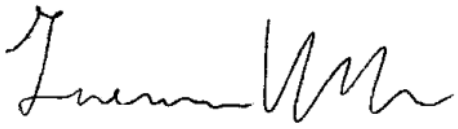
PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

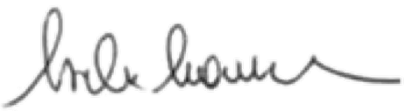
PhD Student: Francesco Vitale

Cycle: XXXVII

Training and Research Activities Report

Year: First

Student signature: 



Tutor: Prof. Nicola Mazzocca *Tutor signature:*

Co-Tutor: Ing. Roberto Nappi

Date: October 31, 2022

1. Information:

- **PhD student: Francesco Vitale**
- **DR number: DR995865**
- **Date of birth: 30/06/1997**
- **Master Science degree: Computer Engineering University: Università degli Studi di Napoli Federico II**
- **Doctoral Cycle: XXXVII**
- **Scholarship type: Funding company**
- **Tutor: Prof. Nicola Mazzocca**
- **Co-tutor: Ing. Roberto Nappi**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Modeling Complex Systems	Course	44	6	08/11/21 – 17/12/21	Prof. Mario Di Bernardo	Y
Real-Time Industrial Systems	Course	48	6	28/09/21 – 16/12/21	Prof. Marcello Cinque	Y
Neural Networks and Deep Learning	Course	62.5	6	11/01/22 – 06/04/22	Prof. Giorgio Buttazzo	N
Virtualization technologies and their applications	Course	20	5	17/01/22 – 18/02/22	Dr. Luigi De Simone	Y
Statistical Data Analysis for Science and Engineering	Course	12	4	22/03/22 – 07/04/22	Prof. Roberto Pietrantuono	Y
Imprenditorialità Accademica	Course	8	4	26/05/22 – 20/06/22	Prof. Pierluigi Rippa	Y
Big Data Analytics and Architectures	Course	18	5	06/04/22 – 28/06/22	Prof. Giancarlo Sperlì	Y
Hitachi Rail Workshop	Seminar	8	1.6	17/11/21	Prof. Nicola Mazzocca	Y
Complexity And The City: Transitioning Towards The Smart Cities Of The Future	Seminar	1	0.2	23/11/21	Prof. Mario Di Bernardo	Y

Introduction to Intellectual Property Management	**Seminar**	**2**	**0.4**	**19/07/22**	**5G Academy**	**Y**
Privacy-Preserving Machine Learning	**Seminar**	**2**	**0.4**	**14/10/22**	**Prof. Simon Pietro Romano**	**Y**
Presentation of a seminar in one of "Computer Systems Design" course lectures	**Tutorship**	**2**	**0.4**	**07/06/22**	**Prof. Nicola Mazzocca**	**N**
Study on the state-of-the-art about process mining and machine learning	**Research**	**30**	**6**	**-**	**-**	**N**
Study on the state-of-the-art about anomaly detection	**Research**	**30**	**6**	**-**	**-**	**N**
Study on the state-of-the-art about virtualization	**Research**	**15**	**3**	**-**	**-**	**N**
Application of process mining and machine learning for anomaly detection in industrial case studies and public datasets	**Research**	**45**	**9**	**-**	**-**	**N**
Preparation of the paper "Online Anomaly Detection through Conformance Checking: Methodology and Proof-of-Concept"	**Research**	**15**	**3**	**-**	**-**	**N**
Publication of the paper "Using Log Analytics and Process Mining to enable Self-Healing in the Internet of Things"	**Research**	**15**	**3**	**17/05/22**	**-**	**N**
Preparation of the paper "A Methodology for Anomaly Detection in IIoT Applications based on Machine Learning and Process Mining"	**Research**	**30**	**6**	**-**	**-**	**N**

Preparation of the paper "A Digital Twin Architecture for Anomaly Detection in the Industrial Internet of Things"	**Research**	**30**	**6**	**-**	**-**	**N**
Preparation of the paper "Evaluating Virtualization for Fog Monitoring of Real-time Applications in Mixed-Criticality Systems"	**Research**	**30**	**3**	**-**	**-**	**N**
Participation to the workshop "National Workshop for Technology Transfer and Higher Education"	**Research**	**14**	**2.8**	**16-17/06/22**	**Università di Verona**	**N**
Participation to the workshop "7th Italian Workshop on Embedded Systems"	**Research**	**14**	**2.8**	**23-24/09/22**	**Politecnico di Bari**	**N**
Held a seminar entitled "Process Mining use cases for Cyber-Physical Systems"	**Research**	**2**	**0.4**	**10/10/22**	**Università Campus Bio-Medico di Roma**	**N**

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	**0**	**1.8**	**6**	**0**	**7.8**
Bimonth 2	**6**	**0**	**5**	**0**	**11**
Bimonth 3	**11**	**0**	**10**	**0**	**21**
Bimonth 4	**4**	**0**	**12**	**0.4**	**16.4**
Bimonth 5	**9**	**0.4**	**6**	**0**	**15.4**
Bimonth 6	**0**	**0.4**	**12**	**0**	**12.4**
Total	**30**	**2.6**	**51**	**0.4**	**84**
Expected	**30 - 70**	**10 - 30**	**80 - 140**	**0 – 4.8**	

3. Research activity:

My research activity during the past year has first and foremost dealt with studying the state-of-the-art about key topics to pursue my PhD end goal, that is the application of data-driven techniques for anomaly detection in industry out of field-collected data to innovate industry practices and prove the feasibility and effectiveness of anomaly detection techniques based on process mining for Industrial Internet of Things (IIoT) systems.

The main topics I have studied are: machine learning; process mining; and anomaly detection. As these topics are often intertwined with other research areas, I have also got in touch with modern computing paradigms (edge/cloud computing paradigms), Industry 4.0 applications (smart cities, smart transportation systems, etc.), embedded systems technology (MultiProcessor System-on-Chips, FPGAs, etc.), and virtualization technology (hypervisors, containerization, etc.). Moreover, I also have studied research work about and standards from the railway domain, as my PhD is funded by Hitachi Rail S.T.S., a company that develops software and hardware solutions for service provision in railway systems.

In order to prove the feasibility and effectiveness of process mining for anomaly detection in IIoT systems, I have researched in collaboration with Hitachi and academia to:

1. Survey the state-of-the-art attempts in using process mining in IIoT systems for anomaly detection
2. Develop methodologies and prototypes integrating machine learning and process mining for anomaly detection
3. Assess performance results linked to the deployment of process mining-based anomaly detection techniques with respect to timing and detection quality

Concerning point 1., I have prepared and published a paper entitled “Using Log Analytics and Process Mining to enable Self-Healing in the Internet of Things”, which: surveys the state-of-the-art machine learning techniques to detect, manage, and solve anomalies found in log files monitored from IoT systems; surveys process mining main use-cases (i.e., process discovery and conformance checking) and their use in IoT systems to analyze log files; and experiments conformance checking techniques to check whether they are able to classify control-flow anomalies when anomalous process instances linked to normative Petri nets are analyzed. Point 2. and 3. have been addressed by the body of work I have developed in the context of several papers and projects I have been, and still am, working on. I will here detail two of the projects I’ve been following until now, as they well represent the effort so far spent throughout this first year. The first of them is the “Catenary Inspection System” (CIS) project Hitachi is working on. My main duty in this project is developing an anomaly detection technique whose primary goal is diagnosing a train’s conditions throughout its journey by analyzing time-series data. Until now, a prototype dealing with this goal has been developed, whose main stages are termed “training” and “monitoring”: the former characterizes normal behavior using process discovery by analyzing time-series data collected as the train runs in controlled conditions, whereas the latter checks behavior compliance to normal behavior through conformance checking, thus providing diagnoses about the train’s state. The second project is a research work developed in collaboration with the University of Messina that is entitled “A Methodology for Anomaly Detection in IIoT Applications based on Machine Learning and Process Mining”. In this research work me and my colleagues have developed a methodology integrating machine learning and process mining techniques to improve results that would be obtained by the plain application of process mining. Indeed, results have proven that data

preprocessing through unsupervised machine learning techniques, such as autoencoders, improve the detection accuracy, precision, and recall, outlining a promising step forward in studying the feasibility and effectiveness of process mining for anomaly detection. Other projects that are worth mentioning are: the research work entitled "A Digital Twin Architecture for Anomaly Detection in the Industrial Internet of Things", whose goal is proving the use of process mining in IIoT by means of Digital Twin technology for anomaly detection brings promising results, both in terms of detection quality and self-adaptation capability; the research work that I'm carrying out in collaboration with a Master's student that deals with the use of deep learning to classify time-series data by handling raw time-series by means of process mining techniques; and the research work I am carrying out with other colleagues of the DESSERT research group, whose main goal is studying the feasibility and effectiveness of process mining-based monitoring in virtualized embedded systems.

The main technologies I had the opportunity to learn and experiment with are: the python programming language; several process mining tools (e.g., ProM); and embedded hypervisors (e.g., Xen).

4. Research products:

Journal papers:

- P. Singh, M. Saman Azari, F. Vitale *et al.*, "Using log analytics and process mining to enable self-healing in the Internet of Things" *Environ Syst Decis* 42, 234–250 (2022). https://doi.org/10.1007/s10669-022-09859-x
- A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, F. Vitale, "A Digital Twin Architecture for Anomaly Detection in the Industrial Internet of Things" (submitted to IEEE Transactions on Industrial Informatics, currently under revision)
- F.Vitale, F. Flammini, M. Caporuscio, N. Mazzocca, "Online Anomaly Detection through Conformance Checking: Methodology and Proof-of-Concept" (submitted to IEEE Transactions on Dependable and Secure Computing, currently under revision)

Preprints:

- D. Bruneo, F. De Vita, N. Mazzocca, F. Vitale, "A Methodology for Anomaly Detection in IIoT Applications based on Machine Learning and Process Mining"
- M. Cinque, L. De Simone, N. Mazzocca, D. Ottaviano, F. Vitale, "Evaluating virtualization for fog monitoring of real-time applications in Mixed-Criticality Systems"

Prototypes:

- A prototype for the CIS project to diagnose the train pantograph's behavior throughout its journey (the source code cannot be disclosed)
- A Java open-source implementation of the token replay conformance checking algorithm providing control-flow and performance diagnostics[1]

[1] https://github.com/francescovitale/TokenReplay

5. Conferences and seminars attended

Conferences:

- *National Workshop for Technology Transfer and Higher Education*, Ordine degli Ingegneri di Verona e Provincia, 16-17/06/22; presented the poster "Anomaly Detection in IIoT through machine learning and process mining"
- *7th Italian Workshop on Embedded Systems (IWES '22),* Politecnico di Bari, 23-24/09/22; presented the technical contribution "Characterizing monitoring solutions for real-time embedded applications using virtualization"

Seminars:

- *Process Mining use cases for Cyber-Physical Systems*, Università Campus Bio-Medico di Roma, 10/10/22; I held the seminar

6. Activity abroad:

I have attended no activities abroad this year.

7. Tutorship

I conducted a seminar entitled "Virtualizing CPUs: Intel x86-64 architectures" in the "Computer Systems Design" course held by Prof. Nicola Mazzocca