PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Antonio Montieri

XXXII Cycle

Training and Research Activities Report – Second Year

Tutor: Prof. Antonio Pescapè

Information

I am **Antonio Montieri** and I received a M.Sc. degree cum laude in Computer Engineering from the University of Napoli Federico II in July 2015. Currently, I am a PhD Student attending the XXXII Cycle of the Information Technology and Electrical Engineering (ITEE) PhD program at the Department of Electrical Engineering and Information Technology (DIETI) of the University of Napoli Federico II. My fellowship is financed by a university ITEE grant. My tutor is prof. **Antonio Pescapè** and I am a member of the TRAFFIC research group, part of the larger COMICS, whose activities are carried out in the field of Computer Networks.

Study and Training activities

During the second year of PhD program, I attended the courses and seminars reported in the following. In June 2018, I also attended the 8th TMA PhD School on Traffic Management & Analysis at the AIT Austrian Institute of Technology in Vienna (Austria) that has been included in the list of seminars.

Courses

1. *Tecnologie Digitali e Scienze Umane*, ad hoc module, Prof. Guglielmo Tamburrini, 12/01/2018; 01/02/2018; 23/02/2018; 16/03/2018; 13/04/2018; 11/05/2018, 3 Credits.
2. *Morphic Sensing*, ad hoc module, Prof. Gaetano D. Gargiulo, 04/07/2018 – 05/07/2018, 2.4 Credits.
3. *Author Seminar: How to Publish a Scientific Paper*, ad hoc module, Dr. Aliaksandr Birvkov; Dr. Elisa Magistrelli, 26/11/2018, 0.4 Credits.
4. *Ciberconflitti*, ad hoc module, Dr. Gian Piero Siroli; Dr. Francesco Vestito; Prof. Simon Pietro Romano; Prof. Daniele Amoroso, 28/11/2018, 0.8 Credits.

Seminars

1. *Logic-based Languages and Systems for Big Data Applications*, Prof. Carlo Zaniolo, 13/03/2018; 15/03/2018, 0.8 Credits.
2. *IBM Q: Building the First Universal Quantum Computer for Business and Science*, Dr. Federico Mattei; Dr. Najla Said, 16/05/2018, 0.4 Credits.
3. *How does MathWorks accelerate the pace of engineering and science?*, Dr. Francesco Alderisio, 01/06/2018, 0.2 Credits.
4. *THE NAPOLI FEDERICO II IEEE STUDENT BRANCH*, Ing. Stefano Marrone, 17/07/2018, 0.2 Credits.
5. *Parallel and Distributed Computing with MATLAB*, Ing. Stefano Marrone, 21/11/2018, 0.4 Credits.
6. *PhD School: 8th TMA PhD School on Traffic Management & Analysis*, Frank Brockners; Joseph Allemandou; Idilio Drago, 25/06/2018 – 26/06/2018 (35 hours), 3.5 Credits.

Credits Summary

Finally, I provide a table reporting a summary of the credits obtained attending modules and seminars and doing research activities.

	Credits year 1								Credits year 2							
	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
Modules	20	4	3	0	6	4	6	23	10	0	3	2,4	0	1,2	0	6,6
Seminars	10	4,9	1,8	2	0,5	0,3	0,3	9,8	5	0,8	0,4	3,9	0	0,4	0	5,5
Research	30	1,1	5,2	8	3,5	5,7	6,7	30,2	45	9,2	6,6	3,7	10	8,4	10	47,9
	60	10	10	10	10	10	13	63	60	10	10	10	10	10	10	60

Research Activity

Techniques for mobile and encrypted traffic classification

In my PhD, I am working in the context of monitoring and management of computer networks, with specific focus on mobile and encrypted traffic classification. Specifically, with Traffic Classification (TC) we refer to the process of associating network traffic with specific applications generating it. This research activity aims at shed light on the changing nature of the traffic generated by smartphones (and other handled devices) whose deep usage in everyday life is growing more and more. Indeed, various tools and middle-boxes, such as security/quality-of-service enforcement devices and network monitors, rely on the knowledge of the application generating the traffic and thus are limited (or impaired) when this requirement is not completely satisfied.

During my second year of PhD, I have worked (and I am currently working) on Deep Learning (DL)-based TC techniques, which allows to train classifiers directly from input data by automatically learning structured (and complex) feature representations, overcoming the limitations of "traditional" Machine Learning (ML) classifiers based on handcrafted (domain-expert driven) features.

Moreover, I have also tried to apply hierarchical ML-based classification scheme to the classification at different levels of anonymous traffic (i.e. the traffic generated by means of anonymity tools such as Tor, I2P, and JonDonym). Hierarchical classification allows fine-grained tuning and design of classifiers, potentially leading to performance gains, and it also brings a number of "practical" benefits by design, at cost of moderate complexity increase.

Data Rate Coverage along Transport Routes in Mobility

This research activity is made in collaboration with the Simula Research Laboratory, Oslo (Norway) and it concerns the analysis and characterization of the Quality of Service and Quality of Experience parameters of communication networks in high-mobility scenarios. We are working on this topic with the aim of paper submission in Q2 2019 [C2].

Research Description

The efficacy of security and quality-of-service enforcement devices, as well as network monitors, is limited (or qualitatively hampered) when there is no accurate knowledge of the application generating the traffic. The process inferring such information, known as network Traffic Classification (TC), has a long-standing application in many fields [1] and is facing unprecedented challenges due to the users' massive shift toward mobile devices (as witnessed by recent Internet traffic evaluations [2]), leading to a multi-faceted and evolving composition of network traffic [3].

Hence, the appeal of mobile TC has bloomed nowadays, nurtured (other than usual TC drivers, e.g. service differentiation) by valuable profiling information (e.g., to advertisers, security agencies, and insurance companies), while also implying privacy downsides (e.g., recognition of context-sensitive applications, such as dating and health ones, and bring-your-own-devices policies). Equally important, the effort towards the protection of privacy and security has fueled the widespread adoption [4, 5] of encrypted protocols (TLS). This shift, together with the use of dynamic transport ports or the clustering on a few well-known (and commonly unblocked) ports, resulted in the hampering of accurate TC, as both Packet Inspection (DPI) and port-based techniques become ineffective [1]. Moreover, other than the ET issue, mobile TC comes with exacerbated challenges and requirements due to a large number of apps to discriminate from and the automatic frequent updates of the apps—leading to inadequate number of training samples per app and hindering the achievement of targeted performance.

In this context, ML classifiers have proved to be a good fit, since they suit also Encrypted Traffic (ET) while not expressly relying on port information [6, 7, 8]. However, their usual form resorts to the process of obtaining handcrafted (domain-expert driven) features (e.g., packet sequence statistics), which is time-consuming, unsuited to automation, and it is becoming rapidly outdated when compared to the evolution and mix of mobile traffic, being a constantly moving target, and precluding the design of accurate and up-to-date mobile-traffic classifiers [8, 9] with "traditional" ML approaches. Therefore, DL is emerging as the stepping stone toward the fulfillment of high performance in the dynamic and challenging (encrypted) TC contexts, allowing to train classifiers directly from input data by automatically distilling structured (and complex) feature representations [10].

This activity has led to the publication of one conference paper [C1] and one journal paper [J2], the latter currently under third round of review, both made in collaboration with other members of the research group. Specifically, in these works we proposes the design of mobile traffic classifiers (able to operate with ET) via the adoption of DL umbrella. To this end, this work resorts on the development of a systematic framework for the design of novel DL-based TC architectures and comparison of existing ones. We applied this framework for a critical analysis of several non-mobile-specific DL classifiers recently appeared in TC literature [11, 12, 13, 14, 15, 16]. In detail, the proposed framework dissects the DL-based TC problem from different viewpoints: (A) the TC object adopted, (B) the type (and the amount) of input data fed to the DL classifier, (C) the DL architecture employed, and (D) the required set of performance measures for an objective and comprehensive evaluation. The outcomes of this work underline the deficiencies of current DL-based traffic classifiers and the need for: (i) unbiased, informative, and heterogeneous inputs extrapolated from traffic data, (ii) sophisticated DL architectures, and (iii) a rigorous and multifaceted performance evaluation. Indeed, this study represents a first attempt to address (i) and (ii) issues, being also a “safe” groundwork for paving the way to the design of accurate DL-based classifiers coping with highly-diverse mobile traffic, whereas it provides designers with a fine-level performance evaluation workbench (iii).

However, most of these DL-based efforts have focused on one type of input information (e.g., payload bytes or header fields), despite traffic data being naturally “multimodal”. Indeed, the main asset of multimodal DL is the ability to automatically learn a hierarchical representation exploiting jointly all the available modalities, instead of handcrafting modality-specific features for a given ML approach [17]. For such reasons, in [J3] (currently under first round of review), we propose to address the above challenging issue via a novel MultImodal DL-based MobilE TraffIc Classification (MIMETIC) framework, having the capability of exploiting effectively the heterogeneous nature of the different views of a TC object, by capturing both intra- and inter-modalities dependence. Results, based on three datasets collected by human users, highlight a performance improvement (in terms of both concise and fine-grained measures) while reporting a lower training time (more than three times) with respect to existing (single-modality) DL-based traffic classifiers.

Starting form the expertise gained with the above-mentioned studies, in [J5] we first give an overview of the key network traffic analysis problems where DL is foreseen as attractive, due to their common capitalization of network-level data in automatic fashion. Secondly, we categorize the state-of-the-art in DL-based TC toward its effective application in mobile and encrypted context. To pinpoint and overcome the limitations found in literature, we propose a general framework for DL-based mobile and encrypted TC, based on a rigorous definition of its milestones: the choice of the traffic object, the definition of the input(s), the number of (simultaneous) TC tasks required and the corresponding DL architecture. We validate two concrete instances of our framework on three recent mobile traffic datasets, also surfacing future directions toward sophisticated DL-based mobile TC.

Finally, in [J6] we are planning to systematically describe and release to the research community a mobile-traffic dataset that could be used for traffic analysis and monitoring tasks, such as (ML- and DL-based) traffic classification, malware and anomaly detection, traffic prediction, etc. The dataset has been collected by more than 150 human users on a volunteer basis and contains the traffic generated by more than 50 Android apps. To the best of our knowledge, no dataset with such characteristics has been released to the research community to date.

I have also applied ML-based classification techniques for the classification of the traffic generated by Anonymity Tools (ATs). Specifically, among the several ATs developed in recent years, I have considered the Onion Router (Tor) [18], the Invisible Internet Project (I2P) [19], and JonDonym [20] being the most popular ones. Starting from the work in [J2], in collaboration with other members of the research group, we have submitted a journal paper [J4] (under second round of review) in which we propose TC of anonymity tools (and deeper, of their running services and applications) via a truly hierarchical approach. Capitalizing the Anon17 public dataset [21] released in 2017 containing anonymity traffic, we provide an in-depth analysis of TC and we compare the proposed hierarchical approach with a flat counterpart [J2]. The proposed framework is investigated in both the usual TC setup and its “early” variant (i.e. only the first segments of traffic aggregate are used to take a decision). Results highlight a general improvement over the flat approach in terms of all the classification metrics. Moreover, fine-grain performance investigation allows to (a) demonstrate lower severity of errors incurred by the hierarchical approach (as opposed to the flat case) and (b) highlight poorly-classifiable services/applications of each anonymity tool, gathering useful feedback on their privacy-level.

Collaborations

We strictly collaborate with **Huawei Technologies Co. Ltd.** in the context of the research activity related to mobile and encrypted traffic classification.

Furthermore, I am collaborating with the **Simula Research Laboratory**, Oslo, Norway, and the **IMDEA Networks Institute**, Leganes (Madrid), Spain, on the research topics defined during the periods of study and research abroad detailed in the following appropriate section.

Products

Journal Papers

1. [J1] **Antonio Montieri**, Domenico Ciuonzo, Giuseppe Aceto, Antonio Pescapè, *Anonymity Services Tor, I2P and JonDonym: Classifying in the Dark (Web)*, IEEE Transactions on Dependable and Secure Computing, Early Access.

Conference Papers

1. [C1] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *Mobile Encrypted Traffic Classification using Deep Learning*, in IEEE/ACM 2018 Network Traffic Measurement and Analysis Conference (TMA), pp. 1-8, 26-29 June, 2018, Vienna, Austria.

Papers Under Review

1. [J2] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *Mobile Encrypted Traffic Classification using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges*, IEEE Transactions on Network and Service Management, under third round of review, 2019.
2. [J3] **Antonio Montieri**, Domenico Ciuonzo, Giuseppe Aceto, Antonio Pescapè, *A Dive into the Dark Web: Hierarchical Traffic Classification of Anonymity Tools*, IEEE Transactions on Network Science and Engineering, under second round of review, 2019.
3. [J4] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *MIMETIC: Mobile Encrypted Traffic Classification using Multimodal Deep Learning*, IEEE Transactions on Mobile Computing, under first round of review, 2019.

Papers in Preparation

4. [J5] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *Toward Effective Mobile Encrypted Traffic Classification through Deep Learning*, IEEE Communications Magazine, expected submission Q1 2019.

5. [J6] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Fabio Palumbo, Valerio Persico, Antonio Pescapè, *REMBRANDT: REproducible MoBile tRaffic Analysis aND moniToring*, ACM Computer Communication Review, expected submission Q1 2019.
6. [C2] **Antonio Montieri**, Cise Midoglu, Özgü Alay, Antonio Pescapè, *Characterizing Data Rate Coverage along Transport Routes in Mobility*, TBD, expected submission Q2 2019 (in collaboration with the Simula Research Laboratory, Oslo, Norway).

Conferences and Seminars

I attended the following conferences:

1. *The 2018 Network Traffic Measurement and Analysis Conference* (TMA 2018), June 26-29, 2018, Vienna, Austria.
2. *The 16th Italian Networking Workshop 2019* (INW 2019), January 16-18, 2019, Bormio, Italy.

I also made the following presentations:

1. *Mobile Encrypted Traffic Classification using Deep Learning*, presented at the 2018 Network Traffic Measurement and Analysis Conference (TMA 2018), June 26-29, 2018, Vienna, Austria.
2. *Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark (Web),* presented at the 16th Italian Networking Workshop 2019 (INW 2019), January 16-18, 2019, Bormio, Italy.

Activity Abroad

During my second PhD year I have carried out a period of study and research abroad **from 19/07/2018 to 08/09/2018** at the **Simula Metropolitan Center for Digital Engineering**, Oslo, Norway, under the supervision of the Prof. Özgü Alay. Study and research activity concerned the analysis and characterization of the Quality of Service and Quality of Experience parameters of communication networks in high-mobility scenarios.

Moreover, **from 24/01/2019** I have started a period of study and research abroad at the **IMDEA Networks Institute**, Leganes (Madrid), Spain, under the supervision of the Prof. Narseo Vallina-Rodriguez. The activity abroad will last **until 18/04/2019**. Study and research activity concerns the mobile traffic privacy and the 3rd-party tracking ecosystem: cross-device tracking, transparency, and compliance with GDPR.

Tutorship

Teaching assistant at the B.Sc. course of **Reti di Calcolatori I** and M.Sc. course of **Analisi e Prestazioni di Internet**.

Tutorship grant for A.A. 2017/2018 at DIETI for the B.Sc. courses of **Fondamenti di Informatica** and **Calcolatori Elettronici I**.

References

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

[2] F. Jejdling et al., Ericsson mobility report. Ericsson AB, Business Area Networks, Stockholm, Sweden, Tech. Rep. EAB-18, 4510, 2018.

[3] H. Shi and Y. Li, Discovering periodic patterns for large scale mobile traffic data: Method and applications, IEEE Transactions on Mobile Computing, 2018.

[4] Sandvine, Global Internet Phenomena Spotlight: Encrypted Internet Traffic., 2016.

[5] A. Razaghpanah, A. A. Niaki, N. Vallina-Rodriguez, S. Sundaresan, J. Amann, and P. Gill, Studying TLS usage in Android apps, in 13th ACM CoNEXT, 2017.

[6] B. Saltaformaggio, H. Choi, K. Johnson, Y. Kwon, Q. Zhang, X. Zhang, D. Xu, and J. Qian, Eavesdropping on fine-grained user activities within smartphone apps over encrypted network traffic, in USENIX Workshop on Offensive Technologies (WOOT), 2016.

[7] Y. Fu, H. Xiong, X. Lu, J. Yang, and C. Chen. Service usage classification with encrypted internet traffic in mobile messaging apps, in IEEE Transactions on Mobile Computing, 15(11), 2016.

[8] V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, Robust smartphone app identification via encrypted network traffic analysis, in IEEE Trans. Inf. Forensics Security, 13(1), 2018.

[9] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, Multi-classification approaches for classifying mobile app traffic, Journal of Network and Computer Applications, 103, 2018.

[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.

[11] Z. Wang, The Applications of Deep Learning on Traffic Identification., 2015.

[12] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, Malware traffic classification using convolutional neural network for representation learning, in IEEE International Conference on Information Networking, 2017.

[13] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, End-to-end encrypted traffic classification with one-dimensional convolution neural networks, in IEEE International Conference on Intelligence and Security Informatics (ISI), 2017.

[14] M. Lotfollahi, R. Shirali, M. J. Siavoshani, and M. Saberian, Deep packet: a novel approach for encrypted traffic classification using Deep Learning, arXiv, 2017.

[15] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, Network traffic classifier with convolutional and recurrent neural networks for Internet of Things, IEEE Access, 5, 2017.

[16] S. E. Oh, S. Sunkam, and N. Hopper, Traffic analysis with deep learning, arXiv, 2017.

[17] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, Multimodal deep learning, in 28th International Conference on Machine Learning (ICML), 2011.

[18] P. Syverson, R. Dingledine, and N. Mathewson, Tor: the second generation onion router, in USENIX 13th Security Symposium (SSYM), 2004.

[19] The Invisible Internet Project (I2P), [Online] https://geti2p.net/en/, Jul. 2017.

[20] Project: AN.ON - Anonymity, [Online] http://anon.inf.tu-dresden.de/index_en.html, Jul. 2017.

[21] K. Shahbar and A. N. Zincir-Heywood, Packet momentum for identification of anonymity networks, Journal of Cyber Security and Mobility, vol. 6, no. 1, pp. 27–56, 2017.