

Unveiling Radio Resource Utilization Dynamics of Mobile Traffic through Unsupervised Learning

Arcangela Rago^{1,2}, Giuseppe Piro^{1,2}, Hoang Duy Trinh³, Gennaro Boggia^{1,2}, Paolo Dini³

¹*Dept. of Electrical and Information Engineering, Politecnico di Bari, Italy*

Email: {arcangela.rago, giuseppe.piro, gennaro.boggia}@poliba.it

²*CNIT, Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Italy*

³*CTTC, Centre Tecnològic de Telecomunicacions de Catalunya, Spain*

Email: {hoangduy.trinh, paolo.dini}@cttc.es

Abstract—Understanding mobile traffic dynamics is a key issue to properly manage the radio resources in next generation mobile networks and meet the stringent requirements of emerging heterogeneous services, such as enhanced mobile broadband, autonomous driving, and extended reality (just to name a few). However, radio resource utilization patterns of real mobile applications are mostly unknown. This paper aims at filling this gap by tailoring an unsupervised learning methodology (i.e. K-means), able to identify similar radio resource utilization patterns of mobile traffic from an operating mobile network. Our analysis is based on datasets referring to residential and campus areas and containing wireless link level information (e.g., scheduling, channel conditions, transmission settings, and duration) with a very precise level of granularity (e.g., 1 ms). Obtained results reveal the properties of groups of sessions with similar characteristics, expressed in terms of bandwidth demands and application level requirements.

Index Terms—Mobile Traffic Analysis, Radio Resource Utilization Dynamics, Unsupervised Learning

I. INTRODUCTION

The widespread and growing usage of smartphones and machine-type communications are deeply changing the type of traffic that traverses the mobile networks. Next generation of mobile systems (5G and beyond) have to be designed to fulfill the high performance requirements of such applications in terms of latency, capacity and context awareness [1]. Understanding the dynamics of mobile traffic demands is then of utmost importance for proper and secure management of network resources (e.g., spectrum, energy, computation) [2] [3]. Specifically, traffic classification at the radio link-level and forecast may enable advanced Quality of Service (QoS) and Quality of Experience (QoE) enforcement policies based on a priori knowledge of application behaviors.

The scientific literature addressed traffic classification in mobile networks through machine or deep learning methodologies [4]. At the time of writing, the majority of works in this direction try to classify the applications running on mobile terminals by means of the Random Forest algorithm [5] [6] [7]. A cross comparison of different classifiers is widely argued in [8]. All of these contributions, however, rely on supervised approaches and target the classification of mobile traffic at the application layer. Unfortunately, details about radio resource utilization dynamics are completely neglected. The work presented in [9] uses Support Vector Machine

and k-Nearest Neighbors algorithms to reliably identify a smartphone, i.e., through a user fingerprint scheme learned from the traffic patterns produced by background applications. Nevertheless, also in this case, the developed solution cannot be used to classify traffic flows according to their behavior at the radio-link level.

Recently, researchers are given access to Call Detail Records (CDRs) of mobile operators and the analysis is more oriented on extrapolating spatio-temporal characteristics of the mobile user traffic [10] [11]. However, the considered datasets miss characterizing radio access level dynamics, which, instead, are considered crucial for mobile network optimization. In fact, CDRs would help to identify user service requests and throughput, but do not offer any detail on wireless link level information, channel conditions, retransmission, packet fragmentation, and so on.

To bridge this gap, the goal of this paper is to unveil radio resource utilization dynamics of mobile traffic. From a methodological perspective, the mobile traffic analysis is treated as an unsupervised learning problem, which aims at identifying spatio-temporal radio resource utilization patterns of mobile sessions. The Online Watcher for LTE (OWL) tool [12] [13] is used for monitoring the unencrypted Physical Downlink Control CHannel (PDCCH) of an operating Long Term Evolution (LTE) network deployed in Spain. The advantage of this tool is the richness of the gathered information (i.e., link level data) and the temporal granularity of the data (i.e., 1 ms). Obtained datasets, referring to residential and campus areas, are properly processed to group monitored sessions according to the achieved data rate, the adopted transmission settings, the radio resource usage, and the duration. The outcomes of the conducted study clearly highlight the properties of groups of sessions with similar characteristics, expressed in terms of bandwidth demands and application level requirements.

The rest of this paper is organized as follows: Section II introduces the background on LTE and the collected datasets together with the proposed unsupervised learning methodology; Section III discusses the numerical results; Section IV provides lessons learned and highlights related opportunities for mobile network optimization, and, finally, Section V concludes the paper and draws future research activities.

II. THE PROPOSED APPROACH

With reference to the radio interface, LTE embraces two main communicating entities: the base station and the mobile terminal. At the physical layer, radio resources are distributed among mobile terminals in a time-frequency domain and the Resource Block (RB) represents the smallest assignable radio resource unit. It lasts 1 ms in the time domain, namely Transmission Time Interval (TTI), and 180 kHz in the frequency domain. Every TTI, the base station allocates RBs to mobile terminals according to a specific scheduling algorithm. Transmission settings, expressed in terms of Modulation and Coding Scheme (MCS), are dynamically defined through a link adaptation mechanism. Moreover, the resulting amount of data to send during one TTI is fixed and depends on both the selected MCS and the number of RBs assigned to a given mobile terminal, as reported by the Transport Block Size (TBS) table. The scheduling decisions are shared with mobile terminals through control messages exchanged, at the beginning of the TTI, by using the PDCCH. Immediately after, data packets are exchanged through the Physical Downlink Shared CHannel (PDSCH).

The possibility of accessing to mobile traffic data represents a challenging task to accomplish. For security reasons, in fact, mobile operators avoid sharing their logs and data packets sent across the radio interface are encrypted. Nevertheless, control messages exchanged through the PDCCH are transmitted in clear. This represents a valid opportunity to extract key information related to mobile traffic data, as well as generating reference datasets to be used for traffic analysis. In this context, data collection from the PDCCH has been already presented in [12] [13]. In those papers, an online decoder based on Software Defined Radio (SDR), namely OWL, is used to decode PDCCH messages sent by the base station within a given coverage area, thus collecting scheduling decisions during the time (i.e., every TTI). OWL generally produces a raw file. Then, a Python script can be used to generate a usable dataset that summarizes the main data of interest, associated with each captured traffic session.

The methodology proposed in this contribution evaluates mobile traffic sessions by investigating radio resource utilization dynamics in the downlink. Specifically, a multivariate analysis has been conceived to make a classification of mobile traffic sessions at the radio link level, according to their properties (the average data rate, the average MCS, the average number of RBs, and the duration of sessions). To this end, K-means [14], which is a well-known unsupervised machine learning scheme, is used to map sessions with similar properties into K clusters. The variables of interest are firstly normalized within the range]0,1]. Then, each session is represented as a point in a hyperplane, whose dimensions refer to the variables of interest of the conducted analysis. At this point, the dissimilarity associated with two sessions is defined as the Euclidean distance between the two related points in the aforementioned hyperplane. Indeed, the optimal value of K is calculated in order to ensure that the intra-cluster

TABLE I: Distribution of sessions among clusters

Dataset	cluster id	Number of sessions per cluster [#]	Percentage of sessions per cluster [%]
residential area	1	235	45.11
	2	187	35.89
	3	82	15.74
	4	17	3.26
campus area	1	4942	99.92
	2	4	0.08

distances are minimized and the inter-cluster distances are maximized [14] (note that, according to K-means terminology, this means that the silhouette [15] is maximized). Finally, the clustering process provides in output the sessions of each cluster and a special point of the hyperplane, namely centroid, that identifies the cluster itself. Their coordinates are obtained by averaging the value of variables associated with the sessions belonging to the considered cluster. By studying the obtained groups of sessions, it is now possible to extract statistical details associated with each cluster and finalize the traffic characterization.

III. MOBILE TRAFFIC ANALYSIS AND DISCUSSION

Two LTE base stations in a residential and campus area of Barcelona, operating in a bandwidth of 20 MHz, are monitored to collect mobile data. The residential area has been monitored from 6 February 2018 to 5 March 2018 and the resulting dataset contains 521 sessions. Instead, the campus area has been monitored from 22 March 2017 to 26 April 2017 and the resulting dataset contains 4946 sessions. The analysis proposed next discusses the properties of the gathered mobile data for each base station considering (i) the dataset as a whole and (ii) the dataset divided into 4 time-slots that are morning, afternoon, evening, and night.

A. Study of the datasets as a whole

The two monitored base stations show different behavior in terms of radio resource usage patterns, as detailed below. The first difference refers to the output of the silhouette analysis, which groups the residential and campus traffic in four and two clusters, respectively. Figures 1 and 2 show the outcome of the K-means clustering process, carried out for the residential area and the campus area, respectively. For each variable of interest, the figures highlight the identified clusters, their centroids (i.e., the red dots), the 25th and the 75th percentile (i.e., the bottom line and the top line of the blue rectangle), as well as the minimum and the maximum measured value (i.e., the edges of the vertical red line) of the variables of interest. The number of sessions belonging to every single cluster and the related percentage, instead, are reported in Table I.

Comments for the residential area. It is important to note that there is a strict relation between the number of sessions belonging to the cluster and the average data rate experienced by its traffic sessions. About 45% of sessions report an average data rate equal to 0.46 Mbps. Instead, only 3.26% of sessions register an average data rate equal to 5.74 Mbps. Intermediate average data rates refer to intermediate groups of sessions (i.e.,

35.89% and 15.74% of sessions present an average data rate equal to 1.33 Mbps and 2.60 Mbps, respectively).

Similar behavior is observed for MCS indexes and the allocated RBs. Figure 1(b) shows that the selected average MCS index is lower than 4 for about 45% of sessions. Only 3.26% of sessions use an average MCS index close to 10. Moreover, 35.89% and 15.74% of sessions use an average MCS index approximately equal to 6 and 8, respectively. Only intermediate clusters register peaks of MCS up to nearly 26. Considering that LTE allows a maximum MCS index equal to 31, obtained findings clearly highlight that the channel quality experienced by mobile terminals during the monitoring is relatively scarce.

Very interesting details related to the distribution of radio resources among mobile terminals are depicted in Figure 1(c). About 45% of sessions, which have the average data rate equal to 0.46 Mbps, consume the lowest amount of physical resources. Considering that 100 RBs per TTI are available in 20 MHz bandwidth, an average number of RBs per TTI approximately equal to 23 means that sessions belonging to the first cluster occupy less than 1/4 of the overall amount of resources available within a cell. On the other hand, only 17 sessions consistently use a larger amount of resources per TTI, thus obtaining higher data rates.

A quite different behavior emerges from the analysis of the average session duration. Sessions that register the average data rate equal to 5.74 Mbps remain active for about 40 s, which is the lowest amount of time among the four clusters. For other clusters, instead, the duration increases with the number of sessions belonging to the cluster. It is also important to note that the session duration always presents a very high variability: the actual duration of 75% of sessions in each cluster is lower than the one associated with the related centroid.

Comments for the campus area. The campus area presents a number of sessions extremely higher than the residential case, but the reported bandwidth requirements are extremely lower. As expected, there is a strict relation between the number of sessions per cluster and the average data rate. Nevertheless, almost all the sessions monitored in the campus area (i.e., 99.92%) fall within the same cluster and register a very low average data rate equal to 0.14 Mbps. Only 0.08% of sessions register an average data rate of 5.47 Mbps.

The study of MCS indexes provides a reverse relation, as shown in Figure 2(b). The former group of sessions experiences variable channel conditions, translating into the usage of all the admitted transmission settings. While the average MCS index is 13, the maximum value is equal to 31. The second group of sessions (4 out of 4946) registers worse channel conditions. In this case, the average and the maximum MCS indexes are about 7 and 10, respectively.

Figure 2(c) confirms what observed for the residential area: the higher the average number of RBs used per TTI, the higher the achieved data rate. Reported results still show that 4942 sessions use about 1/4 of the bandwidth per TTI. On the contrary, only 4 sessions use a larger amount of resources

per TTI (i.e., more than 52).

As depicted in Figure 2(d), the campus area hosts sessions with very short durations. Apart from one exception (e.g., the graph reports one session duration equal to 1465 s), the former group of sessions registers an average session duration of 5 s. The duration of sessions belonging to the second cluster, instead, is lower than 2 s.

B. Time-slot Analysis

The analysis of mobile traffic on time-slots basis leads to a detailed characterization of sessions, with a consequent deeply recognition of resource usage and QoS requirements that a mobile network has to address during different parts of the day. The outcomes of the proposed clustering methodology on time-slots basis, applied to both residential and campus areas, are summarized in Tables II and III, respectively.

Comments for the residential area. The relation between the number of sessions belonging to the cluster and the average variable registered by related traffic sessions still exists. About 30% of morning sessions report an average data rate equal to 0.22 Mbps, while about 40% have an average data rate equal to 0.65 Mbps. During the afternoon, that is the time-slot with the highest number of residential sessions, the data rate starts growing. In fact, about 40% of afternoon sessions have an average data rate equal to 0.48 Mbps and the other 40% of sessions register an average data rate equal to 1.26 Mbps. The data rate still grows during the evening. The average data rate is 0.79 Mbps and 2.48 Mbps for about 60% and 35% of evening sessions, respectively. Considering night sessions, whose number is limited because people tend to sleep, the average data rate goes down: about 87% of night sessions report an average value equal to 0.30 Mbps.

The average MCS index is lower than 5 for about 70% of morning sessions. In particular, around 40% use an average MCS index close to 5 and around 30% even use an average MCS index approximately equal to 3. During the afternoon, average MCS indexes increase. In fact, about 40% of afternoon sessions have an average MCS index close to 4 and a further 40% register an average value close to 6. The MCS indexes still grow during the evening, as the data rate. The average MCS index is approximately 5 and 8 for about 60% and 35% of evening sessions, respectively. As regards night sessions, MCS indexes tend to reduce. In fact, about 87% of night sessions report an average value close to 3.

The distribution of radio resources follows a similar pattern. About 30% of morning sessions report an average number of RBs per TTI equal to 1/6 of the overall amount of resources available within a cell, while about 40% have an average value equal to 1/4. During the afternoon, about 40% of sessions use an average number of RBs per TTI close to 1/4 of bandwidth per TTI and a further 40% register an average number equal to 1/3. During the evening, the average amount of resources per TTI is more than 1/4 and about 1/2 of the overall bandwidth for about 60% and 35% of sessions, respectively. Then, bandwidth consumptions decrease during the night: about 87% of night sessions consume less than 1/4 of bandwidth per TTI.

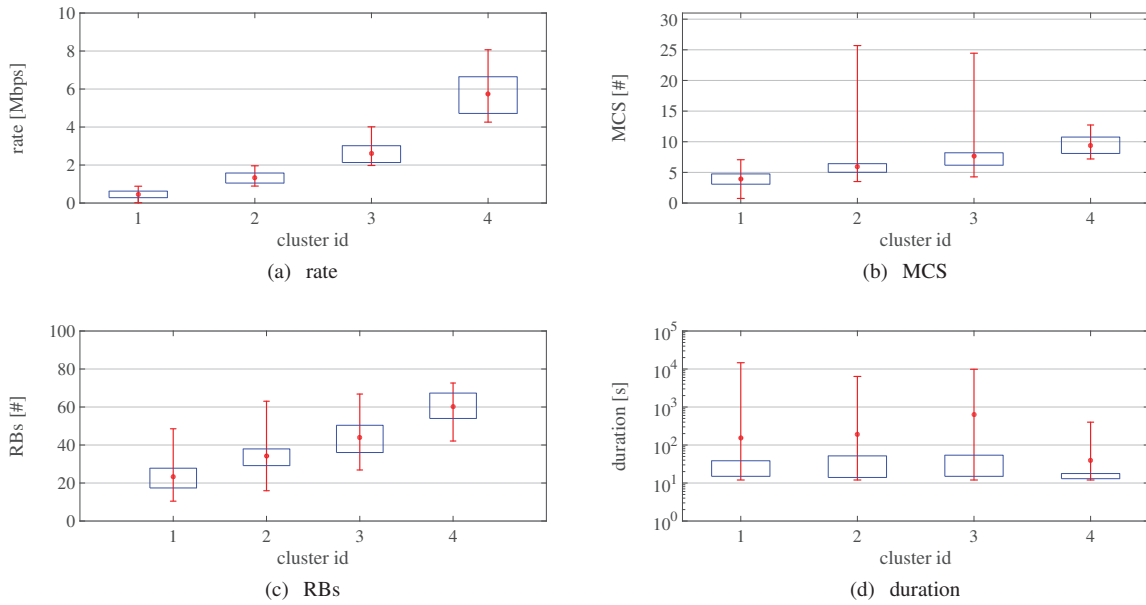


Fig. 1: Study of dataset related to the residential area, as a whole.

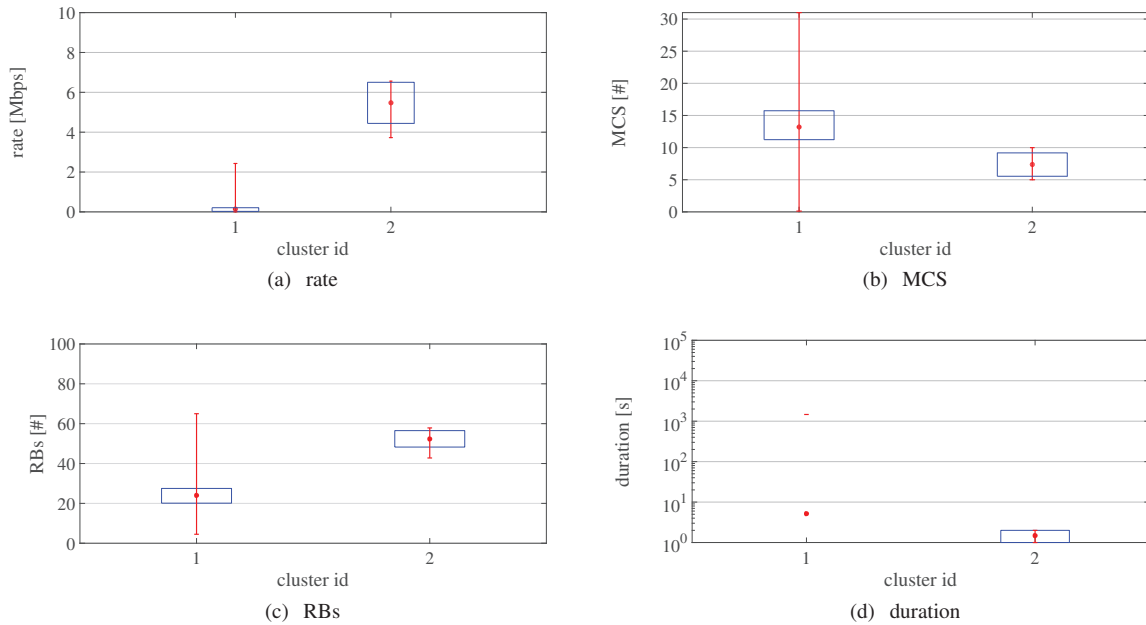


Fig. 2: Study of dataset related to the campus area, as a whole.

The average duration, which varies greatly, has different behavior. Sessions generally register a short duration (i.e., 600 s), except for those available in the night time-slot. In fact, about 87% of night sessions last about 800 s. Moreover, around 6.5% have an average duration equal to 1718 s.

Comments for the campus area. The campus area reports a more balanced distribution of sessions among the whole-day slots. As expected, traffic classification on time-slots basis offers a better characterization of sessions. For example, up to

7 clusters are identified for the afternoon time-slots, against the only two clusters reported for the analysis of the dataset as a whole.

Regarding the data rate in the campus area, more than 55% of morning sessions report an average data rate equal to 0.06 Mbps, while about 40% have an average data rate equal to 0.25 Mbps. During the afternoon, the data rate tends to decrease. In fact, about 40% of afternoon sessions report an average data rate equal to 0.02 Mbps and about 26% register an average data

TABLE II: Study of the dataset related to the residential area, on time-slots basis.

cluster id	# of sessions	rate [Mbps]			MCS [#]			RBs [#]			duration [s]		
		min	max	avg	min	max	avg	min	max	avg	min	max	avg
morning (from 6:00 am to 12:59 pm, 58 sessions)													
1	17	0.08	0.37	0.22	1.59	3.56	2.68	11.85	24.29	17.90	12	69	19.88
2	23	0.46	0.97	0.65	2.96	7.07	4.59	19.11	37.75	26.14	12	2696	324.39
3	18	0.99	2.17	1.31	4.28	6.83	5.75	26.75	47.93	34.64	12	3304	367.11
afternoon (from 1:00 pm to 7:59 pm, 297 sessions)													
1	120	0.06	0.86	0.48	1.15	6.95	4.11	12.08	48.55	23.72	12	478	51.67
2	116	0.87	1.80	1.26	3.51	25.69	5.80	19.09	63.05	33.79	12	3716	158.72
3	53	1.84	3.39	2.39	4.27	21.28	6.99	27.78	61.86	40.99	12	9845	598.09
4	8	3.56	6.12	4.47	7.39	11.96	9.49	45.81	72.62	61.34	12	20	15.38
evening (from 8:00 pm to 12:59 am, 134 sessions)													
1	78	0.09	1.58	0.79	1.98	7.45	4.67	12.98	57.24	27.12	12	135	26.14
2	46	1.64	4.31	2.48	5.14	24.44	8.07	26.86	66.82	44.36	12	8600	571.87
3	10	4.80	8.07	6.26	7.19	11.74	9	42.10	70.52	57.83	12	41	17.60
night (from 1:00 am to 5:59 am, 31 sessions)													
1	27	0.02	0.71	0.30	0.75	6.72	3.38	10.45	44.89	23.19	12	14630	793.81
4	2	1.15	1.60	1.38	3.97	6.87	5.42	34.77	48.56	41.67	20	3416	1718
3	1	2.85	2.85	2.85	7.83	7.83	7.83	46.86	46.86	46.86	126	126	126
2	1	5.91	5.91	5.91	12.73	12.73	12.73	66.62	66.62	66.62	399	399	399

TABLE III: Study of the dataset related to the campus area, on time-slots basis.

cluster id	# of sessions	rate [Mbps]			MCS [#]			RBs [#]			duration [s]		
		min	max	avg	min	max	avg	min	max	avg	min	max	avg
morning (from 6:00 am to 12:59 pm, 1225 sessions)													
1	681	<0.01	0.16	0.06	0.79	29	12.54	5	47.33	24.18	<1	292	8.45
2	538	0.16	0.79	0.25	1.55	20.93	14.44	10.83	38.25	24.99	<1	11	<1
3	6	0.88	2.43	1.42	3.70	14.19	8.60	20.73	33.34	27.73	<1	908	151.33
afternoon (from 1:00 pm to 7:59 pm, 1134 sessions)													
1	425	<0.01	0.07	0.02	1	29	11.35	5	49	24.18	<1	1114	12.04
2	298	0.07	0.17	0.12	6.27	27	13.80	11.65	54.50	22.66	<1	<1	<1
3	278	0.17	0.27	0.21	2.65	21.12	14.32	11.66	40.08	24.60	<1	3	<1
4	99	0.27	0.44	0.32	2.34	20	14.38	13.62	36.95	25.43	<1	1465	23.10
5	32	0.45	1.10	0.57	3.44	17.65	13.72	18.73	31.36	24.86	<1	<1	<1
6	1	2.23	2.23	2.23	8.09	8.09	8.09	47.64	47.64	47.64	16	16	16
7	1	3.73	3.73	3.73	6.09	6.09	6.09	42.78	42.78	42.78	1	1	1
evening (from 8:00 pm to 12:59 am, 848 sessions)													
1	846	<0.01	0.82	0.13	0.58	31	13.09	4.50	52	23.41	<1	234	3.04
2	2	5.16	6.45	5.80	5	9.99	7.49	55.16	57.86	56.51	2	2	2
night (from 1:00 am to 5:59 am, 1738 sessions)													
1	627	<0.01	0.09	0.03	0.12	26.67	11.22	4.75	52	22.37	<1	695	12.39
2	690	0.09	0.22	0.16	0.72	23.11	14.08	8.77	45.25	23.62	<1	1	<1
3	363	0.22	0.42	0.28	1.90	21.29	14.66	11.77	36.63	25.31	<1	11	<1
4	57	0.42	0.98	0.56	2.92	16.82	14.07	16.14	31	25.01	<1	1080	19.14
5	1	6.56	6.57	6.56	8.36	8.36	8.36	53.74	53.74	53.74	1	1	1

rate equal to 0.12 Mbps. The data rate is still low during the evening. In fact, the average value of 0.13 Mbps is measured for more than 99% of sessions. During the night, that is the time-slot with the highest number of sessions, the average data rate tends to increase. From Table III, it is 0.03 and 0.16 Mbps for about 36% and 40% of night sessions, respectively.

The average MCS index is similar among the time-slots. In particular, about 55% of morning sessions have an average value close to 13. About 40% of afternoon sessions register an average MCS index close to 11, while about 26% and 25% use an average MCS index approximately equal to 14 and more than 14, respectively. During the evening, the average MCS index is approximately 13 for 99.76% of sessions. Lastly, it is close to 11 and 14 for about 36% and 40% of night sessions,

respectively.

Also the allocated RBs per TTI have similar behavior. In particular, they slightly increase and decrease during the morning and the afternoon and during the evening and the night, respectively. About 56% of morning sessions and 38% of afternoon sessions report an average number of RBs per TTI close to 1/4 of the overall amount of resources available within a cell. Instead, the average amount of resources per TTI is more than 1/4 of the overall bandwidth for 99.76% of evening sessions and about 76% of night ones.

The average duration is extremely low during all the considered time-slots. In particular, about 56% of morning sessions last about 9 s. Furthermore, about 38% of afternoon sessions last longer than 10 s (i.e., about 12 s), while more than 50%

(the clusters 2 and 3 in the afternoon) last less than 1 s. The average duration is approximately equal to 3 s for about 99% of evening sessions. As the last report, about 36% of night sessions last longer than 10 s (i.e., about 12 s), while around 60% last less than 1 s.

IV. LESSONS LEARNED AND NETWORK OPTIMIZATION OPPORTUNITIES

The proposed study clearly shows that the analysis of mobile traffic on time-slots basis gives a deep insight into radio resource utilization dynamics. Interesting outcomes are summarized in what follows. As far as the residential area is concerned, a high number of sessions are measured for the afternoon time-slot and peaks of bandwidth requirements are registered in both afternoon and evening time-slots. By observing data related to the night time-slot, it is possible to understand that a residential area significantly reduces its traffic load when people usually go to sleep. Nonetheless, differently from daily time-slots, the few sessions active during the night present very high durations. As far as the campus area is concerned, sessions use a higher MCS index than residential sessions, but a very low rate: analyzed campus sessions do not transmit a lot of data, even if the quality of channel could be good, because the traffic load is not significant.

Now, by knowing the radio resource utilization patterns of mobile traffic, it will be possible to conceive novel methodologies that aim at optimizing mobile networks. Interesting research activities to address in the future may include:

- Advanced QoE/QoS management through dynamic radio resource scheduling algorithms exploiting the deep properties expected for mobile flows at the radio level;
- Dynamic and fine-grained management of slices and virtual functionalities offered through the radio access networks in upcoming 5G architectures;
- Optimal energy savings mechanisms (e.g. sleep mode of base stations and discontinuous reception in mobile terminals) and opportunistic handover management procedures that leverage the predicted behavior of classified traffic flows;
- Planning for new base stations deployments in geographical regions where higher traffic load is expected;
- Massive usage of mobile base stations (i.e., deployed as drones), chasing the actual radio resource utilization dynamics.

V. CONCLUSIONS

This work investigates the radio resource utilization dynamics of mobile traffic by means of an unsupervised learning methodology. Two datasets, collected from a real operating mobile network and referring to residential and campus areas, have been investigated. Specifically, a multivariate analysis revealed the properties of groups of sessions with similar characteristics, expressed in terms of bandwidth demands and application level requirements. Obtained results report a clear heterogeneity among traffic sessions, whose clustering offers key instruments for the optimal management of the radio

resources for mobile operators. Thus, further research activities may extend the presented methodology and support mobile network optimization at the radio access level.

ACKNOWLEDGMENT

This work was partially funded by Italian MIUR PON projects Pico&Pro (ARS01_01061), AGREED (ARS01_00254), FURTHER (ARS01_01283), RAFAEL (ARS01_00305), by Apulia Region (Italy) Research projects E-SHELF (OSW3NO1) and INTENTO (36A49H6), by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675891 (SCAVENGE), and by Spanish MINECO grant TEC2017-88373-R (5G-REFINE).

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [2] P. Torres, P. Marques, H. Marques, R. Dionísio, T. Alves, L. Pereira, and J. Ribeiro, "Data analytics for forecasting cell congestion on LTE networks," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 2017, pp. 1–6.
- [3] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1. ACM, 2013, pp. 17–28.
- [4] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Mobile encrypted traffic classification using deep learning," in *Proc. of IEEE Network Traffic Measurement and Analysis Conference (TMA)*, 2018, pp. 1–8.
- [5] Y. Liu, S. Zhang, B. Ding, X. Li, and Y. Wang, "A cascade forest approach to application classification of mobile traces," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.
- [6] Q. Wang, A. Yahyavi, B. Kemme, and W. He, "I know what you did on your smartphone: Inferring app usage over encrypted data traffic," in *2015 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2015, pp. 433–441.
- [7] V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, "Robust smartphone app identification via encrypted network traffic analysis," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 63–78, 2018.
- [8] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Mobile Encrypted Traffic Classification Using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges," *IEEE Transactions on Network and Service Management*, vol. PP, no. 99, 2019.
- [9] T. Stöber, M. Frank, J. Schmitt, and I. Martinovic, "Who do you sync you are?: smartphone fingerprinting via application behaviour," in *Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks*. ACM, 2013, pp. 7–12.
- [10] D. Naboulsi, R. Stanica, and M. Fiore, "Classifying call profiles in large-scale mobile traffic datasets," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 1806–1814.
- [11] A. Furno, M. Fiore, and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.
- [12] N. Bui and J. Widmer, "OWL: A Reliable Online Watcher for LTE Control Channel Measurements," in *Proc. of ACM Workshop on All Things Cellular: Operations, Applications and Challenges (ATC)*, New York, NY, USA, 2016, pp. 25–30.
- [13] H. D. Trinh, N. Bui, J. Widmer, L. Giupponi, and P. Dini, "Analysis and modeling of mobile traffic using real traces," in *Proc. of IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–6.
- [14] A. J. Izenman, "Modern multivariate statistical techniques," *Regression, classification and manifold learning*, 2008.
- [15] J. Garcia and A. Brunstrom, "Clustering-based separation of media transfers in DPI-classified cellular video and VoIP traffic," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.