PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Luigi Gallo

XXXIV Cycle

Training and Research Activities Report – First Year

Tutor: Prof. Alessio Botta

Add the following items according to the meeting we had today.

Concerning the structure of the document, use the Section number as is. Use the sub-contents indicated with a letter only as a suggestion for your content (a free form text is preferable)

1. Information
    a. Name Surname, MS title – University
    b. XXIX Cycle- ITEE – Università di Napoli Federico II
    c. Fellowship type
    d. Tutor
2. Study and Training activities
    a. Courses
    b. Seminars
    c. External courses
3. Research activity
    a. Title
    b. Study
    c. Research description
    d. Collaborations
4. Products
    a. Publications
        i. Books, Book Chapters, Journal papers, Conference papers (mark international products)
        ii. List those in preparation
    b. Patents
5. Conferences and Seminars
    a. Details (Conference name, place, dates, number of papers)
    b. Presentations made
6. Activity abroad
    a. Details (Place, dates, number of papers, contact persons)
7. Tutorship
    a. Type, subjects, hours

1 – INFORMATION

Luigi Gallo

Computer Engineering

Università di Napoli Federico II

PhD in Information Technology and Electrical Engineering

Università di Napoli Federico II

No Fellowship

Tutor: Prof Alessio Botta

2 – STUDY AND TRAINING ACTIVITIES

The study and training part includes the following short courses and seminars.

- "How to publish a scientific paper" Springer at Federico II
- "Ciberconflitti: sicurezza informatica, difesa, stabilità internazionale e diritto umanitario" at Federico II
- "L'accademia delle Startup – Le Startup dell'Accademia" at Federico II
- "Parallel and distributed computing with MATLAB" Ing. Stefano Marrone at Federico II
- "MATLAB and Embedded Systems" Ing. Stefano Marrone at Federico II

Moreover, I attended the following internal courses and doctoral schools.

- "Big Data" PhD ad hoc Module, Prof. Picariello, Federico II
- "Lipari PhD Summer School on Network and Computer Science" J.T. Schwartz International School for Scientific Research, in Lipari
- "Machine Learning and Cyber Security Phd School" at University of Padua
- "AI: towards a critical utopia" Nexa PhD School 2019, at Politecnico di Torino

Finally, I attended the following courses and their final examination is scheduled for the next few weeks. (the credits for these courses have NOT been included in the calculation)

- "Data Mining", Prof. Carlo Sansone, MS Course
- "Cloud e Datacenter Networking", Prof Roberto Canonico, MS Course

	Credits year 1								Credits year 2								Credits year 3									
		1	2	3	4	5	6			1	2	3	4	5	6			1	2	3	4	5	6			
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Total	Check
Modules	20	1,6	0	3	0	6	5	16	14							0								0	16	30-70
Seminars	5	0,4	0	0,4	1,5	0	0	2,3	6							0								0	2,3	10-30
Research	35	8	10	6,6	8,5	4	5	42	40							0								0	42	80-140
	60	10	10	10	10	10	10	60	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	180

3 – RESEARCH ACTIVITY

Analysis and identification of cyber threats and frauds in the context of large companies

Internet is revolutionizing our society and the reduction of network access costs and the development of ultra broadband will lead to a further growth of cyberspace, making it an increasingly crucial factor for economic and social growth. Increasing dependence on cyberspace and advancing new technologies (for example, the Internet of Things), on the one hand, offer new opportunities and, on the other hand, introduce new threats and vulnerabilities, for which Cyber Security is considered the second emergency in Europe, after climate change and before immigration.

Big Data and Machine Learning technologies are of primary importance in order to dominate the problem, and they have been at the centre of the study and training activities for the whole of the first year.

In this research project the goal is to contribute to the fulfilment of the necessary study and experimentations, for the design, development and testing, on real contexts, of the defense systems of the future. The "real context" is made available

by the collaboration with the Cyber Security Lab of TIM S.p.A (Telecom Italia Lab, Turin), which provides real data and environments (in compliance with current regulations on privacy).

The scientific starting point are the results provided by current Firewalls, Antiviruses, network monitoring tools etc. mostly based on heuristics, signatures and patterns recognized in network traffic and device logs over the last decade. The scientific results obtained from the first Cyber Security applications of Artificial Intelligence and Big Data technologies are also being used on the market. In a very dynamic and complex area such as Cyber Security, with a huge expected increase in the amount of data to be monitored, the observability of internal processes and the consequent reliability of the results that generate these approaches, are still experimental.

The expected scientific results concern the application of Machine Learning and Big Data to the Cyber Security context, with specific focus on approaches, models, techniques, and tools for detecting anomalies in network traffic, for detecting high-risk spam campaigns, for classifying new types of malware and vulnerabilities. It is interesting in a real context to use XAI approaches to obtain observable results that are accepted and understood by human operators (SOC analysts) and generalizable to real industrial contexts.

The research activity focused on the continuation of those begun earlier during the Master's thesis (Anomaly Detection, also considering novel Architectures for Robotics applications). The main idea is to use Big Data Analytics (Hadoop, Spark) to analyze network traffic to detect attacks (e.g. DDoS, NetScan), making it possible to perform these kinds of analysis without sampling the traffic (loss of information).

Moreover, thanks to the collaboration with TIM, the research activity was also focused on the mitigation of the problem of threats through email, which is still one of the most used channels to execute (start) a cyber attack in business contexts. The loss of money for this type of fraud is considerable (FBI and Data Protection Entities raised alarms to highlight the problem).

The first big effort was to design an adequate structured data collection in the real environment: spam emails, SOC actions, interviews etc. The goal is to extract from

these datasets useful information to train automatic classifiers to detect spam emails that really can create a security incident.

We have started testing several machine learning algorithms and feature sets. The most critical procedures for obtaining the best results were data preparation (normalization) and feature selection. The research has produced interesting preliminary results, but it has to be continued with further analysis and experiments, also considering, e. g., the cognitive and psychological aspects of the average tricked recipient of a fraudulent email.

4 – PRODUCTS

Products about Anomaly Detection and Architectures for Robotics Applications:

- Conference Paper: A. Botta, L. Gallo, G. Ventre. Cloud, Fog, and Dew Robotics: architectures for next generation applications, IEEE Mobile Cloud 2019
- Poster : A. Affinito, A. Botta, L. Gallo, M. Garofalo, G. Ventre. SPADA: SPark-based AnomalyDetection Ace, ACM CoNEXT 2018
- Conference Paper (Under Review): A. Affinito, A. Botta, L. Gallo, M. Garofalo, G. Ventre. Spark-Based Port and Net Scan Detection, ACM SAC Conference 2020

Products about the Critical Spam Detection:

- Conference Paper: L. Gallo, A. Botta, G. Ventre. Identifying threats in a large company's inbox, Big-DAMA'19: ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks

Journal papers for the integration of the above mentioned conference papers are being prepared.

5 – CONFERENCE AND SEMINARS

Next December (2019), participation in ACM CoNEXT 2019 Conference and Big Dama workshop, during which the first research results will be presented, is scheduled.

The same results have already been mentioned in a short presentation at Machine Learning and Cyber Security PhD School in Padua.

6 – ACTIVITY ABROAD

Not Yet.

7 – TUTORSHIP

I have been involved as assistant to the exercises of the courses of “Fondamenti di Informatica” and “Computer Networks” (20 + 20 hours), for which I have also prepared some course materials.

In addition, I follow as co-advisor the preparation of the thesis by the MS student Alessandro Maiello.