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

PhD Student: Alfredo Nascita

Cycle: XXXVII

Training and Research Activities Report

Academic year: 2022-23 - PhD Year: Second

Alfredo Nascita

Tutor: prof. Valerio Persico Valerio Persico

Date: October 23, 2023

1. Information:

PhD student: Alfredo Nascita
PhD Cycle: XXXVII
DR number: DR995853
Date of birth: 01/10/1994
Master Science degree: Computer Engineering
University: Università degli Studi di Napoli Federico II
Scholarship type: UNINA
Tutor: Prof. Valerio Persico

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	21-30/11/2022	DIETI ITEE PhD	Y
Data mining the output of quantum simulators - from critical behavior to algorithmic complexity	Seminar	1	0.2	11/11/2022	Proff. P. Lucignano, D. Montemurro, D. Massarotti, V. D'Ambrosio, F. Cardano and M. Esposito	Y
Cybercrime and Information Warfare: National and International Actors	Seminar	2	0.4	18/11/2022	Proff. S. P. Romano, R. Natella, DIETI - Unina	Y
Privacy and data Protection	Seminar	2	0.4	22/11/2022	Proff. S. P. Romano, R. Natella, DIETI - Unina	Y
Digital Forensics, Group-IB	Seminar	2	0.4	06/12/2022	Proff. S. P. Romano, R. Natella, DIETI - Unina	Y

From Cyber Situational Awareness to Adaptive Cyber Defense: Leveling the Cyber Playing Field	Seminar	2	0.4	13/12/2022	Prof. G. Sperlì, DIETI - Unina	Y
Threat Hunting & Incident Response	Seminar	2	0.4	13/12/2022	Proff. S. P. Romano, R. Natella, DIETI - Unina	Y
Iot Data Analysis	Course	12	4	09-13-16-20-23-27/01/2023	DIETI ITEE PhD	Y
Industry 4.0 Fundamentals in Bosch Applications	Seminar	10	2	23-26/01/2023	National Doctoral program in Autonomous Systems, in collaboration with Bosch	Y
MLOps: Achieving Operational Velocity with Faster Delivery and Collaboration	Seminar	1	0.2	02/03/2023	Prof. Carlo Sansone and Dr. Stefano Marrone, DIETI - Unina	Y
How to Publish Under the CARE-CRUI Open Access Agreement with IEEE	Seminar	1.5	0.3	05/04/2023	CARE-CRUI, IEEE	Y
Co-supervisor Master Thesis, student: Andrea D’Arco	Tutorship	8	0.32	01-04/2023	-	-
Co-supervisor Master Thesis, student: Salvatore Santella	Tutorship	8	0.32	01-04/2023	-	-
Orientamento - Porte Aperte Scuola Politecnica e delle Scienze di Base 2023	Tutorship	2	0.08	13/02/2023	-	-
Lecture on Explainable Artificial	Tutorship	2	0.08	05/2023	-	-

Intelligence for Network Traffic Analysis in the *Internet Data Analysis* course (Prof. Antonio Pescapé)						
Using Deep Learning properly	**Course**	**10**	**4**	**10-12-17-19-24/01/2023**	**DIETI** **ITEE PhD**	**Y**
Migration of legacy IT infrastructures into the cloud: approaches and strategies	**Seminar**	**2**	**0.4**	**23/05/23**	**Prof. R. Canonico, DIETI, Unina**	**Y**
Traffic Engineering with Segmented Routing: optimally addressing popular use cases	**Seminar**	**1**	**0.2**	**23/06/23**	**Prof. V. Persico, DIETI - Unina**	**Y**
TMA PhD School	**Doctoral School - Seminar**	**12**	**2**	**26-27/06/23**	**Università di Napoli Federico II**	**Y**
BGP & Hot-Potato Routing: graceful and optimal convergence in case of IGP events	**Seminar**	**1**	**0.2**	**30/06/23**	**Prof. V. Persico, DIETI, Unina**	**Y**
Laboratory activities in the "Internet Data Analysis" course (Prof. Antonio Pescapé)	**Tutorship**	**14**	**0.56**	**01/03 - 31/10/2023**	**-**	**-**
Lectures in the *Computer Networks* course (Prof. Antonio Pescapè)	**Tutorship**	**6**	**0.24**	**09-10/2023**	**-**	**-**

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	2.2	5	0	10.2
Bimonth 2	4	2	4	0.5	10.5
Bimonth 3	0	0.5	8	0.3	8.8
Bimonth 4	4	2.8	7	0.2	14
Bimonth 5	0	0	8	0.3	8.3
Bimonth 6	0	0	8	0.3	8.3
Total	11	7.5	40	1.6	60.1
Expected	30 - 70	10 - 30	80 - 140	0 – 4.8	

3. Research activity:

During my second PhD year, I continued my research activities in the field of network traffic analysis (NTA) with a particular emphasis on enhancing the explainability of AI models and tools for various tasks in this domain.

NTA is a crucial process that involves gathering and examining network data to comprehend and optimize network performance. Various practical activities are essential to this process and are becoming increasingly significant in today's complex network environments. Network traffic has undergone substantial transformations in terms of its composition and volume, presenting novel challenges that render traditional approaches no longer viable. For instance, the widespread use of cryptographic protocols hampers the efficacy of previously prevalent signature-based methods.[1].
Additionally, modern network traffic is notably diverse, encompassing different applications, services, users, and devices, and it continually evolves. These factors underscore the necessity for the development of new tools to manage traffic effectively and meet more stringent requirements.

In recent years, Artificial Intelligence (AI) solutions, particularly Deep Learning (DL) approaches, have emerged as the cutting edge of traffic analysis tools. These tools stand out for their capacity to work directly with raw traffic data and adapt to dynamic traffic conditions. However, their data-driven nature raises concerns about the interpretability of their results. To overcome such limitations, eXplainable Artificial Intelligence (XAI) has recently emerged to elucidate the decision-making processes of AI methods and provide insights into the rationale behind their decisions.

In line with these considerations, my research is centered on interpreting current NTA tools and implementing more effective and secure network tools by harnessing the insights derived from explainability analyses. The aim is to ensure that the decisions provided by AI models are well-founded and based on solid, objective evidence, fostering user trust and facilitating their adoption in real network scenarios. To attain these objectives, I have devised a theoretical framework that encompasses several key stages. Starting with network traffic data, this framework involves the training of AI-based models to address specific NTA tasks. Subsequently, in the ensuing phase, XAI techniques are applied to interpret these models from diverse perspectives (including input importance on predictions, internal operations, and components). The insights gained from this phase can then serve as input for the final stage where the AI models are potentially refined and optimized along chosen enhancement directions, ultimately yielding improved versions of the tools.

During this year, I adhered to the outlined theoretical framework to address the challenge of Traffic Classification (TC), recognized as a pivotal prerequisite for efficient network management.
This research activity led to the publication [J1], where we conducted XAI analyses to interpret and enhance a multimodal DL model (e.g., fed with raw bytes of payload data and informative fields of packet sequences) designed to address multiple TC tasks in a multitask fashion (e.g., predicting the average packet length of each flow and detecting mice/elephant flows).

In particular, we leveraged insights gained from XAI analyses to optimize a state-of-the-art classifier, named DISTILLER [2], along dimensions of reliability, performance, and feasibility.
We started with the training of the DL model, employing the public ISCX VPN-NONVPN dataset [3], which consists of human-generated traffic categorized according to three distinct TC tasks: encapsulation, traffic type, and application recognition. Then, we proceed to analyze the model through two interpretability methods, Deep SHAP [4] and Integrated Gradients [5], to investigate the importance of each modality in addressing the three tasks, as well as the importance of inputs within each single modality. Leveraging interpretability results, we reduced the input to include only the subset highlighted as important by XAI techniques, thus reducing training times by 58%, and obtaining reduced time-to-insight as fewer packets needed to be collected to obtain classification outcomes. This input reduction led also to a reduction in model size. We then analyzed the model from a reliability perspective, specifically examining the level of confidence we could place in its predictions through calibration analysis. By employing label smoothing, we halved the calibration error for all three tasks, significantly enhancing reliability without substantial performance loss. Finally, we explored three compression techniques to further reduce the model's size, with the aim of enhancing model feasibility: knowledge distillation, pruning, and quantization. Among these methods, pruning emerged as the most effective compression technique, reducing memory usage by 50% without affecting the other improvement directions.
In this way, we were able to propose a new traffic classifier, DISTILLER EVOLVED, obtained through XA-driven refinements of the original model.

Another need in Internet traffic analysis is addressing the continuous traffic growth and the management of new services and traffic categories. This translates into a strong demand for the development of adaptable classification tools capable of recognizing new traffic types. The Class Incremental Learning (CIL) paradigm responds to this need [6], extending existing classifiers with new traffic classes without forgetting the old knowledge and avoiding complete retraining, as happens in the traditional training-from-scratch approach.

Regarding this topic, we conducted a comprehensive investigation into the effectiveness of recent CIL techniques applied to Internet TC. This research, presented in publication [J3], was carried out within the framework of the Huawei Innovation Lab project on 'Network Traffic and AI-enabled Network Technologies,' a collaboration between Huawei Technologies France SASU and DIETI, University of Napoli Federico II.
This study revealed that CIL methods fall short in terms of performance w.r.t. traditional training approaches and pointed out the crucial need for conducting interpretability analyses to understand the reasons behind unsatisfactory performance and attempt to narrow the performance gap. In line with these considerations, I started investigating the interpretability of incremental traffic classifiers. In particular, in the work [C2], I introduce a methodology to grasp the differences between updated models and the model trained from scratch (scratch) with the data of all applications at once. This methodology involves studying the importance of inputs and analyzing both the base models (the models to be extended) and the models extended with new classes. In detail, from an input importance viewpoint, XAI techniques helped us understand that the two models focus on different aspects of traffic to discriminate apps. The analysis of base models' behavior highlighted some

cases where the introduction of new knowledge is more challenging. The analysis of the updated models revealed an imbalance in responses toward the new class, primarily caused by the alignment between the backbone and the classification layer.
With some of these insights in mind, we proposed a new fine-tuning approach, MEMENTO [J4], and plan to continue our investigation in order to design adaptive and interpretable traffic classifiers.

In addition to these activities, I have conducted another study and research activity with the aim of assessing the robustness of anomaly detection tools in traffic scenarios different from those in which they were trained. This is a fundamental aspect to understand how much we can trust the results provided by these models and evaluate their applicability in new network scenarios. To delve into the details, we evaluated state-of-the-art classifiers in various cross-evaluation contexts (baseline, generalization, and extension) and compared the performances obtained with different inputs: raw data or information fields extracted from packets. We also assessed the effectiveness of multimodal approaches in this context. This activity resulted in the publication [C1].

In parallel with the above-mentioned activities, two collaborations with the Ningbo University (Zhejiang, China) and the University of Salento (Lecce, Italy) have led to the papers [J3] and [J5], respectively. In detail, in the work [J3], we presented a strategy for Mobility-aware Computing Offloading and Task Migration (MCOTM) designed to enhance the average task turnaround time and reduce energy consumption in mobile devices within Industrial Internet of Things settings. In the publication [J5], we examined the practicality of an integrated perspective of IoT and Fog paradigm and explored the motivations behind researchers' investigations into IoT-Fog networks. We provided an in-depth analysis of the IoT-Fog architecture and highlighted its promising applications. We also emphasized the importance of efficiency, security, and privacy, shedding light on the underlying rationales and prerequisites for their integration.

References:

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

[2] G. Aceto, D. Ciuonzo, A. Montieri and A. Pescapé. *"DISTILLER: Encrypted traffic classification via multimodal multitask deep learning"*, Journal of Network and Computer Applications, 183, 102985, 2021.

[3] G. Draper-Gil, A. H. Lashkari, M. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," 2nd International Conference on Information Systems Security and Privacy (ICISSP), pp. 407–414, 2016.

[4] M. Sundararajan, A. Taly, and Q. Yan. *"Axiomatic attribution for deep networks"*, International Conference on Machine Learning (ICML), 2017.

[5] S. M. Lundberg and S.-I. Lee, *"A unified approach to interpreting model predictions"* in NIPS'17 Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.

[6] M. Masana, X. Liu, B. Twardowski, M. Menta, A. Bagdanov, J. Van De Weijer, *"Class-incremental learning: survey and performance evaluation on image classification"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45, pp. 5513--5533, 2022

Research products

Conference Papers:

[C1] Cross-Evaluation of Deep Learning-based Network Intrusion Detection Systems, Ciro Guida, Alfredo Nascita, Antonio Montieri, Antonio Pescapé, accepted for presentation in the 10th International Conference on Future Internet of Things and Cloud (FiCloud 2023)

[C2] Explainable Mobile Traffic Classification: the case of Incremental Learning, Alfredo Nascita, Francesco Cerasuolo, Giuseppe Aceto, Domenico Ciuonzo, Valerio Persico, Antonio Pescapé, accepted for presentation in the 19th International Conference on emerging Networking EXperiments and Technologies (CoNEXT 2023), Workshop on 'Explainable and Safety Bounded, Fidelitous, Machine Learning for Networking'

Journal Papers:

[J1] Improving Performance, Reliability, and Feasibility in Multimodal Multitask Traffic Classification with XAI, Alfredo Nascita, Antonio Montieri, Giuseppe Aceto, Domenico Ciuonzo, Valerio Persico, Antonio Pescapé. Accepted for publication in IEEE Transactions on Network and Service Management (TNSM) 2023

[J2] Benchmarking Class Incremental Learning in Deep Learning Traffic Classification, Giampaolo Bovenzi, Alfredo Nascita, Lixuan Yang, Alessandro Finamore, Giuseppe Aceto, Domenico Ciuonzo, Antonio Pescapé, Dario Rossi. Accepted for publication in IEEE Transactions on Network and Service Management (TNSM), 2023

[J3] MCOTM: Mobility-Aware Computation Offloading and Task Migration for Edge Computing in Industrial IoT Wei Qin, Haiming Chen, Lei Wang, Yinshui Xia, Alfredo Nascita, Antonio Pescapé. Accepted for publication in Elsevier Future Generation Computer Systems (FGCS) Journal

[J4] MEMENTO: A Novel Approach for Class Incremental Learning of Encrypted Traffic, Francesco Cerasuolo, Alfredo Nascita, Giampaolo Bovenzi, Giuseppe Aceto, Domenico Ciuonzo, Antonio Pescapé, Dario Rossi. Submitted to Elsevier Computer Networks

[J5] An Integration Perspective of Security, Privacy, and Resource Efficiency for IoT-Fog Networks, Saeed Javanmardi, Alfredo Nascita, Antonio Caruso, George Loukas, Antonio Pescapè. Submitted to IEEE Communications Magazine

All publication venues are indexed in Scopus and/or ISI Web of Science

5. Conferences and seminars attended

19th Italian Networking Workshop 2023 (INW2023), Ponte di Legno (BS), IT, 16 - 18 January 2023
Presentation of the Contribution: *Extending Traffic Classifiers to New Applications via Class-Incremental Learning*

Italian Conference on Cybersecurity 2023 (ITASEC2023), Bari, IT, 2-5 May, 2023
Presentation of the Article: *Machine and Deep Learning Approaches for IoT Attack Classification*

Network Traffic Measurement and Analysis (TMA) PhD School, University of Napoli Federico II, Napoli, IT, 26-27 June 2023
Presentation of the Poster: *Improving Performance, Reliability and Feasibility in Multimodal Multitask Traffic Classification with XAI*

7th edition of the Network Traffic Measurement and Analysis Conference (TMA Conference 2023), University of Napoli Federico II, Napoli, IT, 28-28 June 2023

6. Periods abroad and/or in international research institutions

I have not carried out any activity abroad during my second PhD year.

7. Tutorship

During this year, I carried out tutorship activities during *Computer Networks* and *Internet Data Analysis* courses for Master's and Bachelor's Degrees in Computer Engineering. Details on my tutorship activities are listed in the following:

- Co-supervisor of two Master's theses in Computer Engineering, *Internet Data Analysis* course (16 hours)
- Teaching activities during the *Internet Data Analysis course,* Master's Degree in Computer Engineering, Prof. Pescapé (2 hours)
- Teaching activities during the *Computer Networks* course, Bachelor's Degree in Computer Engineering, Prof. Pescapé (6 hours)
- Laboratory activities during the *Internet Data Analysis* course, Master Degree in Computer Engineering, Prof. Pescapé (14 hours)
- Support: *Porte Aperte Scuola Politecnica e delle Scienze di Base 2023* (2 hours)

8. Plan for year three

Over the next year, I intend to continue my ongoing research activities and explore new related aspects. In detail, I plan to continue my analyses for incremental scenarios, explore the explainability of other NTA problems, such as Anomaly Detection and Traffic Prediction, and assess Explainable-by-design approaches. Simultaneously, I will continue my tutoring activities for courses in the Master's Degree in Computer Engineering and the Bachelor's degree program in Computer and Mechatronic Engineering.

Additionally, I plan to embark on a research period abroad at *Huawei Technologies France* in Paris. Coordination for this internship is already underway.

Naturally, during this final year, I will allocate a significant amount of time to composing my doctoral thesis, which will be centered on the topic of eXplainable Artificial Intelligence for Internet Network Traffic Analysis.