PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Federico Gargiulo

XXIX Cycle

Training and Research Activities Report – First Year

Tutor: Prof. Pasquale Arpaia – co-Tutor: Prof. Nicola Mazzocca

1. Information

PhD student: Eng. Federico Gargiulo, Master Degree in Computer Engineering at University of Naples “Federico II”. PhD student in XXIX Cycle- ITEE – Università di Napoli Federico II
Tutor(s) Prof. Pasquale Arpaia et Prof. Nicola Mazzocca

2. Study and Training activities

Courses:

Training on Cumminication to become a CERN Guide, Communication Training for Experienced CERN Guide, Self-Rescue Mask, SuperComm French Language Course, Inverted CERN School of Computing

Seminars:

i. Open Source and Containers: A Proverbial Chicken and Egg?
ii. VyPR: a framework for automated performance analysis of Python programs
iii. How we broke the world record for computing digits of Pi
iv. Scalable Applications
v. Rootless containers with Podman and fuse-overlayfs
vi. Speeding up Scientific Codes in HPC Architecuytures by Code Modernization
vii. Deep Dive into the Kubecon Higgs Analysis Demo
viii. Mathematics, Cryptography, Blockchains and Cryptocurrencies: Myths and Realities
ix. Eos Access Patterns
x. Twizzler: Data-Centric Oses for a non-volatile memory future
xi. Generating realistic synthetic workloads using Generative Adversarial Networks

3. Research activity

Fualt detection and prediction.

In-depth study and research on issues relevant to prediction and failure analysis, through supervised and unsupervised machine learning techniques. The research activities were in the field of fault detection and fault prediction.

For the first part of the year, the work focused on the development of fault detection methodologies for electromechanical fluid systems.

The methodology is based on the use of hidden Markovian chains. It is a methodology suitable for all those contexts for which there is reason to believe that the system works in nominal states but there is no possibility of identifying them exactly. For example, it can happen that some devices are inserted in a system composed of a plurality of devices that can work in states-based dynamics. This affects the device in question but is not clearly identifiable.
The proposed methodology makes use of automatic learning techniques for the production of an identification system model based on ergodic Markov chains. Finally it is possible to identify operating anomalies that lead the system to work outside the nominal conditions.
The meotodology has been tested and validated on some electromechanical devices for fluid treatment and has brought excellent results. The methodology studied and developed is highly

generic and easily portable to various systems. This work was carried out in collaboration among the University of Naples Federico II, the University of Sannio, the University of Amiens and CERN in Geneva.
Finally the work was published in Measures - Elsevier magazine.

Having successfully completed the first type of work in fault prediction by means of machine learning methodologies, from the second half of the first year of PhD, the research work focused on another fault detection and prediction problem . Fault detection and prediction in electromechanical data storage devices. Magnetic storage devices have electronic and mechanical components. Their combination makes it particularly difficult to be able to predict failures and therefore operate data recovery actions in time. Currently it is preferred to take advantage of redundancy and fault tolerance techniques. The number of disks is increased, data is replicated, checksum mechanisms, etc. are used. The development of predictive techniques has always been difficult due to the enormous development of new products on the market. A few years ago a standard called S.M.A.R.T. The smart standard makes it possible to have a series of formalized information on the hard disks and the development of new models is decidedly slowed down in recent years. This slowdown has therefore made it necessary to study prediction techniques because the disks are replaced by failures and no longer because they are overtaken by new ones that are more efficient. Currently the research study is focused on the study of Regularized Greedy Forest for fault prediction.
Currently for some models of magnetic hard disk we are able to foresee well 80% of the cases of fault. The methodology needs to be explored further so that it can be generalized but for now the results bode well for the future. The current development has encountered problems due to a massive relocation of Hard Disks to the datacenters. This has caused inconsistencies in the measures that will need to be addressed in the near future. For now, the methodology has only been tested on models for which sufficient measures are available and consistent with the necessary context conditions.
The work has a strategic importance as in the world there are few datacenters that make the measurement data available in sufficient quantities. The reasons for this aura of secrecy are many and often of a commercial nature. But data centers supplied to public bodies are many and often public bodies do not have the possibility to develop a fault prediction system indoors. The current research project aims to develop a ready-to-use methodology for all the computing centers that need it.
This work is in collaboration between the University of Naples Federico II and CERN of Geneva.

4. Products

a. Publications
   i. “Fault Detection on Fluid Machinery using Hidden Markov Models” Measurement Elsevier Journal MEAS-D-19-01806
   ii. “Prediction based on S.M.A.R.T. attributes of Hard Disk Faults in Storage Data Center” in preparation, papare in preparation to be submitted.

5. Activity abroad

a. CERN project associate in the framework of Failure Detection project between CERN and DII Unina

Student: Federico Gargiulo federoco.gargiulo@unina.it

Tutor: Pasquale Arpaia pasquale.arpaia@unina.it

Cycle XXIX

	Credits year 1								Credits year 2								Credits year 3									
	Estimated	bimonth 1	bimonth 2	bimonth 3	bimonth 4	bimonth 5	bimonth 6	Summary	Estimated	bimonth 1	bimonth 2	bimonth 3	bimonth 4	bimonth 5	bimonth 6	Summary	Estimated	bimonth 1	bimonth 2	bimonth 3	bimonth 4	bimonth 5	bimonth 6	Summary	Total	Check
Modules	18	0	1	4.2	0	1	0	**6.2**	9							**0**								**0**	**6.2**	**30-70**
Seminars	13	0	0	0	4.4	0.4	1.8	**6.6**	6	0	0		0		0	**0**								**0**	**6.6**	**10-30**
Research	34	6	7	8	9	9	9	**48**	42	0	0	0	0	0	0	**0**								**0**	**48**	**80-140**
	65	6	8	12.2	13.4	10.4	10.8	**60.8**	57	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	**0**	**60.8**	**180**