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 – First 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 an university ITEE grant. My tutor is prof. **Antonio Pescapè** and I am a member of the COMICS research group whose activities are carried out in the field of Computer Networks.

Study and Training activities

During the first year of PhD program, I attended the courses and seminars reported in the following. In June 2017, I also attended the 7th TMA PhD School on Traffic Management & Analysis at the Maynooth University (Ireland) and the FMAI 2017 workshop that have been included in the list of seminars.

Courses

1. *Le imprese e la ricerca*, ad hoc module, Dr. Marco Frizzarin, 28/02/2017 – 02/03/2017, 4 Credits.
2. *Ethical, legal and social aspects of ICT and Robotics*, ad hoc module, Prof. Guglielmo Tamburrini, 7/03/2017 – 04/04/2017, 3 Credits.
3. *Machine Learning*, ad hoc module, Prof. Carlo Sansone et al., 08/05/2017 – 19/05/2017, 4 Credits
4. *Analisi e Prestazioni di Internet*, M.Sc course, Prof. Antonio Pescapè, March 2017 – June 2017, 6 Credits.
5. *Network Security*, M.Sc course, Prof. Simon Pietro Romano, September 2017 – December 2017, 6 Credits.

Seminars

1. *Cognitive Computing and Da Vinci Robot: Research proposal and discussion*, Prof. Paolo Maresca, 17/02/2017, 0.2 Credits.
2. *IBM Cognitive Computing: Challenges and Opportunities in Building an Artificial Intelligence Platform for business*, Ing. Pietro Leo, 17/02/2017, 0.4 Credits.
3. *Smart Nanodevices for Theranostics*, Ilaria Rea, 24/02/2017, 0.3 Credits.
4. *Fuzzy Logic, Genetic Algorithms and Their Application to Next Generation Networks*, Prof. Leonard Barolli, 10/03/2017; 14/03/2017, 0.8 Credits.
5. *How to Organize and Write a Scientific Rebuttal*, 10/03/2017, Prof. Pasquale Arpaia, 0.4 Credits.
6. *Scaling Adaptive Streaming Systems With Network Support*, Dr. Ali C.Begen, 13/03/2017, 0.3 Credits.
7. *Living Bots and Alter Ego*, Prof. Marco Gori, 04/04/2017, 0.4 Credits.
8. *From Ethernet Switching to Virtual and Programmable Switching*, Prof. Stefano Secci, 10/04/2017, 0.4 Credits.
9. *Embedding Protocols for Virtualized Networks*, Prof. Stefano Secci, 11/04/2017, 0.4 Credits.
10. *Dataflow Supercomputing for Big Data*, Prof. Veljko Milutinovic, 12/04/2017, 0.6 Credits.
11. *Power System Stability and Synchronization Application to the Lossy Power Grid System*, Prof. Navdeep M.Singh, 30/06/2017, 0.2 Credits.
12. *A Shared Memory Parellel Heuristic Algorithm for the Large-scale P-median Problem*, Prof. Vasilyev Igor, 12/09/2017, 0.2 Credits.
13. *Wireless Opportunistic Networking*, Prof. Gunnar Karlsson, 28/09/2017, 0.3 Credits.

14. *Challenges and Opportunities for IT Innovation in the Space Business*, Ing. Ernesto Doelling, 21/11/2017, 0.3 Credits.
15. *Large Scale Integrative Bioinformatics and Systems Biology in Cancer Genomics*, Prof. Michele Ceccarelli, 18/01/2018, 0.3 Credits.
16. *Workshop: FMAI 2017 – Formal Methods in Artificial Intelligence*, Prof. Aniello Murano, 22/02/2017 – 24/03/2017, 2.5 Credits.
17. *PhD School: 7th TMA PhD School on Traffic Management & Analysis*, Özgü Alay; Alberto Dainotti; Emile Aben; John Heidemann, 19/06/2017 – 20/06/2017 (18 hours), 1.8 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							
		1	2	3	4	5	6	
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary
Modules	**20**	4	3	0	6	4	6	**23**
Seminars	**10**	4,9	1,8	2	0,5	0,3	0,3	**9,8**
Research	**30**	1,1	5,2	8	3,5	5,7	6,7	**30,2**
	60	10	10	10	10	10	13	**63**

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. 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, different tools and middle-boxes, such as performance enhancement proxies, network monitors and policy enforcement devices, base their functions on the knowledge of the applications generating the traffic.

Specifically, I am currently working on machine-learning-based traffic classification techniques that try to overcome the limitations given by the encrypted nature of the majority of traffic generated by mobile apps and do not base their operations on the payload content, but take advantage of statistical features extracted from the mobile app traffic.

Moreover, I have also tried to apply similar approaches 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) showing also their effectiveness in accomplishing this task.

Research Description

Several tools, such as security/quality-of-service enforcement devices and network monitors base their operations on the knowledge of the application generating the traffic. As a consequence, their use is limited (or impaired) when this requirement is not (or loosely) satisfied. The process of associating (labeling) network traffic with specific applications or application types is known as Traffic Classification (TC) and has a long-established application in several fields, backed by a wide scientific literature [1, 2, 3, 4].

This process is increasingly challenged by recent evaluations in Internet usage, as the global spread and growing usage of smartphones is profoundly changing the kind of traffic that travels over home and enterprise networks and the Internet. Thereupon, both the necessity and the difficulty of TC of mobile traffic have become very high nowadays. Indeed, other than the traditional drivers for TC, classification of mobile apps' traffic has the potential of providing extremely valuable profiling information (e.g., to advertisers, insurance companies and security agencies). On the other hand, it surely raises privacy issues, especially in regards to recognition of context-sensitive apps (such as health and dating ones) by malicious parties. Unluckily, TC comes with its own challenges and requirements that are even exacerbated in a mobile-traffic context, usually characterized by a large number of apps to discriminate from and an inadequate number of training samples per app, which hinder the achievement of satisfactory performance. Moreover, the increasing adoption of encrypted protocols (TLS) makes the classification even more challenging, defeating established approaches, such as those based on payload inspection. Indeed, approaches based on Machine Learning (ML) classifiers are deemed the most appropriate, especially in this context, since they suit also Encrypted Traffic analysis [5, 6, 7, 8].

As a first step in this direction, I tried to improve the classification performance of mobile apps by proposing a Multi-Classification System (MCS) which intelligently-combines decisions from state-of-the-art (base) classifiers specifically devised for mobile- and encrypted-traffic classification and currently considered the best approaches in such context [5, 6, 8]. Indeed, the MCS framework can potentially overcome the deficiencies of each single classifier and provide improved performance w.r.t. any of the base classifiers, also allowing for modularity of classifiers' selection in the pool [9]. This activity has led to the publication of one conference paper [C3] and one journal paper [J1] made in collaboration with other members of the research group. Specifically, in these works we apply our MCS framework to the traffic collected by a global mobile solutions provider made of true users' activities. The results show that MCS framework can improve classification performance with respect to the best base classifiers considered for the task. In detail, in the extended journal version [J1], we have enlarged the type of combination techniques taken into account (considering both hard and soft combiners), we have given useful hints to perform dataset pre-processing and classifiers/combiners parameter tuning (specifically tailored for mobile TC), and we have shown the results of a more comprehensive experimental analysis.

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) [10], the Invisible Internet Project (I2P) [11], and JonDonym [12] being the most popular ones. Indeed, a key issue is to understand whether their (encrypted) traffic data can be classified and, if so, to which depth. More specifically, it is interesting to ascertain to which degree an external observer can recognize an AT and how fine would be the fingerprinting granularity achievable, that is, whether traffic types and/or services hidden into them could be inferred. This investigation is equally useful to designers of anonymity networks, as it suggests how privacy of anonymity networks could be further robustified.

On the latter topic, I have published one conference paper [C2], in collaboration with other members of the research group, and another journal paper [J2] has been accepted but not yet published. In detail, we have applied five ML classifiers, belonging to different families (i.e., Bayesian approaches and decision trees) and using different types of features, on the publicly released Anon17 dataset [13]. This dataset consists of a collection of traces gathered by different anonymity networks, as well as related services and applications running inside them. Our analysis is carried out at different levels of granularity, as we try to infer whether the Anonymity Network being observed can be classified and, in affirmative case, whether the Traffic Type and Application transported hidden within them could be inferred. The obtained results show that anonymity networks can be easily discerned, and the traffic type and the service running within it can be reasonably inferred as well (by a judicious use of the appropriate classifier and optimized set of features).

Collaborations

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

Products

Journal Papers

1. [J1] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *Multi-Classification Approaches for Classifying Mobile App Traffic*, Elsevier Journal of Network and Computer Applications, vol. 103, pp. 131–145.

Conference Papers

1. [C1] Giuseppe Aceto, **Antonio Montieri**, Antonio Pescapè, *Internet Censorship in Italy: an Analysis of 3G/4G Networks,* 2017 IEEE International Conference on Communications (ICC 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, May 21-25, 2017, Paris, France.
2. [C2] **Antonio Montieri**, Domenico Ciuonzo, Giuseppe Aceto, Antonio Pescapè, *Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark*, 29th International Teletraffic Congress (ITC 29), September 4-8, 2017, Genova, Italy.
3. [C3] Giuseppe Aceto, Domenico Ciuonzo, **Antonio Montieri**, Antonio Pescapè, *Traffic Classification of Mobile Apps through Multi-classification*, 2017 IEEE Global Communications Conference (IEEE GLOBECOM 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, December 4-8, 2017, Singapore.

Accepted Papers

1. [J2] **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.

Conferences and Seminars

I attended the following conferences:

1. *The 2017 Network Traffic Measurement and Analysis Conference* (TMA 2017), June 21-23, 2017, Maynooth, Dublin, Ireland.
2. *The 29th International Teletraffic Congress* (ITC 29), September 4-8, 2017, Genova, Italy.
3. *The 2017 IEEE Global Communications Conference* (IEEE GLOBECOM 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, December 4-8, 2017, Singapore.

I also made the following presentations:

1. *Traffic Classification of Mobile Apps through Multi-classification [POSTER]*, presented at the poster session during the 7th TMA PhD School on Traffic Management & Analysis, June 19-20, 2017, Maynooth, Dublin, Ireland.
2. *Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark,* presented at the 29th International Teletraffic Congress (ITC 29), September 4-8, 2017, Genova, Italy.
3. *Traffic Classification of Mobile Apps through Multi-classification*, presented at the 2017 IEEE Global Communications Conference (IEEE GLOBECOM 2017); Communication QoS, Reliability and Modeling (CQRM) Symposium, December 4-8, 2017, Singapore.

Activity Abroad

I have not carried out any long-term activity abroad during my first PhD year.

Tutorship

Teaching assistant at the B.Sc. courses of **Calcolatori Elettronici I** and **Reti di Calcolatori I**.

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

References

[1] S. Valenti, D. Rossi, A. Dainotti, A. Pescapè, A. Finamore, M. Mellia, Reviewing traffic classification, in: Data Traffic Monitoring and Analysis, Springer, 123–147, 2013.

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

[3] A. Callado, C. Kamienski, G. Szabó, B. P. Gero, J. Kelner, S. Fernandes, D. Sadok, A survey on internet traffic identification, IEEE Communications Surveys & Tutorials 11 (3) (2009) 37–52.

[4] T. T. T. Nguyen, G. Armitage, A survey of techniques for internet traffic classification using machine learning, IEEE Communications Surveys & Tutorials 10 (4) (2008) 56–76.

[5] D. Herrmann, R. Wendolsky, H. Federrath, Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial Naïve-Bayes classifier, in: Proceedings of the ACM workshop on Cloud computing security (CCSW), 31–42, 2009.

[6] M. Liberatore, B. N. Levine, Inferring the source of encrypted HTTP connections, in: Proceedings of the 13th ACM conference on Computer and communications security (CCS), 255–263, 2006.

[7] B. Saltaformaggio, H. Choi, K. Johnson, Y. Kwon, Q. Zhang, X. Zhang, D. Xu, J. Qian, Eavesdropping on fine-grained user activities within smartphone apps over encrypted network traffic, in: Proc. USENIX Workshop on Offensive Technologies (WOOT'16, in conjunction with Security'16), 2016.

[8] V. F. Taylor, R. Spolaor, M. Conti, I. Martinovic, Appscanner: Automatic fingerprinting of smartphone apps from encrypted network traffic, in: IEEE European Symposium on Security and Privacy (EuroS&P), 439–454, 2016.

[9] A. Dainotti, A. Pescapé, C. Sansone, Early classification of network traffic through multi-classification, in: International Workshop on Traffic Monitoring and Analysis (TMA), Springer, 122–135, 2011.

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

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

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

[13] 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.