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

PhD Student: Francesco Cerasuolo

Cycle: XXXVIII

Training and Research Activities Report

Year: First

Francesco Cerasuolo

Tutor: prof. Antonio Pescapè

Date: October 18, 2023

1. Information:

- ➢ PhD student: Francesco Cerasuolo
- ➢ DR number: DR996623
- ➢ Date of birth: 22/02/1997
- ➢ Master Science degree: Computer Engineering
  University: Università degli Studi di Napoli Federico II
- ➢ Doctoral Cycle: XXXVIII
- ➢ Scholarship type: UNINA
- ➢ Tutor: Prof. Antonio Pescapè

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
On the challenges and impact of Artificial Intelligence in the Insurance domain	Course	12	3	21/11-30/11 2022	Dr. Lorenzo Ricciardi Celsi	Y
Crash Course on Data Excellence	Seminar	2	0.4	14/11/2022	Prof. G. Longo	N
Data mining the output of quantum simulator: from critical behavior to algorithmic complexity	Seminar	1	0.2	11/11/2022	Dr. P. Lucignano, Dr. D. Montemuro, Dr. D. Massarotti, Dr. V. D'Ambrosio, Dr. F. Cardano, and Dr. M. Esposito	Y
Connecting the dots: Investigating an APT campaign using Splunk	Seminar	2	0.4	11/11/2022	Proff. D. Cotroneo, S.P. Romano, R. Natella	Y
Complex Network Systems: introduction and open challenges	Seminar	1	0.2	17/11/2022	Dr. Pietro De Lellis	Y

Cybercrime and Information Warfare: National and International Actors	Seminar	2	0.4	18/11/2022	Proff. S.P. Romano, R. Natella	Y
Privacy and Data Protection	Seminar	2	0.4	22/11/2022	Proff. S.P. Romano, R. Natella	Y
Automated Offensive Security: Intelligence is all you need	Seminar	2	0.4	28/11/2022	Prof. G. Longo	N
Digital Forensics	Seminar	2	0.4	6/12/2022	Proff. S.P. Romano, R. Natella	Y
Threat Hunting & Incident Response	Seminar	2	0.4	13/12/2022	Proff. S.P. Romano, R. Natella	Y
From Cyber Situational Awareness to Adaptive Cyber Defense: Leveling the Cyber Playing Field	Seminar	2	0.4	13/12/2022	Prof. Giancarlo Sperlì	Y
IoT Data Analysis	Course	12	4	09/01 - 09/02/2023	Prof. Raffaele Della Corte	Y
Industry 4.0 Fundamentals in Bosch Applications	Seminar	10	2	23/01-26/01/2023	Ing. Martino Bruni	Y
Data Analytics	Course	30	6	21/09 - 21/12/2022	Prof. Domenico Ciuonzo	Y
MLOps: Achieving Operational Velocity with Faster Delivery and Collaboration	Seminar	1	0.2	02/03/23	Prof. Carlo Sansone and Dr. Stefano Marrone	Y

How to Publish Under the CARE-CRUI Open Access Agreement with IEEE	Seminar	1.5	0.3	06/04/2023	Prof. Nino Grizzuti	Y
Statistical data analysis for science and engineering research	Course	12	4	06/02 - 16/02/2023	Prof. R. Pietrantuono	Y
Migration of legacy IT infrastructure into the cloud: approaches and strategies	Seminar	2	0.4	23/05/2023	Prof. R. Canonico	Y
TMA Ph.D. School	Doctoral School	10	2	26/06 - 27/06/2023	Università degli Studi di Napoli Federico II	Y
Traffic Engineering with Segment Routing: optimally dealing with most popular use-cases	Seminar	1	0.2	23/06/2023	Prof. V. Persico	Y
BGP & Hot-Potato Routing: graceful and optimal convergence in case of IGP events	Seminar	1	0.2	30/06/2023	Prof. V. Persico	Y
Using Deep Learning properly	Course	10	4	10/01 - 24/01/2023	Dr. Andrea Apicella	Y

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	3	3.6	4	0	10.6
Bimonth 2	4	2	4	0	10
Bimonth 3	6	0.5	4	0	10.5
Bimonth 4	4	2.8	6	0	12.8
Bimonth 5	4	0	4	0	8
Bimonth 6	0	0	10	0	10
Total	21	8.9	31	0	61.9
Expected	30 - 70	10 - 30	80 - 140	0 – 4.8	

3. Research activity:

During my first PhD year, I conducted in-depth investigations into the motivations, uses, and hurdles related to network traffic analysis (NTA). I focus on classifying network traffic and extending these classifiers to accommodate new network traffic types using Class Incremental Learning (CIL) techniques.

NTA is a crucial process that involves the collection and analysis of network traffic data to gain insights into and enhance the performance of communication networks. In recent times, network traffic has undergone significant transformations in terms of both its composition and volume. Analyzing it now presents unprecedented challenges, rendering previously proposed methods impractical. For instance, the widespread adoption of cryptographic protocols has diminished the effectiveness of deep packet inspection, a widely used approach in the past [1].

In response to these challenges, artificial intelligence-based approaches have emerged as viable solutions for NTA problems. Among these, *Deep Learning* (DL) approaches have taken center stage as the new frontier in traffic analysis tools. A notable feature of these tools is their capability to directly process raw traffic data, making them exceptionally well-suited to handle traffic dynamics [2].

The dynamic nature of internet traffic presents a significant challenge in the field of NTA. With the continuous evolution of communication networks, the types and patterns of network traffic are constantly changing. This dynamic landscape poses several critical issues, including changing traffic composition, increasing traffic volume and velocity, and growing security threats. In this ever-changing scenario, a key issue within networks is the ability to differentiate the various types of traffic circulating on the network. Classification is essential for ensuring and managing *cybersecurity* policies, providing distinct *Quality of Service* (QoS) to different traffic flows, and optimizing network load management.

To cope with these needs, there is a growing need for CIL methodologies in the field of NTA. CIL allows DL models to evolve and adapt to changes in internet traffic patterns and composition over time. This adaptability is crucial for staying up-to-date with evolving network dynamics. CIL offers several key advantages in this context, including: (i) adaptability, enabling DL models to seamlessly incorporate new traffic classes and adapt to emerging traffic patterns without the need for retraining from scratch, (ii) efficiency, since they update models using few data, and (iii) real-time detection, by optimizing and reducing the time to deploy an up-and-running classifier.

Nonetheless, CIL poses challenges since DL models, when trained on new data, often exhibit a tendency to forget what they have previously learned, a phenomenon referred to as *catastrophic forgetting*. On the other hand, models may also struggle to include new classes, leading to what is known as *intransigence*, which is the inability to include new knowledge [3]. To overcome these challenges and enable the effectiveness of CIL approaches in network applications, several methods have been introduced in the literature to mitigate these concerns. These strategies have been combined to propose diverse approaches to tackle these issues.

First, I evaluate state-of-the-art solutions, also designed for other domains (e.g. Computer Vision), tailored to enhance the adaptability of classifiers. Then, starting from prior approaches proposed in traffic classification or adapted to solve this task [4], my research embarked on a comprehensive examination of the components within the incremental model. This endeavor encompassed the assessment of various potential implementations of each building block and the evaluation of their performance for each individual element. Through this process, a novel approach named Memento was developed synthesizing all the outcomes from the conducted analyses. This approach falls under the fine-tuning family and represents an enhancement of existing techniques for traffic classification. It was assessed using two network datasets: MIRAGE19, which consists of mobile traffic app data from the ARCLAB at the University of Naples Federico II, and CESNET-TLS22, a dataset containing traffic collected over two weeks using high-speed probes situated at the CESNET2 network's perimeter. By employing this solution, we attain performance that approaches those of the training-from-scratch model, particularly in scenarios involving a single-app increment and a significant number of app increments.

This research activity led to a journal submission [J1], made in collaboration with other members of the research group, where we propose this new fine-tuning approach outperforming state-of-the-art.

Following that, to gain a more profound comprehension of how incremental models operate and to shed light on the difference between models obtained through incremental learning and those originating training from scratch, I focused on examining models utilizing eXplainable AI (XAI) methods. From these examinations, it becomes apparent that the significance of fields and packets (fed to DL modes) undergoes transformations as a consequence of incremental learning, and the similarity of applications (intrinsic to their network traffic) can impact the performance of the incremental training.

This research activity led to a conference publication [C1], made in collaboration with other members of the research group, where we propose a new XAI methodology to analyze and explore CIL models.

Finally, my focus was directed towards the cybersecurity domain. I delved into the realm of *Intrusion Detection Systems* (IDS) and explored the potential for enhancing them with new categories of attacks (0-day attacks), i.e. novel threats that target networks, prompting network administrators to fortify their defenses. The specific area of interest revolved around investigating how CIL methodologies can bolster DL-based IDS, especially within the intricate landscape of IoT networks. We harnessed the publicly available IoT-23 dataset to enable models to seamlessly adapt to emerging attack patterns while retaining their proficiency in detecting known threats, all without the need for retraining classifiers from scratch. In this context, we undertook an assessment of several cutting-edge CIL techniques. Among these, the most promising approach was BiC [5], although it exhibited suboptimal performance. We conducted a comprehensive analysis to highlight the factors influencing this performance discrepancy and the input impacting the most on predictions. Such work has also highlighted the need for further exploration of the effective application of CIL techniques in the cybersecurity domain

This research endeavor resulted in a conference publication [C3], created in partnership with other members of the research group. In this publication, we evaluated the effectiveness of CIL methods in incorporating 0-day attacks into an IDS.

References:

[1] Dainotti, A., Pescapé, A. and Claffy, K.C., Issues and future directions in traffic classification, IEEE Network 26 (1) (2012) 35–40.

[2] Aceto, G., Ciuonzo, D., Montieri, A., & Pescapé, A. (2019). Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges. IEEE Transactions on Network and Service Management, 16(2), 445-458.

[3] Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A. D., & Van De Weijer, J. (2022). Class-incremental learning: survey and performance evaluation on image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5), 5513-5533.

[4] Bovenzi, G., Nascita, A., Yang, L., Finamore, A., Aceto, G., Ciuonzo, D., Pescapè A. and Rossi, D. (2023). Benchmarking Class Incremental Learning in Deep Learning Traffic Classification. IEEE Transactions on Network and Service Management.

[5] Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., & Fu, Y. (2019). Large scale incremental learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 374-382).

4. Research products:

Journal Papers:

[J1] *MEMENTO: A Novel Approach for Class Incremental Learning of Encrypted Traffic*, Francesco Cerasuolo, Alfredo Nascita, Giampaolo Bovenzi, Giuseppe Aceto, Domenico Ciuonzo, Antonio Pescapè, Dario Rossi*, submitted to Elsevier Computer Networks* - submitted

Conference Papers:

[C1] *Explainable Mobile Traffic Classification: the case of Incremental Learning,* Alfredo Nascita, Francesco Cerasuolo, Giuseppe Aceto, Domenico Ciuonzo, Valerio Persico, Antonio Pescapè, *submitted to International Conference on emerging Networking EXperiments and Technologies Workshop on "Explainable and Safety Bounded, Fidelitous, Machine Learning for Networking"* - submitted

[C2] *Adaptive Intrusion Detection Systems: Class Incremental Learning for IoT Emerging Threats,* Francesco Cerasuolo, Giampaolo Bovenzi, Christian Marescalco, Francesco Cirillo, Domenico Ciuonzo, Antonio Pescapè, *submitted to IEEE International Conference on Big Data Workshop "Machine Learning for Securing IoT Systems Using BigData"* - submitted

5. Conferences and seminars attended

Italian Conference on CyberSecurity (ITASEC) Conference, …, 2-5 May 2023, Bari
Presentation of the Paper*: A Comparison of Machine and Deep Learning Models for Detection and Classification of Android Malware Traffic*

Network Traffic Measurement and Analysis (TMA) PhD Scool, Unversity of Napoli Federico II, 26-27 June 2023
Presentation of the Poster*: Class Incremental Learning for Mobile Traffic Classification*

Network Traffic Measurement and Analysis (TMA) Conference, Unversity of Napoli Federico II, 28-29 June 2023

Presentation of the Poster*: Class Incremental Learning for Mobile Traffic Classification*

6. Activity abroad:

I have not carried out any activity abroad during my first Ph.D. year.

7. Tutorship

I have not carried out any tutorship during my first Ph.D. year