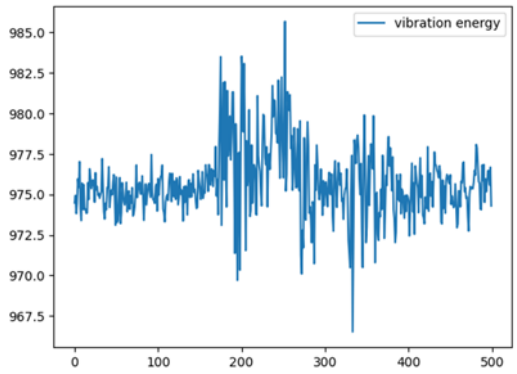


Fig. 3. Example of collected vibration by the device, evolution of the amplitude of the vibration as a function of time.

6. Conclusion

In general, sensors are often installed in heterogeneous and changing environments. Anomaly detection and recognition of different events by these sensors is crucial for the empowerment of any type of sensor. Smarter sensors are more aware of their tasks and avoid measurement and operational errors. A bi-directional learning architecture is proposed in order to learn the global AI model and specific local AI models updated to local environments.

Acknowledgements

This work has been partially funded by French ANRT under the contract n° 2022/0277.

References

[1]. G. He *et al.,* Early classification on multivariate time series, *Neurocomputing*, Vol. 149, Part B, 3 February 2015, pp. 777-787.

[2]. Usue Mori *et al.,* Early classification of time series using multi-objective optimization techniques, *Information Sciences*, Vol. 492, August 2019, pp. 204-218.

[3]. T. S. Brisimi *et al.,* Federated learning of predictive models from federated electronic health records, *Int. J. Med. Inform*, Vol. 112, April 2018, pp. 59-67.

[4]. C. Jiang *et al.,* Machine learning paradigms for next-generation wireless networks, *IEEE Wireless Commun,* Vol. 24, Issue 2, April 2017, pp. 98 – 105.

(4863)

An Application of Fuzzy Competition Graphs at Analog Circuits Analysis

Malinka Ivanova
Technical University of Sofia, Faculty of Applied Mathematics and Informatics, Department of Informatics,
8 Kl. Ohridski Blvd., Sofia, Bulgaria
E-mail: m_ivanova@tu-sofia.bg

Summary: The process of electronic circuits analysis is related to studying their electrical behavior and functionality and to better understanding their specific or common applications. It is important, the engineer analyst tasks to be facilitated and automated. Despite the existing methods for circuit analysis, the researchers are still looking for new approaches, which are capable of saving effort and time for analysis and at the same time to propose suitable solutions. The aim of the paper is to demonstrate the applicability of fuzzy graph theory and analytic hierarchy process at analysis of analog electronic amplifiers and especially to facilitate the engineer analyst at decision making.

Keywords: Circuit analysis, Analog amplifiers, Automation process, Interval-value competition Fuzzy graph, Analytic hierarchy process.

1. Introduction

Process related to analysis of electronic circuits concerns the study of their electrical behavior and functionality mainly in time and frequency domain, but also it could explore their signal distortion or noise introduction. Analysis also is connected to understanding and studying the circuit topology and its building blocks. Each electronic circuit is characterized with a set with parameters, which describe its functionality and workability. The unique combination of values of these parameters leads to different circuit applications. Thus, it is important for the engineer analyst to be supported at analysis of electronic circuits and at automation of some analytical tasks. This can give a more useful explanation about circuit functionality and its application, can reduce the time and effort for analysis.

There are a wide variety of methods and techniques for analyzing the performance of electronic circuits like symbolic network analysis [1], stability analysis through modeling of common denominator [2], semi-topological analysis driven by two-graph theory [3], analysis through artificial neural networks, others.

Among the analytical approaches are also used the advantages of fuzzy set theory and fuzzy logic when uncertainty and fuzziness have to be described. Babanli and Kabaoglu apply fuzzy logic to explain the dynamics of electronic circuits [4] and Ram et al. to diagnose the faults in analog circuits [5]. Fault detection at analog circuits analysis through fuzzy logic is discussed in [6] as well as a fuzzy expert system building for analog circuits diagnostics is presented in [7].

Furthermore, fuzzy graphs and operations upon them are utilized in some areas as the most popular are analysis of systems, economic and operational analysis [8]. Bozhenyuk and Ginis use fuzzy direct graphs to analyze functioning of complex systems [9] and Robert et al. identify and analyze some defects of electronic components in order to predict them [10].

Analytic hierarchy process (AHP) is a method, which is used when decisions have to be made in a complex environment according to multiple objectives and considering different factors. Goh and Kok propose a methodology based on AHP to solve a problem in an electrical power system considering multiple criteria [11]. Zhao et al. discuss the implementation of improved AHP for evaluation the safety of mining equipment and electrical devices [12]. Rajesh et al. apply AHP to optimize the usage and coordination of low power devices in a network of internet of things [13].

It seems that both fuzzy graphs and AHP are suitable for analytical purposes and could be used in assistance of the analyst and in support of decision making. They are applicable in various domains, including for some purposes in electronics.

The aim of the paper is to present a method for analysis of analog electronic circuits based on AHP and competition fuzzy graph theory and to demonstrate an approach for facilitating the engineer analyst.

2. Competition Fuzzy Graphs

Fuzzy graph theory and relations among fuzzy sets are developed to explain and evaluate different kinds of relationships, systems and networks. There are a variety of fuzzy graphs coming to describe their different properties. Anyway, each of them possess common properties like values of edge membership function and vertex membership function, which are calculations taking into account a set of parameters. The competition fuzzy graph is introduced by Cohen to solve an ecology problem at the food chain of competitors [14]. The competition graph represents the competition among biological species regarding their food space. Each competition fuzzy graph is described through a vertex set as each two vertices are connected via an edge. The competition fuzzy graph $CF(\vec{D})$ is explained through a digraph $\vec{D} = (V, \vec{E})$. The Cohen

theory regarding fuzzy graphs is evolved and generalized in the form of fuzzy k-competition graphs [15], p-competition graphs [16], m-step fuzzy competition graph [17], etc. in order to give possibilities for solving diverse engineering problems, specific economics issues and challenging social networks topics.

In this work, the fuzzy competition graph theory is applied to describe and solve a problem at analysis of single stage analog amplifiers and to support the engineer analyst, who has to understand whether a given electronic circuit is suitable for concrete application.

3. Proposed Approach

The proposed method is depicted on Fig. 1 and it is based on:

(1) Analytic hierarchy process [18] for obtaining priority scales by experts regarding the analog amplifiers properties and deriving the membership values of the graph vertices as well as obtaining the weight of edges, representing the relationship among the electronic circuits; and

(2) Interval-valued fuzzy competition graph [19] for evaluating the competition of the analog amplifiers whether they are suitable and can be used for a specific application.

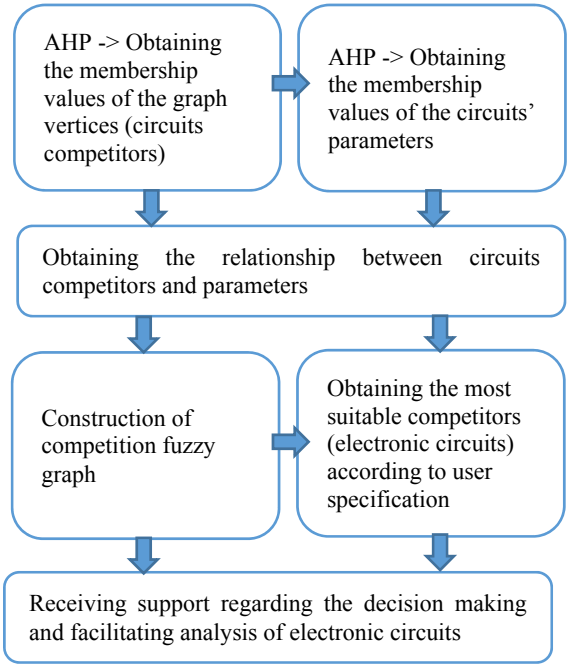


Fig. 1. Proposed method for circuits analysis based on AHP and fuzzy competition graphs.

4. An Application of Fuzzy Competition Graphs at Analysis of Analog Amplifiers

To demonstrate and verify the proposed method three single stage amplifiers are examined and compared: common emitter (CE), common base (CB) and cascode CE-CB, which electronic circuits are presented on Fig. 2, Fig. 3 and Fig. 4 [20, 21].

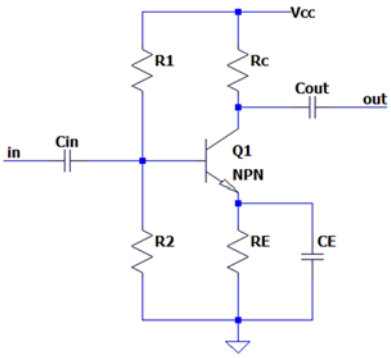


Fig. 2. Common emitter [20], [21].

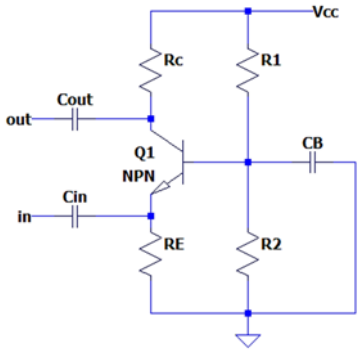


Fig. 3. Common Base [20], [21].

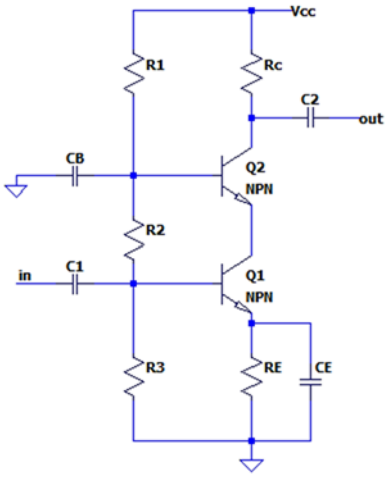


Fig. 4. Cascode CE-CB [20], [21].

The examined circuits possess similar, but not the same properties, which are defined through various parameters like: amplification gain A_v, input and output resistance r_{in} and r_{out}, input and output equivalent capacitance C_{in} and C_{out} responsible for high frequency description of input and output circuits, others (Table 1). These parameters are used for comparing the competitors regarding the user requirements.

Table 1. Main parameters of the competitors [20, 21].

Parameters of CE	
Amplification gain	$A_v \approx -g_m R_C$
Output resistance	$r_{out} \approx R_C$
Input resistance	$r_{in} \approx r_{BE}$
Input equivalent capacitance	$C_{in} = C_{BE} + C_{BC}(1 + \lvert A_U \rvert) + C_M$
Output equivalent capacitance	$C_{out} = C_{BC} + C_M$
Parameters of CB	
Amplification gain	$A_v \approx g_m R_C$
Output resistance	$r_{out} \approx R_C$
Input resistance	$r_{in} \approx \frac{1}{g_m}$
Output equivalent capacitance	$C_{out} \approx C_{BC}$
Cascode CE-CB	
Amplification gain	$A_v \approx -g_m R_C$
Output resistance	$r_{out} \approx R_C$
Input resistance	$r_{in} = r_{BE} \lVert R_2 \rVert R_3$
Input equivalent capacitance	$C_{in} = C_{BE} + 2C_{BC} + C_M$
Output equivalent capacitance	$C_{out} = C_{BC} + C_M$

The engineer analyst strives to understand which schemotechnique variant is the most suitable for a given application concerning known circuit topologies and their main parameters. Let suppose that he/she wishes to understand which circuit CE, CB or CE-CB is a suitable approach for working at high frequency, possesses relatively high input resistance and high amplification gain (user requirements).

There are many competitive topologies of electronic circuits, which could satisfy these requirements taking into account the values of circuit elements (resistors, capacitors), the regime current and voltage, the frequency wideband. So, this problem can be defined as a competition among possible circuit variants, which have capability to satisfy the pre-defined requirements about parameters.

One solution could be an interval-values fuzzy competition graph usage with graph vertices: circuit competitors CE, CB, cascode CE-CB and their main parameters that could be achieved: amplification gain (AG), input resistance (IR) and frequency wideband (FBW). The graph vertices are connected through edges, because the competitors are characterized with similar parameters.

According to the requirements for circuit application: wide frequency band, relatively high input resistance and high amplification gain, the following membership values of the graph vertices of CE 0.75-0.85, CB 0.60-0.65, CE-CB 0.90-0.95 are obtained after applying AHP [18] in R Studio. The importance of the main parameters are also derived through AHP as the values of membership functions are: AG 0.80-0.85, IR 0.70-0.75, FBW 0.85-0.90. The connection among each circuit and related parameters are described with coefficients like these: (CE,AG)=0.80-0.85, (CE,IR)=0.90-0.95, (CE, FBW)=0.70-0.75, (CB,AG)=0.75-0.80, (CB,IR)=0.65-0.70, (CB, FBW)=0.90-0.95, (CE-CB,AG) =0.75-0.80, (CE-CB,IR)=0.80-0.85, (CE-CB, FBW)=0.80-0.85.

Then, the proposed approach of interval-valued fuzzy competition graph [19] applicable for manufacturing industries is used. It is applied in a new context as at the surroundings *S* of the vertices, the intersection of vertices' surroundings *I* and the membership value of the edges Θ are obtained. The results are summarized in Table 2.

Table 2. Obtained results for constructing the fuzzy competition graph.

	CE	**CB**	**CE-CB**
CE	$\boldsymbol{S(CE)}$ $\{AG(0.80,0.85), IR(0.90,0.95), FBW(0.70,0.75)\}$	$\boldsymbol{I(CE \cap CB)}$ $= (0.70,0.75)$ $\boldsymbol{\Theta(CE,CB)}$ $= (0.52,0.63)$	$\boldsymbol{I(CE \cap CE - CB)}$ $= (0.80,0.85)$ $\boldsymbol{\Theta(CE,CE - CB)}$ $= (0.60,0.72)$
CB		$\boldsymbol{S(CB)}$ $\{AG(0.75,0.80), IR(0.65,0.70), FBW(0.90,0.95)\}$	$\boldsymbol{I(CB \cap CE - CB)}$ $= (0.65,0.85)$ $\boldsymbol{\Theta(CB,CE - CB)}$ $= (0.52,0.63)$
CE - CB			$\boldsymbol{S(CE - CB)}$ $\{AG(0.75,0.80), IR(0.80,0.85), FBW(0.80,0.85)\}$

Finally, the competition membership functions for each competitor (analog circuit) are obtained and presented in Table 3. It is seen that the main competitors are CE and CE-CB as the percentages for their suitability for the application, which satisfies the user requirements, are between 60 % and 72 %.

The relationships among the circuits-competitors and their important parameters are presented on Fig. 5 Fig. 6 demonstrates the constructed fuzzy competition graph.

Table 3. Competition membership functions

Competitors	Competition membership function	Competition, %
CE,CB	(0.52,0.63)	52 %, 63 %
CE, CE-CB	(0.60,0.72)	60 %, 72 %
CB,CE-CB	(0.52,0.63)	52 %, 63 %

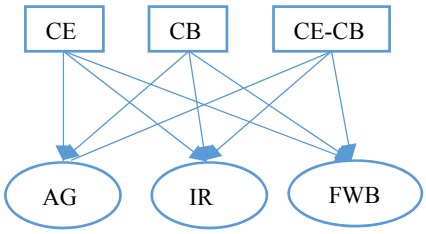


Fig. 5. Relationship among competitors and their important parameters.

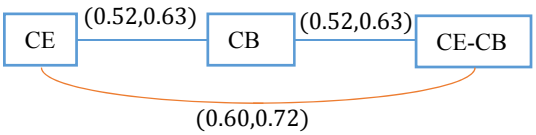


Fig. 6. The constructed fuzzy competition graph.

5. Conclusions

The paper presents an approach for analysis of single stage analog amplifiers through applying analytic hierarchy process and interval-valued fuzzy competition graphs. The demonstration shows the applicability of such method in support of an engineer analyst, who has to find suitable circuit solutions, which are capable to satisfy predefined requirements. Also, this approach could be easily integrated in software tools for automating some analytical engineering tasks.

It is proved that the fuzzy graphs successfully could be used in the case when for one application many circuit variants are possible, e.g. in a case where fuzziness and uncertainty regarding appropriate solutions exists. This can be considered as an advantage in comparison to other well-known analytical methods.

Acknowledgements

This research is supported by Bulgarian National Science Fund in the scope of the project "Exploration the application of statistics and machine learning in electronics" under contract number КП-06-Н42/1.

References

[1]. S. D. Djordjević, P. M. Petković, V. B. Litovski, A new topology oriented method for symbolic analysis of electronic circuits, *Journal of Circuits, Systems and Computers*, 19, 8, 2010, pp. 1781-1795.

[2]. A. Cooman, F. Ferranti, Y. Rolain, G. Vandersteen and E. Louarroudi, Common-denominator modelling for stability analysis of electronic circuits, in *Proceedings of the 11th European Microwave Integrated Circuits Conference (EuMIC' 2016)*, 2016, pp. 361-364, doi: 10.1109/EuMIC.2016.7777565.

[3]. M. Pierzchala, M. Fakhfakh, Two-Graph Based Semi-topological Analysis of Electronic Circuits with Nullors and Pathological Mirrors, in Fakhfakh, M., Pierzchala, M. (eds.), Pathological Elements in Analog Circuit Design, Lecture Notes in Electrical Engineering, Vol. 479, 2018. *Springer*, Cham. https://doi.org/10.1007/978-3-319-75157-3_3.

[4]. K. M. Babanli, R. O. Kabaoglu, Fuzzy Logic-Based Approach to Electronic Circuit Analysis, in Aliev, R., et al. (eds), Advances in Intelligent Systems and Computing, Vol. 1095, *Springer*, Cham, 2020, https://doi.org/10.1007/978-3-030-35249-3_49.

[5]. R. Bharat Ram, V. Prasanna Moorthy and N. Devarajan, Fuzzy based time domain analysis approach for fault diagnosis of analog electronic circuits, in *Proceedings of the International Conference on Control, Automation, Communication and Energy Conservation*, 2009, pp. 1-6.

[6]. M. Merabet and N. Bourouba, DC Hard Faults Detection and Localization in Analog Circuits Using Fuzzy Logic Techniques, 23, 1, 2019, *Electronics*, https://doi.org/10.7251/ELS1923018M.

[7]. D. E. Grzechca, Construction of an expert system based on fuzzy logic for diagnosis of analog electronic circuits, *International Journal of Electronics and Telecommunications*, 61, 1, 2015, pp. 77-82.

[8]. T. Al-Hawary, Complete Fuzzy Graphs, *International J. Math. Combin.*, Vol. 4, 2011, pp. 26-34, https://core.ac.uk/download/pdf/144820905.pdf .

[9]. A. V. Bozhenyuk, L. A. Ginis, Modeling and analysis of complex systems on the basis of fuzzy graph models, *Life Science Journal,* 11, Spec. Issue 7, 37, 2014, pp. 187-191.

[10]. S. G. Robert, N. Bizon, M. Oproescu, Predictive maintenance of electronics systems based on analysis with thermographic camera and fuzzy graphs, in *Proceedings of the 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI 2017)*, 2017, pp. 1-4.

[11]. H. H. Goh and B. C. Kok, Application of Analytic Hierarchy Process (AHP) in load shedding scheme for electrical power system, in *Proceedings of the 9th International Conference on Environment and Electrical Engineering*, Prague, Czech Republic, 2010, pp. 365-368, doi: 10.1109/EEEIC.2010.5489942.

[12]. E. Zhao, H. Jiang, Y. Zhao, X. Ma, Safety Evaluation of mining machinery and electrical system based on improved Analytic Hierarchy Process, in *Proceedings of the 4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA 2016), Series: Advances in Computer Science Research, Vol. 71,* 2017, https://doi.org/10.2991/icmmita-16.2016.135.

[13]. R. Rajesh, C. Annadurai, K. Nirmaladevi, Low power device coordination in internet of things environment using analytic hierarchy process model, *Concurrency*

*and Computation: Practice and Experienc*e, 33, 7, 2021, p. 1.

[14]. J. E. Cohen, Interval graphs and food webs: a finding and a problem, Document 17696-PR, *RAND Corporation,* Santa Monica, CA, 1968.

[15]. S. Samanta, M. Pal, A. Pal, Some more results on fuzzy k-competition graphs, *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, 3, 1, 2014, pp. 60-67.

[16]. S. Eoh, T. Hong, S.-R. Kim, and S. C. Lee, Using p-row graphs to study p-competition graphs, 2019, https://arxiv.org/pdf/1905.10966.pdf

[17]. S. Samanta, M. Akram, M. Pal, m-Step fuzzy competition graphs. *J. Appl. Math. Comput*., 47, 2015, 461–472, https://doi.org/10.1007/s12190-014-0785-2.

[18]. T. L. Saaty, Decision making with the analytic hierarchy process, *Int. J. Services Sciences*, 1, 1, 2008, pp. 83-98.

[19]. T. Pramanik, G. Muhiuddin, A. M. Alanazi, M. Pal, An Extension of Fuzzy Competition Graph and Its Uses in Manufacturing Industries, *Mathematics,* 8, 6, 2020, p. 1008. https://doi.org/10.3390/math8061008.

[20]. M. Ivanova, Analog techniques (in Bulgarian), *TUS*, 2020.

[21]. G. Saggio, Principles of Analog Electronics, *CRC Press*, 2014.

(5244)

A Single Input-Feedback Control of the Generalized Kuramoto-Sivashinsky Equation

R. Al Jamal [1] and N. Smaoui [2]
[1] Australian University of Kuwait, College of Engineering, Department of Mathematics & Physics, P.O. Box 1411, Safat 13015, Kuwait
[2] Kuwait University, Faculty of Science, Department of Mathematics, P.O. Box 5969, Safat 13060, Kuwait
E-mail: [1] r.aljamal@au.edu.kw, [2] n.smaoui@ku.edu.kw

Summary: The control problem of the generalized Kuramoto-Sivashinsky equation (GKS) which is a nonlinear partial differential equation that models the dynamics of flame-front propagation, front propagation in reaction-diffusion systems, and the motion interface of viscous film flows is considered. First, we show that the stability of the equilibria of the GKS equation depends on the value of the parameters $\nu \in \mathbb{R}$ and $\gamma_2 > 0$, and that the set of all constant equilibria is unstable when the instability parameter $\nu < -\gamma_2$. Then, a single bounded input-feedback control is designed to control the GKS equation to any desired constant solution. It is shown that the same controller that stabilizes the linearized GKS equation will locally stabilize the nonlinear system without having a spillover effect. Finally, numerical simulations that illustrate the proposed approach are presented.

Keywords: Generalized Kuramoto-Sivashinsky equation, Input feedback control, Linearization, Partial differential equations, Stability of nonlinear systems.

1. Introduction

The GKS equation has attracted many researchers due to its crucial physical aspects [1-5]. For instance, using spectral methods, the well-posedness of the global and smooth solution of the GKS equation has been proven in [2]. In addition, setting $\gamma_1 = \gamma_2 = 1$, and $\alpha = 1$ in Eq.(1), the GKS equation becomes the Kuramoto-Sivashinsky (KS) equation. The KS equation was derived by Kuramoto [6] to model chemical reaction system, and by Sivashinsky [7] to model the dynamics of flame front propagation. Since then, the KS equation has been investigated analytically as well as numerically by many researchers from different scientific disciplines (see for instance, [8-29]). Al Jamal and Morris [27] showed that stabilizing the linearized form of the KS equation implies local exponential stability of the original nonlinear KS equation.

In this paper and in line with the theory presented in [27, 31], we show that the stabilization of the linearized GKS equation will also imply the local exponential stabilization of the original nonlinear GKS equation. Furthermore, we show that the GKS equation can be stabilized by a single state-feedback control.

The paper is organized as follows. In Section 2, the generalized Kuramoto-Sivashinsky (GKS) model is presented. Section 3 discusses the stability analysis of the set of the constant equilibrium solution to the GKS equation. Section 4 presents a stabilization procedure that locally exponentially stabilizes the GKS equation by using a finite-dimensional controller. The controller scheme provided is used to stabilize the GKS equation to a non-zero constant solution. Finally, Section 5 presents some numerical results using spectral techniques to show the dynamics of the equation with and without a controller followed by concluding remarks in Section 6.

2. The Generalized Kuramoto-Sivashinsky Model

Consider the controlled generalized Kuramoto-Sivashinsky (GKS) equation with a single state-feedback control and periodic boundary conditions

$$\frac{\partial z}{\partial t} + \gamma_2 \frac{\partial^4 z}{\partial x^4} - \nu \frac{\partial^2 z}{\partial x^2} + \gamma_1 z^\alpha \frac{\partial z}{\partial x} = b(x)u(t), \qquad (1)$$

$$\frac{\partial^n z}{\partial x^n}(0,t) = \frac{\partial^n z}{\partial x^n}(2\pi,t), \quad n = 0,1,2,3 \qquad (2)$$

and initial condition

$$z(x,0) = z_0(x), \qquad (3)$$

where $\gamma_1, \gamma_2 > 0$, $\nu \in \mathbb{R}$ and $\alpha \in \mathbb{N}$. The vector $z \in L^2(0,2\pi)$ is the state of the system which represents the flame front velocity at point x and time t. The influence of the actuator is modeled by $b(x) \in L^2(0,2\pi)$. Define the operator $K: L^2(0,2\pi) \to \mathbb{R}$ such that

$$Kz = \langle k, z \rangle, \qquad (4)$$

for some $k \in L^2(0,2\pi)$ and $\langle \cdot,\cdot \rangle$ is the L^2-inner product defined in the space. The input-feedback control $u(t)$ in equation (1) is of the form

$$u(t) = -Kz(t). \qquad (5)$$

Notation 1. Denote $H^n_{per}(0,2\pi)$ for $n \in \mathbb{N}$ to be the Hilbert space $H^n(0,2\pi)$ with proper periodic boundary conditions.

Define the linear operator $A: H^4_{per}(0,2\pi) \subset H^4(0,2\pi) \to L^2(0,2\pi)$

$$Az = -\gamma_2 \frac{\partial^4 z}{\partial x^4} + \nu \frac{\partial^2 z}{\partial x^2}, \tag{6}$$

the nonlinear operator $F: H^1_{per}(0,2\pi) \subset H^1(0,2\pi) \to L^2(0,2\pi)$

$$F(z) = -\gamma_1 z^\alpha \frac{\partial z}{\partial x}, \tag{7}$$

and the bounded linear operator $B: \mathbb{R} \to L^2(0,2\pi)$

$$Bu = b(x)u. \tag{8}$$

The controlled GKS equation (1) can be written in the abstract form

$$\begin{aligned} \dot{z} &= Az + F(z) + Bu, \\ z(0) &= z_0. \end{aligned} \tag{9}$$

Theorem 2. [30] The controlled GKS equation (1) with periodic boundary condition is well-posed. That is, there exists a unique strong solution that satisfies the system

$z \in C\left([0,T]; L^2(0,2\pi) \cap L^2\left((0,2\pi); H^4_{per}(0,2\pi)\right)\right)$, where $0 < T < \infty$.

It is worth mentioning that the uncontrolled GKS equation ((1) with $u(t) = 0$) possesses a conservation law property. That is, the integral $\int_0^{2\pi} z\,(t)dx$ is invariant under time evolution.

The uncontrolled GKS equation ((1) with $u(t) = 0$) has infinitely many equilibrium solutions. In particular, any constant function is an equilibrium solution to the GKS equation. It is worth mentioning that the equilibrium solution to the GKS equation can be determined using its initial condition. This is true due to the conservation law property of the GKS equation.

Define the closed invariant set $Z_e = \{z_e;\ z_e$ is a constant function$\}$ to be the set of all constant equilibrium solutions. Next, we shall linearize the GKS equation at a constant equilibrium solution $z_e \in Z_e$.

3. The Stability Analysis of the GKS Equation

Consider the controlled linearized GKS equation at a constant equilibrium $z_e \in Z_e$

$$\begin{aligned} \frac{dz}{dt} &= Az + F'(z) + Bu, \\ z(0) &= z_0, \end{aligned} \tag{10}$$

where $F': H^1_{per}(0,2\pi) \to L^2(0,2\pi)$ is the Gâteaux derivative of the nonlinear operator F at a constant equilibrium z_e which is given by

$$F'(z) = -\gamma_1 z_e^\alpha \frac{\partial z}{\partial x}. \tag{11}$$

Solving the eigenvalue problem for the operator $A + F'$ yields the eigenvalues $\lambda_n = -\gamma_2 n^4 - \nu n^2 - i\gamma_1 z_e^\alpha n$ and the corresponding eigenfunctions $\phi_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$ for $n = 0, \cdots, \infty$. The real part of the eigenvalue λ_n crosses the imaginary axis when $\nu = -\gamma_2 n^2$. As a result, one can show that the number of unstable eigenvalues to the linearized GKS equation is finite and can be calculated by finding the largest integer N such that $N < \sqrt{-\frac{\nu}{\gamma_2}}$.

Furthermore, if $\nu > -\gamma_2$, then the set of constant equilibrium solutions is globally asymptotically and locally exponentially stable. Moreover, if $\nu = -\gamma_2$, then the set of equilibrium solutions to the nonlinear GKS equation is globally Lyapunov stable. Finally, if $\nu < -\gamma_2$, then the set of equilibrium solution becomes unstable.

In order to analyze the stability/instability of the nonlinear GKS equation using the stability/instability of its linearization, the Féchet differentiability of the nonlinear C_0-semigroup generated by the nonlinear GKS equation is needed. Moreover, the derivative of the nonlinear semigroup has to be the linear C_0-semigroup generated by the linearized GKS equation. In addition, we use the Lyapunov indirect method which can be generalized to infinite dimensional systems defined on a Hilbert space as in [31]. It should be noted that the results presented in this paper can be considered as a special case of the results on the generalized Kortweg-de Vries-Burgers-Kuramoto Sivashinsky equation discussed by Al Jamal and Smaoui [30] which is the key of the stability/instability analysis of the GKS equation.

4. The Stabilization of the GKS Equation

In this section, we shall introduce a procedure to stabilize the GKS equation using a single bounded input feedback control to a constant equilibrium solution. It is worth mentioning that the feedback controller is originally designed for the linearized GKS equation at a desired constant equilibrium, then the same controller will be used to locally stabilize the nonlinear GKS equation to that desired constant equilibrium.

Define $\phi_0 = \frac{1}{\sqrt{2\pi}}$, $\phi_n(\cdot) = \frac{1}{\sqrt{\pi}}\cos(n\,\cdot)$, $\psi(\cdot) = \frac{1}{\sqrt{\pi}}\sin(n\,\cdot)$ for $n = 1, \cdots \infty$, then the set $\{\phi_0, \phi_n, \psi_n, n = 1, \cdots, \infty\}$ forms a basis in $L^2(0,2\pi)$.

Also, define

$$\begin{aligned} b_{\phi_n} &= \langle b(x), \phi_n(x)\rangle, \quad n = 0, \cdots, \infty \\ b_{\psi_n} &= \langle b(x), \psi_n(x)\rangle, \quad n = 1, \cdots, \infty. \end{aligned} \tag{12}$$

Let N be the finite number of unstable eigenvalues to the linearized GKS equation and assume that

$$\begin{aligned} b_{\phi_n} &\neq 0, \quad n = 0, \cdots, N \\ b_{\psi_n} &\neq 0, \quad n = 1, \cdots, N. \end{aligned} \tag{13}$$

Then there exists a finite dimensional controller that globally stabilizes the linearized GKS equation and hence will locally stabilize the nonlinear GKS equation to the constant z_e state.

The theory of designing a stabilizing input-feedback controller for linear infinite-dimensional systems is well established; (see for instance [32-37] and the references within). One well-known approach is carried by using a linear quadratic (LQ) controller [34, 38]. In this paper, it will be assumed that the full state of the solution to the GKS equation in known and LQR controller design is used to construct the controller u.

5. Numerical Simulations

Consider the nonlinear GKS equation (1) with parameters $\nu = -4,\ \gamma_1 = 1,\ \gamma_2 = 3,\ \alpha = 2$ and initial condition

$$z_0(x) = \sin\left(\frac{x}{2}\right). \tag{14}$$

and

$$b(x) = \frac{1}{0.3} \cdot \chi_{[\varepsilon-0.1, \varepsilon+0.2]}, \tag{15}$$

where $\varepsilon > 0$ and $\chi_{[a,b]}$ indicates the characteristic function with support on $[a, b]$. The solution to the GKS equation (9) is done using Galerkin Projection method. That is, for some $\widetilde{N} > N > 0$ define

$$z_N(x,t) \approx a_0(t)\phi_0(x) + \sum_{i=1}^{\widetilde{N}} a_i(t)\phi_i(x) + \sum_{i=1}^{\widetilde{N}} c_i(t)\psi(x) \tag{16}$$

where $a_i(t), c_i(t)$ yields the solution of the ODE system resulting from the Galerkin projection method. Note the approximation is of order $2\widetilde{N} + 1$. Moreover, with the choice of $\widetilde{N} > N > 0$, the solution captures all unstable eigenvalues of the GKS equation.

The solution of the uncontrolled GKS equation is given in Fig. 1. It is clear that the system is unstable which is consistent with the theoretical findings.

Using the stability result discussed earlier, the number of unstable modes to the GKS equation is equal to $N = 1$. This means that there are 3 eigenfunctions in the GKS equation correspond to eigenvalues with positive real parts. All controllers will be designed using only the corresponding eigenfunctions. Simulations are done using the first 51 eigenfunctions ($\widetilde{N} = 25$).

Consider the stabilization to the equilibrium solution $z_e = 1$ with control centred at $\varepsilon = 0$. Since $b1_n =< b, \phi_n > \neq 0$ for $n = 0,1$ and $b2_n =< b, \psi_n > \neq 0$ for $n = 1$, the linearized GKS system (10) is stabilizable. Using LQR scheme, a controller of the form $u(t) = -Kz(t)$ can be constructed to minimize the quadratic cost function

$$J(u) = \int_0^\infty \langle z(t), z(t) \rangle + u^*(t)u(t)dt. \tag{17}$$

This yields the control $u(t) = -\langle k, \mathrm{z(t)} \rangle$ where

$$\mathrm{k} = [-1 \quad 10.0182 \quad 4.3246].$$

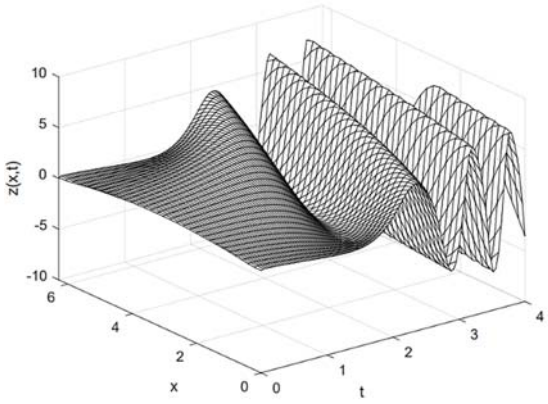


Fig. 1. A 3D landscape of the solution of the GKS equation with periodic boundary conditions, $\nu = -4,\ \gamma_1 = 1,\ \gamma_2 = 3,\ \alpha = 2$ and initial condition $z_0(x) = \sin\left(\frac{x}{2}\right)$. The solution of the GKS equation is unstable.

Fig. 2 shows the 51^{th}-order approximation and the controller designed with only 3 eigenfunctions to the nonlinear GKS equation. Note that the number of eigenfunctions used in the simulations is larger than the number of eigenfunctions used to design the controller, yet the feedback controller achieved the stabilizing goal and there was no spillover.

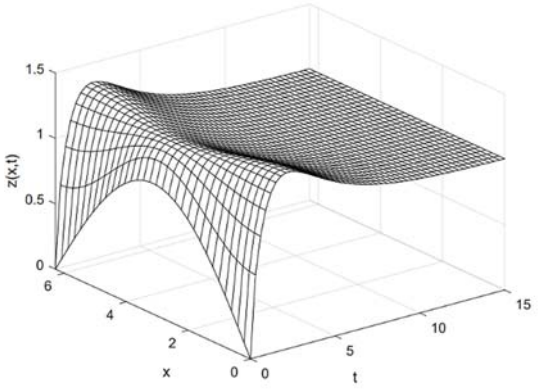


Fig. 2. A 3D landscape of the solution of the controlled GKS equation with periodic boundary conditions, $\nu = -4,\ \gamma_1 = 1,\ \gamma_2 = 3,\ \alpha = 2$ and initial condition $z_0(x) = \sin\left(\frac{x}{2}\right)$. The solution of the GKS equation converges to the desired equilibrium solution $z_e = 1$.

6. Conclusions

A single bounded input-feedback control for the GKS equation with periodic boundary conditions was designed. The Fŕechet differentiability of the nonlinear C_0 -semigroup generated by the nonlinear GKS

equation with its derivative being the linear C_0 -semigroup generated by the linearized GKS equation is the key for the controller design. It was shown that the same controller that globally exponentially stabilizes the linearized GKS equation will locally exponentially stabilize the nonlinear GKS equation. Furthermore, Lyapunov indirect method was used to analyze the stability of the GKS equation where it is shown that the GKS equation is unstable when $\nu < -\gamma_2$. Finally, numerical simulation to illustrate the proposed controller scheme approach was presented to show that the solution of the GKS equation can be driven to a desired constant equilibrium solution $z_e = 1$.

References

[1]. B. Guo, and X. Wu. The Spectral Method for the Generalized Kuramoto-Sivashinsky equation, *Journal of Computational Mathematics,* 9, 4, 1991, pp. 330-336.

[2]. Z. J. Yang. Travelling Wave Solutions to Nonlinear Evolution and Wave Equations, *Journal of Physics A: Mathematical and General,* 27, 1994, pp. 2837-2855.

[3]. C. Li, G. Chen, and S. Zhao. Exact Traveling wave solutions to the generalized Kuramoto-Sivashinsky equation, *Latin American Applied Research,* 34, 2004, pp. 64-68.

[4]. T. Kobayashi. Adaptive Stabilization of a Class of Reaction-Diffusion Systems Dynamics and Control, 11, 2001, pp. 47-56.

[5]. S. N. Gomes, D. T. Papageorgiou, and G. A. Pavliotis. Stabilizing non-trivial solutions of the generalized Kuramoto-Sivashinsky equation using feedback and optimal control. IMA Journal of Applied Mathematics, 82, 2017, pp. 158-194.

[6]. Y. Kuramoto. Diffusion-induced chaos in reaction systems. Progress of Theoretical Physics Supplements (Japan), 64, 1978, pp. 346-367.

[7]. G. I. Sivashinsky. On flame propagation under conditions of stoichiometry. SIAM Journal on Applied Mathematics, 39, 1980, pp. 67-82.

[8]. E. Tadmor. The well-posedness of the Kuramoto-Sivashinsky equation. SIAM Journal of Mathematical Analysis and Applications, 17, 1986, pp. 884-893.

[9]. D. Armbruster, J. Guckenheimer, and P. J. Holmes. Kuramoto-Sivashinsky dynamics on the center-unstable manifold. SIAM Journal on Applied Mathematics, 49 (3), 1989, pp. 676-691.

[10]. G. Sell and Y. You. Dynamics of evolutionary equations. Springer-Verlag, 143, 2000.

[11]. P. D. Christofides and A. Armaou. Global stabilization of the Kuramoto-Sivashinsky equation via distributed output feedback control. Systems and Control Letters, 39, 2000, pp. 283-294.

[12]. W.-J. Liu and M. Krstic, Stability enhancement by boundary control in the Kuramoto-Sivashinsky equation. Nonlinear Analysis, Theory, Methods and Applications, 43(4), 2001, pp. 485-507.

[13]. T. Kobayashi. Adaptive stabilization of the Kuramoto-Sivashinsky equation. International Journal of Systems Science, 33(3), 2002, pp. 175-180.

[14]. A. T. Cousin and N. A. Larkin. Kuramoto-Sivashisky equation in domains with moving boundaries. Portugaliae Mathematica, 59, 2002, pp. 336-349.

[15]. R. Sakthivel and H. Ito. Nonlinear robust boundary control of the Kuramoto-Sivashinsky equation. IMA Journal of Mathematical Control and Information, (UK), 24(1), 2007, pp. 47-55.

[16]. Y. Zhang, L. Song, and W. Axia. Dynamical bifurcation for the Kuramoto-Sivashinsky equation. Nonlinear Analysis, 74, 2011, pp. 1155-1163.

[17]. H. Brown, A. Kevrekidis, Oron, and P. Rosenau. Bifurcations and pattern formation in the 'regularized' Kuramoto-Sivashinsky equation. Physics Letters A, 163, 1992, pp. 299-308.

[18]. J. Elgin and X. Wu. Stability of cellular states of the Kuramoto-Sivashinsky equation. SIAM Journal on Applied Mathematics, 56, 1996, pp. 1621-1638.

[19]. C. Foias, M. Jolly, I. Kevrekidis, G. Sell, and E. Titi. On the computation of inertial manifolds. Physics Letters A, 131 1989, pp. 433-437.

[20]. J. Hyman and B. Nicolaenko. The Kuramoto-Sivashinsky equation: A bridge between PDEs and dynamical systems. Physica D, 18D, 1986, pp. 113-126.

[21]. I. Kevrekidis, B. Nicolaenko, and J. Scovel. Back in the saddle again: A computer assisted study of the Kuramoto-Sivashinsky equation. SIAM Journal on Applied Mathematics, 50, 1990, pp. 760-90.

[22]. A. Armaou and P. D. Christofides. Feedback control of the Kuramoto-Sivashinsky equation. Physica D, 137, 2000, pp. 49-61.

[23]. C. Byrnes, D. Gilliam, C. Hu, and V. Shubov, Zero dynamics boundary control for regulation of the Kuramoto-Sivashinsky equation. Mathematical and Computer Modelling, 52, 2010, pp. 875-891.

[24]. B. Guo and H. Gao. Global attractor for axially symmetric Kuramoto-Sivashinsky equation in annular domains. Communications in Nonlinear Science and Numerical Simulation, 1, 1996, pp. 34-37.

[25]. Y. Lou and P. Christofides. Optimal actuator/sensor placement for nonlinear control of the Kuramoto-Sivashinsky equation. IEEE Transactions on Control Systems Technology, 11, 2003, pp. 737-45.

[26]. B. Nicolaenko, B. Scheurer, and R. Temam. Some global dynamical properties of the Kuramoto-Sivashinsky equations: Nonlinear stability and attractors. Physica D, 16D, 1985, pp. 155-83.

[27]. R. Al Jamal and K. Morris. Linearized stability of partial differential equations with application to stabilization of the Kuramoto-Sivashinsky equation. SIAM journal on Control and Optimization, 56(1), 2018, pp. 120-147.

[28]. N. Smaoui. A hybrid neural network model for the dynamics of the Kuramoto-Sivashinsky equation. Mathematical Problems in Engineering, 3, 20014, pp. 305-321.

[29]. N. Smaoui. Linear versus nonlinear dimensionality reduction of high-dimensional dynamical systems. SIAM journal on Scientific Computing, 25(6), 2004, pp. 2107-2125.

[30]. R. Al Jamal and N. Smaoui. A single bounded input-feedback control to the generalized Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky equation. Mathematical Methods in Applied Sciences, 46(2), 2022, pp. 2222-2248.

[31]. R. Al Jamal, A. Chow, and K. Morris. Linearized stability analysis of nonlinear partial differential equations, in Proceedings of the 21st International Symposium on the Mathematical Theory of Networks and Systems, 2014, pp. 847-852.

[32]. G. Bastin, J. Coron, and B. d'Andrea Novel. On Lyapunov stability of linearized Saint-Ventant

equations for a sloping channel. Network and Heterogeneous Media, 4(2), 2009, pp. 177-187.

[33]. A. Bensoussan, G. Prato, M. Delfour, and S. Mitter. Representation and control of infinite-dimensional systems. Birkhauser, 2nd edition, 2007.

[34]. R. Curtain and H. Zwart. An introduction to infnite-dimensional linear systems theory. Springer-Verlag, 1995.

[35]. I. Lasiecka and R. Triggiani. Control theory for partial differential equations: Continuous and approximation theories. Cambridge University Press, 2000.

[36]. Z. Luo, B. Guo, and O. Morgul. Stability and stabilization of infinite-dimensional systems with applications. Springer-Verlag, London, 1999.

[37]. K. Morris. Control of systems governed by partial differential equations. IEEE Control Theory Handbook, CRC Press, 2010.

[38]. J. Zabczyk. Mathematical control theory: An introduction. Birkhauser, 2007.

(5349)

General Guidance for the Realization of Smart Retrofitting in Legacy Systems for Industry 4.0

Eusebio Jiménez López [1], Gabriel Luna Sandoval [2], Baldomero Lucero Velázquez [3], Francisco Javier Ochoa Estrella [4], Flavio Muñoz [5], Juan José Delfín Vázquez [6] and Francisco Cuenca Jiménez [7]

[1] Universidad Tecnológica del Sur de Sonora-ULSA Noroeste,
Dr. Norman E. Borlaug Km. 14, Obregon City, Sonora, Mexico
[2] Sonora State University, Carretera San Luis Río Colorado - Sonoyta km. 6.5,
Parque Industrial, San Luis Río Colorado, Mexico
[3,4,5,6] TecNM/ Instituto Tecnológico Superior de Cajeme, Carretera Internacional a Nogales Km. 2 s/n, Cd. Obregon, Sonora, Mexico
[7] Universidad Nacional Autónoma de México, Av. Universidad 3004, Col. Copilco Universidad, Coyoacán, Mexico City, CDMX, México
Tel.: + 52 (644) 414-8687, fax: + 52 (644) 414-8687, E-mail: ejimenezl@msn.com

Summary: Many companies around the world are having trouble aligning to the Industry 4.0 (I4.0) philosophy because technology costs are high and out of reach, especially for micro and small companies. Smart retrofitting is a good option for companies as it is possible to access I4.0 systems by modernizing existing machines and production systems. Smart retrofitting seeks technological modernization based on I4.0 objectives and represents an ideal method for companies to use their existing technology and even offline machinery to align to the requirements of the fourth industrial revolution. This article presents a general guide for carrying out a smart retrofitting project to legacy systems. The proposal considers traditional retrofitting as the basis of smart retrofitting and proposes Reverse Engineering as an important analysis method within the study of a retrofitting project for I4.0. The proposed guide is composed of six phases and proposes some recommendations for organizations that are going to implement a smart retrofitting project.

Keywords: Traditional retrofitting, Reverse engineering, Industry 4.0, Companies, Smart retrofitting.

1. Introduction

In recent years, dynamic changes have occurred in the industrial environment as a result of innovations called "Industry 4.0" (I4.0) [1]. The application of digital technologies and the development of new production methods have led to an evolution and reconfiguration of the world in general, including human life, industries, factories and manufacturing processes [2]. I4.0 is an industrial approach that promotes large-scale digitization and optimization of production processes. This makes industrial systems look for the fastest and most efficient way to modernize or reconvert to achieve the I4.0 objectives.

In many parts of the world, technological modernization towards I4.0 is in a transition and represents a serious problem for companies, particularly for micro and small companies, as they face the challenge of deciding whether a technological upgrade (an almost total change of the production systems) that involves a considerable economic investment or a technological modernization that involves a substantial improvement of the production systems they already have and, therefore, a lower economic investment, is relevant [3]. The retrofitting of machinery and systems seems to be a good option to accompany companies that do not have sufficient economic resources to upgrade towards the objectives of I4.0 [4]. Retrofitting is a technical process of modernizing a device, apparatus, machine, factory or system (whether in operation or not) so that it fulfills or satisfies the original functions (e.g. efficiency losses and wear due to use) or new functions or requirements. When retrofitting is performed under I4.0 requirements and objectives, it is known as "smart retrofitting".

Retrofitting is currently applied by several companies and industries, for example, in [5] they retrofitted equipment of an industrial loom, providing the machines with communication functionalities aligned with the I4.0 concept. In [6] a retrofitting process was developed for an old machine used for drilling tasks. External sensors were adapted to this machine for data collection and to feed cloud-integrated databases for analysis/monitoring purposes.

This article proposes a conceptual guide related to smart retrofitting applied to legacy manufacturing machines and systems.

2. Industry 3.5 and Technological Reconversion

Industry 3.5 represents the migration and reconversion phase between Industry 3.0 and I4.0 and sets the tone for various companies seeking to align themselves with the fourth industrial revolution [7]. The transition between two industrial revolutions

generates an important set of scientific and technological challenges, not only for companies, but also for governments, the education sector, trade and strategic competition between regions or countries [3].

Various small and medium-sized companies worldwide face in the Industry 3.5 phase problems such as high investment costs, lack of know-how and personnel, loss of time, lack of a recognizable need, lack of qualified employees, data protection, infrastructure and interfaces or long waiting for market maturity [8]. In the field of investment in modern technology, many companies are faced with making decisions on what kind of strategy to consider in order to align to I4.0, i.e., whether to opt for the direct purchase of technology considering a large investment or whether to take the path of modernization of their production systems seeking good competitiveness with a lower economic investment.

For many companies, industrial retrofitting to I4.0 is no longer optional, especially those that are part of the value chains of industrial consortiums that have more economic power, as they must adjust to new production requirements. A smart retrofitting program can enable company managers to reintroduce older facilities to I4.0. Smart retrofitting offers an opportunity for companies that are not digital at all, to redesign the relationship between man and machine and open new horizons such as predictive maintenance and quality control and efficiency [9].

3. Smart Retrofitting

Retrofitting is to provide a machine with a part, or a place with equipment, that it did not originally have when it was built [10]. The retrofitting process involves a series of modernization actions using new technologies. The goal of retrofitting is to preserve the functionality of existing equipment and adapt it to current needs, as well as to extend its useful life [11]. More than updating a piece of equipment, retrofitting can reinvigorate the entire installed industrial park, allowing, in addition to the inclusion of more advanced management models, the continuously updated knowledge of the role of each asset in the process [12]. Identifying commonalities in retrofitting, such as adaptations, technologies and resources, is a way to facilitate, streamline and increase the number of migrations to I4.0.

Retrofitting is classified into "Traditional" and "Intelligent". The former refers to the replacement of parts or subsystems to achieve the optimization of some process variables such as, for example, the reduction of maintenance time, the speed of machines and processes, as well as the accuracy of various tasks, among others; while the latter focuses on the low-cost adaptation of subsystems, machines or equipment already existing in companies [13], or as the integration of new technologies in legacy systems to enable the transition to I4.0 [14]. The idea of smart retrofitting is to stop thinking about the necessary substitution (complete replacement of a technology by a more modern one) of the old for the new in a company, but rather seeks to be the basis for implementing I4.0 projects taking into account the current equipment and, at the same time, tries to be a guide to support the changes that must occur within the organization. The basic premise of smart retrofitting is that retrofitting existing machines or equipment is the cheapest, most practical and functional strategy that can be implemented to align to I4.0 regardless of the size of the organization.

4. Guidelines for Smart Retrofitting

There is no single methodology to perform smart retrofitting as it can be applied to machines, production lines and cyber-physical systems and even factories, and depends largely on the "I4.0 objectives" that the company intends to achieve. For example, in [15] a smart retrofitting process for an industrial plant is described and the I4.0 objectives to be met were: safety and maintenance. The general steps for the development of smart retrofitting were: 1) Definition of objectives, 2) Definition of new variables, 3) Definition and installation of new software, 4) Acquisition and control of the platform, 5) I4.0 applications (Prediction, simulation, forecasting and safety) and 6) User interface. Steps 3 to 5 refer to the smart retrofitting process. Another work presents a methodology for smart retrofitting that includes four phases: 1) Analysis and Reverse Engineering, 2) Development, 3) Virtual Commissioning and Validation, and 4) Physical Retrofitting [16]. This proposal includes a digital twin developed in an emulator and is used to reduce planning errors and the modernization process can be tested virtually with this twin.

On the other hand, smart retrofitting has its limitations, risks and challenges, since incorporating disruptive technologies into production systems will require changes in the organization, such as hiring specialized personnel to operate the equipment and install information security systems, among others. The installation of modernized systems in production lines also implies upgrading networks and facilities and, above all, configuring old networks with new ones. The operation of modernized systems under I4.0 criteria and traditional machines (without the incorporation of disruptive technologies) on the same production line has significant technical challenges that must be overcome.

4.1. General Guidelines for Smart Retrofitting in Legacy Systems.

Although smart retrofitting will bring significant improvements to production processes, it will also imply important challenges for the entire organization that will operate and manage the modernized system, since the insertion of disruptive technologies to perform the retrofitting and its operation under

Industry 4.0 criteria will demand a new preparation and new functions for operators, managers and owners of the companies and even for suppliers. For this reason it is necessary for companies to be aware that the implementation of the I4.0 vision is not a simple change of technology for another, but involves profound changes in production processes. Smart retrofitting, when guided by the I4.0 objectives, will necessarily bring about progressive and even disruptive changes in the organization. The following is a guide to help organizations seeking to implement a smart retrofitting project in their production systems:

Phase 0: Guide to assist the organization in a smart retrofitting project.

1) The organization must have clarity on the needs of the business to join the I4.0 vision.
2) The tangible benefits that operating under the I4.0 philosophy will bring to the company or corporation must be identified.
3) Evaluate modernization costs and amortization times.
4) Design a strategic plan for the modernization of the plant or factory.
5) Assess the new skills and competencies of operators, managers and executives that will be required to operate and manage the upgraded systems.
6) Identify and select the production line that is a priority to start the intelligent reconversion.
7) Evaluate the device, appliance, machine or system on the production line that will be selected to perform the smart retrofit.
8) Carry out smart retrofitting on the selected system.
9) Train or hire the human resources that will be in charge of the operation of the modernized system.
10) Develop a specialized maintenance plan for the modernized system.

To a certain extent, smart retrofitting requires the integrated application of several existing methodologies, such as reverse engineering (methodology applied to know the characteristics and properties of a reference system [17]), original equipment design, mechanical and mechatronic design, methodology for the design of electronic systems, software engineering methods and information and communication technologies (ICTs), among others. In order to propose a general guideline for smart retrofitting, the following considerations should be taken into account:
In general terms the refurbishment problem depends on the state and technological functionality of the reference machine, device or cyber-physical system, that is:

- Of the operating status of the machine, device or cyber-physical system.
- From the cost of traditional retrofitting and smart retrofitting.
- The number and type of I4.0 requirements.
- The availability of enhancement technologies on the market.
- The status of communications and internet connectivity at the site of operation of the upgraded system.
- The flexibility of the machine, device or cyber-physical system (degree of reconversion).
- From the principle of interoperability.
- Of the optimization capacity required by I4.0.
- The degree of digitalization that the system to be reconditioned supports.

These considerations must be taken into account when retrofitting for I4.0 and also the following aspects:

1) In general, the machines, devices or cyber-physical systems to be retrofitted are physical systems, so Reverse Engineering is an essential tool to determine various characteristics, operations, design and manufacturing unknown to those who are going to perform the retrofitting, since usually the retrofitting tasks are not performed by the companies of origin of the technology.
2) Retrofitting is not only performed on machines, devices or cyber-physical systems, but also extends to energy or power systems or existing computer systems that support the machines before retrofitting.

In order to propose the conceptual guide for smart retrofitting, it is necessary to define the following concepts: *A reference model* refers to the device or system selected for retrofitting. *A representative model* is the reference model generated from a reverse engineering process while a *converted representative conceptual model* refers to the representative model assessed with the I4.0 objectives. The phases and steps proposed for the development of smart retrofitting are:

Phase I: Retrofitting Analysis and Decision Making

11) Analysis of the requirements and objectives of the modernization (Retrofitting).
12) Selection of the general criteria to be met for the I4.0 philosophy.
13) Conceptual analysis of the reference model to be modernized.
14) Analysis of the peripheral technology of the system to be modernized.
15) Preliminary assessment of retrofitting costs.
16) Development of a technical report for decision making.

Phase II: Analysis of traditional retrofitting

17) Reverse engineering of the reference system.
18) Technical analysis of the subsystems to be reconverted: technical improvement of systems, reduction of energy consumption, monitoring of system performance, environmental impact analysis, technical training of personnel and real economic study, among others.

19) Technical documentation of the system to be modernized.
20) Conceptual design of the reference model (representative model).

Phase III: Analysis of retrofitting under the I4.0 objectives)

21) Analysis of the subsystems to be reconverted from the representative model based on I4.0 criteria.
22) Selection of disruptive technologies for retrofitting: smart sensors, Internet of Things technologies, smart boxes, edge and fog computing systems, cloud computing, big data, cybersecurity equipment and digital twins, among others.
23) Study on the migration of the required software.
24) Conceptual design of the representative model + disruptive technologies (converted representative conceptual model).

Phase IV: Retrofitting development

25) Physical reconstruction of the converted representative conceptual model.
26) Development of multiphysics or data analysis models to meet I4.0 requirements (Design of digital twins if needed).
27) Adaptation of the selected disruptive technologies to the representative model (converted representative conceptual model).
28) Migration of the required software.
29) Development of the converted representative physical model.
30) Reconfiguration of the networks (interconnection of the original networks with the networks required for retrofitting).
31) Final configuration of the reconverted system (private networks and cloud).
32) Installation of cybersecurity systems.

Phase V: Testing and validation of retrofitting

33) Verification and validation tests of the reconverted system.
34) Development of operation and maintenance manuals for the reconverted system.
35) Installation and/or industrial transfer of the modernized system.
36) Changes and innovation of the reconverted system.

It is necessary to consider that not all the phases and steps described above should be applied to perform smart retrofitting, since, for example, there are I4.0 applications that do not require the design and implementation of a digital twin or other applications will depend on the state in which the reference system to be improved is. Smart retrofitting is based on traditional retrofitting, therefore it is of utmost importance that the selected machine or system is evaluated in detail in order to make the operational changes for both the correct operation of the machine and the systems that must operate under the I4.0 criteria. Therefore, reverse engineering becomes a priority because with its systematic application it is possible to know in detail the operation and the parts that make up the system or machine to be reconditioned because in general companies have machinery that only operate and maintain them, but do not know their design.

Phase III of the proposed guide is the one that characterizes the smart retrofitting design. In this phase, the incorporation of smart sensors that will provide information on the physical state of the system is analyzed, and the need to use artificial intelligence algorithms and big data, among other technologies, is studied.

In the whole process of installing I4.0 systems, it is necessary to configure all the systems involved which are roughly speaking: old communication and control installations, new IIoT (Industrial Internet of Things) equipment installations and refurbished systems [6]. It is necessary to implement standardized protocols under interoperability schemes and networks operating at the edge (Edge computing) and fog networks (Fog computing) and to install and test cybersecurity equipment for information protection. Fig. 1 shows a general outline of the smart retrofitting roadmap.

Finally, it is worth mentioning that the phases proposed in this work to carry out smart retrofitting only represent a general conceptual guide for the modernization of a machine or system under I4.0 criteria and requirements and should not be considered as a general methodology.

5. Conclusions

In this article a general guide to perform a smart retrofitting to legacy machines and systems was presented. The presented guide can be used as a proposal to follow to carry out a smart retrofitting process to legacy systems and each phase can be expanded in detail according to the needs of each project. The first four phases of the proposed guide refer to planning and conceptual engineering, while the last two phases relate to the physical and technical implementation of retrofitting. Although traditional retrofitting is different from smart retrofitting, the latter depends on the former, so it is of utmost importance to apply techniques to know the precise information of the system selected for modernization. In this part, reverse engineering plays a relevant role since it allows to know the configuration, design principles and operational status of the legacy systems. The I4.0 objectives and costs set the guidelines for the design and development of smart retrofitting, so their selection and analysis must be studied in detail. It is worth mentioning that the main objective of I4.0 is to optimize manufacturing 3.0, therefore, smart retrofitting should seek that the physical and/or digital assets to be modernized, have improvements that allow optimizing their functions and operations.

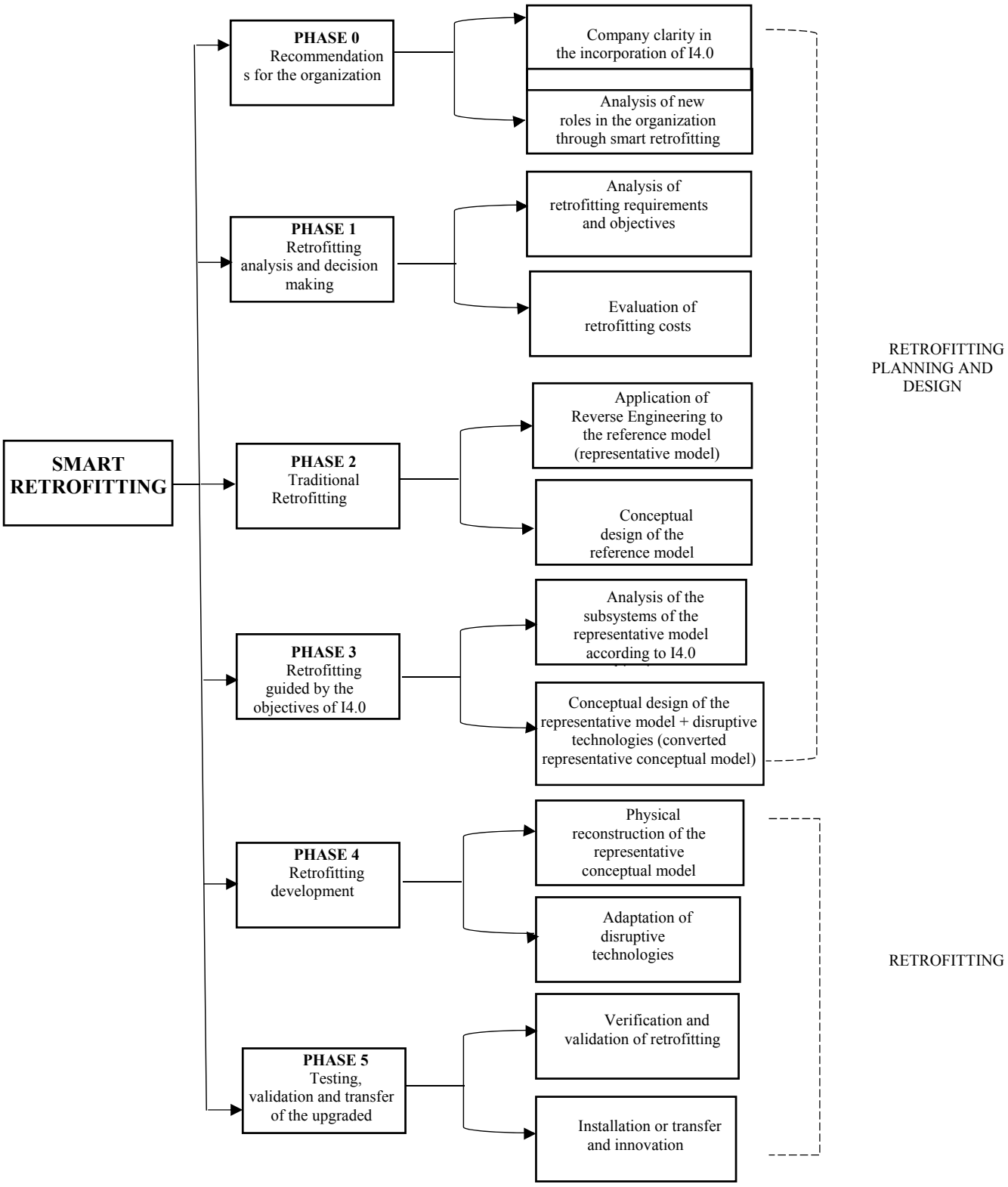


Fig. 1. General scheme of the smart retrofitting guide.

Smart retrofitting involves major technical changes to the production lines and above all requires organizational changes so that the operation and management of the modernized system can be effective and efficient, and meet the requirements of the I4.0 vision.

Acknowledgements

The authors would like to thank the Universidad Tecnológica del Sur de Sonora, the Universidad La Salle Noroeste, the Universidad Estatal de Sonora, the Instituto Tecnológico Superior de Cajeme and the Universidad Nacional Autónoma de México for the facilities granted for the development of this work.

References

[1]. Gajdzik B., Grabowska S., Saniuk S. A Theoretical Framework for Industry 4.0 and Its Implementation with Selected Practical Schedules, *Energies, 14,* 2021, pp. 940. https://doi.org/10.3390/en14040940.

[2]. Chancharoen R., Chaiprabha K., Wuttisittikulkij L., Asdornwised W., Saadi M., Phanomchoeng G. Digital Twin for a Collaborative Painting Robot, *Sensors*, 23, 2023. p. 17. https://doi.org/10.3390/s23010017.

[3]. Jimenez E., Ochoa F.J., Luna, G., Muñoz F., Cuenca F., Maciel M.A. Competency-based Education of the Mechatronics Engineer in the Transition from Manufacturing 3.0 to Industry 4.0, in *Proceedings of the 2nd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 (ARCI' 2022)*, Andorra la Vella, Andorra, 2-3 February 2022, pp. 84-87.

[4]. Jimenez E., Cuenca F.. Luna G., Ochoa F. J., Maciel Monteón M. A., Muñoz F., Limón P. A. Technical Considerations for the Conformation of Specific Competences in Mechatronic Engineers in the Context of Industry 4.0 and 5.0. *Processes.*10, 2022, pp. 1445. https://doi.org/10.3390/pr10081445.

[5]. Torres P., Dionísio R., Malhão S., Neto L., Gonçalves, G., Machinery Retrofitting for Industry 4.0, in: Machado, J., Soares, F., Trojanowska, J., Yildirim, S. (eds.), Innovations in Mechatronics Engineering. icieng 2021. Lecture Notes in Mechanical Engineering. *Springer, Cham*. 2022, pp. 2013-2020 https://doi.org/10.1007/978-3-030-79168-1_20

[6]. Vijay S.S., Machado D., Kumar A. A., Plapper, P. Retrofitting of legacy machines in the context of Industrial Internet of Things (IIoT), *Procedia Computer Science*, Vol. 200, Issue C, 2022, pp 62-70. https://doi.org/10.1016/j.procs.2022.01.205

[7]. Chen C., Ming T., Tanc R. R., Kimhua T., Ondrej V. Industry 3.5 for Sustainable Transition and Total Resource Management, *Resources, Conservation and Recycling*, Vol. 152, 2020, pp. 104482. https://doi.org/10.1016/j.resconrec.2019.104482.

[8]. Niemeyer C. L., Gehrke I., Müller K., Küsters D., Gries T. Getting Small Medium Enterprises started on Industry 4.0 using retrofitting solutions, *Procedia Manufacturing*, Vol 45, 2020, pp. 208-214. https://doi.org/10.1016/j.promfg.2020.04.096.

[9]. Ermini S., Bernabini D., Burresi G., Lorusso M., Rizzo A. Human-Centered Retrofitting, in *Proceedings of the ECCE 2021: European Conference on Cognitive Ergonomics 2021*, Siena, Italy, 2021, pp. 1-6.

[10]. Waite M. Paperback Oxford English dictionary: *Oxford University Press*, 2012.

[11]. Lins T., Rabelo R. A., Cyber-physical production systems retrofitting in context of Industry 4.0, *Computers & Industrial Engineering*, 139, 2020, pp. 1-19.

[12]. Lins T., Oliveira R. A. R., Correia L. H., Silva J. S. Industry 4.0 Retrofitting, in *Proceedings of the VIII Brazilian Symposium on Computing Systems Engineering (SBESC' 2018)*, Salvador, Brazil, 2019, pp. 8-15.

[13]. Al S. S. H., Kuhnhen C., Engel B., Schiller, M. Smart retrofitting of machine tools in the context of industry 4.0, *Procedia CIRP*, 88, 2020, pp. 369-374.

[14]. Jaspert, D., Ebel, M., Eckhardt, A., et al., Smart retrofitting in manufacturing: A systematic review. *Journal of Cleaner Production*, 312, 2021, pp. 127555.

[15]. Di Carlo F., Mazzuto G., Bevilacqua M., Ciarapica, F.E. Retrofitting a Process Plant in an Industry 4.0 Perspective for Improving Safety and Maintenance Performance, *Sustainability,* 13, 2021, pp. 1-18.

[16]. Ayani M., Ganebäck M. H.C. A Ng. Digital Twin: Applying emulation for machine reconditioning, *Procedia CIRP*, 72, 2018, pp. 243-248.

[17]. Jimenez E., Acosta M., Luna G., Lucero B., Delfín J. J, Velásquez L. A., Reverse Engineering and Straightforward Design as Tools to Improve the Teaching of Mechanical Engineering. In: Abdulwahed, M., Bouras, A., Veillard, L. (eds.) Industry Integrated Engineering and Computing Education. *Springer, Cham*. 2019, pp. 93-118. https://doi.org/10.1007/978-3-030-19139-9_7.

(5456)

Federated Learning Model of Multi Key Homomorphic Encryption on the basis of Internet of Things

Ran. Zhai [1,2,3], Xuebin Chen [1,2,3], Ruikui Ma [1,2,3] and Langtao Pei [1,2,3]
[1] North China University of Science and Technology, 063210 TangShan, HeBei, China
[2] Hebei Key Laboratory of Data Science and Application, 063210 TangShan, HeBei, China
[3] Tangshan Key Laboratory of Data Science, 063210 TangShan, HeBei, China
Tel.: 15903159789
E-mail: chxb@ncst.edu.cn

Summary: With the rapid development of the Internet of Things, a large number of mobile devices need to process massive data cooperatively. However, the privacy protection of the data at the participating end is crucial during collaborative training. Federated Learning as a Distributed Machine Learning Framework, the data can be kept locally to complete collaborative training. However, during parameter transmission in federated learning, parameter leakage will lead to raw data leakage. Encrypting parameters using homomorphic encryption solves this problem. However, in homomorphic encryption, multiple participants use the same key, so once any participant disclose the public key, the entire model security will be affected. Therefore, multi key homomorphic encryption is applied to the federated learning model. In multi key homomorphic encryption, participants use their own public keys to encrypt the parameters and jointly use the private keys to decrypt. This process avoid the impact on the model security caused by the key disclosure of a single participant. The experimental results show that the security of federated learning model on the basis of multi key homomorphic encryption technology has better security than traditional federated learning models. Meanwhile, the accuracy of training was not affected.

Keywords: Federated learning, Homomorphic encryption, Cryptography, Multi key homomorphic encryption, Internet of things.

1. Introduction

The rapid development of the Internet of Things, the Internet of Vehicles and 5G technology makes it possible to conduct large-scale joint training of massive intelligent terminal devices in the network [1-2]. However, when massive intelligent terminal devices in the network conduct joint training, data privacy protection is crucial. As a distributed machine learning framework, federated learning has effectively solved this problem, which was proposed by Google in 2016 [3], since this technology can enable data to be saved locally to participate in collaborative training of federated learning model, and effectively solve the problem that data owners are unwilling to share their own data[4]. Nevertheless, in the procedure of transferring parameters in federated learning, there may be the problem of raw data leakage caused by parameter leakage [5-7]. In order to solve this problem, many researchers have studied the privacy protection of parameters. Differential privacy, homomorphic encryption, secure multi-party computing and other technologies are applied to the federated learning model for privacy protection. Researches proposed the federated learning technology on the basis of homomorphic encryption [8]. However, in homomorphic encryption, a unified public key is used for encryption, and the public key may be leaked during the transmission between participants [9]. At the same time, there may be malicious participants obtained the public and private keys [10]. In order to tackle this problem, we propose a federated learning framework on the basis of multi key homomorphic encryption [11]. In 1998, Hoffstein et al. proposed the NTRU scheme [12]. In 2012, López Alt et al. took the homomorphic computation of multi-user ciphertext data in the cloud environment as the application scenario, and gave the MKFHE password for the first time [13].The first MKFHE scheme LTV12 based on NTRU cryptosystem is constructed. Its security is based on RLWE problem and DSPR Question. Because the ciphertext structure of NTRU encryption scheme naturally supports the homomorphic computation of multi key ciphertext, it can be used well to construct the MKFHE scheme. In 2013, Gentry et al. proposed the approximate eigenvector method and constructed the learning with errors based on it [14]. The ciphertext homomorphic addition and multiplication of this scheme only need simple matrix addition and multiplication. It has the advantages of simple homomorphic operation, no need to provide additional calculation keys, etc. At the American Secret Conference in 2015, Clear and McGoldrick extended the GSW13 scheme to multiple identity (multiple key) settings, and proposed the first GSW type MKFHE scheme with selective safety under the standard model CM15 [15]. At the 2016 Eurosecret Conference, Mukherjee and Wichs simplified the ciphertext expansion process of CM15 and proposed the MKFHE scheme MW16 [16], which allows one round of distributed decryption of multi key ciphertext, and can further construct two rounds of MPC protocol. It can resist the attack of any semi malicious adversary, and its security has been proved. CM15 and MW16 need to set the number of users participating in homomorphic computing in advance.

In the process, the number of participants is limited. At the American Secret Conference in 2016, Brakerski and Perlman proposed a fully dynamic MKFHE scheme BP16 [17]. The scheme allows new participants to be dynamically added to homomorphic operations, so the number of participants does not need to be set in advance, and the length of ciphertext is extended. The degree only increases linearly with the number of participants.

In the procedure of data transmission, the multi key homomorphic encryption technology is used to encrypt the parameters of the model. Each participant encrypts the parameters with their own public key, and sends the encrypted parameters to the parameter server. During the calculation of the parameter server, the encrypted parameters are used for operation, and the updated parameters are sent to the participants in the form of ciphertext. The participants jointly use their own private keys for decryption, which can tackle the problem of raw data leakage caused by parameter leakage or the attack behavior among the participants, and realize the model training without disclosing the raw data.

The contributions are as follows:

Federated learning model is used for information transmission in the process of information interaction among participants of the Internet of Things, the parameters are encrypted using multi key homomorphic encryption technology which has the function of ciphertext calculation. After encrypting the parameters, the parameters are transmitted to the parameter server, which effectively avoids data leakage in the transmission procedure and data leakage in the parameter server.

In the multi key homomorphic encryption technology, participants use different public keys to encrypt parameters, and then cooperate to decrypt those using private keys, which effectively avoids the problem of data leakage of participants.

The experimental results show that the security of federated learning model of multi key homomorphic encryption based on internet of things is improved without affecting the accuracy.

2. Basic Theory

2.1 Federated Learning

Federated learning was first proposed by Google in 2016. It is a distributed machine learning framework [18]. The data owner can store the data locally, and then complete the collaborative training of the model without sharing. It effectively tackles the problem that most of the data are scattered in different institutions and the data owners are unwilling to share the data [19].

The federated learning framework can save the data locally, use the raw data locally for model training, return the parameters to the central server after the training, and the central server will return the updated parameters to the local [20]. The specific process is shown in Fig. 3.

The working principle of federated learning is as follows:

1) The parameter server sends the initial model to the participants;

2) Participants use their own data sets locally to train the model;

3) Participants send the training parameters to the parameter server;

4) The central server updates the parameters of the global model;

5) The central server returns the parameters of the global model to the participants;

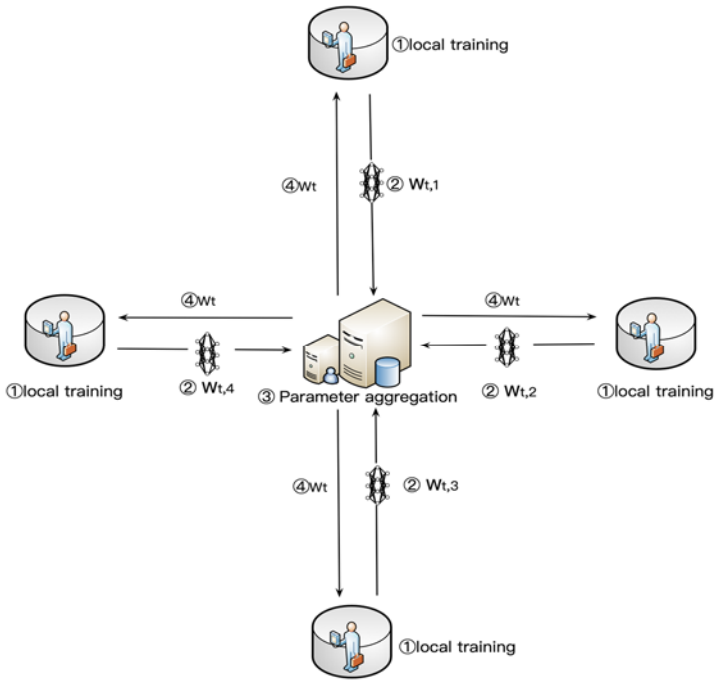


Fig. 1. Federated learning process.

The parameter server first sends the initial model and weight parameters to the participants, as shown in process □. The participants use local data to train on the model and update the model weight parameters, as shown in process □. The participants return the updated global model weight parameters to the parameter server, as shown in process □. The parameter server aggregates the parameters sent by each participant, as shown in process □.

2.2 Homomorphic Encryption

In 1978, Rivest put forward the concept of homomorphic encryption for the first time. Homomorphic encryption refers to computing data first, and then encrypting the result of the operation to obtain result A, and then encrypting the original data, and computing the encrypted data to obtain result B. Result A and result B are the same, and the security of homomorphic encryption depends on the difficulty of large integer decomposition [21]. Homomorphic encryption is divided into additive homomorphic encryption and multiplicative homomorphic encryption [22]. All homomorphic encryption technology is a symmetric encryption algorithm [23] proposed by Craig Gentry. All homomorphic encryption technology is a combination of additive homomorphic encryption and multiplicative homomorphic encryption. Homomorphic encryption plays a very important role in protecting users' privacy. Homomorphic encryption is a widely used privacy protection method using cryptography principles.

The homomorphic encryption process is as follows:

1) First, use the key generation function KeyGen to generate the key;

2) Encrypt the data with the key using the Encrypt function to generate ciphertext;

3) Use the Evaluate function to process the ciphertext with a specific function, so that the result is equal to the result of processing the original data with a function first and then encrypting it;

4) Use the Decrypt function to decrypt.

2.3 Multi Key Homomorphic Encryption

López Alt et al. [13] proposed the concept of multi key fully homomorphic encryption (MKFHE). In multi key homomorphic encryption, ciphertext with multiple keys can be calculated, and users with multiple keys can cooperate to decrypt. At the same time, lattice cryptography has the characteristics of resisting quantum attacks. The emergence of quantum computing makes it possible to crack complex ciphers. Multi key homomorphic encryption based on lattice can effectively resist the cryptography threat brought by quantum computing. However, one limitation of multi key homomorphic encryption is that the number of participants needs to be known in advance. For many cooperative operations, the number of participants will change constantly. To solve this problem, Pratyay Mukherjee et al. [24] proposed a multi key homomorphic encryption algorithm that participants can dynamically join.

The definition of multi key homomorphic encryption [13] is as follows:

Definition 1:

1) Each party encrypts its input with its own key and broadcasts the ciphertext. Then, all parties can homomorphic calculate the output multi key encryption.

2) Each party uses its key broadcast to decrypt the part of the output. Partial decryption can be combined to recover the output in plaintext.

The working process of multi key homomorphic encryption algorithm is as follows:

Initialization process: Given security parameters, return a public parameter.

Key generation process: input public parameters to generate public key, private key and key for homomorphic calculation.

Encryption process: input public key and plaintext to be encrypted, and output ciphertext.

Homomorphic calculation process: input the Boolean circuit, ciphertext, public key set, and key set for homomorphic operation. Output ciphertext after homomorphic operation.

Decryption process: input ciphertext and private keys of multiple participants, and output plaintext.

3. Materials and Methods

Facing the problem of raw data leakage caused by parameter leakage that may occur in the transmission procedure of federated learning, a federated learning model on the basis of multi key holomorphic encryption is proposed. In the training procedure of federated learning, participants use the multi key holomorphic encryption algorithm after using local data to participate in the training of federated learning. Each participant uses its own public key to encrypt the model parameters generated by itself, and sends the encrypted parameters to the parameter server. Because homomorphic encryption has the function of ciphertext calculation, the parameter server aggregates the encrypted parameters, and sends the calculated encrypted data to each participant. Each participant cooperates to decrypt the parameters using his own private key, and uses the decrypted data to participate in new round of iterative training for federated learning.

The central server encrypts the initial model and weight parameters of the initial model using the AES encryption algorithm, and sends them to each participant. After decrypting the initial model and weight parameters, the participants use local data for training. After training, participants encrypt the training parameters with their own public key and send the encrypted parameters to the parameter server. The parameter server performs homomorphic calculation on the encrypted parameters sent by the participants, and sends the calculated results to each participant. Participants cooperate to decrypt the parameters using their own private keys. This model

is applied to the information interaction process of the Internet of Things, which effectively solves the security problem of client cooperation training in the Internet of Things.

The working procedure diagram of MFHEFL (Federated Learning Model on the basis of Multi Key Homomorphic Encryption) model is as follows (Fig. 2).

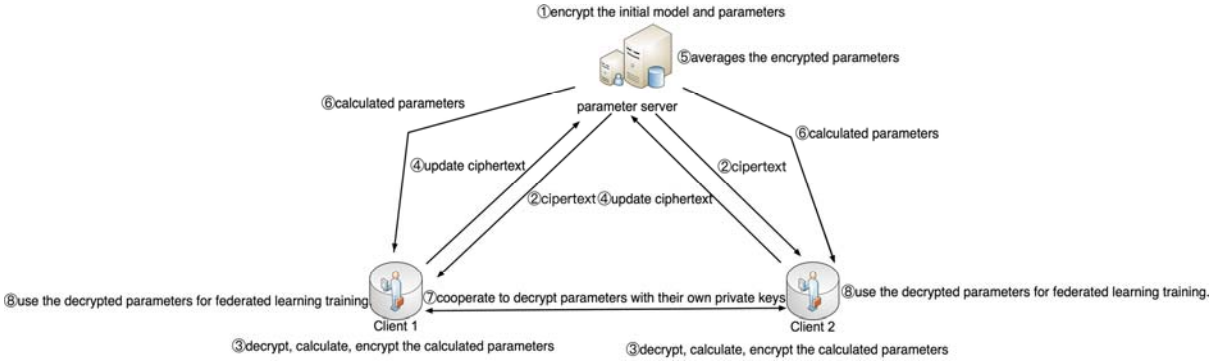


Fig. 2. Working procedure diagram of MFHEFL model.

The procedure of MFHEFL is mainly separated into the following steps.

Step 1: The information transmission between the participants shall be encrypted by AES, parameter server sends AES key to clients;

Step 2: Communication relationship shall be established among the clients to jointly discuss the initial model information suitable for this training, clients and parameter server uses AES encryption to discuss the parameters of federated learning;

Step 3: Parameter server encrypts initial model and parameters using AES key;

Step 4: Parameter server sends ciphertext to clients;

Step 5: Clients decrypt the initial model and parameters using AES key;

Step 6:Clients train on the model;

Step 7: Clients use their own public key to encrypt the training parameters, and send the encrypted training parameters to the parameter server;

Step 8: The parameter server performs homomorphic operation on the ciphertext sent by each participant, and returns the calculated parameters to each participant;

Step 9: Each participant uses its private key to decrypt the ciphertext, and then uses the decrypted parameters for subsequent training;

Step 10: Iterate Steps 6 and 9 until the set termination conditions are met.

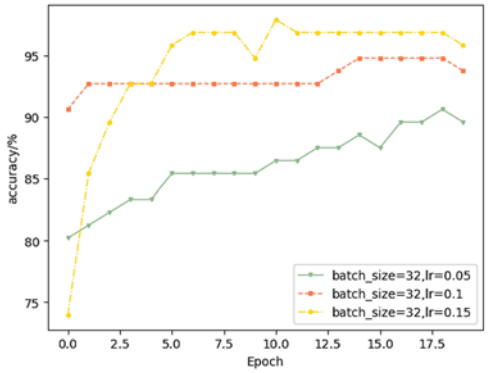
The code in this paper is written in python and C++ in the python and click compilers. The operating system is macOS 10.15.4, and the processor is 1.8 GHz dual core Intel Core i5 with 8GB of memory. The data set adopts cifar10 data set and break data set, and the download address is http://www.cs.toronto.edu/ ~Kriz/cifar.html and https://archive.ics.uci.edu/ml/datasets/Breast +Cancer+Wisconsin+%28Diagnostic%29. The data set includes four parts: training set samples, training set labels, test set samples, and test set labels. The training set contains 50000 samples, and the test set contains 10000 samples. Each picture in the dataset is represented by 32 * 32 pixels, and each pixel is represented by a gray value. The breast data set is Kent post biomedical data set - breast cancer. There are 699 pieces of information in the data set, 458 pieces of benign data and 241 pieces of malignant data in the data set. This data set is a numerical data set.

Table 1. Pseudocode of MFHEFL algorithm.

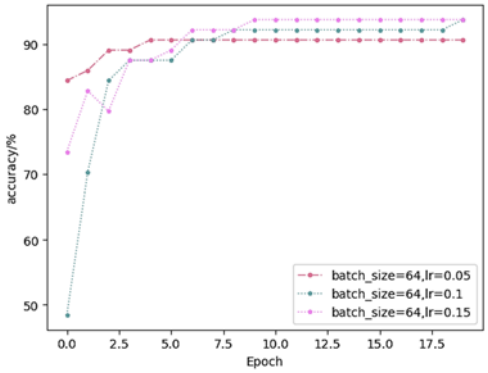
Algorithm 1 MFHEFL algorithm
1:for $i=1$ to n
2: for $j=1$ to n
3: Server executes:
4: initialize $\omega 0$//$\omega 0$ is a model parameter
5: $\omega 0$'=EncAES($\omega 0$);
6: Send $\omega 0$' to clients;
7: for each round t=1,2,...do;
8: S_t =(random set of max$^{(C \cdot K, 1)}$clients);
9: for each client $k \in S_t$ in parallel do;
10: $\omega_{t+1}^k \leftarrow ClientUpdate(k, \omega_t)$;
11: $\omega_{t+1} \leftarrow \sum_{t=1}^{K} \frac{n_k}{n} \omega_{t+1}^k$;
12: Send Wt+1 to clients;
13: ClientUpdate(k,w): //Executed on client k
14: Dec($\omega 0$);
15: for each local epoch i from 1 to E do
16: batches(data P_k split into batches of size B)
17: for batch b in batches do
18: $\omega \leftarrow \omega - \eta \nabla l(\omega; b)$
19: Enc(ω)
20: return ω to server
21: Coorperate to dec
22: end for
23: end for

In federated learning of the Internet of Things, each participant needs to complete the training under the same model conditions. Thus, we need to give each participant preset initial model conditions. In this paper, the federated learning model training parameters are set as follows: the number of iterations of the local model is 3, the number of iterations of the global model is 20, and the number of clients is 10. The Cifar 10 and breast data sets were selected as the experimental data sets.

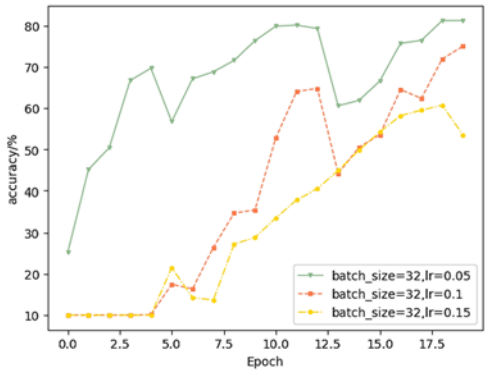
Experiments were conducted with different learning rates lr and batchsize.



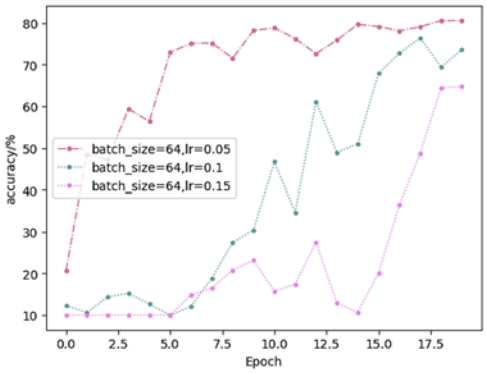
(a)Comparison of prediction accuracy of different LR when batchsize=32 on the Breast data set



(b)Comparison of prediction accuracy of different LR when batchsize=64 on the Breast data set



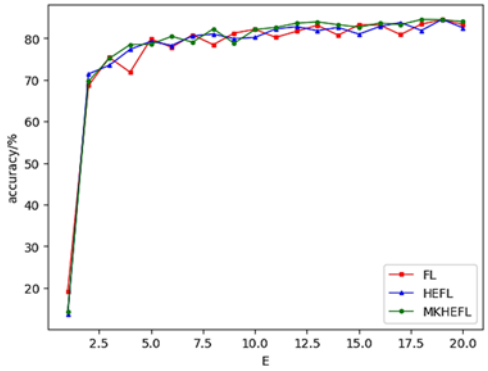
(c)Comparison of prediction accuracy of different LR when batchsize= 32 on the Cifar10 data set



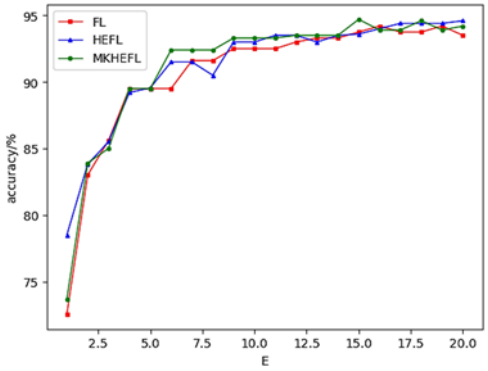
(d)Comparison of prediction accuracy of different LR when batchsize=64 on Cifar10 data set

Fig. 3. Comparison of prediction accuracy of different LR on the Breast data set.

The experimental results show that on the breast dataset, when batchsize=32 and learning rate (LR)=0.15, the model accuracy is higher, so the parameters in the comparison experiment are set as batchsize=32 and LR=0.15 on breast data set. On the cifar10 dataset, when batchsize=64 and LR=0.05, the model accuracy is higher. Therefore, in the comparison experiment, the parameters are set as batchsize=64 and LR=0.05 on cifar10 data set.



(a)Comparison of prediction accuracy on the Cifar10 data set



(b)Comparison of prediction accuracy on the Breast data set

Fig. 4. Comparison of prediction accuracy between FL, HEFL and MKHEFL.

The parameters on the breast dataset are set to batchsize=32, lr=0.15, and the parameters on the cifar10 dataset are set to batchsize=64, lr=0.05. It can be seen from Fig.3. that the prediction accuracy of FL, HEFL and MKHEFL is not much different. At the same time, the security of the federated learning model based on multi key homomorphic encryption technology has been effected improved.

5. Discussion

In the training procedure in federated learning of the Internet of Things, the public and private keys of the participants are generated locally using random numbers, and the public and private keys are only known by the participants themselves, which avoids the leakage of keys.

Step 1: Parameter server encrypts the initial model and the weight parameters of the initial model using the AES encryption algorithm and sends them to each participant, which effectively prevents leakage during the transmission of the initial model and the weight parameters;

Step 2: Participants decrypt the initial model and weight parameters. Participants use local data to train in federated learning locally. After the training is completed, the model parameters encrypted with the public key are sent to the parameter server. This procedure avoid the local data of participants being leaked.

Step 3: The encrypted parameters are used for transmission to avoid parameter leakage caused by attacks in the parameter transmission procedure, so that the raw data is inferred.

Step 4: Parameter server averages the model weight parameters sent by each participant and sends them to each participant. This procedure avoid the parameter server disclose parameters.

Step 5: Participants cooperate to decrypt using their own private keys. This procedure avoids parameter leakage caused by malicious behaviors of participants.

The security of LWE / RLWE problem is on the basis of the difficult problems in the underlying lattice, that is, the security of LWE / RLWE is regulated to the difficult problems in the lattice. If there is a polynomial time algorithm that can tackle the LWE / RLWE problem, the underlying lattice difficulty problem can also be solved. Otherwise, it is not necessarily true. According to the LWE estimator, the security parameter reaches 110 bits. The security of MFHEFL model has been effectively improved.

6. Conclusions

Aiming at the possible data leakage in the procedure of federated learning training of the internet of things, a federated learning model on the basis of improved multi key homomorphic encryption on internet of things is proposed. In the information transmission stage of the participants of the Internet of Things, each participant encrypts the parameter information with his own public key and transmits the encrypted data, which can effectively solve the privacy leakage in the procedure of parameter transmission of federated learning. The improved multi key homomorphic encryption is used to encrypt the parameters. All participants need to cooperate to decrypt the parameters, which effectively solves the problem of data leakage of participants. Experimental results indicate that the MFHEFL model has better security performance than the traditional federated learning model.

References

[1]. Wei, Wei, et al., GI/Geom/1 queue based on communication model for mesh networks, *International Journal of Communication Systems* 27, 11, 2014, pp. 3013-3029.

[2]. Wei, Wei, et al., Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network, *Information Sciences,* 408, 2017, pp. 100-114.

[3]. Jakub Konečný, et al., Federated Learning: Strategies for Improving Communication Efficiency, 2016, arXiv:1610.05492.

[4]. J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang and M. Guizani, Reliable Federated Learning for Mobile Networks, *IEEE Wireless Communications*, Vol. 27, No. 2, April 2020, pp. 72-80.

[5]. N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen and C. S. Hong, Federated Learning over Wireless Networks: Optimization Model Design and Analysis, in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM 2019)*, 2019, pp. 1387-1395.

[6]. Truex, S., et al., A Hybrid Approach to Privacy-Preserving Federated Learning, in *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (AISec'19)*, November 2019, pp. 1–11.

[7]. McMahan, Brendan, et al., Communication-efficient learning of deep networks from decentralized data, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[8]. Gentry, and Craig. Fully homomorphic encryption using ideal lattices, *ACM*, 2009, pp. 169-178.

[9]. Madi, A., et al., A Secure Federated Learning framework using Homomorphic Encryption and Verifiable Computing, in *Proceedings of the Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS),* 2021, pp. 1-8.

[10]. Hz, A, et al., Distributed additive encryption and quantization for privacy preserving federated deep learning., *Neurocomputing,* 463, 2021, pp. 309-327.

[11]. Chen, Hao, Ilaria Chillotti, and Yongsoo Song., Multi-key homomorphic encryption from TFHE, in *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security*, 2019.

[12]. Hoffstein, Jeffrey, Jill Pipher, and Joseph H. Silverman., NTRU: A ring-based public key cryptosystem., Algorithmic Number Theory. ANTS 1998. Lecture Notes in Computer Science, Vol, 1423. *Springer*, Berlin, Heidelberg, 1998.

[13]. López-Alt, Adriana, Eran Tromer, and Vinod Vaikuntanathan., On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption, in *Proceedings of the 44th Annual ACM Symposium on Theory of Computing,* 2012.

[14]. Gentry, Craig, Amit Sahai, and Brent Waters., Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based., in: Canetti, R., Garay, J. A. (eds.) Advances in Cryptology – CRYPTO 2013. CRYPTO 2013. Lecture Notes in Computer Science, Vol. 8042, *Springer*, Berlin, Heidelberg 2013.

[15]. Clear, Michael, and Ciaran McGoldrick., Multi-identity and multi-key leveled FHE from learning with errors., in: Gennaro, R., Robshaw, M. (eds.), Advances in Cryptology -- CRYPTO 2015. CRYPTO 2015. Lecture Notes in Computer Science, Vol. 9216. *Springer*, Berlin, Heidelberg, 2015, pp. 630-656.

[16]. Mukherjee, Pratyay, and Daniel Wichs, Two round multiparty computation via multi-key FHE, in: Fischlin, M., Coron, JS. (eds) Advances in Cryptology – EUROCRYPT 2016. EUROCRYPT 2016. Lecture Notes in Computer Science, Vol. 9666, Springer, Berlin, Heidelberg., 2016, pp. 735–763.

[17]. Peikert, Chris, and Sina Shiehian., Multi-key FHE from LWE, revisited., in *Proceedings of the 14th International Conference on Theory of Cryptography* Part II, Vol. 9986, October 2016, pp. 217–238.

[18]. Brakerski, Zvika, and Renen Perlman., Lattice-based fully dynamic multi-key FHE with short ciphertexts, In: Robshaw, M., Katz, J. (eds.), Advances in Cryptology – CRYPTO 2016. CRYPTO 2016. Lecture Notes in Computer Science, Vol. 9814. *Springer*, Berlin, Heidelberg., 2016, pp. 190–213.

[19]. Li, T., et al., Federated Learning: Challenges, Methods, and Future Directions., *arXiv:1908.07873*, 2019.

[20]. Kim, H., et al., Blockchained On-Device Federated Learning, *IEEE Communications Letters* 24, 6, 2020, pp. 1279-1283.

[21]. Yang, Qiang, et al., Federated machine learning: Concept and applications, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10, 2, 2019, pp. 1-19.

[22]. Paillier, Pascal., Public-key cryptosystems based on composite degree residuosity classes, In: Stern, J. (eds.) Advances in Cryptology — EUROCRYPT '99. EUROCRYPT 1999. Lecture Notes in Computer Science, Vol. 1592, *Springer*, Berlin, Heidelberg, 1999, pp. 223–238.

[23]. ElGamal, Taher, A public key cryptosystem and a signature scheme based on discrete logarithms, *IEEE Transactions on Information Theory,* 31, 4, 1985, pp. 469-472.

[24]. Gentry, Craig., Fully homomorphic encryption using ideal lattices, in *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing,* 2009, pp. 169-178.

(5819)

Distributed Virtual Commissioning Implementation of an Aeronautic Shopfloor

A. Abassi [1], J. H. Lugo [2], R. Balderas and N. Lassabe
Capgemini Engineering, Research and Innovation Department, Toulouse, France
E-mail: ahlam.abassi@capgemini.com, jesus-hiram.lugo-calles@capgemini.com

Summary: Virtual Commissioning uses digital models of devices and processes to verify, validate, and optimize code programming, component selection, and traditional industrial commissioning activities. To perform the validation process, a simulation involving control devices and process digital twins is required, leading to inherent computational complexity. Distributed simulation approach allows for simulation of complex systems by breaking down a large simulation into smaller, manageable parts that can be run simultaneously on separate processors, while still preserving the overall behavior and interactions of the system being simulated. This paper presents a distributed Virtual Commissioning solution for a spray paint process presented in UAV painting shop floor. The methodology for developing the implementation is described in detail: greenfield scenario generation, automation process, software toolchain development, selection of communication protocols, re-use of digital twins for extended applications, and complexity analysis. A set of 3d scenarios is used to demonstrate the result's performance.

Keywords: Virtual commissioning, Automation, Toolchain, Distributed simulation.

1. Introduction

Virtual commissioning (VC) is a method that allows the user to propose a commissioning solution using digital models of industrial systems and devices to accelerate and improve traditional processes [1-4]. The main benefits obtained from VC are the following:

- Early simulation, and verification and validation (V&V) of machine code by using different simulation approaches Hardware-in-the-Loop, Software-in-the-Loop and Model-in-the-Loop (HiL, SiL, MiL) without impacting production.
- Higher quality SW with optimized controls architecture and programming.
- Faster changes implementation while maintaining quality and reducing risks.
- Shortened overall commissioning time.
- Decrease manufacturing system lead time. VC solution methodologies have been widely studied, and standardization efforts regarding the complexity of the involved digital twin specification and V&V process used during the VC activities have been developed. [5] It is worth noting that the more specifications and realistic details are taken into account for developing the solution, the more accurate the representation of the system to be commissioned would be. However, the simulation of highly complex scenarios could lead to poor computer performance, hindering and sometimes, inhibiting a reliable validation process. In order to improve this, the VC developer could evaluate simulation complexity [6] to identify elevated consuming processes and objects in the scenario.

The aeronautic industry is characterized by its highest quality manufacturing and safety standards, leading to a constant demand of precise and accurate automated systems. This situation opens the door to the proposal of multiple Industry 4.0 green-field scenarios where robots can collaborate with humans to improve manufacturing processes.

2. Distributed Simulation Approach

Distributed simulation architectures provide better performance of industrial system representations. Several projects have successfully implemented this simulation scheme in the robotics field with different goals in mind: robot specific behaviors, either individually, as a swarm or within an organism [7]; distributed simulation system to validate the control of the real robot through the connection of GUI and digital twins [8]. It is worth noting that this type of simulation implies issues related to the hardware specifications of each computer in terms of available memory (RAM and ROM) as well as execution time, affecting data synchronization [9]. According to the results obtained from the previous studies, the following outcomes are:

- Improvement of current toolchain simulation parameters (i.e., time slices, cycle times, mathematical solvers, etc.).
- Selection of alternative SW/HW tools.
- Division of complex simulation into simpler ones.
- Communication protocol parameters selection and adjustment.

3. Use Case Development

The use case, which is an aeronautic painting shopfloor taken from our previous work presented in [10], consists of four painting robots (equipped with smart painting nozzles), two smart conveyors, and a

safety system, as shown in Fig. 1 and Table 1. A fairing used as a propeller support of a UAV taxi represents the working piece to be painted. With the development of this scenario, we try to address the need of hybrid automated environments, where robots can coexist and human workers in a safe manner; and robotize legacy industrial processes.

The painting process is described as follows:

1. A fairing is placed on conveyor C1
2. C1 moves the fairing until it reaches the position sensor PS1, stopping C1 and activating the spray painting process for robots LR1 and RR1.
3. After the painting activity of LR1 and RR1 is completed, C1 moves the fairing to conveyor C2.
4. C2 moves the fairing until it reaches the position sensor PS2, stopping C2 and activating the spray painting p After the painting activity of LR2 and RR2 is completed, C2 moves the fairing to its farthest end.
5. The safety system is always running in a parallel thread, allowing it to be activated at any time and provide an emergency stop reaction for a single device or the whole cell.

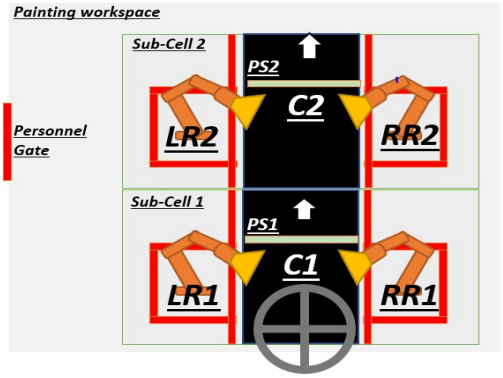


Fig. 1. Painting process schematic.

Table 1. BOM for each simulation scenario.

Component	Description	DELMIA	CoppeliaSim
Industrial robotic manipulator	6 DoF spray paint robotic arm	ABB IRB 6400R/2.5- 150 Quantity: 4	UR10 : Universal Robotics Arms Quantity: 4
Paint nozzle	Customized simulated spray paint nozzle	Quantity: 4	Quantity:4
Conveyor	Customized simulated conveyor	(11 m × 0.3 m) Quantity: 1	(4 m × 3 m) Quantity: 2
Position sensors	Simulated laser beam position sensor	Quantity:2 Diameter:10 mm	Quantity:2 Diameter:10 mm
Safety sensors	Simulated laser beam safety sensor	Quantity:18	Quantity:15
Worker mannequin	Programmable dummy to represent	Quantity:1	Quantity:1
Gates	Entry levels to the cell	Quantity:4	Quantity:1

The scenario was developed simultaneously on two different CAM software: DELMIA and CoppeliaSim, as shown in Fig. 2 and 3. Both software are meant for similar purposes but target different applications: industrial manufacturing for the first one, and robotics programming for the second. In order to cope with the interoperability between our modules, we implemented an OPC-UA network, consisting of a server running on SIMIT SP, and using the client capabilities of the rest of the software to attach to it. OPC-UA supports soft real-time data exchange, event handling, and historical data access, making it a suitable choice for industrial automation, control systems, and other applications that require reliable data exchange. A PLC is used to synchronize the tasks of all smart devices and control the safety system. The PLC can share data using OPC-UA, allowing a simple data mapping between control and semantic layers. The safety system was equally implemented on both scenarios, simple detection of position sensors (laser beams) around the robots and at the entry gates of each scenario in addition to the emergency stop button.

The painting process is defined in different manner according to the CAM used: surface modification using the robot surface simulation application in the first scenario and particles projection using a simple LUA code for the second one.

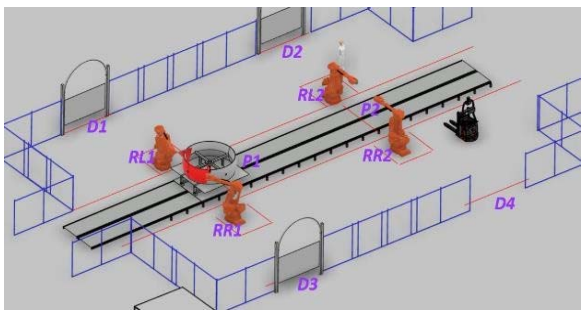


Fig. 2. Aeronautic painting cell 1 developed with Dassault Systems tools: DELMIA 3Dexperience and ControlBuild. This cell is characterized by the use of a single conveyor and four doors used by the shopfloor workers and forklifts.

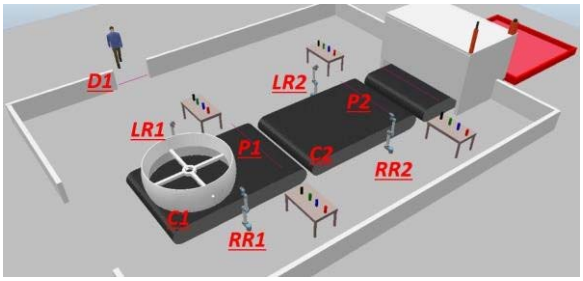


Fig. 3. Aeronautic painting cell 2 developed with CoppeliaSim EDU version and Python code. This scenario has a third conveyor that takes the painted fairing to an electric oven (white box) to accelerate the drying process.

Both scenarios were designed and tested individually using the toolchain shown in Fig. 4, where the digital twin of the process is connected via Control Build as a client to an OPC-UA server for the first scenario and via python in the second digital twin, working as a database, that receives and distributes information to other clients: a S7-1500 PLC (running on Siemens TIA Portal V17 and PLCSim Advanced 4.0) and a supervisory system consisting of a HMI system (Siemens WinCC) and SCADA platform (Fernhill).

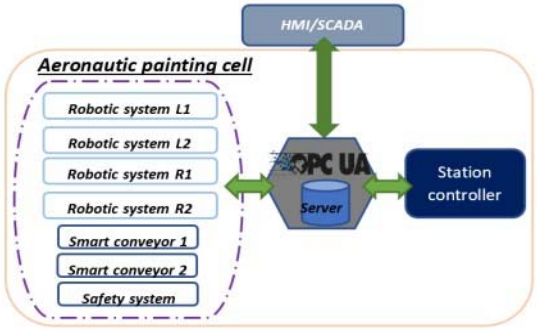


Fig. 4. Single scenario toolchain. Components of the painting system are surrounded by the purple dotted line.

To obtain better performance, we decided to divide the main system into subsystems. Thanks to OPC-UA technology, we could implement a stable industrial network between different computers allowing the interconnection of multiple simulations, see Fig. 5. The proposed toolchain consists of two workstations for the heavy computing processes and a remote PC to display system variables.

The proposed distributed approach, based on the one shown in [6], considers the modelled objects as sets of attributes and a total of 11 measures that needs to be determined for each system or toolchain. The proposed model takes several inputs into account (3D model, Programming parameters, time …). The proposed measured parameters are listed next. - M1: the number of the modelled objects.

- M2: the number of the connections among the modelled objects.
- M3: number of attributes of all the objects (Mass, volume, gravity).
- M4: Number of manually changed attributes (more complex), like detectability.
- M5: Number of attributes that are not inherited from parent object.
- M6: Summarizing the cyclomatic complexity of all the program blocks.
- M7: Total number of lines in a program. - M8: Required time duration
- M9: Number of running software in the toolchain
- M10: Real time estimation (create a weight parameter that effect the complexity) ○ 0: if system is not real time ○ 1: if system soft real time ○ 2: if system is hard real time
- M11: Number of cycles needed to complete the simulation.

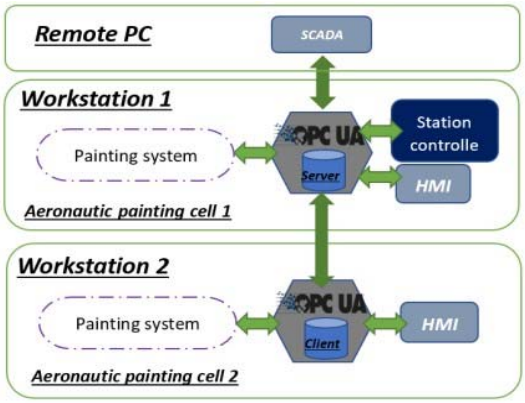


Fig. 5. Extended toolchain implemented for distributed simulation in multiple PC platforms.

We applied this scheme to each software used in the toolchain to have a better vision of the complexity of the desired system, see Table 2. It is noticeable that Delmia toolchain is faster than CoppeliSsim toolchain since the physics engines are different. A comparison between each individual toolchain complexity is shown in Fig 6, while Fig. 7 displays the complexity of each software used for the extended toolchain. It is worth noting that PLCSIM Advanced and SIMIT SP are the only software considered as soft real time. The simulation only requires one cycle to complete the painting tasks, taking 470 seconds in M11.

4. Simulation and Results

To test the efficiency of the safety system, an operator dummy was added to both scenarios: it was programmed to behave as a human worker to verify the reduction of smart components's speed. The results were acceptable since the signals are correctly sent along all the modules: Delmia to Control Build (OPCUA Client) directly connected to the OPC-UA server, while on the other hand, a second toolchain going from CoppeliaSim scenario to the python OPC-UA client, then, connecting to OPC-UA server, which shares the information to the PLC OPC-UA client, containing a simulated HMI for control and monitoring.

Table 2. Complexity analysis of the extended toolchain.

ET	M1	M2	M3	M4	M5	M6	M7	M8
DELMIA	112	203	100	88	188	19	175	450
CB	0	0	0	0	0	0	33	450
CoppeliaS	135	81	236	127	363	24	1569	470
Python	0	0	0	0	0	2	192	470
SIMIT SP	98	49	0	0	0	2	0	470
PLCSim Adv	0	0	0	0	0	0	0	470
TIA PORTAL	55	13	0	0	0	2	0	470
WINCC	52	50	0	0	0	0	0	470
TOTAL	400	346	336	215	551	47	1969	470

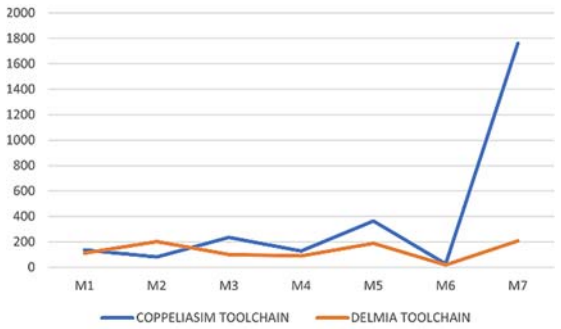


Fig. 6. Complexity comparison of individual toolchains**.**

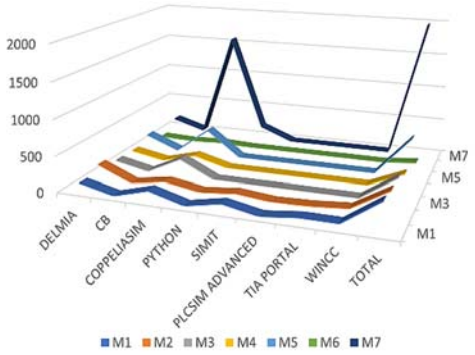


Fig. 7. Complexity analysis of the extended toolchain using the indicators from M1 to M7

Data can also be sent in the other way, i.e., the user can send an emergency stop signal for one of the robots from the virtual HMI and see that the system in the 3d scenario fully stops. To improve simulation performance, the software toolchain was divided into three different workstations connected to a local network: one for simulation layer and OPC-UA server, another one for execution layer and HMI, and the last one for SCADA client configurator and operator interface.

5. Conclusion

A distributed simulation approach can be applied whenever a system or process is considered to be highly complex and high computer resource consumer. For VC applications, the distribution of scenarios on multiple computers allows the developer to achieve high performance during validation and verification of simulations, and even physically test communication protocols. Some of the advantages obtained from this approach are listed below:

1. Resource utilization. By distributing the simulation load among multiple entities, each entity only needs to focus on simulating a smaller portion of the overall system, allowing for more efficient use of available computing resources.
2. Realism. By modeling the interactions between different parts of a system in parallel.
3. Scalability and flexibility: dynamic addition or removal of simulation entities, making it possible to adapt to changing simulation requirements in real-time.
4. Improved accuracy: By breaking down the simulation into smaller parts that can be run in parallel, the overall simulation can be run faster and with greater accuracy, since the errors in each part are independent.

It is worth noting that the use of several computers increases the cost of VC solutions due to the use of extra hardware.

Acknowledgements

This work was conducted and funded with the support of the Research Tax Credit (CIR in French). A special thanks to all the R&I Virtual Commissioning non-permanent collaborators that help us to develop the necessary tools, libraries, and models, allowing us to turn ideas into projects.

References

[1]. R. Balderas Hill, J. Lugo Calles, J. Tsague, T. Master and N. Lassabe, Improving interoperability of Virtual Commissioning toolchains by using OPC-UA-based technologies, in *Proceedings of the 26th IEEE International Conference on Emerging Technologies and Factory Automation,* September 2021, pp. 01 - 07.

[2]. R. Balderas Hill, J. Lugo Calles, J. Tsague, T. Master and N. Lassabe, Virtual Commissioning Implementation of Side Panel Assembly Process for Automotive Shop Floor, *Sensors & Transducers,* Vol. 257, Issue 3, May 2022, pp. 12-19.

[3]. J. Lugo Calles, A. Abassi, R. Balderas Hill, and N. Lassabe, Zero to hero: a generalized virtual commissioning implementation method, in *Proceedings of the Dissemination of Research in the Study of Complex Systems and their Applications Meeting (EDIESCA 2022),* September 2022.

[4]. J. Lugo Calles, A. Abassi, R. Balderas Hill, and N. Lassabe, Virtual Commissioning Implementation of Industrial Aeronautic Assembly Process, in *Proceedings of the Dissemination of Research in the Study of Complex Systems and their Applications Meeting (EDIESCA 2022),* September 2022.

[5]. Albo, Anton & Falkman, Petter. A standardization approach to Virtual Commissioning strategies in complex production environments, *Procedia Manufacturing*, 51, 2020, pp. 1251-1258.

[6]. Gergely Popovics, László Monostori, An approach to determine simulation model complexity, *Procedia CIRP,* 52, December 2016, 257-261.

[7]. Lutz Winkler, Heinz Wörn, Symbricator3D – A Distributed Simulation Environment for Modular Robots, in *Proceedings of the Intelligent Robotics and Applications, Second International Conference (ICIRA 2009),* Singapore, December 16-18, 2009.

[8]. Chen Peihua, Cao Qixin, Yang, Leng Chuntao, An RTM based Distributed Simulation System for Guide Robot, in *Proceedings of the 4th International Conference on Digital Manufacturing and Automation (ICDMA),* 2013, pp. 185 - 189.

[9]. Paul Breugnot, Bénédicte Herrmann, Christophe Lang, Laurent Philippe, Data Synchronization in Distributed Simulation of Multi-Agent Systems, in *Proceedings of the 20th International Conference on Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation (PAAMS 2022),* L'Aquila, Italy, July 13–15, 2022, pp. 50–62.

[10]. Ahlam Abassi, Jesus H. Lugo-Calles, Rafael Balderas Hill, and Nicolas Lassabe, Virtual Commissioning Implementation Of Industrial Painting Process, in *Proceedings of the 2nd International Conference on Computers and Automation (CompAuto 2022).* France. August 18-20, 2022, pp. 65-68.

(5921)

Implementation of a Novel Handheld Endoscopic Operation Platform (EndoGRASP)

Chun Ping LAM [1], Ming Ho [2], Shi Pan SIU [2], Lap Wing CHEUNG [2], Flora Fung LEUNG [1], Tom Man Fung CHAN [1,2], Yeung YAM [1,2] and Ka Chun LAU [1,2]
[1] Department of Mechanical and Automation Engineering
[2] Multi-scale Medical Robotics Center, The Chinese University of Hong Kong
E-mail: cplam@link.cuhk.edu.hk

Summary: Endoscopic procedures have gained widespread use due to their ability to facilitate accurate diagnosis and treatment through minimally invasive surgery. While traditional bimanual platforms are commonly utilized, the use of secondary hand traction-assisted devices offers additional benefits including precise and flexible tissue manipulation. In this work, a novel handheld endoscopic platform, namely EndoGRASP is proposed. EndoGRASP is a handheld endoscopic platform that includes a flexibility-retaining robotic overtube and an integrated actuation unit. This system features dedicated channels for instruments and endoscopes, including a biopsy channel and an active bending section. The biopsy channel enables free translation of instruments, while the active bending overtube compensates for the loss of flexibility in the endoscope and provides rigid support for the instruments and endoscope to bend independently. EndoGRASP improved the safety and efficiency of surgery, reduced the surgeon learning curve, and was suitable for use in small operating theatres.

Keywords: Surgical robot, Continuum joint, Overtube, Endoscope, Handheld robot, Endoscopic submucosal dissection.

1. Introduction

Flexible endoscopes are widely used due to their effective treatment of various conditions with minimal surgery. However, they lack the means to control their shaft trajectory or maintain stability in position during disturbances. Existing ESD platform, such as Incisionless operation platform (IOP) [1] and ANUBIScope [2], while they may offer reliable support for endoscopic bending, they lack the ability to provide precise motorized motion or may potentially restrict the maneuverability of the platform in confined endoluminal environments. Another more advanced operation platform, namely EndoMaster [3], while providing excellent visualization of the surgical area due to its compact design, which allowed for easy docking and prevented interference from the working arms, however, its complex operational requirements, where the surgeon must control the console remotely and manipulate the robotic arms using hand controls, can be time-consuming and challenging for the surgeon to master. Over-the-Scope Clip (OTSC) [4] use a device that places clips over the endoscope to hold tissue before removal, but the effectiveness of these clips in securing the tissue is sometimes inadequate, resulting in incomplete tissue acquisition and suboptimal surgical outcomes. The current operating systems have limitations in mobility, inadequate tissue collection, complex setup, and high cost. To address these problems, a new handheld endoscopic platform called EndoGRASP is introduced. It consists of a flexible robotic overtube and an actuation unit. EndoGRASP is designed to enable surgeons to perform endoscopic submucosal dissection efficiently and effectively, while minimizing the risk of damage to the patient, affording a reliable support for the endoscope and instruments to exert precise motorized control and manipulation of the instrument and overtube, enabling the surgeon to access and work within challenging or hard-to-reach areas of the body.

2. Design & Prototype

EndoGRASP is separated into two different parts, described as overtube system portion and integrated actuation unit portion. The overtube system consists of several components, including the Y-junction adaptor used for the insertion of endoscope, a cable gland and a coil sheath for connection and protection, tendon coils for transmitting actuation force to the overtube, and a biopsy channel tube for guiding the movement of the instrument. The bending section of the overtube is mainly divided into two portions, namely active and supporting sections. The active bending section is controlled by actuating four tendons, allowing for two degrees of freedom with planar bending. When inserting the robotic instrument, it is necessary to ensure that the instrument can easily pass through the continuum joint and reach the outlet. Without a smooth and guiding channel, the instrument would not be able to pass from the proximal end to the distal end of the overtube. To facilitate this, a biopsy channel tube is used as a guide and to provide a smooth translation function. When the overtube is actuated, the robot will bend according to the driven tendon. However, as the overtube stretches during this process, the length will then be prolonged. It is required to reserve some engagement to the biopsy channel for movement, otherwise, limitation of the bending motion of the overtube will occur since the biopsy tube is non-stretchable.

Therefore, to prevent the bending motion of the overtube from being limited by the incompressible biopsy tube, a biopsy channel stopper that located at the distal end of the overtube, has been designed to both allow and limit the movement of the biopsy channel tube, granting the overtube to bend in a constant curvature, and to secures the position of the biopsy channel, prevent slippage and ensure an adequate space for the biopsy channel to move within.

The actuation unit portion of EndoGRASP system shown in Figs. 1 and 2 consists of several key components that allow the surgeon to adjust the bending and movement of the overtube rapidly, based on the specific needs and requirements of the surgical procedure. These include a driving unit equipped with a gearbox, a tools coupler, a linear translation mechanism, and motors. The system is driven by two motors that drive the actuation of the overtube and four motors that control the motion of the robotic gripper. The tendons are attached to the tendon holders that are connected to the motor gear. The linear translation mechanism, which is designed as converting motor rotary motion to linear motion, enables straight-ahead movement of the robotic instrument, while the aluminium platform and the pink pulleys depicted in Fig. 1 provide ample space for the tools to move when it is actuated forward and backward.

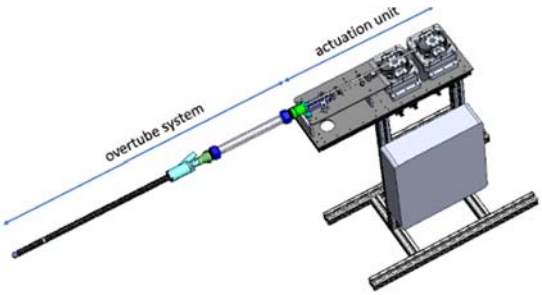


Fig. 1. CAD model of EndoGRASP.

Fig. 2. Prototype of EndoGRASP

3. Result

The bending motion of the endoscopic system can be described using a polar coordinate system into four quadrants. The kinematic model can be derived as follows:

$$\Delta L_i = L - \left[\left(r \pm R_i \sin(\tfrac{\pi}{2} \pm (u_i - \Psi)\right) * \theta\right], \quad (1)$$

where ΔL_i is the change in length of tendons, L is the length of tendons before actuation, r is the bending radius, R_i is the minimum distance between the tendons and the center of the joint, and u_i is the angle between the tendons and central axis of the joint, θ and Ψ are the bending angle and orientation, respectively.

Fig. 3 demonstrated the bending motion of the overtube 45 degrees and 90 degrees with 0 orientation (the reference plane) with constant curvature. It also illustrates 180-degree retroflexion bending, which is a crucial skill and procedure in endoscopic surgery, as the surgeon may need to provide therapy in the fundus portion of the gastrointestinal tract. EndoGRASP also enables the decoupling of the motion of the endoscope, instrument, and overtube itself, allowing for more flexibility and control during surgery. The demonstration of EndoGRASP system's ability to enable comprehensive surgical operations demonstrates its potential for use in a variety of endoscopic procedures.

Fig. 3. Bending performance of EndoGRASP.

References

[1]. Philip Wai-Yan, C., Hon Chi, Y., Ka Chun, L., & Yeung, Y, Techniques and Challenges with the Master–Slave System for Endoscopic Surgery, Innovative Endoscopic and Surgical Technology in the GI Tract, *Springer*, 2021, pp. 289-297.

[2]. Boškoski, I., & Costamagna, G., Endoscopy robotics: current and future applications, *Digestive Endoscopy*, 31, 2, 2019, pp. 119-124.

[3]. Chiu, P. W. Y., Ho, K. Y., & Phee, S. J., Colonic endoscopic submucosal dissection using a novel robotic system (with video), *Gastrointestinal Endoscopy,* 93, 5, 2021, pp. 1172-1177.

[4]. Bartell, N., Bittner, K., Kaul, V., Kothari, T. H., & Kothari, S., Clinical efficacy of the over-the-scope clip device: A systematic review, *World Journal of Gastroenterology,* 26, 24, 2020, p. 3495.

(5965)

Surgical Robotics in Poland – the Future, Opportunities and Threats

M. Sieradzka [1] and J. Wierzbicki [2]
[1] Dr hab., Prof., Head of Department of Public Economic Law, Lazarski University,
Świeradowska 43, 02-662 Warsaw, Poland
[2] Dr, surgeon and gastroenterologist, Head of the Department of Surgical Endoscopy Department of General, Minimally Invasive and Endocrine Surgery, Medical University of Wrocław,
Pasteura 1, 50-367 Wrocław, Poland
Tel.: + 48606996361
E-mail: malgorzata.sieradzka@lazarski.pl; jaroslaw.wierzbicki@umw.edu.pl

Abstract: New technologies and surgical robotics are developing very quickly and bring great benefits to patients. The market for surgical needlework in the world is constantly growing. There is not only an increase in interest in robotics but also in the procedures performed. Global surgical robotics market grows very dynamically with a CAGR of 12% over the forecast period 2022-2027. As the report on the robotics market shows, in Poland the number of surgical procedures performed with the assistance of da Vinci is still growing more intensely. The estimated growth rate of the number of treatments in 2022 compared to 2021 is 48 %. The greatest dynamics are shown by general surgery - 54%, followed by gynecology – 49% and urology – 46 %.

Keywords: Technological progress, Surgical robotics, Robotic surgery.

1. Robotics market in Poland

From estimates based on the results of a study conducted specifically for the PMR and Upper Finance Group report, it shows that in a country the size of Poland by 2025 should be around 50 robots with the help of which will be carried out close to 15 thousand surgeries annually. According to the report between 2021-2026, the compound annual growth rate (CAGR) of the robotic surgery segment will amount to as much as 23% [1]. Medical robots are intended to facilitate operations in a minimally invasive, precise and safe manner for the patient and the surgeon. Robotics permanently changed the image of surgery. Video laparoscopy seemed to be a breakthrough. It is an alternative to open surgery. The patients benefit immensely from minimally invasive surgery, notably fewer complications, hospitalizations and shorter periods of convalescence.

According to the estimates, the global surgical robotics market will reach 19.3 billion dollars by 2030 compared to 5.4 billion dollars in 2020, which means an annual increase of 13.6% between 2020-2030. Surgical robots generate the highest revenues of all medical robot segments. The global market will rise, propelled by the fast adaptation of surgical robotics to various operations and surgeries worldwide. It is expected that in the forecast years, surgeons, patients, hospitals and treatment facilities will increasingly look for advanced robotic solutions for modern surgical procedures [2].

According to the Ministry of Health's data from December 2021, only 12 healthcare providers used the surgical system da Vinci in the public health system.

2. Types of Surgical Robotics

Currently, surgical robots are used around the world, especially in the US. Worldwide, 3% of surgeries are currently performed with surgical robots. The most widespread and dominant robotic system used in surgery is Intuitive Surgical's da Vinci system. In Poland, among 32 hospitals performing procedures assisted by a surgical robot - which Modern Healthcare Institute lists in its report "Development of Robotic Surgery in Poland" – No. 26 is Vinci [3].

However, what should be noted – in 2021 only one healthcare provider – Military Medical Institute – has performed more than 200 procedures using a robot. It is indicated, however, that a minimum of 400 procedures performed annually guarantees the profitability of investment in robotics technology for the hospital. The Supreme Audit Office's report in Poland showed the results of inspections of hospitals with highly specialized medical apparatus, including da Vinci robot: "The majority of the inspected medical entities did not use the full potential of the purchased highly specialized medical equipment".

The first Versius system in the country from British company CMR Surgical has been launched in Lodz - by the end of July 2022, seven facilities had started using it. The Versius surgical robot consists of a surgeon's console and a mobile endoscopic column with a camera. The technology allows the operator to use only as many arms as needed to successfully perform a specific procedure [4].

In Poland, Robin Heart is an original structure designated for cardiac surgeries. Robin Heart includes a planning system, training system, expert program,

telemanipulators and automated surgical instruments. Currently, advanced scientific work continues on the development of force coupling and remote surgery. Robin Heart robots have the potential to become a widely used, technically advanced and telemedical system, facilitating minimally invasive, precise, safe operations. Their future, however, depends on finding an investor [5].

3. Medical Robotics – Competition for Laparoscopy?

It is believed that some robotic surgical procedures will even replace laparoscopic procedures in the future. Surgical robots used for radical prostatectomy procedures (removal of prostate tumours) will become a very serious competitor to laparoscopy [6].

Considering the minimal invasiveness and the benefits of performed procedures, it is hard to imagine a return in these cases to surgical procedures.

Minimally invasive surgery – however? When identifying obstacles, the first thing to point out is cost. These are not only technological costs associated with the production process. It's also the cost of purchasing equipment and training doctors. Although, it is worth noting that a part of these obstacles was erased. In Poland, as of 1 April 2022, the Ministry of Health has introduced separate pricing for radical prostatectomy, a procedure performed with the assistance of a surgical robot. It is now double that of laparoscopic or open (classic) procedures. Considering the above, a very large number of facilities have already volunteered for these treatments, while at the same time committing to performing a minimum of 100 treatments annually [7].

Financing, which until now has been pointed out as one of the obstacles to the widespread use of robots, is currently losing its importance. However, it is necessary to increase the competition in the market because a higher number of available robots will result in a price reduction.

4. Smart Robot's Legal Liability

It is also worth considering whether robotics will also affect liability issues. An example of cardiac surgery performed in the UK using a robot illustrates this issue well. In Freeman Hospital in Newcastle as a result of post-operative complications with the da Vinci robot, 69-year-old music teacher Stephen Pettitt died. It was the first cardiac surgery in the UK performed using a robot. Its main performer was Dr Sukumaran Nair, who operated on the mitral valve. During the procedure, the artery burst. It was reported in the media that "blood gushed as far as the robot's camera." In Poland, not all surgeries performed with the da Vinci robot also ended well. Patients had to be operated on again, but the hospital stressed that this, however, was not directly related to the use of the da Vinci robot itself [8].

In this case, whether the robot is to blame or the operator (made a mistake) or perhaps it is the risk of this procedure. In accordance with the European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics [9], the smart robot will be a robot fulfilling all the following 5 criteria: 1) acquires autonomy through sensors or by exchanging data with its environment (by inter-connectivity) and by trading and analysing of those data; 2) possesses at least a minor physical support; 3) adapts its behaviour and actions to the environment; 4) does not have life in the biological sense; 5) optional criterion – can self-learn from experience and by interaction with the environment. Although in this case, it is obvious - what was noted above - that the purpose of robotic procedures is to minimize complications and errors. A surgical robot, on the other hand, has no legal personality, meaning that it cannot be held liable in the event of a medical error.

While an autonomous robot can detect the changing internal and external environment using software and a minimum of one sensor, its work is controlled via software. Additionally, the robot's software has to respond to variables in a minimum of 1 way: the machine must have its own power source, the robot must understand the language that programmed it, the robot must execute the software independently and must not be a living organism [10].

When considering the responsibility of robots, legal subjectivity is pointed out. The referenced resolution of the European Parliament proposes giving smart robots the status of electronic persons. Thus, we see here a clear reference to autonomy. In this case, it would be appropriate to seek redress for the damage they have caused. In particular, there are ideas of granting them legal personality [11]. Another idea is the construction of mandatory insurance, proposed in the Resolution on Robotics. European Parliament recommends the Commission and the insurance sector, creating a new category of obligatory insurance for technologically advanced robots, further strengthened by a fund from which compensation would be paid for damages not covered by insurance. Poland, however, in the document "Policy for the Development of Artificial Intelligence in Poland for 2019-2027", opposes granting artificial intelligence the status of legal personality in Polish and international law [12].

5. Autonomous Surgical Robot – the Future of Surgery in Space?

The future of surgical robotics is not only a duplication of existing da Vinci. It is crucial to introduce a solution that can replace humans, just as in the automotive industry, the autonomous car. Under this term, we mean an unmanned car that moves without human intervention and is controlled by a computer. Also, for surgical robots, solutions must be put in place that allows them to perform operations autonomously. Prof. Shane Farritor, an engineer at the University of Nebraska in the US, has constructed a mini-robot which would perform surgical procedures

using remote control [13]. Aboard the International Space Station (ISS), NASA wants to test a miniaturized surgical robot. For now, it will automatically cut through a rubber membrane, but in the future, similar robots could perform procedures on astronauts. As part of a NASA-sponsored grant, the University of Nebraska-Lincoln is working on a surgical robot that is expected to arrive on the ISS in 2024 [14]. It is worth pointing out that Prof. Stanislaw Czudek became a member of an elite group of one hundred surgeons from around the world selected by NASA to possibly operate remotely on astronauts in space.

6. Conclusion

Robotic surgery is successfully used in many countries. The benefits of its implementation are primarily, patient's safety - minimization of complications, short hospitalization time, and quick return to normal function. The introduction of public funding for these procedures would certainly increase the interest of medical facilities in purchasing medical robots. In the future, legal liability in situations of harm to the patient should also be regulated. However, this requires the development of an appropriate concept of liability and the granting or not of legal subjectivity to robots.

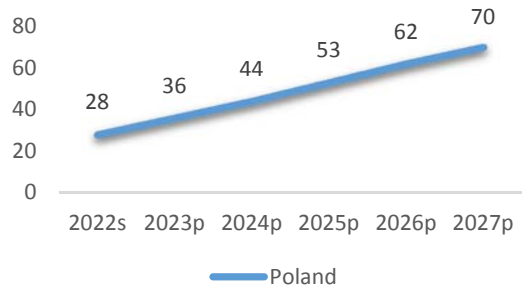


Fig. 1. Number of authorised surgical robots in Poland. Forecast for 2022-2027.

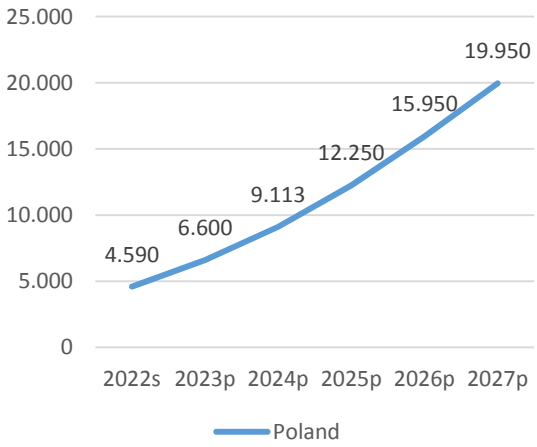


Fig. 2. % Number of procedures and dynamics in the surgical robotics market in Poland. Forecast for 2022-2027.

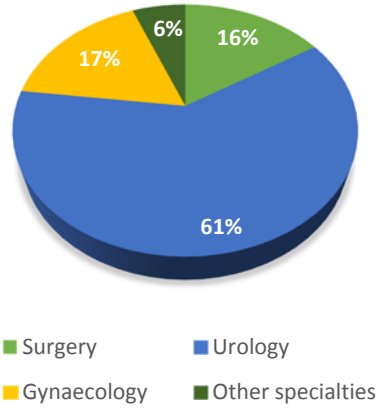


Fig. 3. % Number of procedures performed using da Vinci devices in Poland 2022.

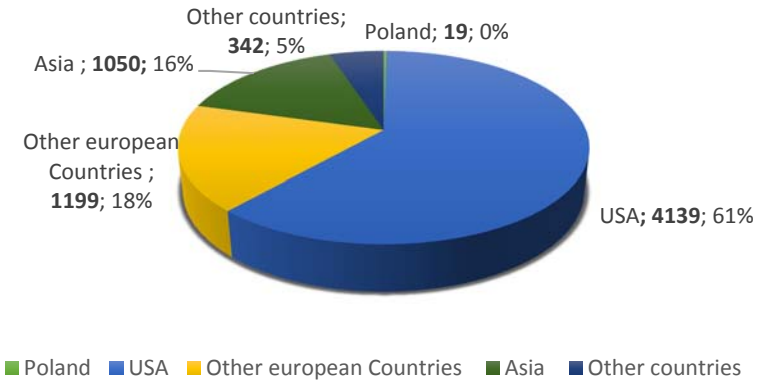


Fig. 4. % Number of Da Vinci instruments in Poland and worldwide in 2022.

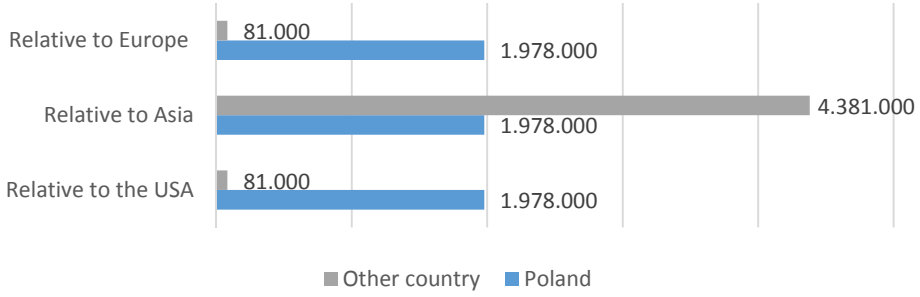


Fig. 5. Population per da Vinci machine in Poland, compared to selected regions of the world.

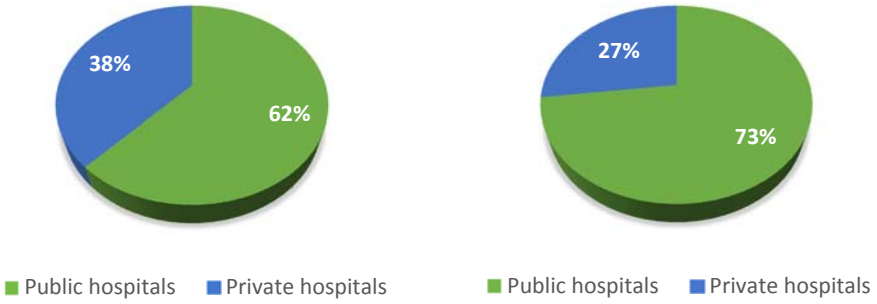


Fig. 6. Number of procedures using da Vinci in Poland. Comparison 2021 and 2022.

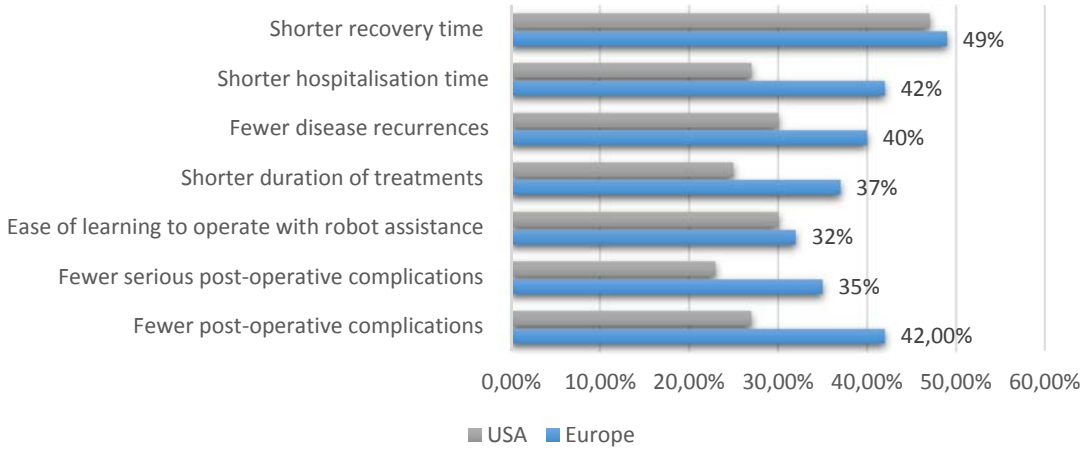


Fig. 7. Key advantages of robotics according to operator surgeons in the US and Europe in 2020.

References

[1] PMR, Upper FINANCE Group: Rynek robotyki chirurgicznej w Polsce 2021, *Prognozy rozwoju na lata 2021-2026* (in Poland).
[2] Report: Global Surgical Robotics Market 2020-2030. 2021. (https://www.researchandmarkets.com).
[3] Modern Healthcare Institute, Rozwój chirurgii robotowej w Polsce (in Poland).
[4] Pierwszy w Polsce robot chirurgiczny Versius (https://www.zwrotnikraka.pl) (in Poland).
[5] Roboty chirurgiczne w Polsce. Jak wygląda polski robot chirurgiczny (https://www.fakt.pl).
[6] MHI, Raport – Rozwój chirurgii robotowej w Polsce. Placówki. Świadczenia. Rentowność. Perspektywy rozwoju rynku (in Poland),
[7] Rozporządzenie Ministra Zdrowia z dnia 25 stycznia 2022 r. zmieniające rozporządzenie w sprawie gwarantowanych świadczeń z zakresu leczenia szpitalnego (in Poland).
[8] Miało być dużo operacji, żeby nie marnować publicznych pieniędzy. Jak szpital wykorzystuje robota za 15 milionów? (https://dziennikwschodni.pl) (in Poland).
[9] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), *Dz.U. UE. C.*, No. 252, 2018, p. 239.
[10] C. Hughes, T. Hughes, Robot Programming: A Guide to Controlling Autonomous Robots (K. Matuk, Trans.), *Helion*, 2017, p. 20.

[11] D.V. Vladeck, Machines without principals: liability rules and artificial intelligence, *Washington Law Review*, Vol. 89, Issue 1, 2013, p. 89.

[12] MC, MPiT, 26 lutego 2020, Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019–2027. Warszawa: Ministerstwo Cyfryzacji, *Ministerstwo Przedsiębiorczości i Technologii,* https://www.gov.pl (in Poland).

[13] Operacje w kosmosie minirobot ma to umożliwić (https://www.nauka.tvp.pl) (in Poland).

[14] Na Międzynarodową Stację Kosmiczną poleci robot-chirurg (https://scienceinpoland.pap.pl) (in Poland).

(6039)

Gamification for Machinery and Infrastructure Learning in Industries: A Review

M. Fiore and M. Mongiello
[1] Polytechnic University of Bari, Via E. Orabona, 4, 70121, Bari, Italy
E-mail: Marco.Fiore@poliba.it, Marina.Mongiello@poliba.it

Summary: Gamification has the aim of making work lighter and more pleasant. The purpose of this research is to analyze current literature related to the state of the art of gamification process applied to machinery and infrastructure learning in industries. We conduct a literature review defining a search query and applying inclusion and exclusion criteria based on the PRISMA workflow. After reviewing these criteria, we consider 94 papers. Results show that gamification is growing in popularity in a variety of areas also thanks to the recent COVID-19 pandemic that created a need for companies to move to fully online recruitment and training solutions. To address these challenges, organizations are using gamified solutions. We found several application areas such as virtual training environment, gamified e-health, self-management interventions, and holography. Gamification can reach new levels of accessibility, interactivity, and lifestyle integration, facilitated by the use of immersive technologies such as augmented reality (AR), virtual reality (VR), and mixed reality (MR). The implementation of gamification, however, may also have some downsides, attributed to the potentially high costs to switch to this new approach.

Keywords: Gamification, Industries, Machines, Infrastructures, Literature review, PRISMA.

1. Introduction

Gamification methods are strategies, procedures, and mechanisms that help consistently determine how to incorporate game basics in a non-gaming setting. Gamification was born for educative purposes as a technique using approaches similar to the ones used in videogames world. Gamification is based on incorporating game features to improve humans, their motivation, and their purposes. In interactive gaming and digital selling, the concept of using game design features in non-game ways inspires and improves user activity.

Gamification is expanding in various fields of interest. Industries are adopting gamification approaches to let employees know how to use machinery and infrastructure. The use of a virtual environment grants safety during the training phase and more engagement for workers. Gamification also comes in hand after the COVID-19 pandemic created new needs in the online training topic [1].

2. Research Methodology

We use the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines to conduct the review [2]. We firstly define Research Questions (RQs), then we perform a complete search with a papers screening phase. Finally we extract relevant information. The PRISMA flow diagram is shown in Fig. 1.

2.1. Research Questions

We define five Research Questions to accomplish our goal.
RQ1: When was gamification for industry born?
RQ2: What are the trend topics on gamification for machinery and infrastructure?
RQ3: What are the application fields?
RQ4: What are the main problems in the application of gamification for industries and what prevents stakeholders to use this approach?
RQ5: Can gamification track the level of acquired skills during training?

2.2. Data Gathering and Filtering

The data gathering phase is conducted using a search string:

gamification AND (learning OR training) AND (infrastructure OR machinery) AND industry

The obtained papers are filtered using some inclusion (language, topic) and exclusion criteria (importance of the topic, open access availability, published after 2020). The PRISMA process results in 94 main papers. The number could be increased considering a wider range for the publication date or including some works not published in open access.

3. Results

Below are presented the main results of the literature review. One of the key benefits of gamification-based training software is its ability to provide real-time and actionable data. Companies can track the progress of their employees at any point in their training program, identifying the causes of malfunctioning and making ad-hoc adjustments. There are some downsides to analyze. Most of the gamification applications apply together with holograms and Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) techniques.

The initial capital commitment necessary to get digital and holographic instruments is quite expensive. This technique is generally costly, with high prices mainly attributed to hardware parts and the need for specific equipment.

Over time, safety training has improved together with playful solutions and augmented reality and virtual reality simulations aimed at improving teaching effectiveness. Applying the principles of gamification and using serious games to train workers in their workplace offers important benefits in terms of teaching effectiveness [4]: a) it increases worker engagement by promoting friendly competition and reward for completing tasks; b) it spreads a culture of safety that is strongly linked to risk awareness; c) it allows to create a safe environment where workers can test their knowledge of safety procedures (e.g., company evacuation plans, firefighting activities, etc.) in realistic scenarios; d) it makes workers more alert to their behaviors and aware of the risks; e) it allows workers to practice repeatedly until they have mastered notions and procedures; f) it makes training more enjoyable and facilitates memorization and retention of information learned.

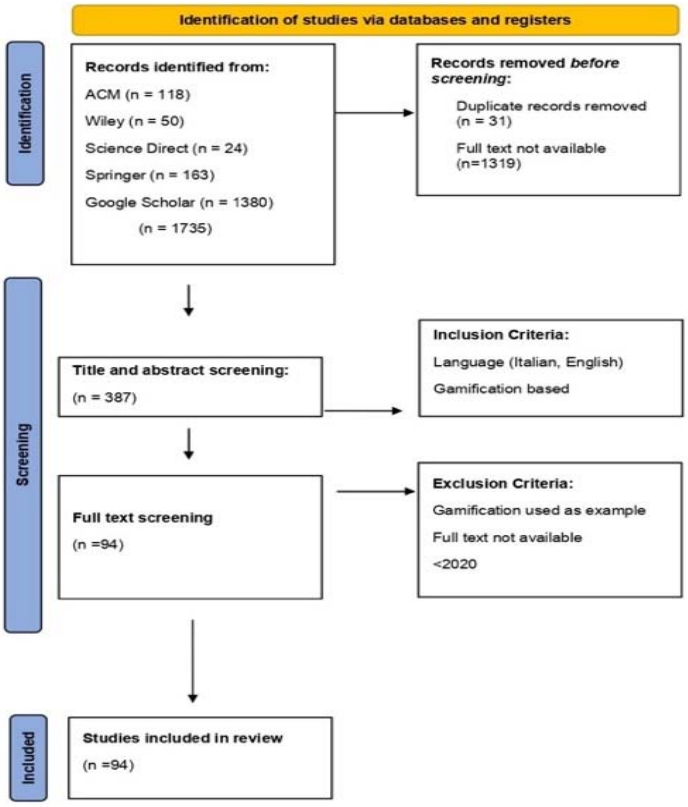


Fig. 1. PRISMA flow diagram.

4. Conclusions

Our literature review aims to become a resource for academics and professionals trying to understand the impact of gamification in teaching machinery and infrastructures in industries and the gap between its benefits and boundaries. Gamification is nowadays used in industries but it still lacks on best practices and standards, so it could be difficult to let it spread because of high hardware costs. The usage of such an approach lets employers improve safety in their workplace, as well as guarantee workers engagement thanks to friendly competition and rewards strategies.

Acknowledgements

This work was funded by the TRACECOOP Research project (Italy) (B96G21000060005).

References

[1]. Oliveira, Renata Pereira, et al., A study on the relation between industry 4.0 technologies and gamification in e-learning, *Interactive Technology Smart Education*, 2022.

[2]. Biolchini, Jorge, et al., Systematic review in software engineering, *System engineering and computer science department COPPE/UFRJ, Technical Report ES*, 2005, 679.05: 45.

[3]. Ulmer, Jessica, et al., Gamified virtual reality training environment for the manufacturing industry, in *Proceedings of the 19th IEEE International Conference on Mechatronics-Mechatronika (ME'2020)*, 2020. p. 1-6.

[4]. Ponis, S. T., et al., Augmented reality and gamification to increase productivity and job satisfaction in the warehouse of the future, *Procedia Manufacturing*, 51, 2020, pp. 1621-1628.

(6327)

PIN Theft Attack against FIDO2 Security Keys

Emin HUSEYNOV
Azerbaijan Technical University, H. Cavid 25, 1073 Baku, Azerbaijan
Tel.: +994125383383
E-mail: emin@huseynov.com

Summary: This paper discusses the potential vulnerabilities of the FIDO2 protocol, specifically with regard to the Passwordless systems. The paper proposes that even with the highest security mechanisms in place, user awareness is still a crucial factor in cybersecurity. The structure of an attack using a phishing page to steal the PIN code for a FIDO2 security key is described, along with possible implementation scenarios and mitigation methods.

Keywords: Strong security, Multifactor authentication, User awareness, Phishing, Social engineering, Password theft, Passwordless access, FIDO2.

1. Introduction

The FIDO ("Fast IDentity Online") Alliance is an open industry association launched in February 2013 whose stated mission is to develop and promote authentication standards that "help reduce the world's over-reliance on passwords" [1]. FIDO keys are considered the most secure solution for protecting online accounts [2]. However, we believe that even this solution is not fool proof. The goal of this article is to demonstrate that even with the highest security mechanism implemented, user awareness still plays a very important role in the cybersecurity. The attack surface of FIDO2 PIN theft attack given as an example in this article is specifically Azure Passwordless [3], but the same or similar techniques can be applied to other authentication systems with passwordless methods implemented using FIDO2 keys.

In Section 2, a brief overview of FIDO2 protocol will be proposed to give the readers an idea of the attack surface described. Then, the paper will present the structure of the attack and then cover the possible implementation scenarios in Section 3. Later, in Section 4 a summary of findings will be discussed, and mitigation methods proposed.

2. FIDO2 Authentication Mechanism and Risks

FIDO2 offers developed authentication possibilities including secure single factor (passwordless), strong two factor, and multi-factor authentication. With these new capabilities, the FIDO2 security keys can fully replace weak classic username/password credentials with powerful hardware-powered public key infrastructure-based credentials.

While being promoted as highly secure MFA method, there are research papers describing vulnerabilities discovered both in certain FIDO2 devices as well as the protocols used in the authentication flow itself. Examples of such cases are:

Security Issue with Bluetooth Low Energy (BLE) Titan Security Keys (Google) [4].

Side-Channel Attack on the Google Titan Security Key (NinjaLab) [5].

Provable Security Analysis of FIDO2 (University of Porto (FCUP) and INESC TEC) [6].

The attack vectors, risks and requirements' similarity of the examples above give us a ground to consider this paper to be at least of the same research value.

3. Structure of the Attack

In this section, we will describe the techniques and components used to prepare and implement the FIDO2 security key PIN stealer attack. The main component of this attack, the phishing page, will be demonstrated using screenshots of the real phishing page created as a proof-of-concept.

A. Prerequisites

The following prerequisites are needed for the attack to commence:

- A user account with Azure AD Passwordless enabled;
- A regular USB or NFC FIDO2 key enrolment (the attack has less success if biometric keys are used);
- A domain name similar to the ones used by Microsoft to host a phishing page used as a part of the attack.

B. Prerequisites

The anatomy of a FIDO2 PIN phishing page is the same as with a regular phishing attack. The page should look as close as possible to the real one. In addition, the components emulating standard windows dialogs, should be mimicking the interface of the

dialogs of the operating system. In the figures below, we will show how the different parts of the legitimate login page are cloned to achieve the same look and feel.

Fig. 1 demonstrates the real sign-in page in comparison with the fake page show in Fig. 2 used as the first step of the phishing attack. The page design elements and the layout were simply copied from the legitimate website.

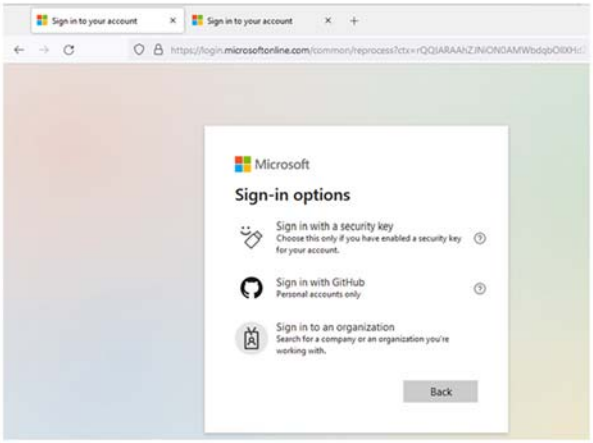


Fig. 1. The legitimate login page.

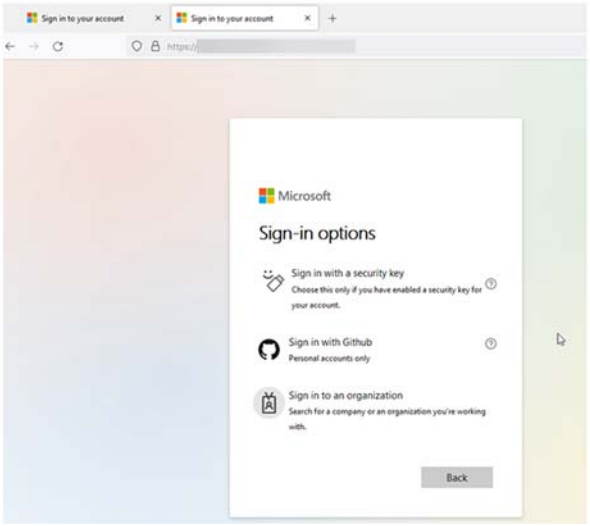


Fig. 2. The fake login page.

Once the potential victim clicks on the "Sign in with a security key" option on the fake login page, a pop-up window appears prompting the user to enter the PIN code for the security key. In our phishing page we used WinStrap template [7] to make the user interface elements as visually similar as possible to the standard interface dialogs of Windows operating system.

The results can be compared using the Fig. 3 (the standard Windows system dialog page asking for the PIN code) and Fig. 4 (the replica of the same dialog generated using the HTML/CSS on our phishing page).

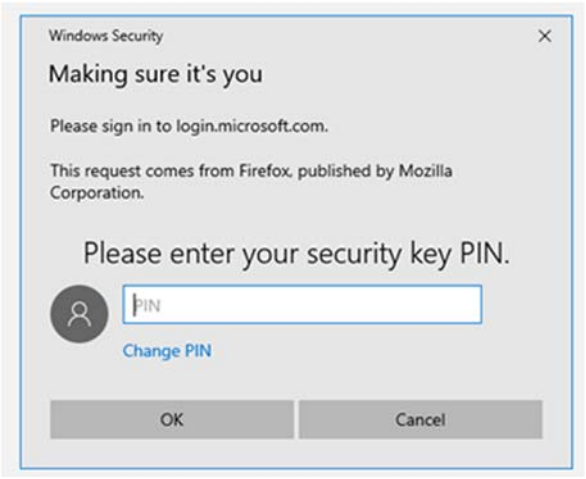


Fig. 3. Windows system dialog for FIDO2 PIN entry.

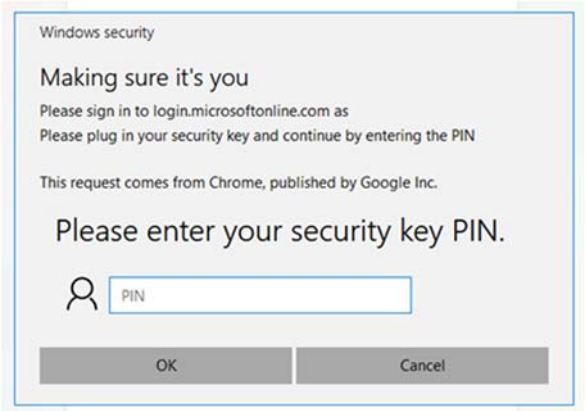


Fig. 4. PIN Entry dialog generated using HTML and CSS on the phishing page.

C. Attack Implementation

The methods of convincing the victim to navigate to this page and, more importantly, to trust the phishing page enough to provide the PIN code are similar to the techniques used by attackers leveraging regular phishing attacks. This was already described in a number of publications, including ours.

For the attack to be successful, the victim will navigate to this page, have the security key plugged in (this is not needed for our attack, but a security key PIN requested without the key itself plugged in will definitely be suspicious), then enter the PIN into the fake dialog box. The page will save the entered PIN, and then redirect the user to the real login page to minimize the time user spends on the phishing page after the PIN was successfully obtained. This technique is also similar to regular phishing attacks; the difference in this case is that after a successfully stolen PIN, the attacker needs to have physical access to the security key to complete the attack and compromise the account.

4. Summary

The paper described a method of creating a phishing page that mimics the legitimate login pages of the FIDO2 protocol. The page uses HTML and CSS to create components that mimic the standard dialog boxes of the operating system, making it visually similar to the real login page. This makes it easier to convince users to enter the PIN code for their security keys. The paper includes screenshots of a proof-of-concept page to show how similar the replica can be to the legitimate login page.

Although the attack is relatively difficult to implement because it requires physical access to the security key, it demonstrates that even with technically well-implemented security mechanisms such as FIDO2 keys, there are still risks of compromising accounts protected by security keys.

The FIDO2 protocol is considered to be a highly secure method of protecting online accounts, but this paper shows that even the most robust security systems can be vulnerable. By creating a phishing page that mimics the legitimate login page and using components that mimic the standard dialog boxes of the operating system, attackers can convince users to enter their PIN codes for their security keys. This type of attack is relatively difficult to implement, as it requires physical access to the security key, but it highlights the importance of user awareness in cybersecurity.

To protect against this type of attack, it is crucial for users to be cautious when entering sensitive information online. This includes being aware of phishing attempts and being careful not to enter sensitive information on pages that do not appear legitimate. Additionally, it is important for organizations to educate their users on cybersecurity best practices and provide them with the tools they need to stay safe online.

Acknowledgements

This paper is supported by Token2 Sàrl, a Swiss company specializing in multifactor authentication solutions.

References

[1]. Password - The Security Issue That the Big Leaders Want to Eliminate, 30 June 2020, [Online], Available: https://envzone.com/why-big-tech-is-striving-for-the-world-without-password/

[2]. Lyastani, S. G., Schilling, M., Neumayr, M., Backes, M., & Bugiel, S., Is FIDO2 the Kingslayer of User Authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication, in *Proceedings of the IEEE Symposium on Security and Privacy*, 2020.

[3]. Microsoft, Passwordless authentication options for Azure Active Directory, *Microsoft | Learn,* 15 09 2022. [Online]., Available: https://learn.microsoft.com/en-us/azure/active-directory/authentication/concept-authentication-passwordless. [Accessed 15 12 2022].

[4]. Google, Inc., Advisory: Security Issue with Bluetooth Low Energy (BLE) Titan Security Keys, *Google Security Blog*, 15 05 2019. [Online].Available: https://security.googleblog.com/2019/05/titan-keys-update.html. [Accessed 15 12 2022].

[5]. Lomne, V., & Roche, T., A Side Journey to Titan, *Cryptology ePrint Archive*, 2021, Paper 2021/028, https://eprint.iacr.org/2021/028

[6]. Barbosa, M., Boldyreva, A., Chen, S., & Warinschi, B., Provable security analysis of FIDO2, in *Proceedings of the International Cryptology Conference*, Virtual, 2021.

[7]. Lijanto M. et al, The Bootstrap theme for Microsoft design language, winstrap 0.5.13, 01 01 2018. [Online]. Available: https://winstrap.azurewebsites.net/ [Accessed 15 12 2022].

(6987)

A Single Actuator vs. Multi Actuators Design of an Input-Feedback Control to the Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky Equation

R. Al Jamal [1] and N. Smaoui [2]
[1] Australian University of Kuwait, College of Engineering, Department of Mathematics & Physics, P.O. Box 1411, Safat 13015, Kuwait
[2] Kuwait University, Faculty of Science, Department of Mathematics, P.O. Box 5969, Safat 13060, Kuwait
E-mail: r.aljamal@au.edu.kw, n.smaoui@ku.edu.kw

Summary: The Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky (KdVBKS) equation is a nonlinear partial differential equation. It models propagation of waves in a thick elastic tube filled with a viscous fluid. The KdVBKS equation has infinitely many equilibrium solutions. In particular, any constant is an equilibrium solution to the KdVBKS equation. It is shown that the stability of the equilibria depends on the value of the parameters of the equation. A bounded input-feedback controller design to stabilize the KdVBKS equation is presented using a single actuator and multi actuators. The controller design is done through its linearized system and a local result is successfully obtained for the nonlinear system. Moreover, numerical results to stabilize any solution to the zero-equilibrium solution are presented using a single actuator and three actuators approaches. Finally, the two numerical results are compared with one another.

Keywords: Exponential stability, Fréchet differentiable, Galerkin Projection, Input-feedback control, Linearization, Lyapunov direct/indirect method, Spillover, Stability of nonlinear systems.

1. Introduction

In this paper, the Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky (KdVBKS) equation with periodic boundary conditions is considered

$$\frac{\partial z}{\partial t} - \nu \frac{\partial^2 z}{\partial x^2} + \gamma_2 \frac{\partial^4 z}{\partial x^4} + \mu \frac{\partial^3 z}{\partial x^3} + \gamma_1 z \frac{\partial z}{\partial x} = 0,$$

where $x \in (0,2\pi)$, $t > 0$, the parameters $\gamma_1, \gamma_2, \mu > 0$ and $\nu \in \mathbb{R}$. The independent variable x represents the medium of propagation, t is proportional to elapsed time, and $z(x,t)$ is a velocity of the wave at the point x and time t.

The KdVBKS equation can be transformed to different well-known partial differential equations for different values of the parameters ν, γ_1, γ_2 and μ. For instance, it reduces to the Korteweg-de Vries-Burgers (KdVB) equation when $\gamma_2 = 0$ and $\nu > 0$, which is a model that captures the features of both dispersion and dissipation [16, 27]. The KdVBKS equation becomes the Korteweg-de Vries (KdV) equation when $\nu = 0$, $\gamma_1 = 1$ and $\gamma_2 = 0$ which is a nonlinear dispersive partial differential equation that presents a model of propagation of small amplitude along water waves in a uniform channel [5, 26, 28-30]. In addition, if $\mu = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$ and $\nu > 0$, the KdVBKS equation becomes the well-known Burgers equation which models turbulent liquid flow through a channel [7], [33-35]. Finally, if $\gamma_1 = \gamma_2 = 1$, $\mu = 0$ and $\nu < 0$, then the KdVBKS equation becomes the Kuramoto-Sivashinsky (KS) equtaion which models the propagation of a flame front of a combustor. The KS equation has attracted many researchers and was investigated analytically as well as numerically (see for instance, References [1, 2, 9, 10, 12 – 15, 21, 31, 40]).

The boundary and distributed control problem of the Burgers equation, the KdV equation, the KdVB equation, the KS equation, KdVBKS equation were treated in [1, 6, 8, 22 – 25, 35 – 37]. The input-feedback control design to nonlinear infinite-dimensional systems is commonly done through its linearization. However, in general this method does not usually work unless some conditions hold for the system (see for example [1, 15, 17, 38]. The key is to show the Fréchet differentiability of the C_0-semigroup corresponding to the nonlinear system and its derivative is the semigroup generated by the linearized system.

In this paper, the input feedback control will be designed through the linearized system. Two approaches to stabilize the system will be used: In the first approach, we take the number of actuators to be equal to the number of unstable modes of the system, and in the second approach, we only use one actuator. In both approaches, we show that the solution of the KdVBKS equation can be stabilized to a desired equilibrium solution. However, the rate of convergence of the stability with one actuator is slower than the rate of convergence when the number of actuators is equal to the number of unstable modes.

The paper is organized as follows: In Section 2, the controlled KdVBKS equation with periodic boundary conditions is presented. Section 3 investigates the stability of the set of constant equilibrium solutions to the KdVBKS equation using Lyapunov direct and indirect methods. It should be noted that the Fréchet differentiability of the C_0-semigroup generated by the KdVBKS equation is the key for using Lyapunov indirect method. In Section 4, a bounded input-

feedback control scheme is presented to control the nonlinear infinite-dimensional KdVBKS equation regardless of the number of actuators used in the model. Furthermore, in Section 5, the Galerkin projection method was introduced to approximate the solution with the needed conditions to guarantee the correct design of the controller without having any spillover effect in the infinite dimensional system. Moreover, numerical simulations are presented in Section 6 to illustrate the control approach using a single actuator and three actuators and the results of the two controllers design are compared with one another. Finally, concluding remarks are presented in Section 7.

2. The Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky Equation Model

Consider the controlled Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky (KdVBKS) equation

$$\frac{\partial z}{\partial \mathrm{t}} = \nu\frac{\partial^2 z}{\partial x^2} - \gamma_2\frac{\partial^4 z}{\partial x^4} - \mu\frac{\partial^3 z}{\partial x^3} - \gamma_1 z\frac{\partial z}{\partial x} + \sum_{n=1}^{m} b_n(x)u_n(t)\,, \tag{1}$$

with periodic boundary conditions

$$\frac{\partial^n z}{\partial x^n}(0,t) = \frac{\partial^n z}{\partial x^n}(2\pi,t), \;\; n = 0,1,2,3 \tag{2}$$

and initial condition

$$z(x,0) = z_0(x), \tag{3}$$

where $\gamma_1, \gamma_2, \mu > 0$ and $\nu \in \mathbb{R}$. The state of the system $z \in L^2(0,2\pi)$ represents the velocity of a wave at point x and time t. The actuators $b_n(x) \in L^2(0,2\pi)$ and the controllers $u_n(t) \in \mathbb{R}$ and $m \geq 1 \in \mathbb{N}$. We shall design input-feedback controllers $u_n(t)$ of the form

$$u_n(t) = -K_n z(t), \; n = 1, \cdots, m, \tag{4}$$

where $K_n : L^2(0,2\pi) \to \mathbb{R}$ is defined by $K_n z = \langle k_n, z\rangle$, with $k_n \in L^2(0,2\pi)$.

Notation. The space $H^n_{per}(0,2\pi)$ for $n \in \mathbb{N}$ denotes the Hilbert space $H^n(0,2\pi)$ with proper periodic boundary conditions.

The KdVBKS equation can be written in the state-space form

$$\begin{aligned} \dot{z} &= F(z) + \textstyle\sum_{n=1}^{m} B_n\, u_n(t), \\ z(0) &= z_0, \end{aligned} \tag{5}$$

where $F : H^4_{per}(0,2\pi) \subset H^4(0,2\pi) \to L^2(0,2\pi)$,

$$F(z) = \nu\frac{\partial^2 z}{\partial x^2} - \gamma_2\frac{\partial^4 z}{\partial x^4} - \mu\frac{\partial^3 z}{\partial x^3} - \gamma_1 z\frac{\partial z}{\partial x}, \tag{6}$$

and $B_n : \mathbb{R} \to L^2(0,2\pi)$ is a linear bounded operator for $n = 1, \cdots, m$ given by

$$B_n u_n(t) = b_n(x)u_n(t). \tag{7}$$

The controlled KdVBKS equation (1) has a unique strong solution [25, Theorem 2.3]. Using the Galerkin method it can be shown its solution

$$z(t) \in C\left([0,T]; L^2(0,2\pi) \cap L^2\left([0,T]; H^4_{per}(0,2\pi)\right)\right), 0 < T < \infty.$$

3. The Stability Analysis of the KdVBKS equation using Lyapunov Approach

The KdVBKS equation (1) has infinitely many equilibrium solutions. In particular, any constant function is an equilibrium to the system. Let Z_e to be the set of all constant equilibrium solutions to the KdVBKS equation

$$Z_e = \{z_e : z_e \text{ is a constant function}\}. \tag{8}$$

Then we shall investigate the stability of the equilibrium set Z_e using Lyapunov direct and indirect methods. It is worth mentioning that the stability of the equilibrium set Z_e depends on the value of the parameters ν and γ_2. In particular, it can be shown that it is globally exponentially stable if $\nu > -\gamma_2$, Lyapunov stable if $\nu = -\gamma_2$ and unstable if $\nu < -\gamma_2$.

Theorem 1. [25, Theorem 2.4] Consider the KdVBKS equation ((1) with $u_n(t) = 0$ for $n = 1, \cdots, m$). If $\nu > -\gamma_2$, then the set of equilibrium solutions Z_e given in (8) is globally exponentially stable.

Next, we shall consider the linearized KdVBKS equation at a constant equilibrium. Without loss of generality, the zero-equilibrium solution will be considered in this paper. That is, the linearized KdVBKS equation at $z_e = 0$ is given by

$$\dot{z}(t) = F'z(t), \tag{9}$$

where $F' : H^4(0,2\pi) \to L^2(0,2\pi)$ is the Gáteaux derivative of the nonlinear operator F in (6) at $z_e = 0$. That is,

$$F'z = \nu\frac{\partial^2 z}{\partial x^2} - \gamma_2\frac{\partial^4 z}{\partial x^4} - \mu\frac{\partial^3 z}{\partial x^3}. \tag{10}$$

Theorem 2. [25, Theorem 2.7] Consider the linear operator F' in (10), then it is a Riesz-spectral operator that with eigenvalues

$$\lambda_n = -\nu n^2 - \gamma_2 n^4 + i\mu n^3 \tag{11}$$

and corresponding eigenfunctions $\phi_n = \frac{1}{\sqrt{2\pi}} e^{inx}$, $n \in \mathbb{Z}$.

The real part of the eigenvalues (11) of the operator F' crosses the imaginary axis when $\nu = -\gamma_2 n^2$ for $n = \pm 1, \cdots, \pm\infty$. Furthermore, the operator F' is a Riesz-spectral operator (Theorem 2), this implies that the spectrum determined growth assumption [11, Theorem 2.3.5c] holds and the growth bound $w_0 =$

$\sup_{z\in\mathbb{Z}} Re\{\lambda_n\}$. That is, the linearized KdVBKS equation at $z_e = 0$ in (9) is asymptotically stable if $\nu > -\gamma_2$, Lyapunov stable when $\nu = -\gamma_2$ and finally unstable when $\nu < -\gamma_2$. Note that the linearized equation at $z_e = 0$ is always stable when $\nu > 0$. Using the above analysis along with the eigenvalues in (11), the number of unstable eigenvalues to the linearized KdVBKS equation at $z_e = 0$ is finite. In particular, the number of unstable eigenvalues is given by the largest integer N such that $N < \sqrt{\frac{-\nu}{\gamma_2}}$.

Theorem 3. Let the nonlinear operator $S(t)$ be the C_0-semigoup generated by the KdVBKS equation (5), then $S(t)$ is Fréchet differentiable at $z_e = 0$ and the derivative is the C_0 -semigroup generated by the linearized KdVBKS equation at the same equilibrium.

Since the number of input feedback controllers in the model are finite, then the proof of the above theorem can be done using a similar approach to [25, Theorem 2.8]. Note that this result is the key to generalize Lyapunov indirect method to infinite dimensional systems [1, Theorem 2.7 & Corollary 2.9]. That is, the stability analysis for the equilibrium set Z_e in (8) to the KdVBKS equation can be done through its linearization.

Theorem 4. Consider the KdVBKS equation ((5) with $u_n(t) = 0$ for $n = 1, \cdots, m$). If $\gamma_2 < -\nu$, then the set of constant equilibrium solutions Z_e in (8) is unstable, and if $\gamma_2 > -\nu$, then the set Z_e is locally exponentially stable.

4. Input-Feedback Control Design for the KdVBKS Equation

In this section, an input-feedback control design for the KdVBKS equation with periodic boundary conditions (5) is presented. Without loss of generality the controller design presented will stabilize the KdVBKS equation to the zero-equilibrium solution $z_e = 0$. However, the same control scheme can be applied to all equilibrium solutions of the KdVBKS equation.

Consider the controlled linearized KdVBKS equation at $z_e = 0$

$$\begin{aligned} \dot{z} &= F'z + \textstyle\sum_{n=1}^{m} B_n\, u_n(t) \\ z(0) &= z_0, \end{aligned} \tag{12}$$

where F' is defined in (10). First, the controller will be designed for the above linearized system, then the same controller will be used to the nonlinear KdVBKS equation (5) to achieve local stability results. This is true due to the Fréchet differentiability of the C_0-semigroup generated by the nonlinear controlled KdVBKS equation and that its derivative is the C_0-semigroup corresponding to the linearized system (12) [1].

Designing an input feedback control to linear infinite dimensional systems is well known (e.g. [3], [4, 11, 18, 19, 21, 39). In this paper, Linear Quadratic Regulator Controller (LQR) will be used to stabilize the linearized KdVBKS equation. In summary, if the controlled linearized KdVBKS equation (12) is exponentially stable, then the same controller will locally exponentially stabilize the nonlinear KdVBKS equation (5).

5. The Approximate Solution to the KdVBKS equation using Galerkin Projection Method

The idea of using the Galerkin projection method is used to produce a finite-dimensional system of ordinary differential equations (ODEs) that mimics the dynamics of both the KdVBKS and the linearized KdVBKS equations at $z_e = 0$. The controllers are designed for the finite dimensional approximation and hence, some conditions on the approximation scheme are needed to guarantee the reliability of the method and that the approximated finite-dimensional system is correctly capturing the dynamics of the original infinite dimensional system.

The approximate solution to the KdVBKS equation (5) with periodic boundary conditions can be written using

$$\begin{aligned} z_N(x,t) \approx\ & a_0(t)\phi_0(x) + \\ & + \sum_{n=1}^{N} a_n(t)\phi_n(x) + \sum_{n=1}^{N} d_n(t)\psi_n(x), \end{aligned}$$

where

$$\begin{aligned} \phi_0(x) &= \tfrac{1}{\sqrt{2\pi}} \\ \phi_n(x) &= \tfrac{1}{\sqrt{\pi}} cos(nx) \quad n = 1, \cdots, N \\ \psi_n(x) &= \tfrac{1}{\sqrt{\pi}} sin(nx) \quad n = 1, \cdots, N \end{aligned} \tag{13}$$

for a sufficiently large N.

When designing an input-feedback controller for infinite dimensional systems, approximation of the system is needed as most controller designs including the LQ controller scheme cannot be implemented using the full state. However, approximations are not always reliable; an example can be found in [21]. Some conditions are needed to guarantee that the approximations yield stabilizing controllers that correctly predict the closed-loop dynamics of the system. In particular, using the result [15, Theorem 3.3] as well as the Fréchet differentiability of the semigroup generated by the nonlinear system and the fact that its derivative is the semigroup generated by the linearized system which holds for the KdVBKS equation will guarantee a design of a controller that stabilizes the system.

Define an orthogonal projection $P_n: L^2(0,2\pi) \to V_n \subset H^1(0,2\pi)$ where for every $z \in L^2(0,2\pi)$, $\lim_{n\to\infty} \| P_n z - z \| = 0$. The space V_n is equipped with the norm inherited from $L^2(0,2\pi)$. This leads to a sequence of finite-dimensional Galerkin approximations. Using [25, Theorem 4.1], if there

exists a set of controllers K_n that exponentially stabilizes the linearized system approximation (P_nF', B), then for sufficiently large n, the controllers K_n locally stabilize the nonlinear KdVBKS equation. That is, any sequence of stabilizing controllers for the finite-dimensional linearized KdVBKS equation will locally stabilize the nonlinear infinite-dimensional system without any spillover effect [20, 21].

Theorem 5. Consider the linearized KdVBKS equation at $z_e = 0$ in (12). Assume that the sequence of approximations (F'_n, B_n) are stabilizable. Let K_n be a convergent sequence of controllers for the approximating systems such that the limit K exponentially stabilizes (F, B). Then for sufficiently high order n, the controllers K_n locally stabilize the nonlinear KdVBKS equation.

In summary if one can find a sequence of stabilizing controllers for the finite dimensional linearized KdVBKS equation, then controlling the nonlinear KdVBKS equation will follow; see [20, 21] for the full design approach.

6. Numerical Simulations

Consider the nonlinear KdVBKS equation (1) with $\nu = -4, \gamma_2 = 3, \mu = 0.2, \gamma_1 = 1$, and with $u_n(t) = 0$ for $n = 1, \cdots, m$ and the initial condition

$$z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} \big(\sin(nx) + \cos(nx)\big). \quad (14)$$

Using the Galerkin projection method, we approximate the solution of the uncontrolled KdVBKS equation using 51 modes (see Figs. 1 and 2). It is clear that the solution is unstable which is consistent with the theoretical findings. Moreover, the number of unstable eigenvalues is finite and is equal to 3.

Now we shall construct input feedback controllers using a single actuator and multi actuators scenarios and compare the results of both scenarios with one another. Since the number of unstable modes to the KdVBKS equation is 3, The first control design will be done using three actuators. That is, the parameter m in equation (1) is chosen to be 3, with

$$\begin{aligned} b_1(x) &= \frac{1}{\sqrt{2\pi}} \cdot \chi_{[0,2\pi]}, \\ b_2(x) &= \frac{1}{2\varepsilon} \cdot \chi_{\left[\frac{\pi}{2}-\varepsilon, \frac{\pi}{2}+\varepsilon\right]}, \\ b_3(x) &= \frac{1}{2\varepsilon} \cdot \chi_{\left[\frac{7\pi}{6}-\varepsilon, \frac{7\pi}{6}+\varepsilon\right]}, \end{aligned} \quad (15)$$

where $\varepsilon = 0.01$ and $\chi_{[a,b]}$ indicates the characteristic function with support on $[a, b]$. When designing an input feedback control, only unstable eigenvalues will be considered. In particular, the number of modes considered to approximate the solution is $N = 25$. That is, there will be 51 modes considered. However, the controller design will be done only on the 3 unstable modes. The LQR controller synthesis was used to construct an input feedback control of the form given in (4).

$$K = \begin{bmatrix} 0.9963 & 1.2511 & -0.5799 \\ 0.0703 & -1.8511 & 3.4911 \\ -0.0502 & -4.9984 & -0.0538 \end{bmatrix} \quad (16)$$

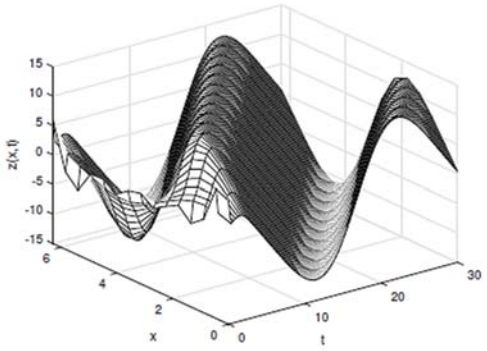


Fig. 1. A 3D landscape of the dynamics of the uncontrolled KdVBKS equation; $\boldsymbol{\nu = -4}$, $\boldsymbol{\mu = 0.2, \gamma_1 = 1, \gamma_2 = 3}$ and $\boldsymbol{z_0(x) = z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} \big(\sin(nx) + \cos(nx)\big)}$.

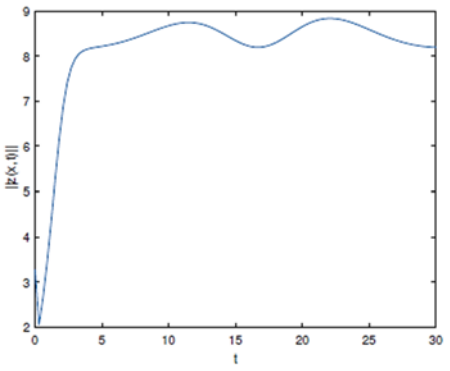


Fig. 2. The $\boldsymbol{L^2}$-norm of the solution of the KdVBKS equation; $\boldsymbol{\nu = -4}$, $\boldsymbol{\mu = 0.2, \gamma_1 = 1, \gamma_2 = 3}$ and $\boldsymbol{z_0(x) = z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} \big(\sin(nx) + \cos(nx)\big)}$.

Fig. 3 shows the controlled 51^{st} -order approximation to the nonlinear KdVBKS equation. Note that the controller was initially designed to stabilize the linearized system at $z_e = 0$. Furthermore, the controller is only designed using the 3 unstable modes, then the same controller is applied on the nonlinear system. The solution of the controlled system is successfully converging to the zero-equilibrium solution with no spillover effect.

The control problem can be improved by using less actuators. That is, one single actuator can be used to stabilize the KdVBKS equation and drive the solution to the desires zero equilibrium. Let

$$b(x) = \frac{1}{2\varepsilon} \cdot \chi_{\left[\frac{7\pi}{6}-\varepsilon, \frac{7\pi}{6}+\varepsilon\right]}, \quad (17)$$

be a single actuator for the controlled problem (1). That is, the parameter $m = 1$. Then using LQR controllers,

an input feedback control can be constructed using the three unstable modes to the system. The controller is designed for the linearized system, and it is applied to the nonlinear system where the convergence is successful without the spillover effect.

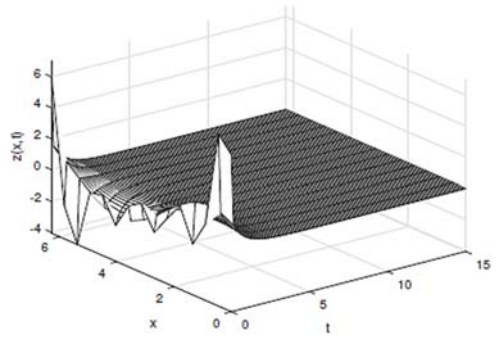


Fig. 3. A 3D landscape of the dynamics of the controlled KdVBKS equation; $\boldsymbol{\nu = -4, \mu = 0.2, \gamma_1 = 1, \gamma_2 = 3}$ and $\boldsymbol{z_0(x) = z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} (\sin(nx) + \cos(nx))}$ using three actuators.

$$K = [-1.0000 \quad -5.896 \quad -7.2283] \qquad (18)$$

Fig. 4 shows the controlled 51^{st} -order approximation to the nonlinear KdVBKS equation using only one single actuator. Note that the controller was initially designed to stabilize the linearized system at $z_e = 0$. Furthermore, the controller is designed only using the 3 unstable modes, and the same controller is applied on the nonlinear system. The solution of the controlled system is successfully converging to the zero-equilibrium solution with no spillover effect.

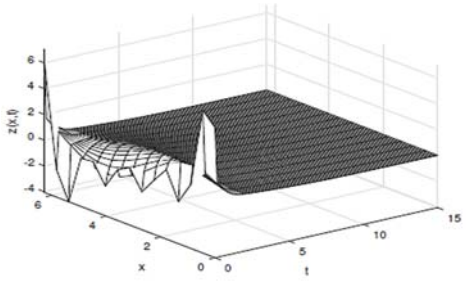


Fig. 4. A 3D landscape of the dynamics of the controlled KdVBKS equation; $\boldsymbol{\nu = -4, \mu = 0.2, \gamma_1 = 1, \gamma_2 = 3}$ and $\boldsymbol{z_0(x) = z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} (\sin(nx) + \cos(nx))}$ using a single actuator..

The L^2 -norm of the solution $z(x,t)$ to the controlled KdVBKS equation is presented in Fig. 5 using the single and the three actuators approaches. It is clear that both approaches have successfully controlled the system and the solution has converged to the desired zero-equilibrium solution. However, the convergence rate for the three-actuator model is faster than the one for the single actuator model. Of course, this is expected since the design cost of three actuators is more expansive than the design cost of one single actuator model. Finally, it should be noted that the choice of the number of actuators taken according to the number of unstable eigenvalues to control the KdVBKS equation is not a necessary condition to stabilize the system to the zero-equilibrium solution.

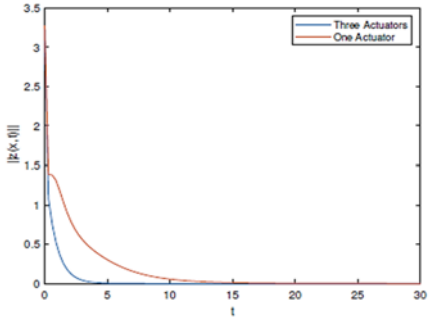


Fig. 5. The $\boldsymbol{L^2}$-norm of the solution to the controlled KdVBKS equation; $\boldsymbol{\nu = -4, \mu = 0.2, \gamma_1 = 1, \gamma_2 = 3}$ and $\boldsymbol{z_0(x) = z_0(x) = \frac{1}{\sqrt{2\pi}} + \frac{2.5}{\sqrt{\pi}} \sum_{n=1}^{5} (\sin(nx) + \cos(nx))}$ using a single and three actuators.

7. Conclusions

Designing an input-feedback controller for the infinite dimensional KdVBKS equation with periodic boundary conditions was presented. Two control design cases were analyzed using three actuators and then one single actuator approach. It is shown that in both approaches, the KdVBKS equation was controlled to the zero-equilibrium solution. This is done by designing an input feedback controller for the linearized system at the desired zero equilibrium solution. The controller was then used to locally stabilize the nonlinear KdVBKS equation. This approach holds due to the Fréchet differentiability of the nonlinear C_0-semigroup generated by the nonlinear system with its derivative being the linear C_0-semigroup generated by the linearized system. The two approaches of the control synthesis worked in stabilizing the system to the zero-equilibrium solution. However, using more than one actuator, although it is more expensive to design, will lead to a faster convergence rate. Finally, numerical results were presented to illustrate the control design.

References

[1]. R. Al Jamal and K. Morris. Linearized stability of partial differential equations with application to stabilization of the Kuramoto-Sivashinsky equation, *SIAM journal on Control and Optimization*, 56, 1, 2018, pp. 120 – 147.

[2]. D. Armbruster, J. Guckenheimer, and P. J. Holmes. Kuramoto-Sivashinsky dynamics on the center-unstable manifold, *SIAM Journal on Applied Mathematics*, 49, 3, 1989, pp. 676–691.

[3]. G. Bastin, J. Coron, and B. d'Andrea Novel, On Lyapunov stability of linearised Saint-Ventant

equations for a sloping channel, *Network and Heterogeneous Media,* 4, 2, 2009, pp. 177 – 187.

[4]. A. Bensoussan, G. Prato, M. Delfour, and S. Mitter, Representation and control of infinite-dimensional systems, 2nd edition, *Birkhauser*, 2007.

[5]. P. Bracken, Some methods for generating solutions to the Korteweg-de Vries equation, *Physica A,* 335, 2004, pp. 70–78.

[6]. H. Brown, I. Kevrekidis, A. Oron, and P. Rosenau, Bifurcations and pattern formation in the 'regularized' Kuramoto-Sivashinsky equation, *Physics Letters A,* 163, 4, 1992, pp. 299 – 308.

[7]. T. M. A. Burgers. Mathematical Model Illustrating the Theory of Turbulence, *Advances in Applied Mechanics,* 1, 1948, pp. 171–199.

[8]. P. Christofides and A. Armaou, Global stabilization of the Kuramoto-Sivashinsky equation via distributed output feedback control, *Systems and Control Letters,* 39, 2000, pp. 283-294.

[9]. P. D. Christofides and A. Armaou, Global stabilization of the Kuramoto-Sivashinsky equation via distributed output feedback control, *Systems and Control Letters,* 39, 2000, pp. 283–294.

[10]. A. T. Cousin and N. A. Larkin, Kuramoto-Sivashisky equation in domains with moving boundaries. *Portugaliae Mathematica,* 59, 2002, pp. 336–349.

[11]. R. Curtain and H. Zwart, An introduction to infinite-dimensional linear systems theory, *Springer-Verlag,* 1995.

[12]. J. Elgin and X. Wu. Stability of cellular states of the Kuramoto-Sivashinsky equation, *SIAM Journal on Applied Mathematics,* 56, 1996, pp. 1621–1638.

[13]. C. Foias, M. Jolly, I. Kevrekidis, G. Sell, and E. Titi, On the computation of inertial manifolds, *Physics Letters A,* 131, 1989, pp. 433–437.

[14]. J. Hyman and B. Nicolaenko, The Kuramoto-Sivashinsky equation: A bridge between PDEs and dynamical systems, *Physica D,* 18D, 1986, pp. 113–126.

[15]. R. A. Jamal, A. Chow, and K. Morris. Linearized stability analysis of nonlinear partial differential equations, in *Proceedings of the 21st International Symposium on the Mathematical Theory of Networks and Systems,* 2014, pp. 847-852.

[16]. G. Karch. Self-similar large time behavior of solutions to Korteweg-de Vries-Burgers equation, *Nonlinear Analysis,* 35, 1999, pp. 199-219.

[17]. T. Kato. A principle of linearized stability for nonlinear evolution equations, *Transactions of the American Mathematical Society,* 347, 8, 1995.

[18]. I. Lasiecka and R. Triggiani, Control theory for partial differential equations: Continuous and approximation theories, *Cambridge University Press*, 2000.

[19]. Z. Luo, B. Guo, and O. Morgul. Stability and stabilization of infinite dimensional systems with applications, *Springer-Verlag*, London, 1999.

[20]. K. Morris. Design of finite-dimensional controllers for infinite dimensional systems by approximation, *Journal of Estimation and Control,* 4, 2, 1994, pp. 1–30.

[21]. K. Morris, Control of systems governed by partial differential equations, IEEE Control Theory Handbook, *CRC Press*, 2010.

[22]. N. Smaoui, B. Chentouf, and A. Alaa, Modelling and non-linear boundary stabilization of the modified generalized Korteweg-de Vries-Burgers equation. *Adv. Differ. Equ.,* 2019.

[23]. N. Smaoui and R. Al Jamal, Dynamics and control of the modified generalized Korteweg-de Vries-Burgers equation with periodic boundary conditions, *Nonlinear Dynamics,* 103, 6, 2021.

[24]. B. Nicolaenko, B. Scheurer, and R. Temam, Some global dynamical properties of the Kuramoto-Sivashinsky equations: Nonlinear stability and attractors, *Physica D,* 16D, 2, 1985, pp. 155 – 183.

[25]. R. Al Jamal and N. Smaoui, A single bounded input-feedback control to the generalized Korteweg-de Vries-Burgers-Kuramoto-Sivashinsky equation, *Mathematical Methods in Applied Sciences,* 46, 2, 2022, pp. 2022-2048.

[26]. L. Rosier, Exact boundary controllability for the Korteweg-de Vries equation on a bounded domain, *ESAIM Control, optimisation and Calculus Variations,* 2, 1997, pp. 33–55.

[27]. L. Rosier, Exact boundary controllability for the linear Korteweg-de Vries equation - A numerical study. ESAIM Proceedings, 4, 1998, pp. 255 - 267.

[28]. D. Russell and B. Zhang, Smoothing and decay properties of solutions of the Korteweg-de Vries equation on a periodic domain with point dissipation, *Journal of Mathematica, Analysis and Applications,* 190, 2, 1995, pp. 449–488.

[29]. D. Russell and B. Zhang, Exact controllability and stabilizability of the Korteweg-de Vries equation, *Transactions of American Mathematical Society,* 348, 9, 1996, pp. 3643–3672.

[30]. D. L. Russell and B. Zhang, Controllability and stabilizability of the third-order linear dispersion equation on a periodic domain, *SIAM Journal on Control and Optimization, 3*1, 3, 1993, pp. 659–676.

[31]. R. Sakthivel and H. Ito, Nonlinear robust boundary control of the Kuramoto-Sivashinsky equation, *SIMA Journal of Mathematical Control and Information,* (UK), 24, 1, 2007.

[32]. G. Sell and Y. You, Dynamics of evolutionary equations, *Springer-Verlag*, 143, 2000.

[33]. D. Senouf, Dynamics and condensation of complex singularities for Burgers' equation, *SIAM Journal in Mathematical Analysis,* 28, 6, 1997, pp. 1457–1489.

[34]. N. Smaoui, Nonlinear boundary control of the generalized burgers equation, *Nonlinear Dynamics Journal,* 37, 1, 2004, pp. 75–86.

[35]. N. Smaoui, Boundary and distributed control of the viscous Burgers equation, *Journal of Computational and Applied Mathematics,* 182, 2005, pp. 91–104.

[36]. N. Smaoui and R. Al Jamal, Distributed control of the generalized Korteweg-de Vries-Burgers equation, *Math. Probl. Eng.,* 2008.

[37]. N. Smaoui, E. El-Kadri, and M. Zribi, Nonlinear boundary control of the unforced generalized Korteweg-de Vries-Burgers equation, *Nonlinear Dynamics,* 60, 2010, pp. 561–574.

[38]. J. Smoller, Shock waves and reaction-diffusion equations, *Springer-Verlag*, New York, 2nd edition, 1994.

[39]. J. Zabczyk, Mathematical control theory: An introduction, *Birkhauser*, 2007.

[40]. Y. Zhang, L. Song, and W. Axia, Dynamical bifurcation for the Kuramoto-Sivashinsky equation, *Nonlinear Analysis*, 74, 2011, pp. 1155–1163.

(7516)

VPN Helper Application for using FIDO2 Security Keys with Legacy VPN Systems

Dr. Emin Huseynov
Azerbaijan Technical University, H. Cavid prospekti 25, 1074 Baku, Azerbaijan
Token2 Sàrl Chemin Pre-Colomb 10, 1290 Versoix, Switzerland
Tel.: +41 79 695 26 00
E-mail: emin@huseynov.com

Summary: Secure VPN connections are essential for protecting sensitive data and maintaining privacy and security. A VPN helper app based on NW.js that includes FIDO2 support is proposed in this paper as a solution to the lack of FIDO2 support in most VPN clients. This app would allow users to log in with a physical FIDO2 security key instead of a password, improving security by making it resistant to attacks such as password cracking or phishing. FIDO2 security keys can also enhance the user experience by eliminating the need to remember complex passwords and reducing the risk of account lockouts due to forgotten passwords. The VPN helper app would provide an easy-to-use solution for VPN users looking to take advantage of the added security and convenience of FIDO2 authentication.

Keywords: VPN (Virtual private network), FIDO2 (Fast identity online 2), MFA (Multifactor authentication), Security keys, Public key cryptography, Biometric authentication, Phishing resistance.

Summary

The paper starts with an introduction to the concept of VPN connections and the importance of maintaining security and privacy. Proposes a concept for a VPN helper app based on NW.js that includes FIDO2 support. Then it provides an overview of current VPN client options and their limitations, including the concept of multi-factor authentication (MFA) and its role in protecting against security breaches. Section 2: discusses the various methods of MFA that can be used with VPN systems, including the use of FIDO2 for strong, phishing-resistant authentication and describes the proposed VPN helper app based on NW.js and its features and capabilities, including the role of the FIDO2 Security key in the authentication process. The conclusion summarizes the main points of the paper and suggests potential future developments and directions for the app.

1. Introduction

Secure VPN connections are essential for protecting sensitive data and maintaining privacy and security. A VPN creates an encrypted connection between a device and a network, helping to protect against attacks such as hacking and data interception [1]. This is especially important when accessing the internet from public locations or using unsecured networks. FIDO2 security keys [2] can enhance the security of VPN connections by allowing users to log in with a physical device instead of a password, which can be easily compromised through attacks like password cracking or phishing. FIDO2 security keys can also improve the user experience by eliminating the need to remember complex passwords and reducing the risk of account lockouts due to forgotten passwords. Despite the increasing adoption of FIDO2 security keys, most VPN clients do not support them, making it difficult for users to take advantage of this enhanced security measure. This paper proposes a concept for a VPN helper app based on NW.js [3] that includes FIDO2 support to address this issue and provide an easy-to-use solution for VPN users.

1.1. Overview of Current VPN Client Options and Their Limitations

A VPN is a technology that creates a secure connection over a less-secure network, such as the internet. When using a VPN, internet traffic is routed through an encrypted tunnel, which helps to protect data from being intercepted by unauthorized parties. VPNs are often used to protect sensitive data, such as login credentials and financial information, when accessing the internet from a public location like a coffee shop or airport. They can also be used to bypass internet censorship and access restricted content in certain countries. In addition to providing security and privacy, VPNs can also be used to obscure one's IP address and appear as if the user is accessing the internet from a different location. This can be useful for accessing content that is only available in certain regions or for protecting privacy online. There are various VPN services available that can be used on various devices, including computers, smartphones, and tablets.

Legacy VPN systems and clients have not natively supported multi-factor authentication (MFA) [4], which leaves organizations vulnerable to security breaches. MFA, also known as two-factor authentication, is a security measure that requires users to provide additional verification beyond just a password in order to access protected systems. This

can include using a physical device, such as a security key or smartphone, to receive a one-time code or using biometric authentication methods, such as fingerprint or facial recognition. The lack of MFA support in legacy VPN systems and clients means that users are only required to provide a single factor, such as a password, to gain access, leaving the system vulnerable to attacks like password cracking or phishing. To address this issue, organizations may need to upgrade their VPN systems and clients to ones that natively support MFA to provide an additional layer of protection and prevent unauthorized access to sensitive systems and data.

1.2. Multifactor Authentication Methods

There are multiple methods for MFA that can be used with VPN systems, even if the VPN system does not natively support MFA. These methods involve adapting the MFA method to be used as part of the authentication process for the VPN. However, many of these methods rely on one-time passwords (OTPs) for authentication, which are not resistant to phishing attacks.

FIDO2 (Fast Identity Online 2) [2] is a set of open standards that provide a secure and convenient way to authenticate users online. It enables the use of strong, phishing-resistant authentication methods, such as security keys and biometric authentication, to replace weak and easily compromised password-based authentication.

MFA based on FIDO security keys can meanwhile be considered as phishing resistant. These are physical devices that use public key cryptography to confirm the identity of the user. Security keys can be used with a variety of online services, including VPNs, and are highly resistant to phishing attacks because they require physical possession of the key to access the service.

1.3. Existing VPN Protocols

There are various types of protocols that can be used for VPN (virtual private network) connections. Some of the most common legacy VPN protocols include [4]:

- PPTP (Point-to-Point Tunneling Protocol): PPTP is a widely used protocol that is supported by most VPN clients. It is relatively fast and easy to set up, but it is not as secure as some other protocols.
- L2TP (Layer 2 Tunneling Protocol): L2TP is a combination of PPTP and L2F (Layer 2 Forwarding Protocol). It provides stronger security than PPTP, but it is slower and requires more processing power.
- SSTP (Secure Socket Tunneling Protocol): SSTP is a Microsoft-developed protocol that is designed to provide secure and reliable connections. It uses SSL (Secure Sockets Layer) to create an encrypted connection and is generally considered to be more secure than PPTP and L2TP.
- IKEv2 (Internet Key Exchange version 2): IKEv2 is a secure and efficient protocol that is designed for use with mobile devices. It is able to quickly establish and maintain VPN connections, even in situations where the network connection is unstable.

Each of these protocols has its own unique features and benefits, and the best choice will depend on the specific needs of the user and the resources available. However, none of the existing clients supporting these protocols natively support the use of FIDO2 Security keys.

2. Summary of the Proposed Solution

2.1. Addressing the Operating Systems Limitations

FIDO2 Security keys are designed to operate inside browsers and most of the available API are web oriented. It is, therefore, quite hard to adapt the VPN clients to serve as FIDO2 Security key compatible software. The idea of this paper is to use existing solutions based on a runtime environment that allows developers to create native applications for the desktop using web technologies such as JavaScript, HTML, and CSS. In addition to regular web technologies, such environment gives access to methods built-in to browser engines that will allow to use security keys for authentication.

We will be leveraging NW.js to achieve our goal in this project. NW.js (previously known as node-webkit) is one of examples of such an envrionment. NW.js combines the power of the Node.js runtime with a web browser engine (Google Chromium) [3] to provide a platform for building desktop applications using web technologies.

2.2. Principle of the Solution

Using NW.js we will create a native desktop application which will have the possibility to operate as a web browser, as it is using a browser engine and in the same time to the operating system's API, as it a desktop application. The diagram of the application is shown on Fig. 1.

A web page, responsible for the authentication, will be set as the starting point of our VPN Client application. As the FIDO2 protocols are based on the verification of the web host's certificates, the web application has to be running on a web server. The web application will prompt for a FIDO2 Security key based authentication, and once authentication is successful a temporary, unique, and allowed to use only once username and password will be created and passed back to the VPN Client application.

At the final phase, the VPN Client application will launch a system command using PowerShell responsible for establishing a VPN connection and pass the username and password to it as parameters [5].

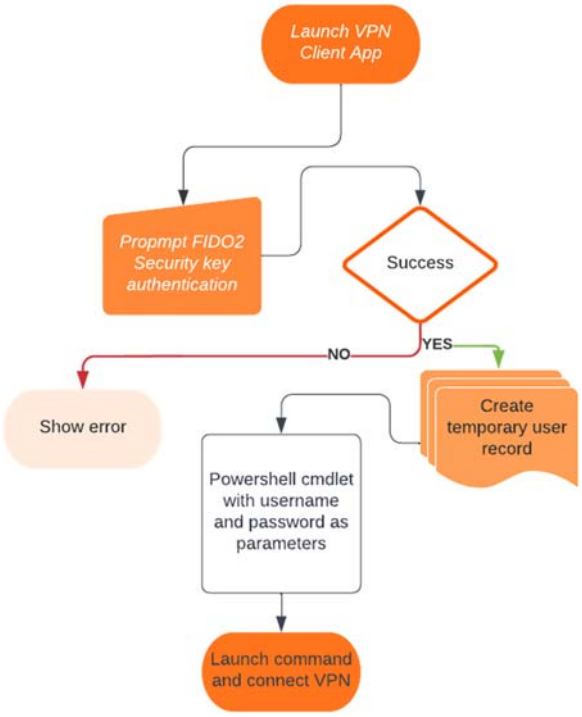


Fig. 1. VPN Client application diagram.

3. Conclusion and Future Work

3.1. Conclusion

This paper is proposing a concept for a VPN helper app based on NW.js that includes support for FIDO2 authentication. The goal of this app is to provide an easy-to-use solution for VPN users who want to take advantage of the added security and convenience of FIDO2 authentication. Most VPN clients do not currently support FIDO2, making it difficult for users to access this enhanced security measure. The app would allow users to log in with a physical FIDO2 security key instead of a password, which can be easily compromised through attacks like password cracking or phishing. FIDO2 security keys can also improve the user experience by eliminating the need to remember complex passwords and reducing the risk of account lockouts due to forgotten passwords. The app would be a native desktop application that could operate as a web browser and also have access to the operating system's API. The app would open a web page for authentication hosted on a web server, and the user would authenticate using a FIDO2 Security key. The web application would generate a temporary, unique username and password, which would be passed back to the VPN Client application. The VPN Client application would then launch a system command to establish a VPN connection and pass the username and password as parameters.

3.2. Future Work

Although the paper is describing a VPN helper app that would secure the authentication process itself, there may be other risks in the process of connecting to a VPN system. Even if the authentication phase is secure, the system it is used for has to be resistant to other risks such as DNS spoofing attacks [6] or Man-in-the-Middle (MITM) attack risk [7] – therefore, when using this method in production, those additional risks have to be taking into account, assessed and mitigated accordingly where possible.

Acknowledgements

This paper is supported by Token2 Sàrl, a Swiss company specializing in multifactor authentication solutions.

References

[1]. Sharma, Y. K., & Kaur, C, The Vital Role of Virtual Private Network (VPN) in Making Secure Connection Over Internet World., *International Journal of Recent Technology and Engineering (IJRTE)*, Vol. 8. , 2020, pp. 2336-2339.

[2]. Lyastani, S. G., Schilling, M., Neumayr, M., Backes, M., & Bugiel, S., Is FIDO2 the Kingslayer of User Authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication, in *Proceedings of the EEE Symposium on Security and Privacy*, 2020, pp. 268-285.

[3]. Benoit, A. , NW. js Essentials, *Packt Publishing Ltd,* Birmingham – Mumbai, 2015.

[4]. Matsui, K., Ota, K., & Kurita, H, The authentication platform for VPN services, *IEICE Technical Report*, Vol. 110, No. 224, 2010, pp. 25-30.

[5]. Microsoft Corp., Add-VpnConnection, Microsoft Learn, 01 11 2020. [Online]. Available: https://learn.microsoft.com/en-us/powershell/module/vpnclient/add-vpnconnection?view=windowsserver2022-ps. [Accessed 15 12 2022].

[6]. Steinhoff, U., Wiesmaier, A., & Araújo, R., The state of the art in DNS spoofing, in *Proceedings of the 4th Intl. Conf. Applied Cryptography and Network Security (ACNS),* 2006.

[7]. Sun, Y., Wang, B., Wang, C., & Wei, Y., On Man-in-the-Middle Attack Risks of the VPN Gate Relay System, *Security and Communication Networks,* Volume 2021, Article ID 9091675.

(7554)

Generative Anomaly Detection in Multivariate Time Series

M. Hoh, A. Schöttl, H. Schaub and N. Leuze
University of Applied Sciences, Dept. of Electrical Engineering and Information Technology,
Institute for Applications of Machine Learning and Intelligent Systems (IAMLIS),
Lothstrasse 64, 80335 Munich, Germany
Tel.: +49891265-3480, fax: +49891265-3403
E-mail: maximilian.hoh@hm.edu

Summary: Multivariate time series can reveal a lot about their present circumstances. Such data may be used for a variety of purposes, including tracking patient health information, recognizing server hacking attempts or spotting unusual activity in industrial plants. In this work we introduce a novel approach for detecting anomalies in multivariate time series. We use a Generative Adversarial Network (GAN) that is conditioned on the input signal by a convolutional autoencoder consisting of one encoder and two decoders. The generator of the GAN reconstructs the input twice with different quality, which helps the discriminator to distinguish better between the original sample and the generated ones. While other approaches are using smaller time windows as input, we focus on longer sequences to be able to identify and evaluate long-term dependencies in the time series. The anomaly score for each input is calculated as a weighted combination of the discriminator output and the latent space differences. Experiments show that our setup returns better results in terms of F1 scores for common anomaly detection datasets than networks with comparable input sample length.

Keywords: Machine learning, Multivariate time series, Anomaly detection, Conditional GAN.

1. Introduction

The recent growth of data collection in the industrial environment is generating an increased demand for automatic anomaly detection systems. In the era of Industry 4.0, sensors are built into every machine or device to monitor processes. This data can also be used to detect anomalous behavior or cyber attacks on systems at an early stage such that further damage or danger can be averted.

Although anomaly detection algorithms have been extensively studied, further research is still required as solutions are application specific and can not be adapted to different situations. *Thresholding* is one of the easiest methods to find data points that are above or below a predetermined range. Without exceeding the threshold, an abnormal behavior can also arise in a setting with values that appear normal in other channels when there is a multivariate data stream from different sensors.

Deep learning approaches [1] proved to be a reliable alternative to statistical [2] and classical machine learning-based methods like clustering [3] for managing contextual correlation. Recent approaches focus on unsupervised learning due to a lack of labeled data. Data can be recorded fast by multiple sensors but labeling requires expert knowledge. It is assumed when using unsupervised approaches that a training dataset primarily represents the nominal state and that anomalies are infrequent occurrences. For the purpose of validating the algorithm and doing performance measurements, testing data are labeled by experts.

Generative Adversarial Networks (GANs) are widely used for anomaly detection [4]. In general, a GAN consists of two networks. A generator learns to generate real-looking samples from a random vector. A discriminator is trained to decide whether its input is a sample from the training dataset or a generated one.

In our work, a novel conditional GAN is introduced that is able to detect anomalies in multivariate time series data. The generator is thereby conditioned on the input signal and generates two outputs of different qualities. This structure helps in stabilizing the training for both the generator and discriminator. The anomaly detection takes place in the latent space.

2. Related Work

Reconstruction-based methods like autoencoders [5] learn a latent representation of the input signal. By mapping anomalous signals during testing to a lower dimensional space, information about the anomaly is lost and the input signal cannot be reconstructed properly. Thus, a higher likelihood of being anomalous is indicated by a large reconstruction error. Depending on the composition of the dataset, autoencoders may attempt to fit anomalies since it cannot be guaranteed that an unlabeled dataset will contain normal data points only. This might lead to an increase in false negatives.

GANs do not have this kind of issues due to adversarial training. Through the interaction of a generator and a discriminator, both networks train each other and can generate samples of similar distribution as the input or distinguish generated from real samples. A set of hyperparameters must be found to balance both components of the architecture such that none performs significantly better than the other and thus prevent it from learning. Either one or both trained networks can then be further used for an anomaly detection.

A convolution based vanilla GAN structure is used by Schlegl et al. [6] to detect anomalies in medical images. Based on a BiGAN architecture [7] Zenati et al. [8] improved the results in GAN-based anomaly detection. In [9] the generator is structured as an autoencoder that tries to reconstruct the input image. In [10] Li et al. implemented long-short term memory (LSTM) layers in their GAN and benefited from the ability to learn long-term dependencies between time steps. Audibert et al. [11] introduced a network with one encoder and two decoders that are trained in an adversarial fashion. The authors reconstruct the input signal twice, as the first reconstruction is looped through the encoder again and then decoded by the second decoder. Their approach measures the anomaly score using the reconstruction loss. An LSTM-based GAN using two discriminators and an autoencoder as a generator is presented in [12]. Real and generated signals may be distinguished by the first discriminator, while the second evaluates how well the latent space is mapped. While [10, 11] use smaller time series (5-30 time steps) as an input to the model, [12] is using an input size of 100 time steps. In [13], high-sampled audio files from a malfunctioning industrial plant were classified correctly by an autoencoder GAN. The anomaly score is calculated in the latent space.

We also make use of the benefits of autoencoders and GANs. We introduce a novel GAN architecture to identify unusual sections of multivariate time series. Our GAN consists of a twofold convolutional autoencoder as a generator and a single discriminator with the signals as an input. Anomalies are detected by comparing latent representations and by the output of the discriminator.

3. Algorithm

A multidimensional time series **X** is described by the data points $\{x^1, x^2, \dots, x^T\}$, where $x^t \in \mathbb{R}^d$ are the measurements at time step t. The training samples are acquired by splitting the original time series **X** into $N = \frac{T - s_w}{s_s}$ sub-sequences $X = \left\{x_i^{1 \dots s_w}\right\}_{i=1}^{N}$ with a window size of s_w and step size s_s. The proposed architecture is a GAN that is conditioned on the input signal x by an autoencoder structure. For each window an anomaly score is calculated that indicates the chance of the sub-sequence to be anomalous.

3.1. Architecture

The first part of our architecture represents the autoencoder-like generator of the GAN. The signal x is first encoded to a latent $z = G_{\text{en}}(x)$ and then decoded twofold by the two different decoders to $x' = G'_{de}(z)$ and $x'' = G''_{de}(z)$. The generated signals are brought back into their encoded representations z' and z'' by using G_{en} once more in order to calculate loss values during training and the anomaly score when using the testing dataset. This structure supports the training of the generator, as the input of the decoders is a latent vector of the training sample and not a random vector like in a traditional generator. The second part of the GAN is a vanilla discriminator D. We were able to demonstrate that the autoencoder is capable of both signal reconstruction and signal generation by using a combination of reconstruction-based and adversarial training. Fig. 1 gives an overview of all the different modules and their connections.

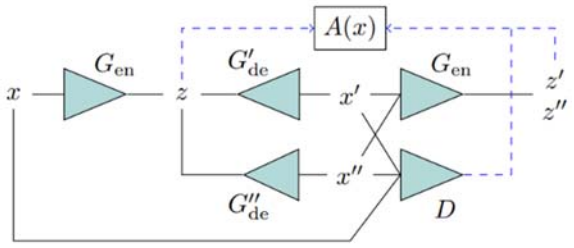


Fig. 1. The generator G consists of an encoder G_{en} and two decoders G'_{de} and G''_{de}. The discriminator D represents the counterpart to G. $A(x)$ is the calculated anomaly score.

We designed G_{en} to be a convolutional network and G'_{de} as the associated decoder. G''_{de} is dimensioned as a smaller network, which therefore has lower representational power. A comparison of the loss values during training demonstrates the different qualities of the structures as the loss curves of G'_{de} converge earlier. Each of the decoders output a signal with the shape of x. D is used as the counterpart of the generator and thus also a convolutional network. The discriminator can learn a distinction more quickly when x'' is reconstructed more poorly.

The proposed model combines several modules. The two autoencoders are sharing the same encoder and form the generator of a conditional GAN. Each of the autoencoders has its own loss function consisting of a reconstruction loss and an adversarial loss

$$\begin{aligned} \mathcal{L}\left(G_{en}, G'_{de}\right) &= \mathcal{L}_{re} + \mathcal{L}_{adv}, \\ \mathcal{L}\left(G_{en}, G''_{de}\right) &= \mathcal{L}_{re} + \mathcal{L}_{adv}. \end{aligned} \tag{1}$$

The reconstruction loss of the encoder-decoder network is a weighted combination of the reconstruction losses in the time, frequency, and latent domain and can be summarized as

$$\mathcal{L}_{re} = \lambda_t \mathcal{L}_t + \lambda_s \mathcal{L}_s + \lambda_l \mathcal{L}_l, \tag{2}$$

with $\mathcal{L}_t$ and $\mathcal{L}_l$ being the mean squared error between the original and the reconstructed signal and latent space respectively. The signal spectrogram discrepancies are represented by the spectral loss $\mathcal{L}_s$, which forces the model to pay attention to mid-to-high frequencies as described in [14]. We apply the approximated Wasserstein distance with gradient penalty [15] as $\mathcal{L}_{adv}$ for the generator and discriminator.

3.2. Anomaly Detection

The anomaly score A of an input x is calculated as a weighted sum of loss terms for this window. The latent representations$z$, z' and z'' are compared as an L2 norm similar to [9, 13] and the output of the discriminator D is taken into consideration like in [10]. The anomaly detection is window-based, so every window obtains one score. If it exceeds a predefined threshold, the window is labeled to be anomalous [11].

4. Experiments

4.1. Datasets

We trained our network on the public datasets SMAP (Soil Moisture Active Passive satellite) and MSL (Mars Science Laboratory rover) [16]. The time series represent multidimensional spacecraft telemetry values. Each dataset consists of several recordings, each of which requires separate training. Table 1 gives an overview of the dataset details.

Table 1. Dataset details.

	MSL	SMAP
Number of recordings	27	53
Number of channels	55	25
Average length of a training recording	2160	2503
Average length of a testing recording	2730	8070
Anomalous testing data	13.13 %	10.72 %

4.2. Results

Before training we normalized the data [-1,1]. We compared our results to [12] as we chose a similar window length $s_w = 128$. We chose $s_s = \frac{s_w}{16}$ to extract the most data from the limited time series. The samples thus created allow us to observe both long-term dependencies as well as punctual events.
In both testing datasets every time step is labeled to be normal or anomalous. We define an input sample to be anomalous, if at least 10% of the time steps are labeled as anomalous. The resulting F1-scores are summarized in Table 2. The scores for [10, 12] are taken from [12]. For evaluating the datasets and calculating anomaly scores, we use non-overlapping windows, so $s_s = s_w$.

Table 2. F1-scores of different datasets and networks.

	MSL	SMAP
[10]	0.11	0.13
[12]	0.62	0.70
Ours	0.67	0.71

While a vanilla GAN as proposed by Li et al. [10] seems not to be suitable for multivariate time series anomaly detection, the performance of the modified adversarial network in [12] is significantly better. They highly benefit from conditioning the generator on the input and the use of a second discriminator that operates in the latent space. In the SMAP and MSL datasets, our design outperforms [10] and [12] without using LSTM layers and instead relying on a convolutional structure. Also the introduction of a second decoder helps in stabilizing the training and resulting in a better performance.

According to [17] not all the recordings in the datasets are appropriate for anomaly detection. For instance, contrary to the idea that anomalies occur rarely, some time series are excessively impacted by anomalies. According to the authors of [17], 30 of the 81 recordings of SMAP and MSL do not meet their requirements. As can be seen in table 1, a closer look at the dataset reveals a high level of contamination. We trained and tested our network using all of the available data because the specific outcomes for each recording are not published and there is no more information in [12] if they filtered the time series prior, as far as we know.

Fig. 2 illustrates exemplarily sub-sequences of two recordings of the SMAP dataset. The top graphs show the first four channels of the signal. The highlighted area is the labeled anomaly. The lower section of the visualizations show how the signal is splitted into non-overlapping windows and which window is predicted to be anomalous by our algorithm. The examples demonstrate that the primary anomaly was discovered, but it is possible that the labels also reveal non-normal elements in the neighboring windows, even if they are not immediately obvious. The experiments show that the anomaly mostly occurs in the first channel of the signal. The first anomaly in Fig. 2b indicates a correlation between the first and other channels.

5. Conclusions

We proposed a novel GAN-based approach for anomaly detection in multidimensional time-series data which outperforms comparable state-of-the-art models in terms of F1-scores. We choose a window size that is up to 60 times bigger than what is used by other publishers. With a big window size, a model can learn not just punctual characteristics but also long-term dependencies.

We added a second, less effective decoder to our autoencoder generator, which aids the discriminator in making a clearer distinction between generated samples and training samples from the dataset. The generator also benefits from the encoder-decoder structure, as a real-looking sample can be generated from the compressed signal and not from a random vector. The output of the discriminator demonstrates that the generator not only has the ability to reconstruct multidimensional data but also to produce new signals with a similar distribution to the input samples.

An anomaly score determines whether the input is normal or abnormal by weighing the disparities between the encodings of the signal and its reconstructions, and the discriminator output. The anomaly scores are the basis of the evaluation, as testing datasets contain labeled data.

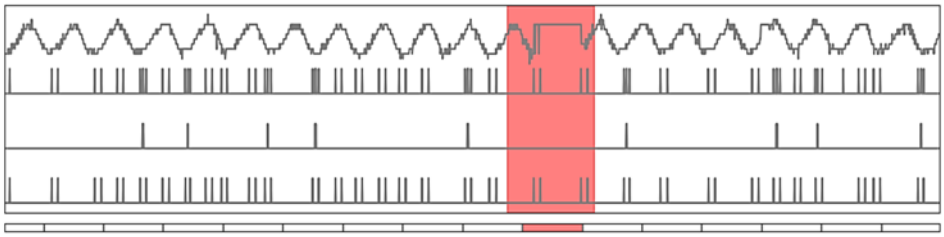


Fig 2a. Example signal A-3 from SMAP.

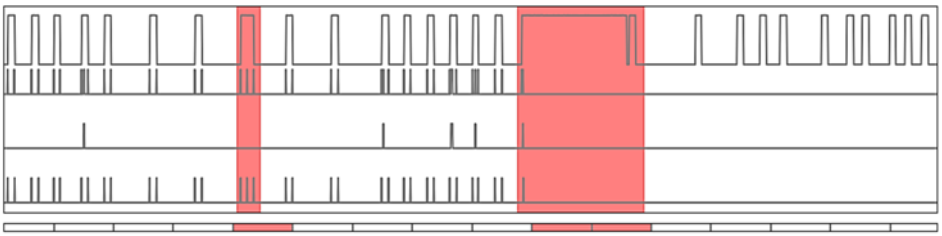


Fig 2b. Example signal E-10 from SMAP.

Fig. 2. Top: The first four channels of an example signal. The anomalous part is highlighted. Bottom: The input windows of the network. A red window indicates an exceeding of the threshold and therefore a detected anomaly.

References

[1]. D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim and K. Kim, A survey of deep learning-based network anomaly detection, *Cluster Computing,* Vol. 22, 2019, pp. 949-961.

[2]. D. Zheng, F. Li and T. Zhao, Self-adaptive statistical process control for anomaly detection in time series, *Expert Systems with Applications,* Vol. 57, 2016, pp. 324-336.

[3]. M. Amer and Markus Goldstein, Nearest-Neighbor and Clustering based Anomaly Detection Algorithms for RapidMiner, in *Proceedings of the 3rd RapidMiner Community Meeting and Conference*, 2012.

[4]. X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang and N. Ding, GAN-based anomaly detection: A review, *Neurocomputing,* Vol. 493, 2022, pp. 497-535.

[5]. J. An and S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, *Special Lecture on IE,* Vol. 2, 2015.

[6]. T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth and G. Langs, Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery, in *Information Processing in Medical Imaging*, 2017, pp. 146-157.

[7]. J. Donahue, P. Krähenbühl and T. Darrell, Adversarial Feature Learning, *International Conference on Learning Representations*, 2017.

[8]. H. Zenati, C. S. Foo, B. Lecouat, G. Manek and V. R. Chandrasekhar, Efficient GAN-Based Anomaly Detection, 2018, https://arxiv.org/abs/1802.06222.

[9]. S. Akcay, A. A. Abarghouei and T. P. Breckon, GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training, in *Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018),* Perth, Australia, 2019, pp. 622-637.

[10]. D. Li, D. Chen, B. Jin, L. Shi, J. Goh and S. Ng, MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks, *Artificial Neural Networks and Machine Learning -- ICANN 2019: Text and Time Series,* 2019, pp. 703-716.

[11]. J. Audibert, P. Michiardi, F. Guyard, S. Marti and M. A. Zuluaga, USAD: UnSupervised Anomaly Detection on Multivariate Time Series, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 2020.

[12]. A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante and K. Veeramachaneni, TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks, in *Proceedings of the IEEE International Conference on Big Data (Big Data' 2020),* 2020, pp. 33-43.

[13]. M. Hoh, A. Schöttl, H. Schaub and F. Wenninger, A Generative Model for Anomaly Detection in Time Series Data, *Procedia Computer Science*, Vol. 200, 2022, pp. 629-637.

[14]. P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford and I. Sutskever: Jukebox: A Generative Model for Music, *CoRR abs/2005.00341*, 2020.

[15]. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, Improved Training of Wasserstein GANs, *CoRR abs/1704.00028*, 2017, pp. 5769-5779.

[16]. K. Hundman, V. Constantinou, C. Laporte, I. Colwell and T. Söderström, Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding, *CoRR abs/1802.04431*, 2018.

[17]. S. Schmidl, P. Wenig and T. Papenbrock, Anomaly Detection in Time Series: A Comprehesive Evaluation, *Proceedings of the VLDB Endownment,* Vol. 15, 2022, pp. 1779-1797.

(8006)

Industry 4.0 and the Covid-19 Pandemic: Literature Review

M. Guennoun and F. Bennouna
National School of Applied Sciences, Engineering, Systems and Applications Laboratory, Fez, Morocco
E-mail: guennounmephtaha@gmail.com, bennouna.ensa@gmail.com

Abstract: The Covid-19 containment period has plunged the industrial sector into a serious crisis, as many factories have indeed been closed for several weeks to protect the health of employees.

At present and in the era of globalization, the majority of countries are undergoing a rapid technological change that has forced them to make new reflections to keep up with this pace. Given that the industrial sector is among the basic pillars of each country, it is necessary to find a way to adapt and accelerate the process of digitalization (Industry 4.0). All this in order to continue production in the face of a new Pandemic in the future. Industry 4.0 plays a very important role in this context.

In this article, we will make a detailed study on the evolution of Industry 4.0, its advantages and disadvantages and the technologies associated with this industry. Then, we will list the effects of Covid-19 on the Moroccan industrial sector. Finally we will show the use of industry 4.0 during the covid-19 crisis.

Keywords: Covid-19, Industry 4.0, Industrial revolution, Digitalization, Impact.

1. Introduction

In todays globally interconnected world, increasingly exposed to digitalization, national and international competition between companies is increasing on the one hand, but also the diversity and complexity of customized production on the other. This digitalization is illustrated in the integration of the fourth industrial revolution "Industry 4.0" both in terms of information management and production. This industry is based on the application of new advanced technologies such as artificial intelligence, the Internet of Things, additive manufacturing, etc.

At the beginning of 2020, the whole world experienced a Covid-19 pandemic that slowed down the global economy. Some companies found themselves having to shut down or scale back their operations. While other companies tried to survive, resorting to massive investments in other alternatives. Moroccan companies, in turn, have implemented these solutions to confront and resist the pandemic.

2. Evolution of the Industry

The industry has undergone a major revolution, it has gone from the first production system that is based on agriculture essentially (Industry 1.0) to an intelligent industrial system (Industry 4.0).

2.1. Industry 1.0 "Mechanization"

Between 1680 and 1720, there was an agricultural and demographic revolution which imposed an industrial revolution. Indeed, in 1765 James WATT exploited the coal and invented the steam engine in order to radically transform the general organization of the production. This production is essentially agricultural. This increased the production capacity, and forced the market to accommodate a production with different outlets [1].

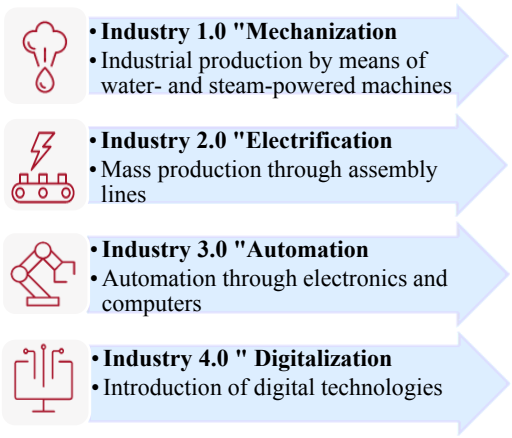


Fig. 1. The evolution of the industry over time

This revolution allows the creation of factories that use the steam engine to drive their devices, which increases the rate and leads to a more important manufacture. But, with the appearance of these factories the society begins to be interested in environmental problems [2].

2.2. Industry 2.0 "Electrification"

The second industrial revolution began at the beginning of the 20th century with the discovery of Electricity, motors and the invention of assembly line work. The primary goal of this revolution was to introduce mass production and bring innovations in chemistry and related fields [3].

Henry Ford (1863-1947) came up with the idea of mass production in a Chicago slaughterhouse and invented the assembly line for mass production of vehicles, which made the designed machines more mobile. Thanks to technology, vehicles, means of transportation and some devices embedded in machines or even machines consume oil, so paying attention to the environment becomes an obstacle.

2.3. Industry 3.0 "Automation"

It was not until the second half of the 20th century that the computer appeared. Since 1970, electronic devices have been invented and used in industrial environments. Vacuum tubes were replaced by transistors and then immigrated electronic circuits in integrated circuit chips [1].

At that time, production was further advanced by the automation of mechanical machines, computers, programmable logic controllers and telecommunications.

Since the introduction of these technologies, it was now possible to record more and more data but with a time lag or after the fact.

Faced with this revolution, the problem of pollution caused by these machines and the non-treatment of industrial waste became a very big undesirable effect. [4].

2.4. Industry 4.0 "Digitalization"

Industry 4.0 is the last step of the industry development process. The term Industry 4.0 was first introduced at the Hannover Fair event in Germany in 2011 [5].

This latest revolution is also called the Industrial Internet of Things or the factory of the future. This revolution aims to implement smart factories that are able to produce productions using greener and more efficient processes.

Industry 4.0 is characterized by highly developed automation and digitalization processes and the use of electronics and information technology (IT) in manufacturing and services [6].

One of the main goals of Industry 4.0 is to connect the physical world and the virtual, combining assembly-line and custom manufacturing.

The benefits of this technology include increased productivity, reduced errors and rework, and the execution of high-risk tasks [1].

3. Advantages and Disadvantages of Industry 4.0

The benefits of Industry 4.0 can be classified into three key dimensions [1]:

- Technical dimensions: increase in productivity and operational efficiency, ease of interconnection and information of traditional industry, communication between machines and humans via IoT, analysis and processing
- Logical dimensions: industry 4.0 can reduce logistics costs, optimize lead times, deliver orders quickly
- Ecological and human dimensions: this industry can reduce production waste, energy consumption, errors, worker complaints and it can facilitate the execution of high risk tasks.

After analyzing the following articles: [7-10] we can synthesize a series of barriers to: High cost; Limited enterprise structure; Hacking by cyber attacks; Mindset of the staff and adapting to new changes and Lack of skilled manpower.

4. Technologies Associated with Industry 4.0

The technologies associated with Industry 4.0 are diverse and the list grows over time. According to PFEIFFER [11] and KAHMANN [12], four dimensions of technologies exist in industrial applications. However their implications differ in terms of human-machine interaction, qualification, work organization and personnel representation:

- Mobile and internet-based communication;
- Intensified networking of production;
- New robotic and production technologies;
- Wearable computing objects.

All these applications are based on nine technological pillars (Cloud and Cybersecurity, Blockchain, Augmented reality, Additive manufacturing, Advanced simulation tool, Vertical/Horizontal Integration, Big Data / Artificial Intelligence, IOT, Collaborative robots and smart machines). These innovations bridge the physical and digital worlds and make intelligent and autonomous systems possible.

5. Impact of the Covid-19 Crisis on Industrial Moroccan Companies

Morocco, like all the countries in the world, could not escape the Covid-19 crisis and suffered very heavy economic and health consequences. Faced with this situation, the Moroccan government has made great efforts to ensure that the impact of this crisis is not too severe. Anticipated and monitored the direct and indirect impact of the Covid-19 pandemic on national economies and identified support for the most affected sectors [13].

According to the survey conducted by the High Commission on the effects of covid-19 at the end of April 2020, among 4,000 Moroccan companies, 57 % had stopped their activity temporarily or permanently. Of the surveyed companies that were temporarily or permanently out of business, 72% were very small companies, 26 % were small and medium-sized companies and 2 % were large companies [14]. These are spread over four main sectors (Service, Industry, Construction and Trade).

6. The use of Industry 4.0 in the Management of the Covid-19 Crisis

The Covid-19 pandemic has changed the way businesses around the world operate and forced them to think about innovative use of technology to survive and combat its challenges. Personal accessibility is limited due to lockdown and social distancing, but the pillars of Industry 4.0 were a support to overcome these problems. Because of lockdown, most employees have been under the compulsion to work

remotely in a collaborative manner which makes telecommuting the right solution for companies to continue much of their remote production methods such as:

- Instant remote communication between teams;
- Remote meetings;
- Remote management of computers and machines;
- Organization and management of projects;
- Document management with access at any time (PLM).

Indeed, Industry 4.0 through its technologies has been able to ensure this continuity. The connectivity between men and machines and connected objects based on the Internet of Things (IoT) have improved the automation of production sites and greatly reduce human intervention. A Gartner survey found that 47% of businesses plan to increase their investment in IoT due to restrictions, and a McKinsey report found that IoT could increase productivity by 10% to 30%, depending on the industry. 3D printing or additive manufacturing has also allowed companies to create complex parts in a few hours, which can be essential in the production chain, instead of weeks using traditional methods (the production of respiratory aid devices for example).

Furthermore and because of their critical positions, other employees have been obliged to work locally, which requires them to follow the precautions of the World Health Organization to prevent the spread of covid-19.

- Wear a mask;
- Ventilate for 10 minutes and wash hands every hour;
- Observe physical distancing;
- Test for covid-19 if necessary.

According to studies, Industry 4.0 has been found to have a positive impact on the application of covid-19 precautions. For example, the IoT is an effective tool that manages personnel not wearing the face mask or not respecting the necessary physical distance. The latter has also been controlled by computer vision and the use of sensors. The machine learning has in turn played a role in the need to test. [15].

7. Conclusion

Before Pandemic Covid-19, the goal of implementing Industry 4.0 within companies focused on competitiveness, increased productivity, cost reduction, sustainability and innovation. The main goal, then, was to improve the operation of well-functioning companies. Currently, many manufacturers are focused primarily on survival and reducing the damage caused by this pandemic. They have based their focus on applying one or a fusion of the nine technology pillars of Industry 4.0.

The situation raises two main questions:

- Is Industry 4.0 still relevant?
- If Industry 4.0 is relevant, what role has it played in order to deal with this pandemic?

References

[1]. A. Samadi & H. Achelhi, Industry 4.0 in The Economic Activity Zones in Morocco: Tangier-Tetouan-Alhoceima Region Case, *International Journal of Accounting, Finance, Auditing, Management & Economics,* 2, 6-1, 2021, pp. 327-338..

[2]. W. Shin, Y. Lee, J. J. Seok, Dahlgaard, A pattern-based decision framework in the era of Industry 4.0, *Total Quality Management and Business Excellence*, 30, 3, 2019, pp. 1-24.

[3]. M. Javaid, A. Haleem, Industry 4.0 applications in medical field: A brief review, *Current Medecine Research and Practice*, 2019, pp. 102-109.

[4]. K. Tantawi, A. Sokolov, O. Tantawi, Advances in Industrial Robotics: From Industry 3.0 Automation to Industry 4.0 Collaboration, in *Proceedings of the 4th Technology Innovation Management and Engineering Science International Conference TIMES-iCON 2019*, 2019.

[5]. A. Grieco, P. Caricato, D. Gianfreda, M. Pesce, V. Rigon, L. Tregnaghi, & A. Voglino, An Industry 4.0 case study in fashion manufacturing, *Procedia Manufacturing,* 11, 2017, pp. 871-877.

[6]. Y. Lu, Industry 4.0: A survey on technologies, applications and open research issues, *Journal of Industrial Information Integration,* 2017, pp. 1–10.

[7]. A. Raj, G. Dwivedi, A. Sharma, A. Jabbour, & S. Rajak, Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective, *International Journal of Production Economics,* 2020, 224, 107546.

[8]. V. Alcacer, & V. Cruz-Machado, Scanning the industry 4.0: A literature review on technologies for manufacturing systems, *Engineering Science and Technology,* 2018, pp. 899-919.

[9]. D. Kiel, C. Arnold, & K. Voigt, The infuence of the Industrial Internet of Things on business models of established manufacturing companies—A business level perspective, *Technovation*, 68, 2017, pp 4–19.

[10]. M. Mamad, Challenges and Benefits of Industry 4.0: An overview, *International Journal of Supply and Operations Management (IJSOM)*. Vol. 5, Issue 3 - Serial Number 3, 2018, pp. 256-265.

[11]. S. Pfeiffer, Technisierung von Arbeit, in Böhle F., Voß G.G., Wachtler G. (eds.), Handbuch Arbeitssoziologie, 2. ed., Wiesbaden, *Springer*, 2018, pp. 321-357.

[12]. M. Kahmann. Allemagne. L'Industrie 4.0 : vers la digitalisation concertée de l'industrie manufacturière ?, *Chronique Internationale de l'IRES,* Numéro 173, 2021, pp. 33-40.

[13]. S. El Amine, N. Bouayad Amine, Moroccan companies in the area of Covid19, Impacts, measures and recovery policies of tourism, automotive and real estate sectors, *Revue Internationale des Sciences de Gestion,* Vol. 3, Numéro 4, 2020, pp. 506-519.

[14]. S. Labrar, A. Maarouf, L. Zouiri, Moroccan industry in the covid-19 era : the case of the automotive sector. Congrès du réseau international de recherche sur les organisations et le développement durable: La "société entrepreneuriale" à l'épreuve de la crise sanitaire, 2021, ffhal-0340738.

[15]. R. Agrawal, M. Mittal, L. Mohan Goyal, Sustainability Measures for COVID-19 Pandemic, *Springer*, 2021, pp. 237.

(8085)

The Industrial Digital Energy Twin as a Tool for Comprehensive Optimization of Industrial Processes

Alejandro Rubio [1], Fernando Mengod [1], Andrés Lluna and Vicente Fuster-Roig [2]
[1] Instituto Tecnológico de la Energía (ITE), Avda. Juan de la Cierva, 24, 46980, Valencia, España
[2] Instituto de Tecnología Eléctrica, Universitat Politècnica de València,
Camino de Vera s/n Edificio 6C, 46022, Valencia, España
Tel[1].: +34961366670, fax: +34961366680
E-mail: alejandro.rubio@ite.es

Summary: Industrial manufacturing processes have evolved and been improved all since the disruption of the Industry 4.0 paradigm, but energy has become an strategic resource needed to maintain industrial competitiveness while maximizing quality and minimizing environmental impacts. In this context of global changes mainly linked to social an economic impact in the short terms and an unprecedented climate crisis, Energy Digital Twins provide companies with a tool to deal with this complex situation. Combining a multi-paradigm approach to industrial process modeling, project GENERTWIN proposes the development of a whole Digital System that empowers the participating companies to develop an energy Digital Twin of their processes. This System also includes a production model of the target process, allowing users to predict energy consumption and costs and, at the same time, evaluating the behavior of the process under certain productive changes in order to maximize consumption optimization, production efficiency and process flexibility.

Keywords: Digital twin, Smart manufacturing, Energy, Energy efficiency, Productive flexibility.

1. Introduction

In the context of evolution from "climate change" to "climate crisis", the industrial sector must face a series of changes and evolve to new, more efficient productive models, under the umbrella of the revolution of the industry 4.0 paradigm and digitalization. Many different tools are sufficiently developed and tested to be applied with success and help maximize productivity and minimize energy consumption and costs, concepts related most of the time.

Digital Twins (DT) are one of the main tools developed in this context. Appeared in the early 2000s [1], the concept represents the idea of developing a digital copy of a physical object. This copy, technically known as a twin, allows implementing process improvements and optimize the way in which it is operated in a fully controlled and risk-free environment.

1.1. The Real Potential of Energy Digital Twins

The optimization of energy consumption in industry has traditionally been based on two trends; 1) The introduction of new process technologies to improve efficiency, and 2) The introduction of specific improvements in plant operations [2]. However, this second approach has never been sufficient to understand the causality inherent in the operation of manufacturing processes in sufficient detail to propose improvements that dynamically adapt to the process context and changes. Some energy efficiency analysis techniques, such as performance measurement and verification protocols, come close to the level of detail needed to achieve such intuition. These kinds of tools, complemented by internationally recognized implementation support protocols such as ISO50001 (Energy Management Systems) or energy audits (in Spain, according to Royal Decree 56/2016 following Energy Efficiency Directives), provide a development framework, which, however, cannot be adapted to the particularities of each improvement case [3]. In consequence, the state of the art of energy efficiency analysis of industrial processes is usually quite far from the implementation of more advanced techniques than those described, seeking in any case high-impact technological changes, not so much an optimization methodology based on dynamically updated measures.

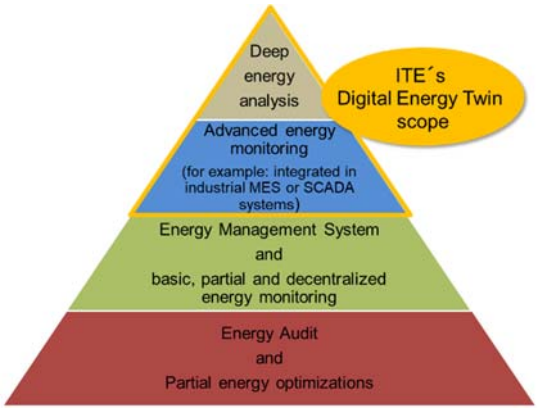


Fig. 1. Classical energy tools and Digital Energy Twin. Source: ITE

The Digital Energy Twin (DET) of an industrial process provides the company with the ability to predict the behavior of its process in different

scenarios. These scenarios can be oriented to analyze different dimensions of the process, such as predictive maintenance, performance monitoring and control, better design of the layout and resources planning, or even facilitate renewable energy integration, to mention a few examples [4]. The DET can, in fact, be designed bearing one or more of these functionalities in mind, and it may even evolve to include new ones at a later stage.

The digital twin as a tool for linking production with energy in industrial plants is currently not sufficiently developed. Some representative references propose an analysis by means of nonlinear planning models [5]. Others integrate both energy and production data through cyber-physical systems, focused on sustainable production [6], but basically framed within the traditional concept of an energy monitoring system. In fact, most DET approaches employ discrete modeling of the process or production chain to introduce discretized energy consumption concepts [7, 8]. However, these approaches lack a dynamic design intrinsic to the operation of machinery subject to possible internal and external changes. They are models focused on plant and supply chain logistics and optimization of material resources in most cases. However, the energy resource presents a characteristic dynamic behavior that this type of models does not contemplate. The most advanced models aligned with the concept of DET are supposed to be designed to allow studying the final and disaggregated impact of the variation of specific operational control parameters while improving technical service and maintenance. All of which should be performed in a digital environment, controlled, connectable and with a much lower risk.

This paper is mainly oriented to describe the concept of an Energy Digital Twin whose main functionalities are oriented to simulate different productive scenarios where situations such as energy price variation or parameters optimization could be studied. Then proposed concept is broader than that of a traditional digital twin, since it also focuses on studying its integration in the plant and some additional end-user functionalities. Therefore, the result to be achieved is called Digital System (DS).

1.2. Project GENERTWIN's Description and Objectives

The objective of the GENERTWIN "*Sistema Digital de Análisis de procesos industriales para GENERación de escenarios alternativos bajo consideraciones productivas y de eficiencia energética*" is the development and implementation of a Digital System applicable to different industrial sectors considering the optimization needs and energy dependence of these processes. The system is based on the operation of a Digital Twin (DT) developed following the identified needs after a deep situation analysis, including the identification of the use cases and the functionalities that this twin must provide to the company.

The DT is developed to a low level of detail, and considers the internal dynamics of the machinery involved in the process, its operation, the relationship with the necessary resources and the impact of the different possible variations on the quality and type of products obtained. By considering it as a DT, the model consequently scales in complexity and capabilities, allowing not only to analyze the operation and interactions of the aforementioned elements that are part of the process, but also to add value to the company's decision making process. The DS in which it is located represents the architecture of elements and technologies necessary to maximize the representativeness and guarantee the implementation of the model in the final process.

1.3 Constraints and Challenges

The development and implementation of a DS incorporating a DT such as the one proposed is not without its difficulties. The first of them is the identification of the real expectations and needs of the companies to which the development of the tool is focused. In many occasions, a bad design of the DT can lead to an excess of work in the development that entails a higher final cost. On the other hand, an insufficient development of the model can result in the DT losing representativeness of its results. As a consequence, an optimal balance between both extremes must be found in order to achieve a balanced, accessible system that meets the needs of the company. In order to mitigate as much as possible the potential design problem of the digital system, a previous design phase has been developed in which the company using the process participates by receiving the appropriate feedback. In addition, a multisectoral analysis of requirements and possible interests of other follower companies is carried out in the project, in order to maximize the representativeness and orient the solution to the most effective development possible.
Another major challenge to solve when developing a model is data accessibility. It is common to find a lack of information, even among process specialists, about the details of its operation and internal mechanisms. In addition, the lack of digital monitoring systems (collecting energy and production data) is also a handicap that hinders development.

2. Digital System Development

2.1. Basic Architecture

The Digital System is structured in different layers. The main ones refer to the Energy productive Model, which represents the simulation core taking into account dynamic models of the processes together with the simulation of discrete events to reproduce the manufacturing activity. The Digital Twin environment, which gathers the interaction and analysis functionalities in a digital environment, is developed using a specialized hybrid and multi-

paradigm simulation software, including multi-agent modeling, which also allows integrating and testing the models and interfaces. And finally, the upper layer of the system itself, which is designed to be integrated with other types of plant systems and with specific interconnectivity capabilities, to connect the simulation models, as analytical advanced digital tool, with Internet of Thing layers and other automation architectures [9].

A future interconnectivity challenge to be met is to link the system to the plant via key automation industry systems like Manufacturing Execution System (MES) or Supervisory Control and Data Acquisition (SCADA) systems, as appropriate. The connectivity of the application with direct Machine to Machine (M2M) field elements will also be addressed by analyzing the options of developing customized communication drivers such as Modbus TCP or making use of OPC UA.

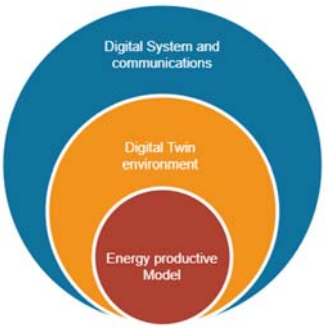


Fig. 2. Division of the DS in basic high-level layers. Source: ITE.

2.3. Deployment Process: Models and DT Interfaces

Summarizing all of the above, the system follows an implementation process based on stages, in which clear objectives are defined with the end user. It is important to agree on the functionalities that the system must achieve before launching the development, and to continue working on these functionalities during the development of the model.

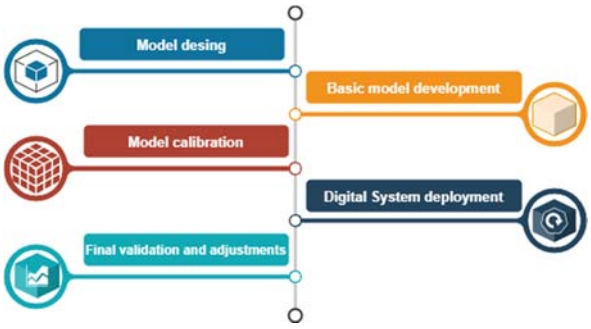


Fig. 3. Necessary steps to accomplish in GENERTWIN. Source: ITE.

It is important to emphasize that the above methodology developed in the project aims to structure a highly iterative process such as a development that includes a Digital Twin. For this reason, the model must be constantly re-examined and re-evaluated to ensure that it meets the required representativeness, allowing it to be sufficiently flexible to add necessary new elements or functionalities. However, the project seeks a balance between sufficient representativeness in terms of energy performance and the feasibility of a time-bound development.

2.4. Working Functionalities and Estimated Impacts

The DS is designed with the clear functionality of predicting process behavior in the face of substantial production changes, especially focused on the impact on energy consumption of both productive and non-productive variables. In this sense, the Digital Twin is in charge of simulating the different states of the system, displaying predictions, and integrating them into the data environment in which the process is developed. This allows optimization adjustments to be made and even optimal production sequencing algorithms to be executed according to the energy cost. The following table shows the potential for improving the economic cost of energy by shifting the batches according to two scenarios of different industrial flexibility.

Table 1. Impact of scheduling optimization based on previous results. Source: ITE.

Application case	Energy cost improvement (%)	Considerations
Ceramic industry process	4	Low flexibility
Glass industry process	37	High flexibility

Another possible use of the DS involves modifying the design or the use of the resources of the process. This approach may appear to be similar to precedent works in point 1.1, but the main strength of the project is that all these approaches are included in a single multi-paradigm development that combines discrete and continuous dynamics. In this way, the model can be easily adapted to include new functionalities once the base knowledge of the process is secured.

4. Conclusions

The Digital System proposed in the GENERTWIN project aims to simulate the energetic behavior of the process under different productive and contextual changes. For this purpose, a multi-paradigmatic model is developed that encompasses the complexity of the production environment and the impact of the energy consumption of the process to offer advanced analysis

possibilities such as: scenario generation, energy cost optimization or energy impact assessment of production actions. The proposed Digital System model of energy – production advanced analysis is implemented in two demonstration experimental pilots: a) ceramic tiles furnace process and b) break friction materials manufacturing process in automotive industry.

Acknowledgements

GENERTWIN "*Sistema Digital de Análisis de procesos industriales para GENERación de escenarios alternativos bajo consideraciones productivas y de eficiencia energética*", IMDEEA/2022/16, has been co-financed by IVACE (Instituto Valenciano de Competitividad Empresarial) and FEDER funds (Fondo Europeo de Desarrollo Regional). Project in which the Instituto Tecnológico de la Energía (ITE) and the Instituto Tecnológico de la Cerámica (ITC) participate.

References

[1]. Grieves, M., Digital twin: manufacturing excellence through virtual factory replication, White paper, 2014, p. 1-7.

[2]. May, G., Stahl, B., Taisch, M., & Kiritsis, D., Energy management in manufacturing: From literature review to a conceptual framework, *Journal of Cleaner Production*, 167, 2017, pp. 1464-1489.

[3]. Dörr, M., Wahren, S., & Bauernhansl, T., Methodology for energy efficiency on process level, *Procedia CIRP*, 7, 2013, pp. 652-657.

[4]. Yu, W., Patros, P., Young, B., Klinac, E., & Walmsley, T. G., Energy digital twin technology for industrial energy management: Classification, challenges and future, *Renewable and Sustainable Energy Reviews*, 161, 2022, 112407.

[5]. Zhang, Z., Tang, R., Peng, T., Tao, L., & Jia, S., A method for minimizing the energy consumption of machining system: integration of process planning and scheduling, *Journal of cleaner production*, 137, 2016, pp. 1647-1662.

[6]. Ma, S., Zhang, Y., Lv, J., Yang, H., & Wu, J., Energy-cyber-physical system enabled management for energy-intensive manufacturing industries, *Journal of Cleaner Production,* 226, 2019, pp. 892-903.

[7]. Keshari, A., Sonsale, A. N., Sharma, B. K., & Pohekar, S. D., Discrete event simulation approach for energy efficient resource management in paper & pulp industry, *Procedia CIRP*, 78, 2018, pp. 2-7.

[8]. Ma, S., Zhang, Y., Lv, J., Yang, H., & Wu, J., Energy-cyber-physical system enabled management for energy-intensive manufacturing industries, *Journal of Cleaner Production,* 226, 2019, pp. 892-903.

[9]. J. Holler and others, From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence, *Elsevier Science*, 2014.

(8279)

Advanced Control of a Distributed Parameter System

Xueru FAN [1], Cheng-Zhong XU [2] and Chunhai KOU [3]
[1] Donghua University, College of Information Science and Technology, 201620 Shanghai, PR China
[2] University of Lyon, Université Claude Bernard Lyon 1, Department of Mechanics, LAGEPP, Bâtiment CPE, 43 Boulevard du 11 Novembre 1918, 69100, Villeurbanne Cedex, France
[3] Department of Mathematics, Donghua University, 201620, Shanghai, PR China
Tel.: 04 72 43 18 90
E-mail: cheng-zhong.xu@univ-lyon1.fr

Summary: This paper deals with stabilization of an infinite-dimensional beam system faced with boundary uncertainty by using the active disturbance rejection control approach (ADRC). The ADRC method, proposed by Han [1] using finite dimensional extended state observer and extended by Feng [2] to infinite-dimensional disturbance estimators, is used here to estimate the total disturbance for constructing a dynamic output feedback control law. In fact, a novel infinite-dimensional estimator, constituted of two auxiliary systems and completely determined by the output measurements, is designed to estimate online the total disturbance (including both external disturbance and internal model uncertainty). The one auxiliary system is to separate disturbance effect from control influence, while the other is to estimate the total disturbance. Based on the estimated total disturbance, a boundary dynamic output feedback control law is proposed such that (i) the controlled beam system is exponentially stabilized to the null state and (ii) all the internal signals in the closed-loop system are maintained to be bounded. The performance of the designed control law is validated by simulations of a numerical example.

Keywords: Euler-Bernoulli Beam, Exponential stabilization, Observer, Disturbance rejection, Dynamic output feedback.

1. Introduction

In engineering practice, the modeling uncertainties and external disturbances usually appear, which may result in some negative effects such as the increase of control error and mechanical failures.

Hence, it is of great significance to study controlling systems with uncertainties from both practical and theoretical perspectives such as shown in [3-6].

In this paper, we consider an Euler-Bernoulli beam equation with boundary uncertainties as follows:

$$\begin{cases} u_{xxxx}(x,t)+u_{tt}(x,t)=0, x\in(0,1), t>0, \\ u_{xx}(0,t)=u(1,t)=u_{xxx}(0,t)=0, \\ u_{xx}(1,t)=U(t)+f(u(\cdot,t),u_t(\cdot,t))+d(t), \\ u(x,0)=u_0(x), u_t(x,0)=u_1(x), \\ y_{out}=\{u_t(0,t),u_x(1,t),u_{xt}(1,t)\}, \end{cases} \quad (1)$$

where $U(t)$ is the bending moment input, y_{out} is the measured output, (u_0,u_1) is the initial condition and $d(t)$ is an unknown external disturbance satisfying $d\in L^{\infty}[0,\infty)$. Set

$$H_L^2(0,1)=\{\varphi\in H^2(0,1)\mid \varphi(1)=0\}.$$

Consider $f: H_L^2(0,1)\times L^2(0,1)\to R$ is an unknown interior nonlinear model uncertainty. We set $F(t)=d(t)+f(u(\cdot,t),u_t(\cdot,t))$ and this signal is noted as the total disturbance. Our goal is to design a boundary bending moment control $U(t)$ constituted of dynamic feedback of the output y_{out} such that the partial state (u,u_t) of the closed-loop system (1) is exponentially stabilized in Hilbert space $X=H_L^2(0,1)\times L^2(0,1)$ with the inner product induced norm

$$\|(\varphi,\psi)\|_X^2=\|\varphi''\|_{L^2(0,1)}^2+\|\psi\|_{L^2(0,1)}^2+\gamma_1|\varphi'(1)|^2,\ \gamma_1>0,$$

where $\forall\ (\varphi,\psi)\in X$.

For the purpose of suppressing the vibration and attenuating the uncertainty, the ADRC control law proposed by [1] and extended by [2] is used to estimate the total disturbance and to construct the dynamic output feedback controller.

2. Main Results

In this section, we consider the stabilization of the Euler-Bernoulli beam system (1). As well-known the static feedback controller

$$U(t)=-\gamma_1 y_{out,2}(t)-\gamma_2 y_{out,3}(t),\gamma_1,\gamma_2>0, \quad (2)$$

exponentially stabilizes the system (1) to $(0,0)$ [8], if $F\equiv 0$. However, the stabilizer (2) is not robust to the

uncertainties. In fact, if $f \equiv 0$, $d(t) = d$ is a constant, then each solution of the closed-loop system (1) exponentially converges to the state

$$(u(x,t),u_t(x,t)) = \left(d(x-1)/\gamma_1, 0\right).$$

So, the equilibrium state $(0,0)$ is no longer asymptotically stable. Thus, the stabilizer has to be re-designed. To this end, we introduce output feedback controller as follows:

$$U(t) = -\gamma_1 y_{out,2}(t) - \gamma_2 y_{out,3}(t) + U_0(t), \gamma_1, \gamma_2 > 0,$$

where $U_0(t)$ will be designed as a new control input for rejecting the total disturbance.
For estimating the total disturbance, we design the following disturbance estimator:

$$\begin{cases} \alpha_{xxxx}(x,t) + \alpha_{tt}(x,t) = 0, \\ \alpha_{xx}(0,t) = \alpha(1,t) = \alpha_{xxx}(0,t) = 0, \\ \alpha_{xx}(1,t) = U(t) - \gamma_1(\alpha_x(1,t) - y_{out,2}(t)) \\ \qquad\qquad - \gamma_2(\alpha_{xt}(1,t) - y_{out,3}(t)), \\ \beta_{xxxx}(x,t) + \beta_{tt}(x,t) = 0, \\ \beta_{xx}(0,t) = \beta(1,t) = 0, \\ \beta_{xxx}(0,t) = -\gamma_3\left(\beta_t(0,t) + \alpha_t(0,t) - y_{out,1}(t)\right), \\ \beta_x(1,t) = u_x(1,t) - \alpha_x(1,t), \\ \alpha(x,0) = \alpha_0(x), \alpha_t(x,0) = \alpha_1(x), \\ \beta(x,0) = \beta_0(x), \beta_t(x,0) = \beta_1(x), \end{cases} \tag{3}$$

where γ_3 is a positive parameter. It should be noted that the input and output of the original plant (1) completely determine this estimator (3). Set

$$\hat{\alpha}(x,t) = u(x,t) - \alpha(x,t),$$
$$\hat{\beta}(x,t) = \beta(x,t) - \hat{\alpha}(x,t),$$

then the errors system is governed by

$$\begin{cases} \hat{\alpha}_{xxxx}(x,t) + \hat{\alpha}_{tt}(x,t) = 0, \\ \hat{\alpha}_{xx}(0,t) = \hat{\alpha}(1,t) = \hat{\alpha}_{xxx}(0,t) = 0, \\ \hat{\alpha}_{xx}(1,t) = -\gamma_1\hat{\alpha}_x(1,t) - \gamma_2\hat{\alpha}_{xt}(1,t) + F(t), \\ \hat{\beta}_{xxxx}(x,t) + \hat{\beta}_{tt}(x,t) = 0, \\ \hat{\beta}_{xx}(0,t) = \hat{\beta}(1,t) = \hat{\beta}_x(1,t) = 0, \\ \hat{\beta}_{xxx}(0,t) = -\gamma_3\hat{\beta}_t(0,t). \end{cases} \tag{4}$$

Lemma 1 Suppose that $f: X \to R$ is continuous, bounded and satisfies global Lipschitz condition, and $d \in L^\infty[0,\infty)$ or $L^2[0,\infty)$. For any initial state $\left(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0, \hat{\beta}_1\right)^{\mathrm{T}} \in X^2$, system (4) admits a unique continuous solution satisfying

$$\left(\hat{\alpha}(\cdot,t), \hat{\alpha}_t(\cdot,t), \hat{\beta}(\cdot,t), \hat{\beta}_t(\cdot,t)\right)^{\mathrm{T}} \in C(0,\infty; X^2)$$

And

$$\left\|\left(\hat{\beta}(\cdot,t), \hat{\beta}_t(\cdot,t)\right)^{\mathrm{T}}\right\|_X \le Me^{-\theta t}, \forall t \ge 0,$$

for some $\theta, M > 0$. If $d \in L^2[0,\infty)$, then

$$\lim_{t\to\infty}\left\|\left(\hat{\alpha}(\cdot,t), \hat{\alpha}_t(\cdot,t)\right)^{\mathrm{T}}\right\|_X = 0,$$

and if $d(t) \equiv 0$, we have

$$\left\|\left(\hat{\alpha}(\cdot,t), \hat{\alpha}_t(\cdot,t)\right)^{\mathrm{T}}\right\|_X \le Me^{-\theta t}, \forall t \ge 0,$$

Furthermore, there exists a positive constant ζ such that $e^{\zeta\cdot}\hat{\beta}_{xx}(1,\cdot) \in L^2[0,\infty)$.

Since

$$\hat{\beta}_{xx}(1,t) = \beta_{xx}(1,t) + \gamma_1\beta_x(1,t) + \gamma_2\hat{\alpha}_{xt}(1,t) + F(t),$$

and Lemma 1, we have

$$F(t) \approx -(\gamma_1\beta_x(1,t) + \beta_{xx}(1,t) + \gamma_2\beta_{xt}(1,t)), \tag{5}$$

which implies that the estimation of the total disturbance is obtained if the initial condition in (4) is smooth.
After obtaining the estimated total disturbance (5), we now design the control input

$$U(t) = -\gamma_1 u_x(1,t) - \gamma_2 u_{xt}(1,t) + \gamma_1\beta_x(1,t) + \beta_{xx}(1,t) + \gamma_2\beta_{xt}(1,t).$$

Combined with (1) and (3), the closed-loop system with state $\left(u, u_t, \alpha, \alpha_t, \beta, \beta_t\right)$ is studied in the Hilbert space X^3.

The main result is stated as follows.

Theorem 1 Suppose that $\gamma_1, \gamma_2, \gamma_3 > 0$, $f: X \to R$ is continuous, bounded and satisfies global Lipschitz condition and that $d(t) \in L^\infty[0,\infty)$ or $d(t) \in L^2[0,\infty)$

. For any initial state $(u_0,u_1,\alpha_0,\alpha_1,\beta_0,\beta_1)\in X^3$ with the compatibility condition

$$\beta_x(1,0)=u_x(1,0)-\alpha_x(1,0),$$

there exists a unique solution satisfying

$$(u,u_t,\alpha,\alpha_t,\beta,\beta_t)^{\mathrm{T}}\in C(0,\infty;X^3)$$

and

$$\left\|\left(u,u_t\right)^{\mathrm{T}}\right\|_X\le Me^{-\theta t},\forall t\ge 0,$$

with some $M,\theta>0$ depending on the initial condition. If $d(t)\in L^2[0,\infty)$, we have

$$\lim_{t\to\infty}\left\|\left(\alpha,\alpha_t,\beta,\beta_t\right)^{\mathrm{T}}\right\|_{X^2}=0.$$

Further, when $d(t)\equiv 0$, exist two constants $M',\theta'>0$ depending only on the initial condition such that

$$\left\|\left(\alpha,\alpha_t,\beta,\beta_t\right)^{\mathrm{T}}\right\|_{X^2}\le M'e^{-\theta' t},t\ge 0.$$

The Theorem 1 can be proved by using Riesz basis approach referring to [7] and [8] and admissible control theory referred to [9]. To save space, the proof is omitted here.

3. Numerical Simulation

To verify the effectiveness of the designed control schemes, we present the closed-loop system by the finite difference method.

We take disturbance and nonlinear uncertainty

$$d(t)=2\sin(t),\ f(u(\cdot,t),u_t(\cdot,t))=\sin(u(0,t)),$$

and the initial states

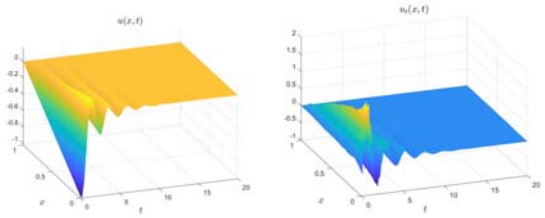
$$u_0(x)=x-1,\beta_0(x)=x(x-1),$$
$$u_1(x)=\alpha_0(x)=\alpha_1(x)=\beta_1(x)=0,\forall x\in(0,1)$$

.

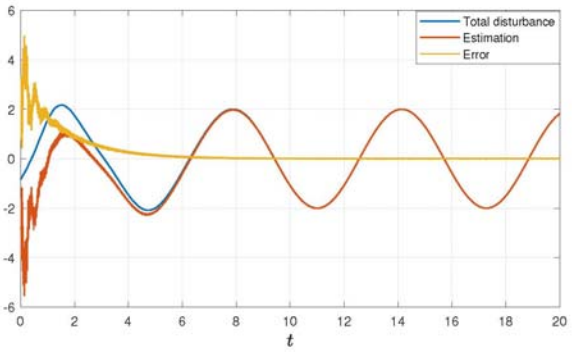
The parameters are chosen as $\gamma_1=\gamma_3=5,\ \gamma_2=2$, $dx=0.05$ and $dt=0.0005$. The state (u,u_t) of the closed-loop system is shown in Fig. 1(a) and Fig. 1(b). Fig. 1(c) shows that the estimated error tends to zero. It is seen that the state (u,u_t) is convergent as $t\to\infty$ and the disturbance is estimated and rejected effectively.

4. Conclusion

The problem of the stabilization of beam system with boundary uncertainties has been resolved by using ADRC control law. Under the assumption that the disturbance satisfies $d\in L^\infty(0,\infty)$, the novel infinite-dimensional disturbance estimator without high gains has been proposed, which improves some previous results such as those in [4, 10, 11]. With our designed controller, the beam state is exponentially stabilized to (0,0) faced with both external disturbance and model uncertainty, and all internal signals of the closed-loop system are maintained to be bounded.



(a) The position $u(x,t)$; (b) The velocity $u_t(x,t)$;



(c) The tracking of disturbance.

Fig. 1. Simulation of beam system.

Acknowledgements

The research work has been done during the stay of the first author at the University of Lyon 1 with the sponsor and financial support of the Chinese Scholarship Council (NO. 202206630060).

References

[1]. J. Han, From PID to active disturbance rejection control. *IEEE transactions on Industrial Electronics*, Vol. 56, Issue 3, 2009, pp. 900-906.

[2]. H. Y. P. Feng, and B. Z. Guo, New unknown input observer and output feedback stabilization for uncertain heat equation. *Automatica*, Vol. 86, 2017, pp. 1–10.

[3]. T. Ahmed-Ali, F. Giri, M. Krstic, L. Burlion, and F. Lamnabhi-Lagarrigue, Adaptive boundary observer for parabolic PDEs subject to domain and boundary parameter uncertainties, *Automatica*, Vol. 72, 2016, pp. 115–122.

[4]. B. Z. Guo, and F. F. Jin, The active disturbance rejection and sliding mode control approach to the stabilization of the Euler-Bernoulli beam equation with boundary input disturbance. *Automatica*, Vol. 49, Issue 9, 2013, pp. 2911–2918.

[5]. Z. Wang, W. Wu, D. Gröges, and X. Lou, Sliding mode vibration control of an Euler-Bernoulli beam with unknown external disturbances, *Nonlinear Dynamics*, Vol. 110, Issue 2, 2022, pp. 1393–1404.

[6]. L. Paunonen, and D. Phan, Reduced order controller design for robust output regulation, *IEEE Transactions on Automatic Control*, Vol. 65, Issue 6, 2019, pp. 2480–2493.

[7]. B. Z. Guo, and R. Y. Yu, The Riesz basis property of discrete operators and application to a Euler-Bernoulli beam equation with boundary linear feedback control, *IMA Journal of Mathematical Control and Information*, Vol. 18, Issue 2, 2001, pp. 241–251.

[8]. X. R. Fan, and C. H. Kou, Output feedback stabilization of Euler–Bernoulli beam equation with general corrupted boundary observation, *Applicable Analysis*, 2022, pp. 1–19.

[9]. G. Weiss, Admissibility of unbounded control operators, *SIAM Journal on Control and Optimization*, Vol. 27, Issue 3, 1989, pp. 527–545.

[10]. B. Z. Guo, and H. C. Zhou, Output feedback stabilization for multi-dimensional Kirchhoff plate with general corrupted boundary observation, *European Journal of Control*, Vol. 28, 2015, pp. 38–48.

[11]. W. Guo, Y. L. Chen, and H. Y. P. Feng, Output feedback stabilization for a Kirchhoff-type nonlinear beam with general corrupted boundary observation, *International Journal of Robust and Nonlinear Control*, Vol. 27, Issue 16, 2017, pp. 3280–3295.

(8296)

Detection of Graphene Oxide in Single HeLa Cells based on MCR-Raman Spectroscopy

Z. Chaloupková, Jan Belza and K. Poláková *
Czech Advanced Technology and Research Institute (CATRIN), Regional Center of Advanced Technologies and Materials, Palacky University Olomouc, Olomouc, Czech Republic
Tel.: +420 585 634 473
E-mail: Katerina.Polakova@upol.cz

Summary: GO is a 2D nanomaterial that has attracted attention in many sectors in recent years, for example in the chemical industry, electronics or medicine. Due to its unique properties such as strength, hydrophilicity, and large specific surface area with the possibility of functionalization, GO is a particularly attractive material in biomedicine as a candidate for use in targeted drug delivery. In such a case, we need information on whether graphene oxide penetrates into cells and whether we are able to detect and monitor it inside the cells. This work focuses on the advanced detection of GO in cervical cancer cells (HeLa) using Raman spectroscopy based on MCR (Multivariate Curve Resolution) analysis enabling detection GO and changes in cellular environment at the molecular level. We demonstrate novel approach for studying of GO fate inside the cells, which can be further utilized as supporting method in any new therapeutic strategies focusing on drug delivery nanosystems. Moreover, Using MCR-Raman spectroscopy, the detection, distribution and chemical stability of GO inside HeLa cells can be monitored over six months.

Keywords: Raman spectroscopy, GO, MCR analysis, HeLa cells, 2D materials.

1. Introduction

Graphene, which was first isolated in 2004 [1], has become one of the best known, most studied and most widely used 2D nanomaterials. The functionalization of graphene by chemical and physical methods is essential for the use of graphene in other fields such as catalysis, detection of substances, carriers, adsorption, etc., which allows to obtain new properties of graphene derivatives according to the needs of their application [2]. An example of a successful derivatization of graphene is the GO. There is still a problem with the detection and monitoring of 2D graphene-based nanomaterials inside cells, since 'the biogenic carbon element in the structure is not easily detectable by any conventional method. On the other hand the fate, biodegradation, or active targeting within cells is extremely important for safety, but also in order to monitor the possible mechanism of action during targeted therapy.

Raman spectroscopy has a number of advantages over other methods in biomedical applications. The low sample preparation requirements and the relatively non-destructive nature of Raman spectroscopy allow it to be used for tissue analysis and, in conjunction with a microscope, individual cells and organelles can be easily analysed. In addition, no need of sample staining allows its use in the study of drug transport and metabolism [3]. 2D nanomaterials used for drug transport into cells must in particular meet the requirements of biocompatibility with the human body, biodegradability and low toxicity [4]. MCR-based Raman spectroscopy can be used to detect GOs penetrating inside HeLa cells and to further study the fate of GOs inside cells. Here, we introduce Raman spectroscopy with cell lineage mapping. The obtained background-adjusted multispectral data are analyzed using the MCR (Multivariate Curve Resolution) algorithm to identify individual components in the sample. The MCR-Raman method offers a promising approach for the determination of GO, its degradation products or the interaction of GO with cellular components (lipids and proteins), organelles or cytoplasmic membrane, showing great potential for expanding the knowledge of anticancer therapy and can be used in personalized diagnostics, therapy and drug transport.

2 Experimental

2.1. Synthesis of GO

GO was prepared by oxidation of graphite powder using the Hummers method. The procedure was as follows: 23 ml of sulphuric acid (dropwise) was added to a mixture of 1 g of graphite powder and 0.5 g of sodium nitrate. The reaction mixture was then cooled to 0 °C and 3 g of potassium permanganate was added successively during further cooling, keeping the temperature below 20 °C during this exothermic reaction. After the addition of the permanganate, the reaction mixture was heated to 35 °C and stirred for 30 minutes. In the next step, 46 ml of distilled water was slowly added and the reaction mixture was further stirred at 98 °C for 15 minutes. After 15 minutes, the mixture was cooled to room temperature, 1 ml of 30 % hydrogen peroxide and 140 ml of distilled water were added. The resulting material was centrifuged at 6000 rpm for 7 minutes and washed with distilled water. This process was repeated several times to remove residual impurities. The obtained material was dried overnight in a vacuum oven and then GO powder

was obtained. From the GO thus obtained, efficient size selection of GO powder dispersed in water was performed by sonication followed by centrifugation.10 mg of GO powder was dispersed in 20 ml of distilled water and sonicated for 1 hour. Then the dispersion was stirred overnight on a magnetic stirrer. This was followed by centrifugation at 13500 rpm for 20 min, and finally decantation of the supernatant was performed, which contained GO particles of the desired size.

2.1. Cell Culture Experiments

The cell culture used for the experiment was human cervical epithelial tumour Hela cells from ATCC, USA. The cells were cultured in DMEM (Dulbecco's Modified Eagle medium) with low glucose concentration from Life Technologies at 37 °C and 5 % CO_2 atmosphere. A calcium fluoride slide (substrate for Raman analysis) was inserted into a well in the culture plate. Standard trypsinization of the cells was performed and a sufficient number of cells (approximately 60 % confluence) were seeded into the well on the substrate. The cells were kept to adhere to the surface of the substrate overnight. The next day, labelling of the cells with GO suspension was performed to achieve the desired concentration of GO in the volume of medium used. The cells with GO were then incubated for a further 24 hours. Control cells were cultured for 48 hours without labelling with GO suspension. After incubation, the culture medium was removed and the cells were gently washed with phosphate buffered saline (PBS). Subsequently, 1 ml of 4 % paraformaldehyde solution in PBS was added to the cells for fixation. The cells were left in the refrigerator for 10 min for fixation, after which the fixing agent was aspirated followed by washing with phosphate buffer and ultrapure water. The cell sample was let to dry before measurement.

2.2. Surface Morphology

The surface morphology of GO was measured by AFM using tapping mode with ACTA-SS (AppNano, USA) tip, a frequency of 200-400 kHz and a force constant of 13-77 N/m. Measurements were performed at room temperature with humidity ranging from 45-55 %. Height profiles were calculated using Gwyddion 2.51 software. The size and overall morphology of samples were analysed by SEM on a Hitachi SU6600 (Hitachi, Japan) with an accelerating voltage of 5 kV. For these analyses, the sample was placed on a copper grid with a carbon layer and dried at room temperature. Zeta potential values were determined by the DLS method using the instrument Malvern Zetasizer Nano.

2.3. Raman Spectroscopy

For detection of Raman signal from GO, Raman spectral data were obtained on a DXR Raman microscopy (Thermos Scientific) with an excitation laser operating at 455 nm. Raman spectra were measured in the range from 500 to 3.500 cm-1. Spectral resolution of the applied combination of laser and optic was 1.0 cm-1. Experimental parameters were as follows: the number of exposures measured was 32 s and the length of exposure was 2 s, laser power 4mW (on the sample surface). A lens with 50x magnification was chosen. Before the CCD detector, a 50 μm pinhole assembly with a 1200 notches/mm grid was chosen. Fluorescence correction was also turned on to co-opt unwanted fluorescence from the cells. The spectra obtained were then adjusted to baseline using OMNIC 8.2.0.403 software (polynomial fit, where n=3).

Presence of GO in cells was confirmed by mapping. During the mapping, an area of 2-4 cells were mapped (usually 100 × 130 μm), in which a total of about 200 spectra were measured. Each spectrum was subjected to 32 exposures and the exposure time was 2 s. A total of 512 exposures were made to collect the background. The measured maps were processed in OMNIC software for background editing and further processing. The obtained background-adjusted multispectral data were further analysed using the Multivariate Curve Resolution (MCR) algorithm to identify individual components in the sample. Two regions of expected components (GO and cell) and four components (only one region and three components for control cells without GO) were given for the calculation. The components corresponding to HeLa cells, GO and substrate were then selected from the algorithmically found components. From these, RGB maps were generated to show the spatial distribution of each component using the assigned colours.

3. Results & Discussion

3.1. Characterization of Prepared GO

The prepared GO was characterized by AFM, SEM, ζ-potential measurements, Raman spectroscopy and the intensity ratio of the D and G bands. Fig. 1A shows that the GO height profile reaches approximately 1.2 nm and exhibits a large number of irregularities and is rugged. The sample has a relatively large area and is mainly composed of regions with fewer layers. The height profile is shown to play an important role in terms of penetration into HeLa cells, where a higher height profile improves penetration. Also, as the height profile stabilizes, GO localizes exclusively in the center of the cells, whereas for surfaces with variable height, the pattern is also found at the cell edges, see section Mapping GO in cells. SEM images of GO with scales of 2 μm and 500 nm are shown in Fig. 1B, 1C. The images show that the surface is more rugose, which is consistent with the AFM results. To characterize the properties of the layers in the dispersion, the zeta potential of GO was measured with a value of -30.5 mV. The negative zeta potential values are due to the presence of electronegative functional groups that are formed on the graphite lattice during oxidation [5].

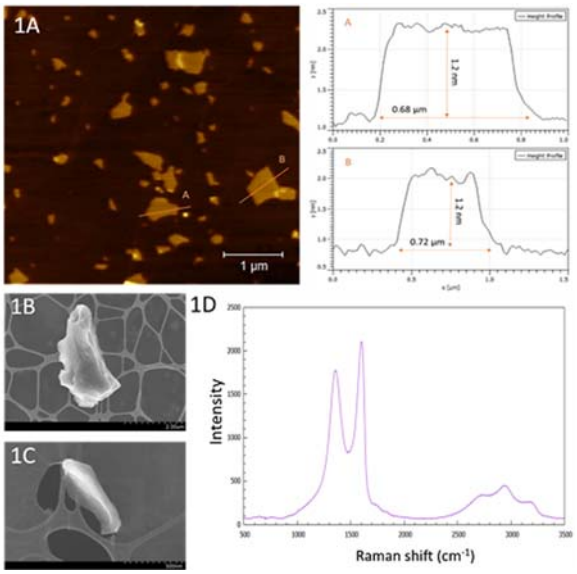


Fig. 1. A - AFM image of the used graphene oxide with height profiles calculated for the marked particles; B-C - SEM images of the used graphene oxide, D - Raman spectrum of graphene oxide used for incubation with cells.

GO was subsequently characterized by Raman spectroscopy; Fig. 1D shows the obtained spectrum typical of GO. The band with the highest intensity at a wavenumber of 1600 cm^{-1} corresponds to the G band, which is typical of both the Raman spectra of graphene and the spectra of its other forms. This band is due to the vibration of the carbon bonds in the hexagonal lattice; in addition, the π electrons in this sp2 bond cause resonance, and the resulting band is therefore often very intense. A second typical but less intense band with a wavenumber of 1350 cm^{-1} belongs to the D band, which occurs in GO spectra in the region 1342-1356 cm^{-1} [6]. This band is the result of distortions and deformations of the hexagonal sp2 lattice [7].

3.2. GO Penetration into HeLa Cells

In a first step, control HeLa cells, i.e. cells without GO, were characterized by Raman spectroscopy. All measured control cells provided spectra typical of a cervical cancer cell line. One representative spectrum is shown in Fig. 2.

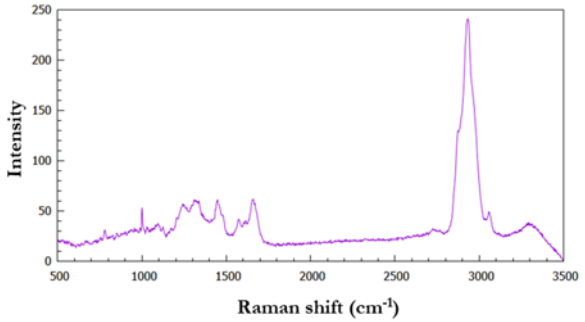


Fig. 2. Representative spectrum.

The bands are shown in the wavelength region 800-1800 cm^{-1}, which is considered the thumbprint region in cells and corresponds to nucleic acid and protein signals. At a wavelength of 780 cm^{-1}, signals from valence vibrations of the heterocycles cytosine and uracil were present, and at a wavelength of 1000 cm^{-1}, a distinct narrow band of the amino acid phenylalanine was observed. Around the 1100 cm^{-1} wavenumber, a band of symmetric vibrations of the O-P-O bonds of the phosphate skeleton of nucleic acids was present. In addition to the vibrations of proteins and nucleic acids, a band of vibrations originating from C-C bonds in lipids was also evident in this region at a wavenumber of 1130 cm^{-1}. Other nucleotides showed bands in the wavenumber range 1180-1400 cm-1 and a band originating from the vibration of amide III - with a wavenumber of 1310 cm^{-1} could be observed. At a wavenumber of 1450 cm^{-1}, a band corresponding to the deformation vibrations of CH_2 and CH_3 bonds in proteins can be clearly observed [8]. At a wavenumber of 1650 cm^{-1}, a signal from the vibration of amide I, which originates from the valence vibrations of peptide bonds in proteins, can be observed. In the 1800-2700 cm^{-1} wavelength region, there was a region without significant signals. This is a typical feature of cell spectra and is referred to as the silent region. The dominant feature of the whole spectrum is a band at 2930 cm^{-1} with high intensity, which corresponds to the symmetric valence vibration of CH3 of the phospholipid membrane. In the region 2854-2900 cm^{-1} a lower intensity signal is evident, corresponding to the asymmetric valence vibration of CH_2 bonds in lipids and fatty acids. At 3060 cm^{-1} a signal originating from the vibration of aromatic nuclei and CH=CH double bonds was observed. In the 3170-3500 cm^{-1} region, a signal originating from low-activity vibrations of water contained in the cell was found [9].

3.3. Mapping of GO Distribution in Cells

GO was detected in the samples by mapping and analyzed by Multivariate Curve Resolution (MCR), Fig. 2A-D. Maps were then generated from these results using RGB analysis [10]. These maps show the relative distribution of graphene oxide in HeLa cells in color. GO was present in all mapped samples, i.e. with concentrations of 18.75 µg/ml, 25 µg/ml, 37.5 µg/ml and 57.6 µg/ml in the resulting maps.

Fig. 3A shows that the sample map with the lowest mapped concentration (18.75 µg/ml) showed a small amount of particles dispersed throughout the cell cross-section. In Fig. 3B-C, the maps show an increase in the number of GO particles in the cells at the higher concentrations (25 µg/ml and 37.5 µg/ml). Clustering of a larger number of particles is also evident in both maps. In contrast, the map of the sample with the highest GO concentration, Fig. 3D, showed a significant predominance of GO. The HeLa cell was completely overwhelmed by it and contained GO in almost the entire cross section.

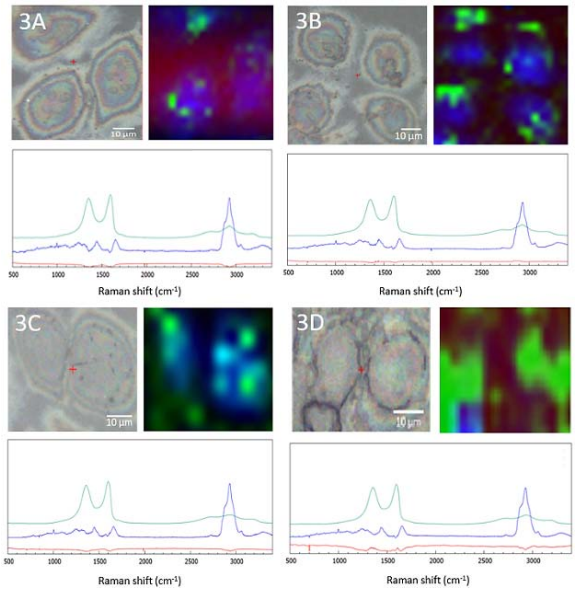


Fig. 3. Map of HeLa cells with GO concentration of 18.75 (A), 25 (B), 37.5 (C) and 57.6 µg/ml (D). The top left shows the microscope image with the mapping area marked, the top right shows the obtained RGB map and the bottom shows the spectra of the components found. Red colour represents the substrate, blue the HeLa cells and green the GO. The spectra are shifted by 100 counts, for better readability of the picture.

4. Conclusion

In conclusion we demonstrated the MCR Raman spectroscopy imaging as a powerful tool for studying graphene oxide inside cervical cancer cells. MCR-Raman spectroscopy could be used as highly complementary to fluorescence labeling-based methods and MCR imaging can become useful for exploratory single-cell studies, as MCR analysis faithfully mimics the structure of the analytical measurement and the MCR-Raman method offers a promising approach to determine GO within single cells at the same time. Since MCR-Raman spectroscopy can be used to observe peak shifts within a cell as a result of GO interaction with single cell components, there is great potential for studying changes in the consequences of this GO interaction within a cell. Based on the mapping, graphene oxide was confirmed inside the all cells, at various concentrations. It was evident from the obtained maps that the number of particles inside each cells increased with increasing concentration of GO used in cell incubation. At highest concentrations, clustering of GO particles was observed and the maps showed GO overwhelming the cells throughout the cross-section. Thus GO as a promising material with potential for drug delivery inside cancer cells can be further investigated using our novel approach in Raman spectroscopy.

Acknowledgements

This work was supported by the ERDF/ESF project "Nano4Future" (grant No. CZ.02.1.01/0.0/0.0/16_019/0000754)

References

[1]. K. S. Novoselov et al., Electric field effect in atomically thin carbon films. *Science,* 306, 5696, 2004, pp. 666-669.
[2]. H. S. Wahab et al., Synthesis and Characterization of Graphene by Raman spectroscopy, *Int. J. Mater. Sci.,* 1(3), 2015, pp. 130–135.
[3]. G. P. S. Smith et al. Raman imaging of drug delivery systems, *Adv. Drug Deliv. Rev*., 89, 2015, pp. 21–41.
[4]. A. Balzerová et al. Cellular Uptake of Graphene Acid by HeLa Cells Studied by Uv Raman Spectroscopy, *NANOCON*, 2018, pp. 566-571.
[5]. K. Krishnamoorthy et al. The chemical and structural analysis of graphene oxide with different degrees of oxidation, *Carbon*, 53, 2013, pp. 38–49.
[6]. A. C. Ferrari, Raman spectroscopy of graphene and graphite: Disorder, electron–phonon coupling, doping and nonadiabatic effects, *Solid State Commun.,* 143, 2007, pp. 47–57.
[7]. E. Aliyev et al., Structural Characterization of Graphene Oxide: Surface Functional Groups and Fractionated Oxidative Debris, *Nanomaterials*, 9, 8, 2019, pp. 1180.
[8]. B. Kann et al., Raman microscopy for cellular investigations — From single cell imaging to drug carrier uptake visualization, *Adv. Drug Deliv., Rev*., 89, 2015, pp. 71–90.
[9]. K. Moto et al., Local change in phospholipid composition at the cleavagefurrow is essential for completion of cytokinesis, *J. Biol. Chem.,* 280, 2005, pp. 37901–37907.
[10]. Ch. Lu et al., A graphene platform for sensing biomolecules, *Angewandte Chemie*, 121, 26, 2009, pp. 4879–4881.
[11]. A. de Juan et al., Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods*, 6, 2014, pp. 4964–4976.

(8882)

Optimizing Unmanned Surface Vehicle Performance for Water Body Monitoring

A. Kukuskins, U. Zaimis and S. Ozolina
Liepaja University, Faculty of Science and Engineering, Institute of Science and Innovative Technologies,
Liela iela 14, LV-3401, Liepaja, Latvia
Tel.: +371 29101515
E-mail: liepu.andrejs.kukuskins@gmail.com, uldis.zaimis@liepu.lv, sintija.ozolina@liepu.lv

Summary: Throughout the research process the work has accumulated in to an Unmanned surface vehicle (USV) that is equipped with sensors that can gather data about the physical, chemical, and biological characteristics of a water body. This data can be used for a variety of environmental monitoring applications, including water body ecological parameter monitoring, mapping of water bodies, species tracking, detection of intrusive objects, and autonomous and remote operation. The Haversine formula can be used to calculate the distance between two GPS coordinates and determine the direction in which the USV should move. It can also be modified to use specific sensors to navigate to the most problematic area and gather GPS data to pinpoint the source of the issue. To improve efficiency and reduce energy consumption, the gait function of the USV can be smoothed using an error-based algorithm or a model-based algorithm.

Keywords: Unmanned surface vehicles (USVs), Environmental monitoring, Water bodies, Sensors, Haversine formula, Navigation, Gait optimization, Autonomous.

1. Introduction

Unmanned surface vehicles (USVs) offer a flexible and cost-effective way for municipalities to monitor and protect their water bodies [1]. The developmental model is equipped with sensors such as potential of hydrogen (pH), oxidation-reduction potential (ORP), total dissolved solids (TDS), an oxygen level sensor, turbidity, an altimeter, and an echosounder. The USV can be equipped with a different array of sensors upon request, after the USVs can gather valuable data about the physical, chemical, and biological characteristics of a water body. The USV may be equipped with marking floats that have anchors attached to them to mark the area with the problematic readings from a specific sensor. This data can be used for a variety of environmental monitoring applications, including:

1. Water body ecological parameter monitoring: By measuring parameters such as pH, dissolved oxygen, and total dissolved solids, USVs can help municipalities identify areas of the water body that may be experiencing changes in water quality. This information can be used to detect pollution or other problems and take appropriate action to protect the water body.
2. Mapping of water bodies: USVs equipped with GPS and mapping software can create detailed maps of water bodies, including features such as depth, shoreline, and underwater obstacles. This information can be used by municipalities to better understand the characteristics of the water body and identify areas that may be of concern.
3. Species tracking in water bodies: USVs equipped with sensors such as hydrophones or cameras can be used to track the movement and behavior of aquatic species in a water body. This information can be used to understand how different species use the water body and how they are affected by human activities or environmental changes.
4. Detection of intrusive objects: USVs equipped with sonar or other sensors can be used to detect objects in the water that may pose a risk to navigation or the environment. For example, USVs could be used to identify abandoned fishing nets or other debris that could harm aquatic life or interfere with water activities.
5. Autonomous and remote operation: USVs can be programmed to operate autonomously, following predetermined routes or responding to changing conditions. This allows them to collect data and perform monitoring tasks without the need for direct human intervention. USVs can also be operated remotely, using telemetry and other technologies to transmit data back to a central location for analysis.

2. Navigating Unmanned Surface Vehicles with the Haversine Formula

To control the movement of the USV, the Haversine formula can be used to calculate the distance between two Global Positioning System (GPS) coordinates and determine the direction in which the USV should move. The Haversine formula is a mathematical formula used to calculate the great-circle distance between two points on a sphere, based on their longitudes and latitudes [2]. It is commonly used in navigation and geography to determine the shortest distance between two points, taking into account the curvature of the Earth. By using the Haversine formula to calculate the distance between two GPS coordinates, the USV can determine the direction in which it should move in order to reach its destination, without the need for an accelerometer or gyroscope.

The Haversine formula can be modified to use one of the specific sensors that is equipped in the boat to navigate to the most problematic area using said sensor instead of using GPS coordinates to navigate, and instead gather GPS data to make pinpoint the source of the problematic area. This approach would help municipalities resolve potential legal disputes with entities that may be polluting the environment.

3. Gait Function Smoothing for Enhanced USV Efficiency

By combining Kalman Filter [3] and Spline Smoothing Fig. 1 - gait function smoothing algorithms, a more accurate navigation path can be achieved compared to the path without any smoothing. The use of gait smoothing algorithms can improve USV navigation efficiency and energy consumption by reducing the amount of deviation from the ideal navigation path. The results provide valuable insights into the potential benefits of gait smoothing algorithms for USV navigation.

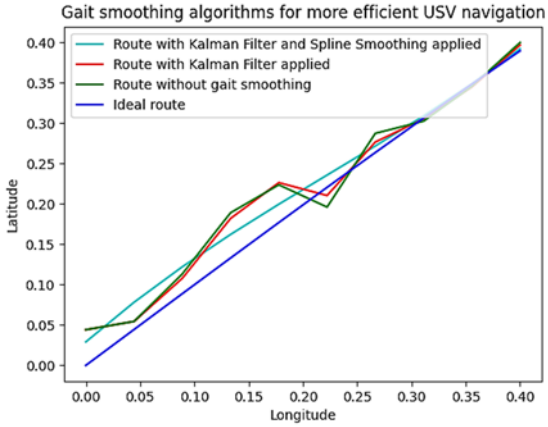


Fig. 1. Gait smoothing algorithms applied to USV Navigation.

4. AI for Enhanced USV Surveys

There are several ways that artificial intelligence (AI) could be used to enable an USV to perform an in-depth survey of problem areas [4]:

1. Object recognition: Machine learning algorithms can enable USVs to recognize objects and map problem areas, such as plants, fish, or pollution, in real-time.
2. Path planning: AI algorithms can help USVs plan their own path while gathering data and avoiding obstacles.
3. Data analysis: It could use AI algorithms to analyze the data gathered by the USV in real-time, looking for patterns or anomalies that might indicate the presence of a problem area. This could allow the USV to identify and prioritize areas that require further investigation.
4. Autonomous operation: It could use AI algorithms to enable the USV to operate autonomously, meaning that it would be able to make decisions and take actions on its own without the need for human intervention. This could allow the USV to perform its survey tasks more efficiently and with greater accuracy.

As an example, for object detection [5]: OpenCV and machine learning can be used to detect oil products on the water's surface, but the task can be challenging due to the variable and cluttered aquatic environment. To improve accuracy, a machine learning model must be trained on a large dataset of images including oil and non-oil products. To enable the USV to detect oil products, it must be equipped with a camera or imaging device and have code to process and classify the images. Factors such as lighting, water clarity, and the size and shape of the oil products can impact the accuracy of the model.

5. Conclusions

USVs provide an efficient and cost-effective solution for monitoring and protecting water bodies. They gather crucial data for environmental monitoring and improve efficiency through the Haversine formula, AI and smoothing algorithms.

Acknowledgements

This work was carried out within partial financial support from the European Regional Development Fund Project No. 1.1.1.5/018 "Promotion of research, innovation and international cooperation in science at Liepaja University"

NATIONAL DEVELOPMENT PLAN 2020
EUROPEAN UNION
European Social Fund
INVESTING IN YOUR FUTURE

References

[1]. Rao, Aravinda S., et al., Design of low-cost autonomous water quality monitoring system, in *Proceedings of the International IEEE Conference on Advances in Computing, Communications and Informatics (ICACCI2013)*, 2013, pp. 14-19.
[2]. PyPI Haversine (https://pypi.org/project/haversine/).
[3]. Welch, Greg, and Gary Bishop, An introduction to the Kalman filter, *ACM*, 1995, 2.
[4]. Gal, Oren, and Eran Zeitouni, Tracking objects using PHD filter for USV autonomous capabilities, in Robotic Sailing 2012: Proceedings of the 5th International Robotic Sailing Conference, *Springer,* Berlin Heidelberg, 2013.
[5]. Anand, Gokulnath, and Ashok Kumar Kumawat, Object detection and position tracking in real time using Raspberry Pi, *Materials Today: Proceedings,* 47 2021, pp. 3221-3226.

(9246)

Relative Exact Controllability of Fractional Linear Systems with Delays in Control

J. Klamka
Institute of Theoretical and Applied Informatics Polish Academy of Sciences,
Street Bałtycka 5, 44-100 Gliwice, Poland
Tel.: + 798191701
E-mail: jerzy.klamka@iitis.pl

Summary: In the paper linear, fractional, continuous time, infinite-dimensional, stationary dynamical control systems with multiple variable point delays in admissible control described by state equations are considered. Using notations, theorems and methods taken directly from functional analysis and linear controllability theory, necessary and sufficient conditions for global relative exact controllability in a given finite time interval are formulated and proved. The main result of the paper is to show, that global relative exact controllability of fractional linear systems with different types of delays in admissible control is equivalent to invertibility of suitably defined relative exact controllability operator. In the proof of main result, methods taken from the theory of linear bounded operators in Hilbert spaces are used. Some remarks and comments on the existing controllability results for linear fractional dynamical system with delays are also presented.

Keywords: Linear control systems, Fractional systems, Delayed systems, Relative controllability, Linear operators.

1. Introduction

In the paper we shall study global relative controllability in a given finite time interval for fractional, infinite dimensional, linear, continuous time dynamical systems with multiple time variable point delays in admissible controls. This is generalizations of controllability concept, which is known in the theory of finite dimensional linear control systems without delays in state variables or admissible control. Using techniques similar to those presented in the monographs [4, 6], and [10] and in the papers [1] and [3] we shall formulate and prove necessary and sufficient conditions for global relative controllability of fractional systems in a prescribed time interval. It should be pointed out, that the main result of the paper are generalizations for fractional and infinite dimensional delayed systems controllability conditions proved in [1-10].

2. System Description

Let us consider linear, fractional, delay dynamical systems containing multiple lumped time varying delays in admissible controls, described by the following differential state equation with fractional derivative [1, 7, 8]

$$D^{\alpha}x(t) = Ax(t) + \sum_{i=0}^{i=M} B_i(t)u(v_i(t)) \qquad (1)$$

for $0 < \alpha < 1$, $t \in [t_0 - h, t_1]$.

Initial data $\{x_0 \mu_{t_0}\}$ forms complete state of the fractional delayed system (1) at initial time t_0.

$$\begin{aligned} &x(t_0) = x_0 \in X, \\ &u(t) = u_{t_0}(t) \ \ for \ \ t \in [t_0 - h, t_0] \end{aligned} \qquad (2)$$

$D^{\alpha}(t)$ denotes fractional Caputo derivative order $0<\alpha<1$. X and U are given Hilbert spaces $x(t) \in X$ is the relative state, A is infinitesimal generator of an analytic semigroup $F(t)$ of uniformly bounded linear operators defined on Hilbert space X. and with resolvent set containing zero.

$B_{i,}$ for i=0,1,2,…,M are given linear bounded operators from U into X. Admissible controls $u \in U_{ad} = L^2([t_0, t_1], U)$.

Initial data $\{x_0 \mu_{t_0}\}$ forms so called complete state of the fractional delayed system (1) at initial time t_0.

The strictly increasing and twice continuously differentiable functions $v_i(t):[t_0,t_1] \rightarrow R$, $i=0,1,2,...,M$, represent deviating arguments in the admissible controls, i.e. $v_i(t)=t-h_i(t)$, where $h_i(t) \geq 0$ are lumped time varying delays for $i=0,1,2,...M$.

Definition 1. [1,8,10]. The system (1) is said to be globally relatively exactly controllable over given time interval $[t_0, t_1]$ if for each pair of vectors $x_0, x_1 \in X$ there exists an admissible control $u \in L^2([t_0, t_1], X)$ such that the solution of (1) with initial conditions (2) satisfies $x(t_1) = x_1$

In order to use results and methods known in the theory of bounded linear operators in Hilbert spaces, let us define linear relative exact controllability operator [10] as follows:

$$C_{\alpha} : L^2([t_0, t_1], U) \rightarrow X \qquad (3)$$

$$C_\alpha u =$$
$$= \sum_{i=0}^{i=m} \int_{t_0}^{v_i(t_1)} \sum_{j=0}^{j=m-i} F_\alpha(t_1 - r_j(s)) B_j r'_j(s) u(s) ds =$$
$$= \sum_{i=0}^{i=m-1} \int_{v_{i+1}(t_1)}^{v_i(t_1)} \sum_{j=0}^{j=m-i-1} F_\alpha(t_1 - r_j(s)) B_j r'_j(s) u(s) ds$$

Using the relative controllability operator C_α and its adjoint operator C_α^* let us define relative exact controllability operator $W_\alpha(t_0, t_1)$ for the linear fractional control system (1) as follows

$$W_\alpha(t_1, t_0) = C_\alpha C_\alpha^* =$$
$$= \sum_{i=0}^{i=m-1} \int_{v_{i+1}(t_1)}^{v_i(t_1)} \Big(\sum_{j=0}^{j=m-i-1} F_\alpha(t_1 - r_j(s)) B_j r'_j(s) \Big) \times$$
$$\times \sum_{j=0}^{j=m-i-1} F_\alpha(t_1 - r_j(s)) B_j r'_j(s))^* ds =$$
$$= \sum_{i=0}^{i=m-1} \int_{v_{i+1}(t_1)}^{v_i(t_1)} \Big(\sum_{j=0}^{j=m-i-1} F_\alpha(t_j - r_i(s)) B_j r'_j(s) \Big) \times$$
$$\times \Big(\sum_{j=0}^{j=m-i-1} B_j^* F_\alpha^*(t_1 - r_j(s)) r'_j(s) \Big) ds$$

Using relative controllability operator it is possible to formulate and prove main result of the paper.

Theorem 1. The following statements are equivalent

(1) Fractional system (1) is globally relatively controllable over $t \in [t_0, t_1]$.
(2) Relative controllability operator $C_\alpha : L^2([t_0, t_1], R^p) \to R^n$ is onto.
(3) Adjoint relative controllability operator $C_\alpha^* : R^n \to L^2([t_0, t_1], R^m)$ is invertible i.e., it is one to one operator.
(4) The bounded linear operator $C_\alpha C_\alpha^* : R^n \to R^n$ is onto and may be realized by *nxn* nonsingular matrix.

4. Conclusion

Using relative controllability operator it is possible to formulate and prove main result of the paper given in theorem 1, which presents necessary and sufficient conditions for global relative controllability in a given time interval. It should be pointed out, that the main result of the paper is the extension for infinite dimensional fractional systems necessary and sufficient conditions for global relative controllability formulated in [1, 3, 8, 10].

Let us observe, that from definition 1 directly follows that the trajectory of dynamical system between initial state, and final state generally is not prescribed. Hence, for globally relatively controllable systems generally there are infinitely many different admissible controls defined on given time interval, which steer dynamical system from initial relative state, to final relative state. Therefore, we may look for admissible control which is optimal in the sense, that it has minimum value of energy (see e.g. monographs [6] and [10]), so, it is so called minimum energy control.

Acknowledgments

The research was founded by Polish National Research Centre under grant "The use of fractional order controllers in congestion control mechanism of Internet", grant number UMO-2017/27/B/ST6/00145.

References

[1]. Kaczorek, T., Rechability and controllability to zero of cone fractional linear systems, *Archives of Control Sciences,* 17, 3, 2007, pp. 357-367.
[2]. Kaczorek, T., Fractional positive continuous-time linear systems and their reachability, *International Journal of Applied Mathematical Computation Science*, 18, 2, 2008, pp. 223-228.
[3]. Kaczorek, T., Positive fractional 2D continuous-discrete time linear systems, Bulletin of the Polish Academy of Sciences, *Technical Sciences*, 59, 4, 2011, pp. 575-579.
[4]. Kaczorek, T., Selected Problems of Fractional Systems Theory, *Springer-Verlag*, Berlin, 2012.
[5]. Kaczorek, T., Positive linear systems with different fractional orders, Bulletin of the Polish Academy of Sciences, Technical Sciences, 58, 3, 2010, pp. 453-458.
[6]. Kaczorek, T., Rogowski, K., Fractional Linear Systems and Electrical Circuits, Studies in Systems, Decision and Control, Vol. 13, *Springer*, 2015.
[7]. Kexue, L., Jigen, P., Laplace transform and fractional differential equations, *Applied Mathematic Letters*, 24, 12, 2011, pp. 2019-2023.
[8]. Kilbas, A. A., Srivastava, H. M., Trujillo, J. J., Theory and Applications of Fractional Differential Equations, *Elsevier*, Amsterdam, 2006.
[9]. J. Klamka, Controllability of dynamical systems. A survey, *Bulletin of the Polish. Academy of Sciences, Technical Sciences,* Vol. 61, No. 2, 2013, pp. 221-229.
[10]. J. Klamka, Controllability and Minimum Energy Control, Studies in Systems, Decision and Control, Vol. 162, pp. 1-175, *Springer Verlag,* 2018.

(9599)

Sliding Mode Control Combined with Neural Networks for a Two-link Robot

A. Belkheir and E. M. Mellouli
Laboratory of Engineering, Systems and Applications (LISA) Sidi Mohamed Ben Abdellah University Fez, Morocco
E-mail: belkheir.ayoub@usmba.ac.ma, mellouli_elmehdi@hotmail.com

Summary: This paper concentrates on modeling and controlling a two-link robot which is a MIMO (Multi-Input and Multi-Output) system using a combination between Sliding Mode Control (SMC) and artificial intelligence (precisely, Radial Basis Function Neural Network (RBFNN)). In the first part, we present the model dynamics of this system in the state space. Then in the second part, we propose a new strategy, where we try to find the best performance and achieve stability in finite time by estimating the non-linear dynamics of the system and also minimizing the disturbance and uncertainties effects on the system using artificial intelligence. By examining the Lyapunov function we can prove the stability of the system. And, based on the simulation shown in the last section of this paper of this new methodology proposed, we demonstrate and improve our methodology's superiority over other approaches, their good results, and efficiency in time tracking, stability, and, robustness.

Keywords: Two-link robot, Sliding mode control, Lyapunov function, Non-linear function, Artificial intelligence.

1. Introduction

In the last few years, automatic systems are becoming very important in all fields of industry. That makes the controlling quality in terms of fastness, security, and robustness very needed, which make the field jump faster in the last twenty years and developing other technics to reduce disturbance, perturbations, and uncertainties to make the system stable.

Two link robot [4, 7] is one of the most used robots in the industry field precisely the car, plane fabrications, and others. For that, fastness and stability are very important for this type of robot. Therefore, some societies try to investigate in this kind of robot to develop other strategies to get the best control and use.

Since SMC has been researched for so long, it is currently one of the most popular technics used in non-linear control theory. The literature has numerous studies on sliding mode control and its extensions, including [1-3, 5].

One of the best estimations is the neural network [6], which can estimate any nonlinearities with arbitrary precision as its size and complexity rise. Due to their effectiveness, neural networks are also widely used to identify and control non-linear systems.

2. Two-link Robot Model

The two-link robot system can be represented in state space as follows [3, 5, 8]:

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= x_3\\ \dot{x}_3 &= f_1(x) + g_{11}(x_1,x_4)u_1 + g_{12}(x_1,x_4)u_2\\ \dot{x}_4 &= x_5\\ \dot{x}_5 &= x_6\\ \dot{x}_6 &= f_2(x) + g_{21}(x_1,x_4)u_1 + g_{22}(x_1,x_4)u_2,\end{aligned} \tag{1}$$

where,

$$\mathrm{x} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6]^T$$

$$\mathrm{U} = [u_1 \quad u_2]^T, f(x) = [f_1(x) \quad f_2(x)]^T$$

$$g(x_1,x_4) = \begin{bmatrix} g_{11}(x_1,x_4) & g_{12}(x_1,x_4) \\ g_{21}(x_1,x_4) & g_{22}(x_1,x_4) \end{bmatrix}$$

3. Proposed Controller

Let's take the two following sliding surfaces [1, 2],

$$\begin{aligned} S_1 &= \ddot{e}_1 + \lambda_1\dot{e}_1 + \lambda_2 e_1\\ S_2 &= \ddot{e}_2 + \lambda_3\dot{e}_2 + \lambda_4 e_2\end{aligned} \tag{2}$$

where, λ_1, λ_2, λ_3 and λ_4 are a positives constants and $e_1$and e_2 are the angle errors which are expressed as following,

$$e_1 = x_1 - x_{1d} \text{ and } e_2 = x_4 - x_{4d} \tag{3}$$

and x_{1d} , x_{4d} are the desired angle trajectories (see Fig. 3 and Fig. 4).

Using the derivative of (2), and based on (1) and (3) the control laws are defined as,

$$u_1 = \frac{1}{g_{11}(x_1,x_4)}[\ddot{x}_{1d} - f_1(x) - g_{12}(x_1,x_4)u_2 - \lambda_1\ddot{e}_1 - \lambda_2\dot{e}_1 - K_1 sgn(S_1)]$$

$$u_2 = \frac{1}{g_{22}(x_1,x_4)}[\ddot{x}_{4d} - f_2(x) - g_{21}(x_1,x_4)u_1 - \lambda_3\ddot{e}_2 - \lambda_4\dot{e}_2 - K_2 sgn(S_2)]$$

Because the system uses non-linear functions to generate the control law, which does not give the desired results for that purpose we propose to estimate the non-linear functions $f_1(x)$ and $f_2(x)$ with artificial intelligence, which are represented as follows,

$$\hat{f}_1(\mathrm{x}) = \hat{w}_1{}^t.[\,h_i(y_1)]\,,\ \hat{f}_2(\mathrm{x}) = \hat{w}_2{}^t.[\,h_i(y_2)], \qquad (4)$$

where $y_k = [e_k \quad \dot{e}_k]^t$ (k=1,2) , is the input of the neural network.

By, replacing $f_k(\mathrm{x})$ by $\hat{f}_k(\mathrm{x})$ in the expression of the control law, then, substitute the new control law (see Fig. 1, and Fig. 2) in the derivative of (2), we find,

$$\begin{aligned} \dot{S}_1 &= -\tilde{f}_1(\mathrm{x}) - K_1.sgn(S_1) \\ \dot{S}_2 &= -\tilde{f}_2(\mathrm{x}) - K_2.sgn(S_2), \end{aligned} \qquad (5)$$

where, $K_1$and K_2 are positives constants, and,

$$\begin{aligned} \tilde{f}_1(\mathrm{x}) &= f_1(\mathrm{x}) - \hat{f}_1(\mathrm{x}) = \tilde{w}_1{}^t.[\,h_i(y)], \\ \tilde{f}_2(\mathrm{x}) &= f_2(\mathrm{x}) - \hat{f}_2(\mathrm{x}) = \tilde{w}_2{}^t.[\,h_i(y)] \end{aligned}$$

Then, $\tilde{w}_1{}^t = w_1{}^t - \hat{w}_1{}^t$ and $\tilde{w}_2{}^t = w_2{}^t - \hat{w}_2{}^t$ are the difference between the ideal and the approximated weight.

To make the system stable we define the Lyapunov functions as,

$$\begin{aligned} V_1 &= \tfrac{1}{2}S_1{}^2 + \tfrac{1}{2}\beta_1\tilde{w}_1{}^t\tilde{w}_1, \\ V_2 &= \frac{1}{2}S_2{}^2 + \frac{1}{2}\beta_2\tilde{w}_2{}^t\tilde{w}_2, \end{aligned} \qquad (6)$$

where, β_1 and β_2 are positive constants.

Based on the derivative of (6) and using (5), we can assume that the terms $\hat{w}_1$ and $\hat{w}_2$ are,

$$\dot{\hat{w}}_1 = -\tfrac{1}{\beta_1}S_1[\,h_i(y_1)] \quad , \quad \dot{\hat{w}}_2 = -\tfrac{1}{\beta_2}S_2[\,h_i(y_2)] \qquad (7)$$

Then, using the results of (7), the system will be stable and the derivative of V_1 and V_2 are satisfied to the Lyapunov theory,

$$\dot{\mathrm{V}}_1 = -\mathrm{K}_1.\mathrm{sgn}(\mathrm{S}_1) \le 0, \dot{\mathrm{V}}_2 = -\mathrm{K}_2.\mathrm{sgn}(\mathrm{S}_2) \le 0$$

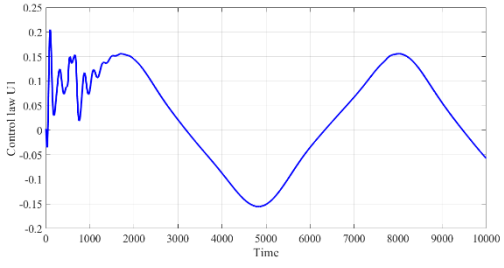


Fig. 1. The control law u_1 generated using the dynamic functions estimates.

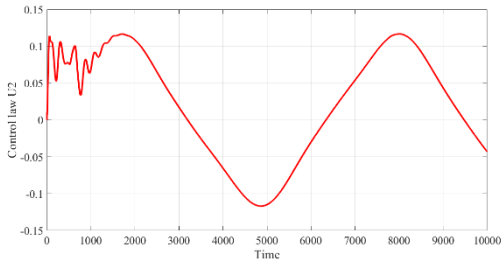


Fig. 2. The control law u_2 generated using the dynamic functions estimates.

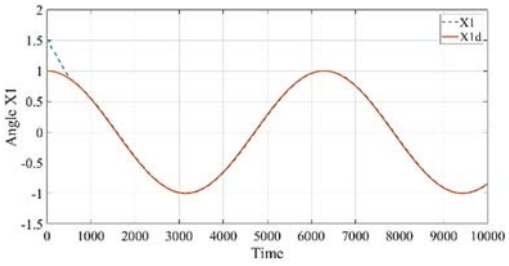


Fig. 3. Desired trajectory x_{1d} and estimate trajectory x_1.

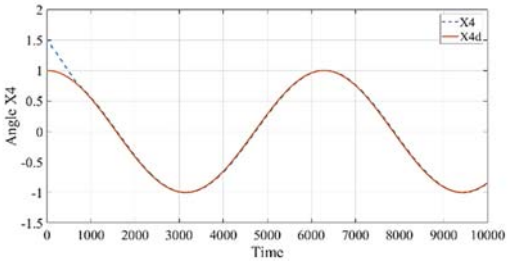


Fig. 4. Desired trajectory x_{4d} and estimate trajectory x_4.

4. Simulation

In this part, we present the results obtained using the developed strategy. By using the model of the system and sliding mode control, we test the force of the control strategy that we use to see their efficacy.

5. Conclusions

The paper presents a controller that estimates the dynamic functions of the two-link robot with two control laws using SMC and neural networks. This controller performs well in terms of time convergence and stability, as demonstrated by simulation.

References

[1]. C. Edwards and S.K. Spurgeon, Sliding Mode Control-Theory and Application, *Taylor & Francis*, London, U.K., 1998.

[2]. E. M. Mellouli, S. Sefriti, I. Boumhidi, Combined fuzzy logic and sliding mode approach's for modeling and control of the two link robot, in *Proceedings of the*

IEEE International Conference on Complex Systems (ICCS), 2012, pp. 1-6.

[3]. D. Cruz-Ortiz, I. Chairez and A. Poznyak, Sliding-Mode Control of Full-State Constraint Nonlinear Systems: A Barrier Lyapunov Function Approach, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 52, No. 10, Oct. 2022, pp. 6593-6606.

[4]. R. Naoual, E. M. Mellouli, S. Sefriti, I. Boumhidi, Fuzzy sliding mode control for the two-link robot, *International Journal of Systems, Control and Communications,* 6, 1, 2014, pp. 84-96.

[5]. E. M. Mellouli, M. Alfidi, I. Boumhidi, Fuzzy sliding mode control for three-tank system based on linear matrix inequality, *International Journal of Automation and Control,* 12, 2, 2018, pp. 237-250.

[6]. S. Zhang, P. Yang, L. Kong, W. Chen, Q. Fu and K. Peng, Neural Networks-Based Fault Tolerant Control of a Robot via Fast Terminal Sliding Mode, *IEEE Transactions on Systems, Man, and Cybernetics*: *Systems*, Vol. 51, No. 7, July 2021, pp. 4091-4101.

[7]. R. Naoual, E. M. Mellouli, I. Boumhidi, Adaptive fuzzy sliding mode control for the two-link robot, in *Proceedings of the 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14)*, 2014.

(9615)

A Rula-based Assessment Framework based on Machine Vision and Zero-Defect Manufacturing 4.0 Initiative

Foivos Psarommatis[1], Victor Azamfirei[2] and Fotios K. Konstantinidis[3]
[1]University of Oslo, SIRIUS, Centre for Scalable Data Access, Gaustadalleen 23B, 0373, Oslo
[2] Department of Product Realisation, School of Innovation, Design, and Engineering, Mälardalen University, Hamngatan 15, 632 20, Eskilstuna, Sweden
[3] Institute of Communication and Computer Systems, National Technical University of Athens, 9 Iroon. Polytechniou Str., Zografou Athens GR-157 73, Greece
E-mail: foivosp@ifi.uio.no

Summary: The Zero-Defect Manufacturing (ZDM) paradigm will drastically change the manufacturing system and its socio-technological interactions. Following the idea that 'quality is free', this is, the cost of appraisals is lower compared to defects, bigger effort will be placed in detect, predict, prevent, and repair. Such activities would require an adequate level of automation. Human operators are the ultimate flexible resource in the manufacturing system, thus should be appreciated and protected. Human Factors and Ergonomics (HF/E) have been studies for decades but not been part of Industry 4.0. To secure a true sustainable growth towards ZDM and Industry 5.0, humans should be the centre of such socio-technological revolution. Utilising state-of-the-art technologies, such as Machine Vision and Artificial Intelligence, and knowledge gained from the ergonomic science field, human-centred ZDM can be secured. In this paper, a RULA-based Machine Vision framework is proposed for the real-time assessment of human ergonomics in shop floor.

Keywords: Zero-defect manufacturing, ZDM, Human ergonomics, Risk assessment, Machine vision, Industry 5.0.

1. Introduction

The contemporary manufacturing world is shifting toward Industry 4.0; traditional production is being replaced by cutting-edge digital technologies that have the potential to completely transform the manufacturing domain [1]. Robotics, artificial intelligence, machine learning, big data, cloud computing, and machinery education are examples of these cutting-edge technologies [2]. The Industry 4.0 revolution has a significant potential to improve sustainability, reduce pollution, increase product efficiency, increase production stability, lower operating costs, and offer the plant a number of other advantages. Sustainable smart manufacturing has the potential to provide better quality, sustainability, and productivity at lower costs across the product life-cycle management [3, 4]. Product inspection, on the other hand, closes this loop by converting physical components into informative information and data (product inspection/measurement reports) [5, 6]. The design and production of products are significantly impacted by these continuing technical advancements. Cyber-Physical Production System (CPPS), which promise a significant increase in flexibility and efficiency during manufacturing, are required for their implementation in the choice of manufacturing equipment. A cyber-physical examination must be realised to guarantee the high quality of the planned and manufactured items [7]. Planning a product inspection is an important step that determines whether the inspection will be useful or not. Poor product quality can have an effect on a variety of levels, including direct financial losses and the environmental impact of resource waste [8]. Therefore, quality management is crucial for manufacturing businesses that wish to maintain or improve their operational and financial performance [9]. Poor quality can have negative social impacts as well, harming a company's reputation through its subpar goods and unsatisfied clients [8]. For manufacturers to create high-quality products with minimum performance loss, they must employ at least one Quality Improvement (QI) technique [10]. There is still more to be done even though firms have been using traditional QI techniques like Six Sigma, Lean manufacturing, Theory of Constraints, and comprehensive quality management for more than three decades. This is because these methods were created based on the production standards of the time without taking into account more recent technical advancements connected to Industry 4.0 [10] or data-driven approaches in general. The goal of Zero-Defect Manufacturing (ZDM), a new paradigm in quality assurance, is to completely eradicate manufacturing faults [10, 11] by utilizing the full potential of Industry 4.0 technology [12].

It can be said that 'without humans, nothing can happen'. Technologies and processes are to assess humans in performing their tasks. Workers in many industries and vocations can be impacted by work-related Musculoskeletal Diseases (MSDs), which continues to be the most prevalent work-related health issue in the European Union. In addition to having an impact on the workers themselves, they are expensive for businesses and society as a whole. All of the risk factors that arise at work must be identified to prevent this occupational disease. Once they have been identified, preventive steps or activities should be implemented to lessen them. Some authors suggest actions like shifting workers between tasks, offering

ergonomic training to staff, creating instruments for workplace MSDs analysis, and changing work equipment from an ergonomic standpoint [15]. The physical ergonomics can be evaluated by means of the Rapid Upper Limb Assessment (RULA) methodology. The RULA approach was created by McAtamney and Corlett in 1993 [16]. Its goal is to ascertain whether employees are exposed to MSDs risk factors in the upper extremities while carrying out their duties. The technique evaluates three elements: the posture of the various body parts, the load or force applied, and the muscle activity (static posture or repetitive movements). The RULA approach presents the advantages of 'user-fiendiness' and 'practical implementation'. It can be applied with the help of software [17].

Due to the exponential advancements of Artificial Intelligence (AI) and Edge Computing (EC), Machine Vision (MV) technology stands as the eyes of the machines in the Smart Factory that capture the environment recognising events without human interaction [18]. Generally, the most common MV applications are object detection, parts counting, surface defect identification, paper reading and locating, quality measurement, and robotic guidance [19]. Nonetheless, the MV systems are not so popular in the occupational, H&S applications, despite the fact that there are some relevant applications, such as the autonomous AROWA (Autonomous RObot framework for Warehouse 4.0 health and safety inspection operations) platform that identifies H&S risks, informing the responsible systems [20].

On the other hand, MV systems that can be mounted on manufacturing cells and assess the ergonomics of the workers in real-time have not been widely explored yet. To foster a safe and healthy environment, the proposed vision-based RULA platform combines modern technologies, vision-based sensors, and intelligent control features. The proposed platform identifies the gesture of the worker and utilises intelligent techniques to extract the significant feature to finally assess the posture based on RULA principles. The contributions of this paper are summarized as follows:

- The introduction of a vision-based system, detecting RULA aspects in real-time within manufacturing cells.
- The proposed architecture utilizes image-based sensors and artificial intelligence to recognise workers' postures.

The rest of this paper is structured as follows. Section II presents related work in the field. Section III introduces the theoretical model for risk assessment of human ergonomics through the usage of machine vision and RULA framework. Section IV discusses the proposed model and its challenges. Section V presents conclusions and suggestions for future work.

2. Related Work

According to Psarommatis et al., [12], ZDM is composed of four strategies 'Detect', 'Predict', 'Repair', and 'Prevent'. The four ZDM strategies are separated into two different categories triggering and action strategies. To identify a quality issue, the Detect and Predict triggering strategies can be used. The Detect strategy represents the product quality inspection, and the prediction uses the data from the inspection and from other sources of data to predict the product quality in the near future [12]. Notably, the triggering ZDM strategies must be applied to every product to achieve true ZDM. The ZDM-Detect strategy can be performed in two ways (i) the physical, where measurements to the physical part are performed, or (ii) the virtual, where data from the process are analysed, and the quality is estimated, in virtual detection, there is no access to the physical part. If a defect is detected, then it can be repaired if possible, or perform preventive actions to avoid defects in the future. On the other hand, if a defect is predicted, then preventive actions are the only solution for avoiding the generation of the defect [12, 21].

Human Factors and Ergonomics (HF/E) has a substantial impact on manufacturing quality; therefore, the human-centric ZDM method could produce new insights and improvements in quality [12, 21]. Along with correction and compensation, human resources have been added as a third crucial policy under ZDM, where "automation must be capable of facilitating the mutual learning of humans and machines" [22]. Research shows that incorporating a human component can boost efficiency despite the trend followed in the past to remove humans from the production environment to reduce human error [23]. Businesses must invest in knowledge and skills, preserving human experience and intelligence for the onsite analysis of Non-Destructive Testing (NDT) [24], in addition to innovative technologies to be competitive [23]. In response to the highly technologically focused Industry 4.0 paradigm, the European Commission produced a study on Industry 5.0 in 2021 that placed a strong emphasis on the human-centric 'revolution' [25]. Industry 5.0's exact shape is still unknown, but the European Commission anticipates a shift towards purposefulness where 'human-centric,' 'resilient,' and 'sustainable' are the key components [26]. With more emphasis being placed on implementing new technologies, the human element might easily be forgotten. The human component of ZDM has not received the same level of attention as lean manufacturing, and more research is needed to understand this vital component of manufacturing [22].

The MV technology stands as a strong tool for the ZDM initiative as replacing human manual visual inspection with autonomous ones, driving high production variability. In the manufacturing environment, a vast variety of solutions such as vision-based tracking system [27, 28], quality-based application [29], augmented reality solutions [30] have been proposed and tested. These applications are made possible by MV and AI technologies. MV also assists in the inspection of humans working with cobots in smart assembly workstations or as part of the

manufacturing [31, 32]. MV observes the actions and movements of humans in order to warn them of potentially risky behaviour or to alter the movements of robotic systems [33]. For the purpose of boosting the well-being of humans performing manufacturing tasks, systems have been created to observe and rate their posture. Fernandez et al. developed a framework to assess RULA scores from snapshots or digital video using MV and machine learning techniques [34]. Using wearable inertial sensors, Huang et al. developed a system to assess work-related musculoskeletal disorders based on RULA and REBA scores [35]. De Rosario et al. [36] presented an ergonomic assessment method using MV and a convolutional neural network to reduce the time of analysis of the videos used in ergonomic evaluations. The authors expressed the possibility of their approach to be extended to other methods like RULA. Paudel et al. [37] proposed a framework to carry out risk analysis of workers' ergonomic postures through 3D human pose estimation from video/image sequences of their actions. In comparison to ergonomic experts' judgement, their approach had a 93 % accuracy for good postures and 82 % accuracy for occluded postures using the RULA method.

RULA is a physical ergonomics assessment method [16]. Physical, cognitive, and organisational ergonomics are deeply interrelated. Stress and MSDs are the two most common causes of work-related ill health. Both conditions share common risk factors, and they frequently occur together, having an influence on each other. Automated RULA surveillance can highly increase resilience and thus aid the ZDM cause, but it comes at a price. Surveillance, depending on the culture and gender, brings competition and distrust. Operators and managers must have an organisational value for ZDM to avoid cognitive stress and, thus, truly aid physical ergonomics without jeopardising the others.

In contrast to the aforementioned advancements, our framework will innovate by proposing a methodology to recognise human gestures, evaluate the RULA score, and notify workers in real-time about potentially hazardous activities, taking ZDM and Industry 5.0 principles into account.

3. Framework

In this section, the flowchart of the proposed RULA-based framework is analysed and presented in Fig. 1. The proposed framework uses vision-based data to capture the workforce, MV techniques (machine- or deep-learning approaches) to understand their gesture, processing units to calculate the RULA score and an integration module to inform humans of their actions.

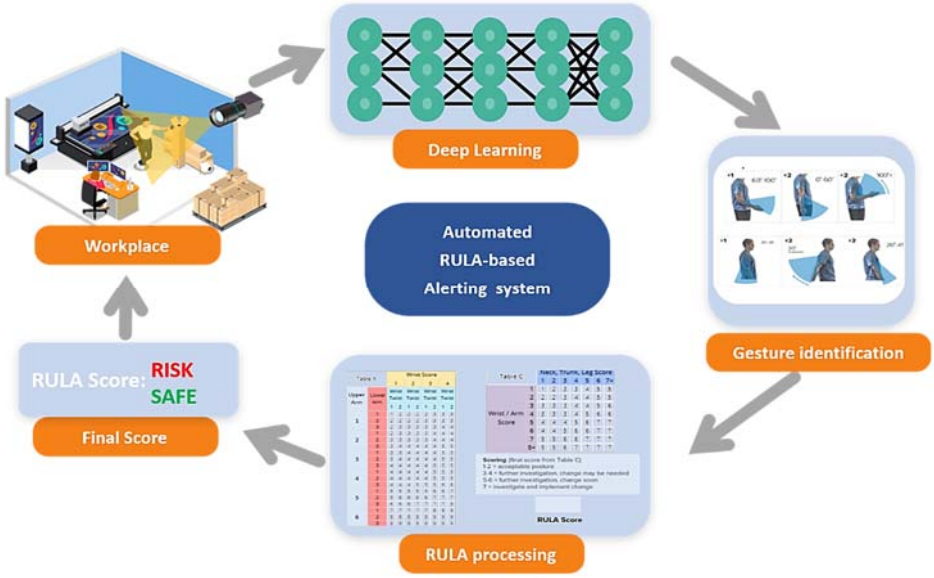


Fig. 1. Flowchart of the architecture of the proposed framework.

3.1. Workplace

Vision-based sensors in smart workplaces are not used to follow the progress of workers or measure their general communication with coworkers but rather to protect humans from potentially hazardous situations, such as ergonomics. For our system, vision-based sensors with a wide lens are employed to collect data about the surrounding area of workplaces. The wide field of view of the camera will aid in monitoring the region and the personnel.

3.2. Semantics and Perception

This module is the most significant component of the proposed framework, as the data of the images are translated into semantics. Segmentation deep learning model is used to identify the workers within the images

and mask them for further processing. The next intelligent layer includes the algorithms that are used to extract the skeleton of the workers from the masked sub-images, as presented in Fig. 2. Specifically, Part Affinity Fields (PAFs) representations based on 2D vector fields that record the direction and placement of limbs are used to identify the gesture of body parts using the publicly accessible code of Cao et al. [38]. Thanks to the multi-person 2D identification method, the suggested framework could be utilised in production environments regardless of the number of individuals. After the body perception, human recognition and tracking methods [39] empowers our framework in order to recognise the identity of the individuals, while tracking them to finally measure the overall time under posture, which is vital characteristic of human-centred RULA ZDM solution.

3.3. RULA Calculation

RULA provide a rapid assessment of the loads on the musculoskeletal system of the operators due to posture muscle function and the forces they exert. The scoring system focuses on two groups: Group A, formed by the upper and lower arm and wrist, and Group B, formed by the neck, trunk and legs [16] and presented in the Table 1.

In the proposed framework, the automated RULA algorithm identifies the angle of the aforementioned body parts and calculates the RULA score.

Fig. 2: This figure presents a worker within a manufacturing environment (designed by Freepik). The second image visualises the mask in the human, while the next one is the key point of the skeleton. Finally, connecting the lines, the angle of the body part is presented.

Table 1. RULA calculation scores.

Group	Body Part	Calculation Scores
Group A	Upper arm	1 for 20◦ extension to 20◦ flexion 2 for extension greater than 20◦ or 20-45◦ flexion 3 for 45-90◦ of flexion 4 for 90◦ or more of flexion +1 if the shoulder is elevated or the upper arm is abducted -1 if the operator is leaning or the weight of the arm is supported
Group A	Lower arm	1 for 60-100◦ flexion 2 for less than 60◦ or more than 100◦ flexion +1 if working across the midline of the body or out to the side.
Group A	Wrist	1 if in a neutral position 2 for 0-15◦ in either flexion or extension 3 for 15◦ or more in either flexion or extension +1 if the wrist is bet away from the midline.
Group B	Neck	1 for 0-10◦ flexion 2 for 10-20◦ flexion 3 for 20◦ or more flexion 4 if in extension +1 if the neck is twisted or side-bending
Group B	Trunk	1 when sitting and well supported with a hip-trunk angle of 90◦ or more 2 for 0-20◦ flexion 3 for 20-60◦ flexion 4 for 60◦ or more flexion +1 if the trunk is twisting or side-bending
Group B	Legs	1 if legs and feet are well supported and in an evenly balanced posture 2 if not.

3.4. Integration

At the integration level, the output of the automated RULA procedure will be sent to a server to track the KPIs on the monthly basis. In addition, integration with other systems is achieved through this module. Taking advantage of Internet of Things (IoT) systems, workers can be instantly informed about the harmful actions that may affect to their well-being. Alarms can be sent to personal devices such as smart watches, mobile phones etc. or screens within the workplace.

4. Discussion

The proposed approach combines the RULA methodology with MV and AI to create a real-time ergonomic evaluation approach for improving workers H&S. Nonetheless, the current model presents the limitation of only assessing the static posture of the RULA methodology. Thus, a less realistic assessment is given as forces and usage of muscles (time under posture and repetitiveness) are not considered. Furthermore, in comparison to similar literature in the field [34-37] the presented approach is connected to human-centred ZDM and Industry 5.0. The developed method is separated into the following steps using the ZDM approach:

- Detect: RULA's observation and analysis techniques using deep learning are used to identify potential ergonomic issues related to upper limb use, such as awkward postures, excessive force, and repetitive movements.
- Prevent: using the RULA final score, various suggested solutions and interventions are proposed to be implemented to prevent ergonomic issues from occurring, such as redesigning equipment or workstations to improve posture, reducing force requirements, and providing training on proper work techniques.
- Repair: RULA's suggested solutions and interventions would be implemented to address existing ergonomic issues, such as providing ergonomic training or modifying the work environment. HF/E assessment methods would also be used to address other existing issues, such as providing rest breaks or modifying the work schedule.

5. Conclusion

The objective of this study is to present a framework for assessing the risk of human ergonomics through the use of the RULA framework, MV and AI to notify the workforce in real-time of potentially hazardous movements. To ensure the H&S of workers, the suggested framework may be included in Industry 5.0 workstations, but it can also be used in other industries, such as construction, oil and gas, machinery, etc. The research concludes the necessary modules for the development of an autonomous system, while MV technology augmented with deep learning is able to recognise human gestures and compute the RULA score in real-time. In contrast to current frameworks, there is a research gap regarding the integration of these systems with workforce management systems or alert systems to transmit allergens in real time to the workforce. Future efforts will concentrate on the physical aspect of HF/E, more precisely, on the development of a vision-based module that exploits the entire RULA methodology, by considering also applied forces and usage of muscles (time under posture and repetitiveness). Furthermore, the vision-based module should be able to recognise several employees in parallel. Nonetheless, for the success of human-centric ZDM, cognitive and organisational ergonomics should also be considered.

Acknowledgements

The presented work was partially supported by the projects RE4DY, Eur3ka, TALON and PLOOTO, EU H2020 projects under grant agreements No 101058384, 101016175, 101070181 and 101092008 accordingly. The paper reflects the authors' views and the Commission is not responsible for any use that may be made of the information it contains.

References

[1]. M. Javaid, A. Haleem, R. P. Singh, S. Rab, and R. Suman, Exploring impact and features of machine vision for progressive industry 4.0 culture, *Sensors International*, Vol. 3, 2022, p. 100132.

[2]. F. Psarommatis and G. May, A literature review and design methodology for digital twins in the era of zero defect manufacturing, *International Journal of Production Research*, 2022, pp. 1–21.

[3]. M. B. Ahmed, F. Majeed, C. Sanin, and E. Szczerbicki, Smart virtual product development (svpd) system to support product inspection planning in industry 4.0, *Procedia Computer Science*, Vol. 176, 2020, pp. 2596–2604.

[4]. S. Ma, W. Ding, Y. Liu, S. Ren, and H. Yang, Digital twin and big data-driven sustainable smart manufacturing based on information management systems for energy-intensive industries, *Applied Energy*, Vol. 326, 2022, p. 119986.

[5]. G. Moroni and S. Petro, Geometric inspection planning as a key` element in industry 4.0, in *International Conference on the Industry 4.0 model for Advanced Manufacturing*, Springer, 2018, pp. 293–310.

[6]. V. Azamfirei, A. Granlund, and Y. Lagrosen, Multi-layer quality inspection system framework for industry 4.0, *International Journal of Automation Technology*, Vol. 15, No. 5, 2021, pp. 641–650.

[7]. O. Anokhin and R. Anderl, Towards design for cyber-physical inspection, *Procedia CIRP*, Vol. 84, 2019, pp. 400–405.

[8]. F. Psarommatis and D. Kiritsis, A hybrid decision support system for automating decision making in the event of defects in the era of zero defect manufacturing, *Journal of Industrial Information Integration*, Vol. 26, 2022, p. 100263.

[9]. P. Kumar, J. Maiti, and A. Gunasekaran, Impact of quality management systems on firm performance,

International Journal of Quality & Reliability Management, 2018.

[10]. F. Psarommatis, S. Prouvost, G. May, and D. Kiritsis, Product quality improvement policies in industry 4.0: characteristics, enabling factors, barriers, and evolution toward zero defect manufacturing, *Frontiers in Computer Science*, Vol. 2, 2020, p. 26.

[11]. D. Powell, M. C. Magnanini, M. Colledani, and O. Myklebust, Advancing zero defect manufacturing: A state-of-the-art perspective and future research directions, *Computers in Industry*, Vol. 136, 2022, p. 103596.

[12]. F. Psarommatis, G. May, P.-A. Dreyfus, and D. Kiritsis, Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research, *Int. J. of Production Research*, Vol. 58, No. 1, 2020, pp. 1–17.

[13]. B. Bidanda, P. Ariyawongrat, K. L. Needy, B. A. Norman, and W. Tharmmaphornphilas, Human related issues in manufacturing cell design, implementation, and operation: a review and survey, *Computers & Industrial Engineering*, Vol. 48, No. 3, 2005,
pp. 507–523.

[14]. F. K. Konstantinidis, N. Myrillas, S. G. Mouroutsos, D. Koulouriotis, and A. Gasteratos, Assessment of industry 4.0 for modern manufacturing ecosystem: A systematic survey of surveys, *Machines*, Vol. 10, No. 9, 2022, p. 746.

[15]. J. De Kok, P. Vroonhof, J. Snijders, G. Roullis, M. Clarke, K. Peereboom, P. van Dorst, and I. Isusi, Work-related musculoskeletal disorders: prevalence, costs and demographics in the eu, *European Agency for Safety and Health at Work*, 2019, Vol. 1.

[16]. L. McAtamney and E. N. Corlett, Rula: a survey method for the investigation of work-related upper limb disorders, *Applied ergonomics*, Vol. 24, No. 2, 1993, pp. 91–99.

[17]. M. Gomez-Gal, A.-J. Callejon-Ferre, J. Perez-Alonso, M. Dìaz-Perez, and J.-A. Carrillo-Castrillo, Musculoskeletal risks: Rula bibliometric review, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 12, 2020, p. 4354.

[18]. F. K. Konstantinidis, S. G. Mouroutsos, and A. Gasteratos, The role of machine vision in industry 4.0: an automotive manufacturing perspective, in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques, IST 2021, Kaohsiung, Taiwan, August 24-26, 2021*, 2021, pp. 1–6.

[19]. F. K. Konstantinidis, I. Kansizoglou, K. A. Tsintotas, S. G. Mouroutsos, and A. Gasteratos, The role of machine vision in industry 4.0: A textile manufacturing perspective, in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques, IST 2021, Kaohsiung, Taiwan,* August 24-26, 2021, pp. 1–6.

[20]. F. K. Konstantinidis, V. Balaska, S. Symeonidis, S. G. Mouroutsos, and A. Gasteratos, Arowa: An autonomous robot framework for warehouse 4.0 health and safety inspection operations, in *Proceedings of the 30th IEEE Mediterranean Conference on Control and Automation (MED)*, 2022, pp. 494–499.

[21]. F. Psarommatis, J. Sousa, J. P. Mendonc¸, and D. Kiritsis, Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: A position paper, *International Journal of Production Research*, Vol. 60, No. 1, 2022, pp. 73–91.

[22]. D. Powell, R. Eleftheriadis, and O. Myklebust, Digitally enhanced quality management for zero defect manufacturing, *Procedia CIRP*, Vol. 104, 2021, pp. 1351–1354.

[23]. L. Cattaneo, M. Rossi, E. Negri, D. Powell, and S. Terzi, Lean thinking in the digital era, in *Proceedings of the IFIP International Conference on Product Lifecycle Management*, 2017, pp. 371–381.

[24]. K. Bakopoulou, G. Michalos, K. Mparis, C. Gkournelos, N. Dimitropoulos, and S. Makris, A human robot collaborative cell for automating NDT inspection processes, *Procedia CIRP*, Vol. 115, 2022, pp. 214–219.

[25]. A. Renda, S. Schwaag Serger, D. Tataj, A. Morlet, D. Isaksson, F. Martins, and E. Giovannini, Industry 5.0, a transformative vision for Europe: governing systemic transformations towards a sustainable industry, *European Commission, Directorate-General for Research and Innovation,* 2022.

[26]. M.-P. Pacaux-Lemoine, D. Trentesaux, G. Z. Rey, and P. Millot, Designing intelligent manufacturing systems through human-machine cooperation principles: A human-centered approach, *Computers & Industrial Engineering*, Vol. 111, 2017, pp. 581–595.

[27]. F.K.Konstantindis, A.Gasteratos, and S. G. Mouroutsos, Vision-based product tracking method for cyber-physical production systems in industry 4.0, in *Proceedings of the IEEE Int. Conf. on Imaging Systems and Techniques (IST' 2018)*, 2018, pp. 1–6.

[28]. V. Balaska, D. Folinas, F. K. Konstantinidis, and A. Gasteratos, Smart counting of unboxed stocks in the warehouse 4.0 ecosystem, in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST' 2022)* , 2022, pp. 1–6.

[29]. F. K. Konstantinidis, V. Balaska, S. Symeonidis, D. Tsilis, S. G. Mouroutsos, L. Bampis, A. Psomoulis, and A. Gasteratos, Automating dairy production lines with the yoghurt cups recognition and detection process in the industry 4.0 era, *Procedia Computer Science*, Vol. 217, 2023, pp. 918–927.

[30]. F. K. Konstantinidis, I. Kansizoglou, N. Santavas, S. G. Mouroutsos, and A. Gasteratos, Marma: A mobile augmented reality maintenance assistant for fast-track repair procedures in the context of industry 4.0, *Machines*, Vol. 8, No. 4, 2020, p. 88.

[31]. S. Nahavandi, Industry 5.0—a human-centric solution, *Sustainability*, Vol. 11, No. 16, 2019, p. 4371.

[32]. K. M. Oikonomou, I. Kansizoglou, P. Manaveli, A. Grekidis, D. Menychtas, N. Aggelousis, G. C. Sirakoulis, and A. Gasteratos, Joint-aware action recognition for ambient assisted living, in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST' 2022)* , 2022, pp. 1–6.

[33]. M. Nardo, D. Forino, and T. Murino, The evolution of man–machine interaction: The role of human in industry 4.0 paradigm, *Production & Manufacturing Research*, Vol. 8, No. 1, 2020, pp. 20–34.

[34]. M. MassirisFernandez, J.´ A. Fernandez, J. M. Bajo, and C. A.´ Delrieux, Ergonomic risk assessment based on computer vision and machine learning, *Computers & Industrial Engineering*, Vol. 149, 2020, p. 106816.

[35]. C. Huang, W. Kim, Y. Zhang, and S. Xiong, Development and validation of a wearable inertial sensors-based automated system for assessing work-related musculoskeletal disorders in the workspace, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 17, 2020, p. 6050.

[36]. H. De Rosario, E. Medina-Ripoll, J. F. Pedrero-Sanchez, M. Sanchis-Almenara, A. Valls-Molist, and P. P. Miralles-Garcera, Ergonomic assessment with a convolutional neural network. A case study with owas, in *Proceedings of the 8th International Ergonomics Conference ERGONOMICS 2020*, 2021, pp. 65–71.

[37]. P. Paudel, Y.-J. Kwon, D.-H. Kim, and K.-H. Choi, Industrial ergonomics risk analysis based on 3d-human pose estimation, *Electronics*, Vol. 11, No. 20, 2022, p. 3403.

[38]. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.

[39]. M. Vrigkas, C. Nikou, and I. A. Kakadiaris, A review of human activity recognition methods, *Frontiers in Robotics and AI*, Vol. 2, 2015, p. 28.

[40]. V. Azamfirei, F. Psarommatis, and Y. Lagrosen. Application of automation for in-line quality inspection, a zero-defect manufacturing approach, *Journal of Manufacturing Systems,* 67, 2023, p. 1-22.

[41]. F. Psarommatis and D. Kiritsis, Identification of the inspection specifications for achieving zero defect manufacturing, in *Proceedings of the Advances in Production Management Systems. Production Management for the Factory of the Future: IFIP WG 5.7 International Conference, APMS 2019,* Austin, TX, USA, September 1–5, 2019, p. 267-273.

(9839)

Backstepping Control Method Combined with Fuzzy Logic for an Anti-Lock Braking System

N. Jennan and E. M. Mellouli
Laboratory of Engineering, Systems and Applications, Sidi Mohammed Ben Abdellah University, Fez, Morocco
E-mail: najlae.sef@gmail.com

Summary: In this work, we are interested in the control of one of the most important safety systems of vehicles, an Anti-Lock Braking system (ABS) of an autonomous vehicle, where we use a nonlinear Backstepping control method combined with the Takagi Sugeno (T-S) Fuzzy Logic method, based on the Lyapunov approach. In the first step, we introduce the Backstepping control method to the ABS system for the aim to control it by regulating the wheel slip at its optimum value, while respecting the stability constraints resulting from the study of the system stability using the Lyapunov approach. Afterwards, we propose the Fuzzy Logic technique for the aim to optimize the objective function that can provide a desired level of stability to the system, based on the Lyapunov approach for the aim to ensure the system stability. The simulation of the results found shows the accuracy and the efficiency of our proposed control strategy.

Keywords: Nonlinear system, Antilock braking system, Backstepping control, T-S Fuzzy logic.

1. Introduction

An Anti-Lock Braking System (ABS) is one of important safety systems of vehicles. It prevents the wheels from self-locking during braking, which offers the possibility of maintaining steering control of the vehicle [1]. For that, the control of this system is very important and must be accurate [2]. Therefore, the backstepping control method is proposed to control the ABS which consists in forcing the system to pursue the desired value in a finite time with certain conditions to respect in order to ensure the system stability [3, 4]. In order to improve the system control and to reach better results, the T-S Fuzzy Logic method is proposed to optimize the objective function that is designed to keep the system in a desired level of stability [5, 6]. Generalizing the control law requires the study of system stability using Lyapunov approach [7]. The efficiency of our control strategy proposed is evaluated and shown in the simulations of results achieved.

This paper is organized as follows: Section 2 concerns the ABS system modelling. Section 3, presents the control strategy proposed. In Section 4, the T-S Fuzzy Logic method is presented. Then, we present in Section 5 the simulations of results found. Finally, a conclusion is presented in Section 6.

2. ABS System Modeling

It is necessary to study the mathematical model of the ABS system in order to control it [1]. Fig.1 represents the wheel model.

The ABS system contains some important units as an electronic control unit, a hydraulic control unit and sensors of speed in each wheel.

Using Newton-Euler formalism, we represent the ABS model by a system of differential equations expressing the velocity v, the angular velocity ω and the slip coefficient λ [1].

$$\begin{cases} \dot{v} = -g.\mu \\ \dot{\omega} = \frac{1}{J}(m.g.\mu.R - T_b) \end{cases} \quad (1)$$

$$\dot{\lambda}(t) = f(\lambda, t) + h(\lambda, t).u(\lambda, t), \quad (2)$$

where $f(\lambda,t) = \left[\frac{\dot{v}}{v}(1-\lambda) - \frac{m.g.R^2}{v}\mu\right]$, $h(\lambda,t) = \frac{R}{v}$ are the system dynamics, $u = T_b$ is the control law that will be studied in the next section which represents the braking torque and $\mu = F(\lambda) = c_1\left(1 - e^{-c_2\lambda}\right) - c_3\lambda$ is the friction coefficient, where c_1, c_2 and c_3 are constant coefficients.

3. Control Strategy

In this work, we are interested in controlling the ABS system using Backstepping control method based on the T-S Fuzzy Logic methodology [3][8].

At first, we define the tracking error as follows:

$$e = \lambda_d - \lambda \quad (3)$$

Then, we define the Lyapunov function to study the system stability [7, 9].

$$V_1 = \frac{1}{2}e^2 + \frac{1}{2\gamma}.\Phi_\alpha.\Phi_\alpha{}^T > 0, \quad (4)$$

where $\Phi_\alpha = \theta_\alpha - \theta_\alpha{}^*$, with θ_α is the optimal parameter using T-S Fuzzy logic, and $\theta_\alpha{}^* = \arg\min(\sup_{\lambda \epsilon R} |\alpha(\lambda,t) - \hat{\alpha}(\lambda,\theta_\alpha)|)$.

The derivative of the Lyapunov function must be negative to ensure the system stability.

$$\begin{aligned} \dot{V}_1 &= e.\dot{e} + \frac{1}{\gamma}.\Phi_\alpha.\dot{\Phi}_\alpha \leq 0 \\ &= \frac{1}{\gamma}.\Phi_\alpha.\left(\dot{\Phi}_\alpha + e^2.\gamma.\xi(\lambda)\right) - \hat{\alpha}(\lambda,\theta_\alpha{}^*).e^2 \end{aligned} \quad (5)$$

We assume that:

$$\dot{V}_1 = -\hat{\alpha}(\lambda, \theta_\alpha{}^*).e^2 \leq 0 \quad (6)$$

Then, we define the control law of the system:

$$u(\lambda, t) = \frac{1}{h(\lambda,t)}\left(\dot{\lambda}_d - f(\lambda, t) + \hat{\alpha}(\lambda, \theta_\alpha).e\right) \quad (7)$$

where $\hat{\alpha}(\lambda, \theta_\alpha)$ is the optimal objective function that will be studied using the T-S Fuzzy Logic technique in the next section.

4. T-S Fuzzy Logic

In this section, we present the T-S Fuzzy Logic method used in the ABS system to optimize the objective function that keeps the system in a desired level of stability respecting the conditions of system stability using Lyapunov approach [10, 11].

We define first the membership functions. Then, the inference phase consists on a collection of IF-THEN rules defined in the following form [11]:

$$\text{If } x_1 \text{is } \alpha_1^j \text{ and..... and } x_n \text{ is } \alpha_n^j \text{ then } \alpha_j \text{ is } \theta_j$$

Finally, it is necessary to deduce the output of the fuzzy system using Lyapunov approach to ensure the system stability, which is expressed as follows:

$$\hat{\alpha}(\lambda, \theta_\alpha) = \frac{\sum_{j=1}^{k} \theta_j (\prod_{i=1}^{n} \mu_{\alpha_i^j}(\lambda_i))}{\sum_{j=1}^{k} (\prod_{i=1}^{n} \mu_{\alpha_i^j}(\lambda_i))} = \theta_\alpha{}^T.\xi(\lambda) \quad (8)$$

After the study of system stability using Lyapunov approach in (4) and (5), we deduce that $\hat{\alpha}(\lambda, \theta_\alpha{}^*)$ must be positive $\hat{\alpha}(\lambda, \theta_\alpha{}^*) \geq 0$ to ensure the system stability. And:

$$\theta_\alpha = \gamma.e^2.\xi(\lambda) \quad (9)$$

5. Simulation of Results

In order to evaluate the performance of our control strategy proposed, we did many simulations of results found. First, we employed the Backstepping control method in the ABS system to control the wheel slip and regulate it to its optimum value, which is chosen as follows: $\lambda_d = 0.15$. Then, we added the T-S Fuzzy Logic method to optimize the objective function that provides a desired level of stability to the system.

Fig. 1 shows the evolution of the control law, the braking torque T_b, obtained using our control strategy proposed. Fig. 2 represents the evolution of the wheel slip λ using the control strategy proposed with and without Fuzzy Logic method, which it can be seen that using the Fuzzy Logic in the system control is efficient; the convergence is faster than without using it.

The results show the efficiency of our proposed control strategy using the Backstepping control method combined with the T-S Fuzzy Logic method.

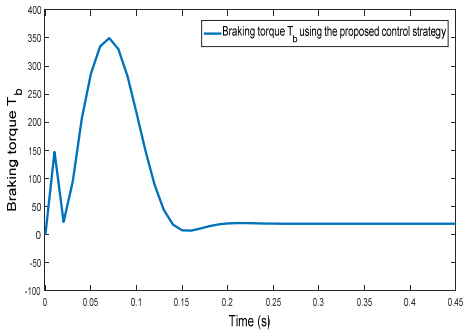


Fig. 1. The control law obtained by the proposed control strategy.

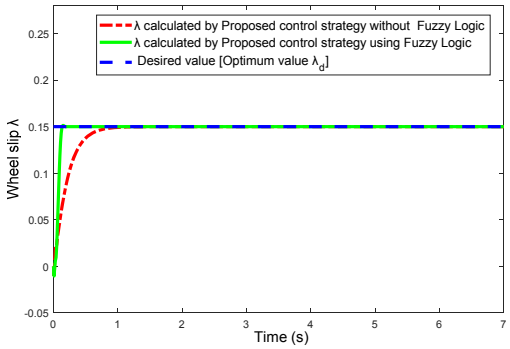


Fig. 2. The evolution of the wheel slip by the proposed control strategy.

6. Conclusion

In this work, we presented the Backstepping control method for controlling an Anti-Lock Braking System, combined with T-S Fuzzy Logic method to optimize the system dynamics in order to reach better results. The generalization of the control law requires the study of system stability using Lyapunov approach. The efficiency of our proposed control strategy is shown in the results simulation, and adding Fuzzy Logic method makes the system control more accurate.

References

[1]. F. Pretagostini, L. Ferranti, G. Berardo, V. Ivanov, B. Shyrokau. Survey on Wheel Slip Control Design Strategies, Evaluation and Application to Antilock Braking Systems. *IEEE Access*, 2020, pp. 10950-10971.

[2]. L. He, W. Ye, Z. He, K. Song, Q. Shi. A Combining Sliding Mode Control Approach for Electric Motor Anti-Lock Braking System of Battery Electric Vehicle. *Control Engineering Practice*, 2020, p. 104520.

[3]. A.M. Fihakhir, Modélisation, Stabilisation et Commande d'un Heliplane, *Thesis, University of science and technology of Oran*, 2012, pp. 43-50.

[4]. Y. Qiu, X. Liang, Z. Dai. Backstepping dynamic surface control for an anti-skid braking system. *Control Engineering Practice*, 2015, pp. 140-152.

[5]. E.M. Mellouli, I. Boumhidi. Direct adaptive fuzzy sliding mode controller without reaching phase for an uncertain three-tank-system, *International Journal of Modelling, Identification and Control*, 2016, pp. 335-342.

[6]. E. M. Mellouli, M. Alfidi, I. Boumhidi. Fuzzy sliding mode control for three-tank system based on linear matrix inequality, *International Journal of Automation and Control*, 2018, pp. 237-250.

[7]. Z. Cai, L. Huang. Generalized Lyapunov approach for functional differential inclusions, *Automatica*, 2020, p. 108740.

[8]. E. M. Mellouli, J. Boumhidi, I. Boumhidi. Using fuzzy logic for eliminating the reaching phase on the fuzzy H∞ tracking control, *International Journal of Modelling, Identification and Control*, 2013, pp. 398-406.

[9]. X. Li, X. Yang. Lyapunov stability analysis for nonlinear systems with state-dependent state delay, *Automatica*, 2020, p. 108674.

[10]. E.M. Mellouli, S. Sefriti, I. Boumhidi, Combined fuzzy logic and sliding mode approach's for modelling and control of the two link robot, in *IEEE International Conference on Complex Systems (ICCS)*, 2012, pp. 1-6.

[11]. R. Naoual, E. M. Mellouli, S. Sefriti, I. Boumhidi. Fuzzy sliding mode control for the two-link robot. *International Journal of Systems, Control and Communications*, 2014, pp. 84-96.

(9891)

Collaborative Robots and Set of Sensors for Learning by Demonstration

A. Rizzotti-Kaddouri [1], M. Kunze [2], L. Jeanneret [1] and L. Depierraz [2]
[1]School of Engineering Neuchâtel, HES-SO University of Applied Sciences and Arts Western Switzerland
[2]School of Engineering and Management Vaud, HES-SO University of Applied Sciences and Arts Western Switzerland
aicha.rizzotti@he-arc.ch

Summary: Nowadays, industrial and collaborative robots are more and more used in the industry. However, deploying a robot to perform a task or reprogramming it is complex and time-consuming. Thus, our solution is a new way to simplify programming of robots. The operator performs the operation holding the tool of the robot. This tool is attached to an HTC VIVE Tracker sensor which allows recording the movement of the operator. This permis to generate a trajectory using the recorded points. The operator can view the recorded points using Microsoft HoloLens 2 glasses. If necessary, the recorded trajectory can be filtered in order to smooth it and/or remove some outliers. Once the trajectory is validated, it can be replayed by the robot. The points of the trajectory are provided to MoveIt, a trajectory generator for robots running on ROS (Robot Operating System). It automatically generates a trajectory for each robot joint. The calculated trajectory is then transmitted to the robot controller which executes it.

Keywords: Collaborative robot, Industrial robot, Human-robot collaboration, Human-robot interaction, ROS, MoveIt, learning from demonstration, HTC VIVE tracker, Microsoft HoloLens.

1. Introduction

The capabilities of robots and humans are more and more complementary allowing them to work together. However solutions need to be developed in order to simplify robot programming. Learning by demonstration is one of them.

According to a recent review [1] in the area of demonstration-based robot learning, our work remains an important area of research. The plethora of existing approaches to learning by demonstration as well as the multiple evaluation criteria make comparisons between methods extremely difficult. The initial approach for this project was to increase the accuracy of motion data using a system composed of multiple sensors [2] and to extract a trajectory from the motion and fingers' position using machine learning. Different sensors including gloves and cameras were tested. However, due to accuracy and occlusion issues, we settled with an innovative solution by using an HTC Vive Tracker sensor. A study [3] has shown that this sensor is accurate enough for robotic applications. The HTC VIVE Tracker sensor is attached to the tool used by the operator. It can then track the movement of the tool, in three dimension, by locating its position relative to a VIVE Base Station. Since the tool position and rotation is directly tracked, it is not necessary to understand the intent and gesture of the user using complex machine learning models. Moreover, our solution is coupled with the Microsoft HoloLens 2 glasses. This allows the user to have an immediate feedback. It also provides tools to correct, in an intuitive way, possible recording errors. The Microsoft HoloLens glasses were developed and introduced in 2016. They are now widely used in medical robotics applications but still relatively limited in industrial robotics applications [4].

2. Methodology

We decided to merge the data from the following sensors: a RealSense D435 (depth camera) to detect objects with precision and the Trackers. The depth camera is placed above the workspace. An algorithm detects the plane of the table and then a threshold is applied to the depth of each pixel. This allows to detect items dimensions with accuracy (±3 mm delta error) on different kind of boxes.

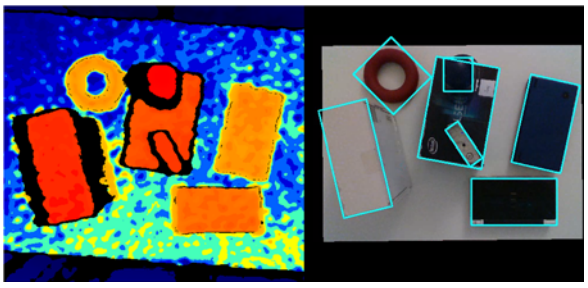


Fig. 1. Left: depth camera data, right: detected objects and color camera.

Two Trackers are used, one to track the tool and one to define the origin. A QR code is used to synchronize the origins of the Tracker and the HoloLens 2. The operator can view the limits of the workspace through the HoloLens and use controls to start and stop recording, as can be seen in Fig. 2.

Once the task is recorded, it can be uploaded to a web server for future usage. The communication between the web server, the HoloLens and the Trackers is done in a modular way.

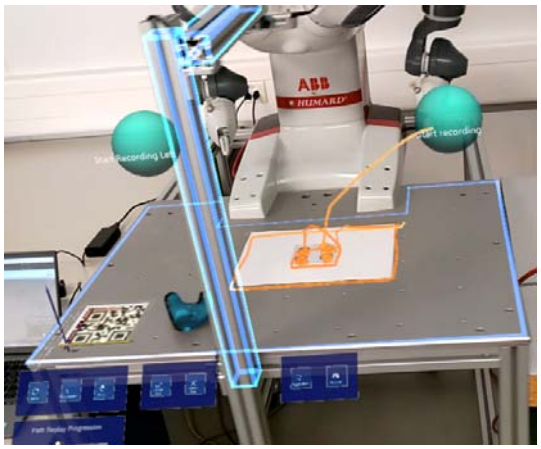


Fig. 2. Workspace viewed through the HoloLens.

3. System Architecture and Protocols

The system is separated into multiple modules, as can be seen in Fig. 3. The modules in red are physical sensors. The module in the middle contains two "sensor aggregator" applications which retrieve and merge data from sensors. A communication bus exchanges data between the HoloLens 2 module, the web server and the sensor aggregators. The server module stores the recorded trajectories. It is then accessed by the ROS (Robot Operating System) module, which converts the operator trajectories into robot trajectories, before sending them to the robot.

4. HTC VIVE Tracker and HoloLens 2

Multiple hand tracking sensors were tested before settling for the HTC VIVE Tracker, such as the Hi5 Gloves, Senso Gloves and more. However, they can be cumbersome to wear and some of them had low tracking accuracy. Additionally, their main purpose is to track fingers and hand position of the operator, which in many cases cannot be easily converted to a tool movement. To solve this in a pragmatic way, we decided to attach the tool to the Tracker. Since we could not use the gripper of the robot directly, we 3D printed a "pen" representing a gluing tool. However, in future work, it will be possible to print custom tips, either to attach existing tools, or to print "smart" tools. For example a gripper with a small electronic circuit, could automatically track the opening of the clamp during the recording, as the HTC VIVE Trackers has programmable pins.

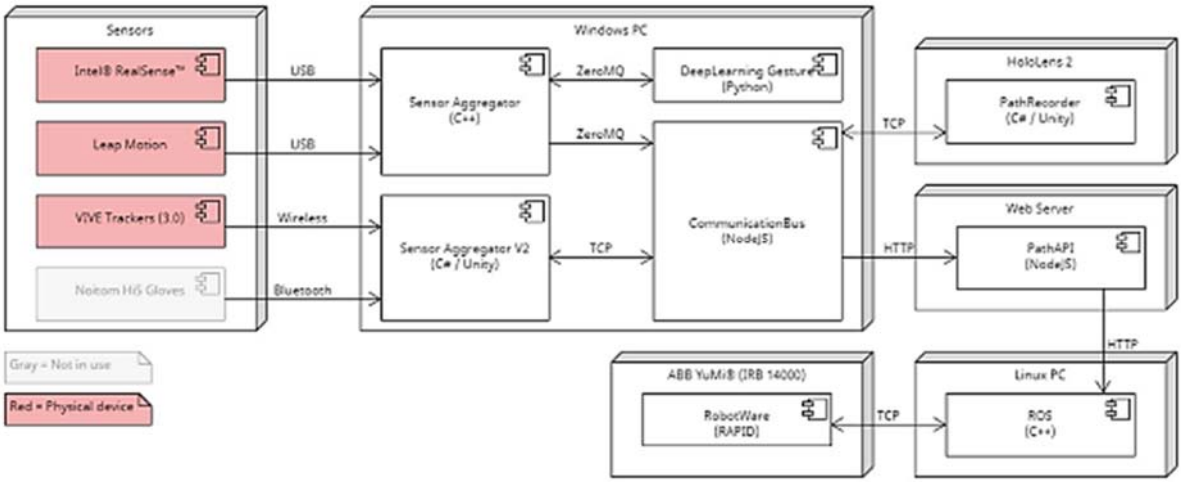


Fig. 3. System Architecture.

The accuracy of the Tracker with the pen has been measured by touching with the tip of the pen, points at a known distance drawn on a sheet of paper. The results of the positions are summarized below in Fig. 4.

The use case for our demonstration is a "gluing task", the operator picks an item and moves it, before applying glue on it and then gluing the other half. Fig. 5 shows the operator during the recording of the trajectory, which is displayed in orange. While this example shows applying glue on a flat paper, the path recorded is three dimensional.

For this use case, an ABB YuMi robot has been used. This robot has two arms. The operator can decide which arm is used in teaching mode, by touching the arm of the robot with the pen. Technically, two Trackers could be used to record trajectories for both arms at the same time.

x	y	z	x	y	z	x	y	z
Expected			Measured			Absolute error		
0	0	0	1	1	1	1	1	1
0	0	53	0	0	54	0	0	1
0	0	143	3	8	145	3	8	2
(More lines truncated...)								
HTC Tracker			Average :			1.875 [mm]		
			Max :			8 [mm]		

Fig. 4. HTC VIVE Tracker measured accuracy.

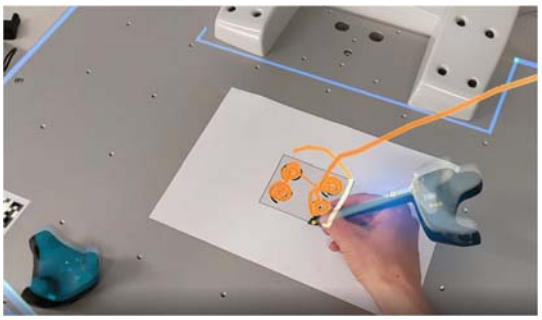


Fig. 5. Recorded path to apply glue on a paper.

5. ROS

The goal is to be able to reproduce the trajectory of the tool that has been recorded with a robot. In our case, we use the ABB YuMi robot . However, the idea is to be able to reproduce the trajectory on any robot. For this reason the framework ROS and in particular its trajectory generator tool MoveIt have been used. This solution is thus robot brand agnostic.

As input, the points recorded, the robot configuration (number of joints, geometry, kinematic limits, …) and the 3D environment have to be given to the tool. As output, we get a trajectory that can be executed by the robot following the path recorded and avoiding collisions with the environment.

Fig. 6 shows MoveIt generating a trajectory from the points recorded in the previous step.

5. Conclusion

New solutions in the robotics field are emerging with new sensors like the HTC Vive Trackers as we could see in the description of our experience. In order to have an industrially exploitable solution we propose to work on the tracking accuracy by exploring methods [5] to improve it.

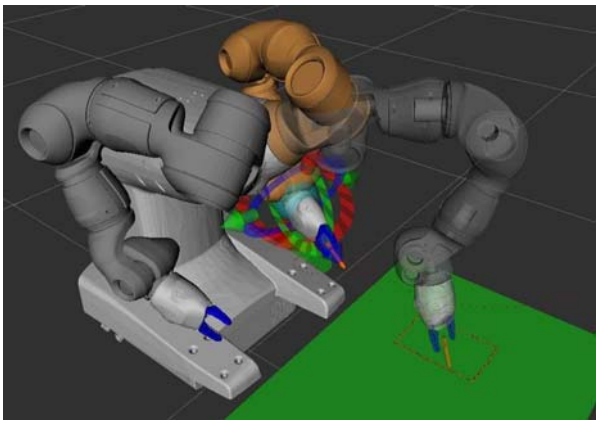


Fig. 6. MoveIt generating a trajectory.

References

[1]. Ravichandar, Harish, et al., Recent advances in robot learning from demonstration, *Annual Review of Control, Robotics, and Autonomous Systems*, 3, 2020, pp. 297-330.

[2]. Ovur, Salih Ertug, et al., Novel adaptive sensor fusion methodology for hand pose estimation with multileap motion, *IEEE Transactions on Instrumentation and Measurement,* 70, 2021, pp. 1-8.

[3]. Soares, Inês, et al. Accuracy and repeatability tests on HoloLens 2 and HTC Vive, *Multimodal Technologies and Interaction,* 5, 8, 2021, pp. 47.

[4]. Makhataeva, Zhanat, and Huseyin Atakan Varol, Augmented reality for robotics: A review, *Robotics,* 9, 2, 2020, pp. 21.

[5]. Borges, Miguel, et al., HTC vive: Analysis and accuracy improvement, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS' 2018)*, 2018.


9 788409 475636