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 – Second Year

Tutor: Prof. Alessio Botta

Template

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.

- "A Dynamic and probabilistic orienteering problem" and "Flexible two-echelon location routing for supply networks", lecturer Prof. Claudia Archetti at Federico II
- "Computational Biology: Large scale data analysis to understand the molecular bases of human diseases", lecturer Prof. Michele Ceccarelli (Virtual due to COVID-19)
- "Elettromagnetismo e Salute" lecturer Prof. Massa (Virtual due to COVID-19)
- "How to get published with the IEEE?" lectuer Dr.ssa Eszter Lukacs (Virtual due to COVID-19)
- "Metasurfaces", lecturers Michele Celebrano, Stefania D'Agostino and Carlo Forestiere (Virtual due to COVID-19)
- "IEEE Xplore - Access the eLearning Library" lecturer Eszter Lukacs (Virtual due to COVID-19)
- "Large scale training of deep neural networks", lecturer Dr. Giuseppe Fiameni (org. Stefano Marrone) (Virtual due to COVID-19)

- "SAS Analytics", lecturer dr. Cinzia GIANFIORI (org. Prof. Picariello) (Virtual due to COVID-19)
- "Planning 5G under EMF constraints: challenges and opportunities" , lecturer Prof. Luca Chiaraviglio (Virtual due to COVID-19)
- "Sensing", lecturers Maria Caterina Giordano, Chiara Novara, Emiliano Descrovi, Riccardo Sapienza (Virtual due to COVID-19)
- "Noninvasive Mapping of Electrical Properties using MRI", lecturer Prof. Riccardo Lattanzi (Virtual due to COVID-19)
- "CyberSecurity nella fabbrica digitale" lecturer Prof. Stefano Zanero (organizer Prof. Gianni Ferretti) (Virtual due to COVID-19)
- "Valutazione dei livelli di esposizione e del rispetto dei limiti Antenne e 5G" (Prof. MD Migliore, Uni Cassino e Lazio Meridionale), "Misure di segnali complessi nell'ambiente: Sistemi 5G" (Dr. D. Franci, Arpa Lazio), "Estrapolazioni su segnali 4G e 5G" (Dr. S. Adda, Arpa Piemonte, Dr. S. Pavoncelli Arpa Lazio) (Virtual due to COVID-19)

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

- "Scientific Programming and Visualization with Python" PhD ad hoc Module, Prof. Alessio Botta, Federico II
- "Virtualization technologies and their applications" PhD ad hoc Module, Dr. Luigi De Simone
- "Innovation management, entrepreneurship and intellectual property" PhD ad hoc Module at Federico II
- "Machine Learning (ML4Health)" PhD ad hoc Module, Prof. Carlo Sansone at Federico II
- Summer School (Virtual due to COVID-19) "RegML 2020 - Regularization Methods for Machine Learning ", Prof. Lorenzo Rosasco, Emanuele Rodolà, Krikamol Muandet

	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	15,6	14	0	3	0	9	2,8	0	14,8	10							0	30,4	30-70
Seminars	5	0,4	0	0,4	1,5	0	0	2,3	6	0,4	0	1,6	2,7	4	1,4	10,1	7							0	12,4	10-30
Research	35	8	10	6,6	8,5	4	5	42,1	40	9,6	7	8,4	3,3	0,7	6,1	35,1	43							0	77,2	80-140
	60	10	10	10	10	10	10	60	60	10	10	10	15	7,5	7,5	60	60	0	0	0	0	0	0	0	120	180

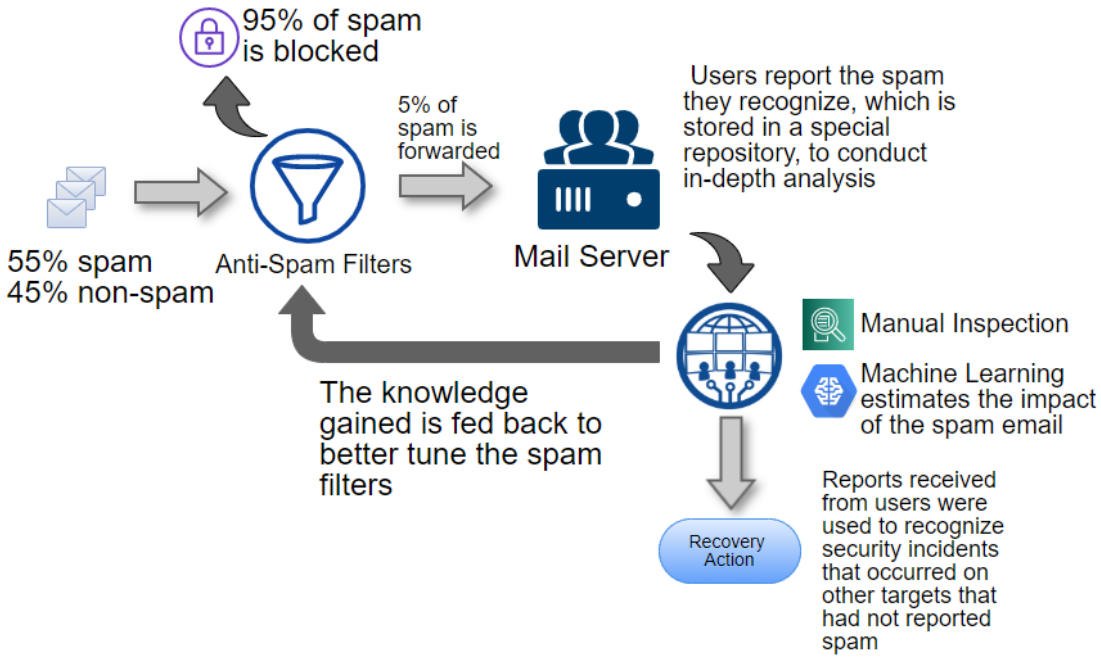
3 – RESEARCH ACTIVITY

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

My research activity concerns the study and experimentation aimed at the design, development and testing, on real contexts, of the defense systems against current and future cyber attacks. 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).

During the first year, the first step was to study the basic knowledges, open issues, and the challenges to be faced by this type of research work. During this phase I have identified a major point to focus on, in order to reach the origin of a large number of cyber attacks hurting people and companies: the identification of cyber threats in the mailboxes. In this second year this branch of research has been strongly developed (in addition to the progress of research activities dating back to the Master Degree's thesis) producing interesting results, which will be refined, validated, and expanded in the third year. To support this work I had to further deepen the studies already started in the first year, about Machine learning, Big Data Analysis and Cyber Security.

The context is the following: the email threat landscape is constantly evolving, making current countermeasures ineffective in protecting companies, especially because actually dangerous emails are able to evade carrier-grade spam filters and also deceive users. For this reason, Email is still one of the most used channels for making cyber attacks. Several law enforcement bodies (e.g. FBI, EUROPOL) and data protection agencies are constantly raising alarms in this regard, as more than 80% of financial fraud is executed by email causing huge monetary losses.
The outcome is that companies typically rely on teams of security analysts to perform manual inspection on such emails. However, spam emails that pass the spam filter check, especially in the case of large companies, are too many for such analysis to be effective. This research project aimed at providing a contribution to this important problem and leverages the collaboration with TIM S.p.A.



I designed a collaborative framework supporting security analysts of the company in collaboration, to analyze the several malicious emails that evade the spam filter and cause security incidents. Thanks to this framework we collected a large labeled dataset, composed of real spam emails received in the company, each classified as critical or not relevant. Using this labeled dataset I have shown that some machine learning algorithms, with a properly designed feature set, can well identify the emails actually threating the security of the company. I have also identified the main features that make a spam email dangerous and the best techniques and technologies to rely on for the defense. I have used both legacy and novel features and evaluated their relevance and correlation with the target. Using the best feature set maximizing the f1-score performance, the supervised approaches reaches 96.3% of precision and 93.4% of recall. I have also identified a reduced feature set that greatly reduce costs with a small impact on the performance.

During the third year the results will be validated with social experiments on the 40000+ employees of the company, in order also to include in the scientific results produced the cognitive aspect of the phenomenon of phishing and fraudulent emails, since they play an important role in this context.

Together with the main activity explained above, for a broader view of cyber security problems, I also conducted research activities on Malware Analysis and Anomaly detection in network traffic.

4 – PRODUCTS

Products of the second year:

- Conference Paper: A. Affinito, A. Botta, L. Gallo, M. Garofalo, G. Ventre. Spark-Based Port and Net Scan Detection, ACM SAC Conference 2020
- Journal Paper (Under review at Elsevier Computers and Security): Luigi Gallo, Alessandro Maiello, Alessio Botta, Giorgio Ventre. "2 Years in the anti-phishing group of a large company"

5 – CONFERENCE AND SEMINARS

Participation in ACM CoNEXT 2019 Conference (December 9-12, 2019) and Big Dama workshop, during which I presented my first research results.

Attended TMA CONFERENCE 2020 (June 10-11, 2020) "Network Traffic Measurement and Analysis Conference"

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 two MS students.