Present-day Verticals and Where to Find Them: A Data-driven Study on the Transition to 5G

Francesco Malandrino, Carla-Fabiana Chiasserini
DET, Politecnico di Torino, Italy
francesco.malandrino@polito.it, chiasserini@polito.it

***Abstract*—Much of the research about 5G networks deals with emerging or upcoming applications, e.g., self-driving cars and virtual reality. In this paper, we focus on *present-day* Internet services and assess which of them can benefit the most integration within 5G, i.e., which of today's service providers are the most likely to become 5G *verticals*. To this end, we leverage a large-scale, real-world, crowd-sourced dataset representing the data required by thousands of smartphone apps, and study the data rate and sparseness associated with each app. We argue that high-data rate, low-sparseness apps have the most to gain from 5G integration, and find that this category includes not only video streaming, but also peer-to-peer file transfer and mobile gaming applications.**

I. Introduction and related work

It is a truth universally acknowledged, that emerging and upcoming applications be the main motivation for 5G. Without 5G, services like virtual reality [1], self-driving vehicles [2], machine-to-machine communication [3], [4] would be impossible or very hard to provide; without said services, 5G networks would be uneconomical to develop and deploy [5], [6]. New services will integrate within the network rather than using it, i.e., they will use custom-made, virtualized *network slices* [7], [8] for their communication and computation needs. The companies offering such services will become *verticals* [9], [10], i.e., stakeholders rather than simple users of 5G networks.

In this context, comparatively little attention has been devoted to *present-day* applications and their relationship with 5G. Will they use 5G networks the same way they use LTE today, i.e., as a mere means to transfer data between mobile users and servers? Will they embrace the 5G paradigm, integrate within the network and become verticals in their turn? Our goal in this paper is to provide an answer to this important question.

To this end, we focus on one of the main innovations of 5G, namely, the multi-access edge computing (MEC) paradigm, providing *localized* support for high-bandwidth services [11], [12]. We then consider the existing mobile applications, and assess to which extent each can potentially benefit from having their demand served within the 5G network, i.e., in a network slice. Intuitively, services requiring high data rates could be better fit for 5G than those that do not; similarly, applications whose demand tends to be concentrated in space are more likely to reap substantial benefits from 5G integration.

In order to characterize the demand of existing mobile applications, we leverage a real-world, large-scale, crowd-sourced dataset, collected by the WeFi mobile app [13]. The dataset includes information about over 120 TByte of mobile traffic in the San Francisco Bay Area, including the position of individual users and the apps they use. We are therefore able to study the spacial features of data demand, e.g., its sparseness, and to perform such analysis on a per-app basis, thus identifying the discrepancies between different applications.

Characterizing present-day mobile apps also helps to identify the stakeholders of 5G. Many popular present-day mobile services are provided by over-the-top (OTT) giants like Facebook and Netflix, and some of these companies have still to define their attitude and strategy on 5G. It is likely that the features of their own demand will be one of the factors OTTs will take into account when deciding their level of integration and involvement in 5G, and such decision will clearly have a substantial impact on the development and deployment of 5G networks.

The remainder of this paper is organized as follows. We start by presenting the real-world dataset we use, in Sec. II. Then, in Sec. III we describe and formalize the metrics we are interested in. Finally, Sec. IV discusses the results we obtain, and Sec. V concludes the paper and summarizes ongoing work.

II. A real-world dataset

The dataset we use for our analysis is a real-world, large-scale, crowd-sourced trace coming from users of the WeFi app [13]. The app provides up-to-date, location-specific information on the available Wi-Fi networks and their features, e.g., throughput and security. At the same time, it collects information about the users and their activity, so as to improve the services offered.

The dataset comes as a collection of records, each containing the following information:

- time (date and hour) and GPS location;
- mobile operator, type of cellular connection (e.g., 3G or LTE), and cell identifier;
- SSID and BSSID of the Wi-Fi network the user is connected to (if any);
- application active on the user's smartphone;
- amount of data transmitted and received.

A new record is generated every time any of the above changes (e.g., the user moves to a new cell) or a one-hour period elapses.

ISBN 978-3-903176-02-7

TABLE I
THE WEFI TRACE.

Metric	Value
Covered area	$216 \times 203\text{km}^2$
Collection time	March 2016
Number of records	641 million
Unique users	45 492
Unique cells	578 157
Unique BSSIDs	478 080
Unique apps	64 297
Total traffic	121.81 TByte
Coverage	2% (WeFi estimate)

Tab. I summarizes the main features of the WeFi dataset, while Fig. 1 shows the area it covers, and the traffic at the different locations therein. As one might expect, the traffic demand is observed in the city of San Francisco, with other high-traffic locations corresponding to urban areas, e.g., Sacramento in the North-East of the map. Lower-density areas, e.g., Marin county north of San Francisco, have lower rates.

The crowd-sourced nature of the WeFi dataset is probably its most important feature, providing it with two main advantages over similar datasets provided by mobile operators [14], [15]. First, it can span different operators and technologies, including Wi-Fi. Even more importantly, it can include information on the individual apps active on users' smartphones, a type of data that technical and legal reasons prevent operators from collecting. The result is a view of mobile data demand that is both wider and deeper, instrumental in understanding not only its global behavior but also its individual components.

III. METRICS OF INTEREST

There are three main aspects of data demand we are interested in: the *total* traffic of each application, the average *data rate* it requires, and how *sparse*[1] such a demand is in space. In the following, we detail how we capture these aspects, either directly from the trace or through appropriate metrics. It is important to stress that all metrics are computed on a per-application basis, i.e., the traffic generated by different applications is considered separately.

Total demand. It is the most straightforward metric, and simply corresponds to the total amount of data downloaded by each application. We use it to identify the most relevant applications.

Peak rate. Our goal is to identify the applications that require high *peak* data rates, i.e., need to download large amounts of data in a short time at a specific location. To this end, we:

1) divide the topology into 100×100 m^2 *tiles*;
2) for each (tile, day, hour) triple, we compute the amount of data downloaded by each application;
3) for each application, we consider the 90th percentile of said quantity.

[1]A sparse demand will look like a sparse population, i.e., small amounts of demand scattered throughout a wide area.

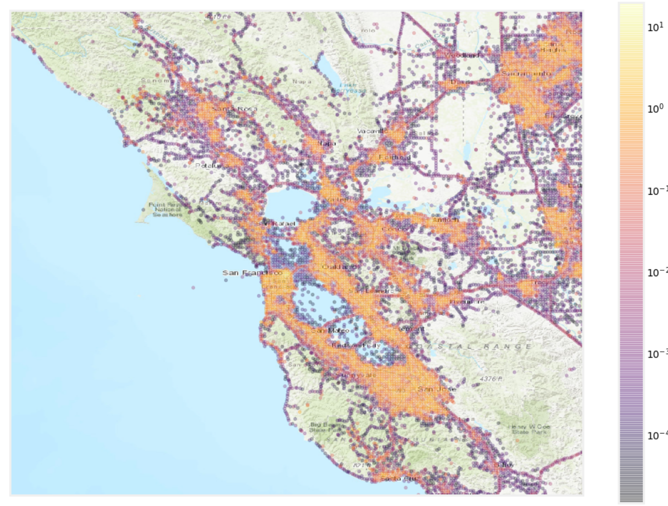


Fig. 1. The area covered by the WeFi dataset. Colors correspond to the total download rate, in kbit/s/km^2.

Sparseness. We are interested in assessing how sparse in space the demand for a certain app is. We therefore consider the same division in tiles as in the peak rate metric and, for each app, compute the total demand in each tile. We then rank the tiles by demand (highest to lowest), and define sparseness as the fraction of tiles that contain 50% of the total demand. As exemplified by Fig. 2, services that are required uniformly throughout the topology will have sparseness values close to 0.5, while services that tend to be required around certain locations will have lower sparseness.

IV. NUMERICAL RESULTS

First of all, we consider the distribution of the three metrics we described in Sec. III with respect to all the applications existing in the WeFi dataset. As depicted in Fig. 3(a) and Fig. 3(b), the distribution of total traffic and peak rate is almost power law-like (notice how the x-axes of both plots are logarithmic). As one might expect, there are many applications with low traffic and peak rate, and a few ones that consume much more data. Moving to Fig. 3(c), it is interesting to observe that the traffic of most applications, large and small, is very concentrated. This confirms that the space locality

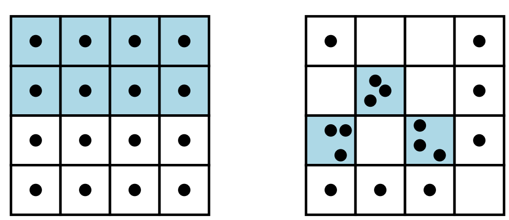
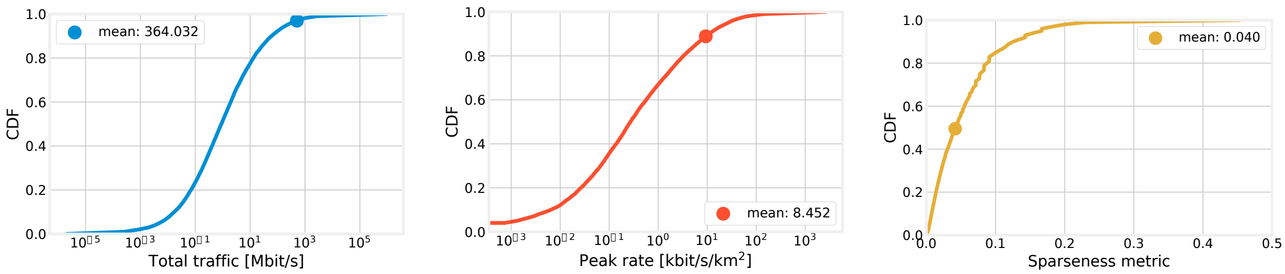


Fig. 2. How the sparseness metric works. In this example, the topology contains 16 tiles and 16 users, all downloading the same quantity of data. If the demand is uniform (left), we need 8 tiles to have 50% of the total demand, and thus obtain a sparseness value of $\frac{8}{16} = 0.5$. In the right case, the demand is not uniform, and the three tiles with the highest traffic contain $\frac{9}{16} > 50\%$ of the total demand, resulting in a sparseness value of $\frac{3}{16} \approx 0.19$.

ISBN 978-3-903176-02-7 © 2018 IFIP



(a) (b) (c)

Fig. 3. Comparison between apps: distribution of the total traffic (a), peak rate (b), sparseness (c).

TABLE II
HIGHEST-TRAFFIC APPLICATIONS, THEIR PEAK RATE AND SPARSENESS. APPLICATIONS WHOSE PEAK RATE AND SPARSENESS ARE (RESPECTIVELY) BELOW AND ABOVE THE AVERAGE ARE MARKED (RESPECTIVELY) IN RED AND BLUE.

Application	Traffic [GBit/s]	Peak rate [kbit/s/km^2]	Sparseness
YouTube	1006.12	103.66	0.04
Facebook	497.56	69.47	0.04
Netflix	487.39	169.27	0.03
XFinity PlayNow	389.26	862.05	0.03
Chrome	219.07	36.18	0.05
Instagram	143.23	46.57	0.04
XFinity CloudTV	63.34	2.10	0.00
Google Music	58.82	1.51	0.02
Google Maps	58.26	7.16	0.05
Snapchat	49.50	51.10	0.04
uTorrent	45.47	6.01	0.04
Hulu	38.60	219.64	0.02
Tumblr	33.68	141.39	0.02
Dice With Buddy	30.09	70.18	0.01
iHeartRadio	27.63	1.07	0.01
Pandora	23.40	4.68	0.04
Spotify	22.78	7.17	0.02
Google Videos	19.46	0.03	0.01
CNN Mobile	18.57	3.23	0.01
NBC LiveExtra	18.28	3.43	0.01
HBO	17.68	58.33	0.01
Weather.com	16.15	3.63	0.04
Twitter	16.03	1.96	0.03
FlipBoard	14.11	0.56	0.02
ESPN Pass	13.65	618.35	0.01
New York Times	12.21	35.97	0.05

principle also holds for present-day mobile applications, i.e., users of the same app are likely to be close to each other.

We now restrict our attention to the 25 apps with the highest total traffic, summarized in Tab. II. As one might expect, YouTube has the highest traffic demand, exceeding one terabit per second throughout the trace – and recall that our trace only accounts for a fraction of the total traffic. Somehow surprisingly, the app with the second highest demand is Facebook, mostly due to its recent decision to increase the presence of video in its users' feeds. Netflix is only third, with roughly half as much traffic as YouTube – but recall that our trace includes mobile devices like smartphones and tablets, but not desktops, laptops, or smart TVs.

Most of the other high-traffic applications fall into the streaming category, either on-demand (e.g., XFinity PlayNow and Google Music) or live (e.g., iHeartRadio and NBC LiveExtra). It is interesting to observe the presence of a browser (Google Chrome), Google Maps, and several social networking applications, including Twitter and SnapChat. Furthermore, both peer-to-peer file exchange (uTorrent) and games (Dice With Buddies) also rank among the highest-traffic application, a sign that already in 2016 the relevance of those services was significant.

A. Possible present-day verticals

We now go back to our original question, i.e., which of the high-traffic applications can gain the most from integration with 5G networks, and thus which providers have the potential to become verticals. To this end, we identify those applications that have a higher peak rate than the average *and* a lower sparseness. The resulting situation is summarized in Fig. 4, where each dot corresponds to a high-traffic app from Tab. II.

Many high-traffic applications have a peak bandwidth that is lower than the average, i.e., they fall in the red area in Fig. 4 (and are marked in red in Tab. II). An important example is Twitter, which has many users and a significant traffic, but is less rich in multimedia content than other social networks. Other applications (marked in blue in Tab. II and falling in the blue area of Fig. 4) have a high peak rate but are sparser than the average. Examples include the New York Times and Google Maps, indeed the type of apps one expects to be used more or less everywhere. Facebook and Snapchat, two other very frequently-used apps, have almost exactly the same sparseness as the average – slightly higher the former, slightly lower the latter.

Finally, the white area in Fig. 4 contains apps that (i) have a high total traffic; (ii) have a peak rate higher than the average, and (iii) are more concentrated in space than the average. The demand profile of these applications suits very well the features of 5G networks, and their owners can potentially obtain significant benefits from becoming 5G verticals. The list includes YouTube, major streaming services such as Netflix and Hulu, social networks like Instagram and Tumblr, along with games like Dice With Buddies and peer-to-peer clients like uTorrent. All together, these apps account for over 27% of the total traffic.

In conclusion, our analysis suggests that there is a significant set of present-day applications that would obtain significant benefits from integration within 5G. Such a set is important from a quantitative viewpoint, as it represents over a

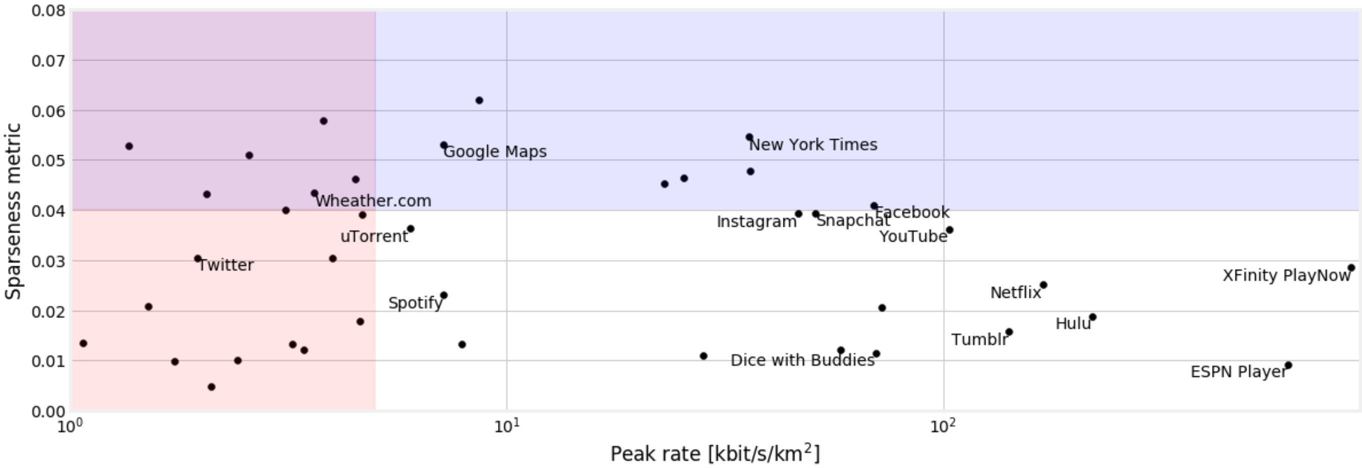


Fig. 4. Peak rate and sparseness of the highest-traffic applications.

quarter of present-day traffic, as well as from a qualitative one, since it includes apps belonging to OTT giants like Google, Facebook (which also owns Instragram), Netflix, and HBO. At a more general level, it suggests that 5G networks will be beneficial for present-day and emerging applications.

V. Conclusion and current work

We endeavored to study the relationship between present-day mobile applications and 5G networks, e.g., to which extent the former can benefit from tighter integration with the latter. To this end, we leveraged a large-scale, real-world, crowd-sourced mobile traffic trace, and classified the existing applications based on their total traffic, peak rate, and sparseness.

We found that a wide set of applications, representing a substantial fraction of the total traffic and belonging to major over-the-top content providers, have the potential to benefit from 5G integration. This suggests that the owners of said applications have an interest in becoming verticals, stakeholders of 5G networks instead of mere customers thereof.

Work is currently ongoing to extend our analysis to *how* data is served, i.e., to the relationship between backhaul infrastructure and the demand of mobile applications.

References

[1] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 110–117, 2017.

[2] X. Cheng, C. Chen, W. Zhang, and Y. Yang, "5g-enabled cooperative intelligent vehicular (5genciv) framework: When benz meets marconi," *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 53–59, May 2017.

[3] G. Y. Lin and H. Y. Wei, "Flexible 5g m2m network access with cognitive ran: Survey and design principles," in *2015 IEEE Conference on Standards for Communications and Networking (CSCN)*, Oct 2015, pp. 36–41.

[4] Y. Mehmood, N. Haider, M. Imran, A. Timm-Giel, and M. Guizani, "M2m communications in 5g: State-of-the-art architecture, recent advances, and research challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 194–201, 2017.

[5] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5g mobile and wireless communications: the vision of the metis project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[6] L. Chiaraviglio, N. Blefari-Melazzi, C. F. Chiasserini, B. Iatco, F. Malandrino, and S. Salsano, "An economic analysis of 5g superfluid networks," in *2017 IEEE 18th International Conference on High Performance Switching and Routing (HPSR)*, June 2017, pp. 1–7.

[7] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," *IEEE Communications Magazine*, 2017.

[8] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks," *IEEE Communications Magazine*, 2017.

[9] G. Biczok, M. Dramitinos, L. Toka, P. E. Heegaard, and H. Lonsethagen, "Manufactured by software: Sdn-enabled multi-operator composite services with the 5g exchange," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 80–86, April 2017.

[10] A. Mayoral, R. Vilalta, R. Casellas, R. Martinez, and R. Munoz, "Multi-tenant 5g network slicing architecture with dynamic deployment of virtualized tenant management and orchestration (mano) instances," in *ECOC 2016; 42nd European Conference on Optical Communication*, Sept 2016, pp. 1–3.

[11] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, "Enabling real-time context-aware collaboration through 5g and mobile edge computing," in *ITNG*, 2015.

[12] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and Its Role in the Internet of Things," in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, 2012.

[13] WeFi. https://wefi.com.

[14] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3G cellular data network," in *2012 Proceedings IEEE INFOCOM*, March 2012, pp. 1341–1349.

[15] P. D. Francesco, F. Malandrino, T. K. Forde, and L. A. DaSilva, "A sharing- and competition-aware framework for cellular network evolution planning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 2, pp. 230–243, June 2015.

Acknowledgment

This work is supported by the European Commission through the H2020 5G-TRANSFORMER project (Project ID 761536).