Beyond Throughput: a 4G LTE Dataset with Channel and Context Metrics

Darijo Raca, Jason J. Quinlan, Ahmed H. Zahran, Cormac J. Sreenan
Department of Computer Science, University College Cork, Cork, Ireland
{d.raca,j.quinlan,a.zahran,cjs}@cs.ucc.ie

ABSTRACT

In this paper, we present a 4G trace dataset composed of client-side cellular key performance indicators (KPIs) collected from two major Irish mobile operators, across different mobility patterns (static, pedestrian, car, bus and train). The 4G trace dataset contains 135 traces, with an average duration of fifteen minutes per trace, with viewable throughput ranging from 0 to 173 Mbit/s at a granularity of one sample per second. Our traces are generated from a well-known non-rooted Android network monitoring application, G-NetTrack Pro. This tool enables capturing various channel related KPIs, context-related metrics, downlink and uplink throughput, and also cell-related information. To the best of our knowledge, this is the first publicly available dataset that contains throughput, channel and context information for 4G networks.

To supplement our real-time 4G production network dataset, we also provide a synthetic dataset generated from a large-scale 4G ns-3 simulation that includes one hundred users randomly scattered across a seven-cell cluster. The purpose of this dataset is to provide additional information (such as competing metrics for users connected to the same cell), thus providing otherwise unavailable information about the eNodeB environment and scheduling principle, to end user. In addition to this dataset, we also provide the code and context information to allow other researchers to generate their own synthetic datasets.

CCS CONCEPTS

• **Information systems** → **Multimedia streaming**; • **Networks** → Public Internet; **Wireless access networks**; **Network measurement**;

KEYWORDS

Dataset, 4G, LTE, ns-3, Mobility, throughput, context information, adaptive video streaming

ACM Reference Format:
Darijo Raca, Jason J. Quinlan, Ahmed H. Zahran, Cormac J. Sreenan. 2018. Beyond Throughput: a 4G LTE Dataset with Channel and Context Metrics. In *MMSys'18: 9th ACM Multimedia Systems Conference, June 12–15, 2018, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/

1 INTRODUCTION

Since the dawn of the first wireless cellular network in late 70's mobile network evolution has exploded, resulting in capabilities and services beyond the original voice communication design. Forty years later, mobile handsets are part of our everyday routine with a wide variety of use cases, including office related tasks (reading and sending emails, making appointments), text messaging, web browsing, playing games and, consuming multimedia content. Mobile device usage has risen from 10% in 2011 to just over 36% by 2018 [20], with mobile data traffic growing 18x over the last five years. Furthermore, cellular data (4G) accounted for 69% of all mobile traffic in 2016, while 3G accounted for 24%, while cellular speeds grew 3x from an average of 2 Mbit/s in 2015 to 6.8 Mbit/s in 2016 [5]. With these rates expected to grow by orders of magnitude when the next iteration of the cellular standard, known as 5G, is deployed in 2020.

However current 4G data throughput rates can fluctuate over a period of few seconds, due primarily to scheduling decisions at the cell tower, and sudden changes in the underlying radio channel. These changes are caused by inter-cell interference, congestion due to a number of devices per cell, and location of the device relative to the cell edge. This throughput variation is inherently a part of the underlying communication system since the first wireless networks and will be further exacerbated in 5G due to technical issues such as non-line of sight and a reduction in overall transmission distance. This variations in throughput can limit the user Quality of Experience (QoE), especially when they cause visible degradation in viewable quality as can occur while streaming audio or video. Underlying network protocols can mitigate these issues, such as TCP whose design reflects throughput variation by embedding an exponential moving average (EWMA) statistic to adapt to rate-distortion [6]. Additionally, adaptation algorithms proposed for HTTP Adaptive streaming (HAS) [21] can further combat the challenge of consistent quality through buffering and graceful adaptation of video quality. One of the main hurdles for these adaptation algorithms is a lack of a broad cellular dataset that captures these throughput variations, especially when combined with channel and context metrics, on which a solution can be designed and compared with other state-of-art algorithms. Recently, researchers have recognised this problem, which resulted in a number of datasets collected over different wireless technologies and video content datasets [18].

In this paper, we present two datasets: the first collected from real 4G production networks and the second a synthetic dataset generated from a large-scale 4G ns-3 [2] simulation. In our production dataset, we collected traces from two major Irish mobile operators, with different mobility patterns (static, pedestrian, car, bus and train). Relative to our research into adaptive video streaming, our initial goal is to provide a standard dataset platform for comparison of various HAS streaming approaches. However, in addition to throughput values, we also collected information about channel condition for the client in respect to serving eNodeB and neighbouring cells, GPS positions of the client and serving eNodeB, client's speed, and handover events. All of this information allows a multi-purpose analysis beyond our original HAS use cases, such

MMSys'18, June 12–15, 2018, Amsterdam, Netherlands
2018.

as handover prediction, coverage analysis, mobility prediction etc. While in our synthetic dataset, we utilise a large-scale 4G ns-3 simulation that includes 100 users randomly scattered across a seven-cell cluster. The purpose of the synthetic dataset is to provide additional information (competing metrics for users connected to the same cell), abstracting eNodeB environment and scheduling principles, and ultimately provide a means of large-scale evaluation of key performance indicators in multi-cell mobility scenarios. To the best of our knowledge, our production dataset is the first publicly available dataset that contains throughput, channel and context information for 4G networks.

The remainder of this paper is organised as follows. Section 2 describes related work. The dataset collection and captured metrics are explained in Section 3, while Section 4 explores statistical traits of the production and synthetic dataset for different mobility patterns. In Section 5 we layout possible use cases, while Section 6 outlines future work and our conclusion.

2 RELATED WORK

Previous datasets in this area, focused primarily on the variance in available bandwidth and typically offered a very limited set of device metrics, such as velocity, GPS and signal strength. We begin with Bokani et al. [3], who offered a dataset, collected from 3G and 4G networks, consists of throughput measurements logged every ten seconds, a timestamp for same and GPS coordinates of the user device itself. The authors utilised a single mobile commute pattern in a metropolitan scenario, and repeated multiple trails within this pattern, warranted by the evidence that network throughput can vary significantly for the same route. They collected a large number of samples across the same path to get statistically significant results on network performance. However, their dataset has a low sampling granularity (ten seconds) and only contains throughput and a very limited set of device values.

Similarly, Xiao et al. and Li et al. collected bandwidth traces over 3G and 4G network respectively [11, 25]. In both papers, the authors use MobiNet[1], a custom developed non-rooted android application for downloading content using TCP. The majority of both datasets are collected in high-speed mobility environments (train) where speeds can rise to 310 kph. The content of the datasets consists of information such as application throughput, signal strength, device velocity and eNodeB id. Riiser et al. [17] obtained bandwidth logs from a 3G network using different mobility patterns; these included tram, train, metro, bus, ferry, and car. The dataset contains a sample granularity in the order of seconds and provides additional information such as timestamp, GPS coordinates of the device, and bandwidth throughput. Also, Hooft et al. [23] used the same approach for collecting 4G network traces for analogous mobility patterns, foot, bicycle, bus, tram, train, and car. However, all these traces focus on acquiring throughput values with high sample granularity. Even though collected in a wireless environment, none of the previous datasets contain any information about the cellular channel. In comparison to these papers, our dataset includes repeated trails as well as one-second sampling granularity from a more diverse set of routes (bus, pedestrian, train and commute routes) coupled with both channel and context metrics for improved cellular and device feedback.

In our research, these datasets provide sufficient throughput information to evaluate the performance of state-of-art HAS algorithms that determine the streamed video quality in response to changes in the operating conditions. These algorithms typically adopt a rate-based and buffer-based strategy. Rate-based algorithms base their decision for the next chunk rate on a series of the previously downloaded chunk's throughput, with FESTIVE [10] being a well-known algorithm utilising a rate-based approach. Buffer-based algorithms map playback buffer levels to the throughput rate for the next segment. BBA-1 and BOLA [8, 19] are one of a number of algorithms that rely on this technique. However, most state-of-art algorithms use a hybrid strategy, combining both rate and buffer-based methods [4, 28].

Xie et al. [26] recently use channel information from the wireless channel in addition to the throughput rate to make a more intelligent decision for the next segment quality. Also, a new strategy emerged recently, relying on throughput forecasting [29] to optimise a quality selection of segments. As a result, there are considerable efforts to accurately predict throughput for the next x seconds in the future by leveraging the channel information in addition to the throughput rate [27], and our own research [16]. Also, context information such as UE's GPS position, velocity, eNodeB GPS position and distance between UE and serving eNodeB can be used for user movement prediction and resource allocation [22]. Wang et al. [24] utilise these metrics for UE movement and direction prediction to minimise the number of handovers. For evaluating and comparing these novel techniques, new datasets are needed containing information beyond throughput, such as the channel and context metrics provided in the dataset presented in the paper.

3 DATASET COLLECTION

For the production dataset collection, we use the Android device G-NetTrack Pro mobile network monitoring tool[2]. This tool enables the capturing of various channel related key performance indicators (KPIs), context-related metrics, downlink and uplink throughput, and also cell-related information. The main advantage of this application is that it does not require a rooted phone. In contrast to G-NetTrack, Li and al. developed an open-source software tool MobileInsight [12] that can capture radio information directly from chipsets in real time. However, the software requires a rooted mobile phone and works with Qualcomm SoCs only. This tool is similar to proprietary Qualcomm's QXDM[3] diagnostic software. While the non-rooted aspect of G-NetTrack is beneficial, there are a number of limitations to the application. Firstly, the minimum granularity of collected samples is one second. Having low-resolution KPIs, e.g., can increase prediction error as reported in [16]. Secondly, the tool uses the standard Android library (*telephony* class) for reporting channel metrics. Implementation of these callback functions depends on the manufacturer of the mobile system on a chip (SoC) chipsets. Also, not all parameters are reported for different cellular technologies (2G/3G/4G). For our dataset, we test mobile

[1] http://www.wandoujia.com/apps/thu.kejiafan.mobinet

[2] http://www.gyokovsolutions.com/

[3] https://www.qualcomm.com/

devices from three major mobile chipsets manufactures, Qualcomm (*Snapdragon*), Samsung (*Exynos*) and Huawei (*Kirin*). Ultimately, the mobile device chosen is a Samsung J5, which provides a means of capturing all 4G network metric KPIs.

For our production dataset, we collected 135 traces for various mobility patterns across two major Irish operators, with different data limit caps. The first provider (*operator A*) gives unlimited 4G data, while the second provider (*operator B*) offers only 15GB per month. However, the second operator provides 60GB on social media including Youtube streaming. For the first mobile operator, we continuously download a file (connection-oriented, TCP) with an average duration of 15 minutes per trace (with a five-second pause after the download completes). We use the same approach for the second operator, but once the data cap is reached, we extend the approach by downloading content from Youtube. We generate a URL for the video from Youtube to exploit the higher data cap for social media. For each trial, regardless of measurement approach, we use large file (> 50MB) to allow the TCP sending window to ramp up to the maximum size. As stated, every sample is logged with one-second granularity. As a result, average trace duration is 15 minutes.

To provide a comparison between operators, we perform measurements trials for both operators at the same time (we use the same mobile device model to limit the impact of device hardware on throughput rate and channel metrics). This subset of traces permits comparison of mobile operators performances across different parameters (throughput and channel KPIs). Competing tests use the same download approach for both cellular operators (file or video download).

The following outlines the various KPIs within our production dataset:

- *Timestamp:* timestamp of sample
- *Longitude and Latitude:* GPS coordinates of mobile device
- *Velocity:* velocity in kph of mobile device
- *Operatorname:* cellular operator name (anonymised)
- *CellId:* Serving cell for mobile device
- *NetworkMode:* mobile communication standard (2G/3G/4G)
- *RSRQ:* value for RSRQ. RSRQ Represents a ratio between RSRP and Received Signal Strength Indicator (RSSI). Signal strength (signal quality) is measured across all resource elements (RE), including interference from all sources (dB).
- *RSRP:* value for RSRP. RSRP Represents an average power over cell-specific reference symbols carried inside distinct RE. RSRP is used for measuring cell signal strength/coverage and therefore cell selection (dBm).
- *RSSI:* value for RSSI. RSSI represents a received power (wideband) including a serving cell and interference and noise from other sources. RSRQ, RSRP and RSSI are used for measuring cell strength/coverage and therefore cell selection (handover) (dBm).
- *SNR*: value for signal-to-noise ratio (dB).
- *CQI:* value for CQI of a mobile device. CQI is a feedback provided by UE to eNodeB. It indicates data rate that could be transmitted over a channel (highest MCS with a BLER probability less than 10%), as the function of SINR and UE's receiver characteristics. Based on UE's prediction of the channel, eNodeB selects an appropriate modulation scheme and coding rate.
- *DL_bitrate:* download rate measured at the device (application layer) (kbit/s)
- *UL_bitrate:* uplink rate measured at the device (application layer) (kbit/s)
- *State:* state of the download process. It has two values, either I (idle, not downloading) or D (downloading)
- *NRxRSRQ & NRxRSRP:* RSRQ and RSRP values for the neighbouring cell.
- *Cell_Longitude & Cell_Latitude:* GPS coordinates of serving eNodeB. We use OpenCellid[4], the largest community open database providing GPS coordinates of cell towers.
- *Distance:* distance between the serving cell and mobile device in metres.

We perform 4G measurement trials (unless otherwise stated) across six different mobility patterns summarised in Table 1.

Table 1: Mobility Patterns

Type	Summary
Static	Static trials (indoor)
Pedestrian	Walking trials around Cork city, Ireland
Bus	Trials include urban and suburban cases
Car	Trials include urban and suburban scenarios
Train	Travelling between Cork - Dublin (240km) and Cork - Farranfore (75km). Combination of 3G and 4G.

4 DATASET OVERVIEW

Production Dataset In this section, we give a short overview of our dataset. We categorise our traces as commute traces as we collected the majority of traces during morning and evening hours while going from home to work and back, and begin with an overview of our trace models:

Static As the name implies, these traces were collected indoors with mobile devices being stationary. This scenario represents how people typically tend to use their smart devices. However, this case has the least appeal as the throughput is quite stable with relatively low variations.

Pedestrian Outdoor traces while walking around Cork city centre using a number of different routes. Characteristics of collected traces (average rate and standard deviation) are similar to the static case with slightly more variation due to channel condition and handovers.

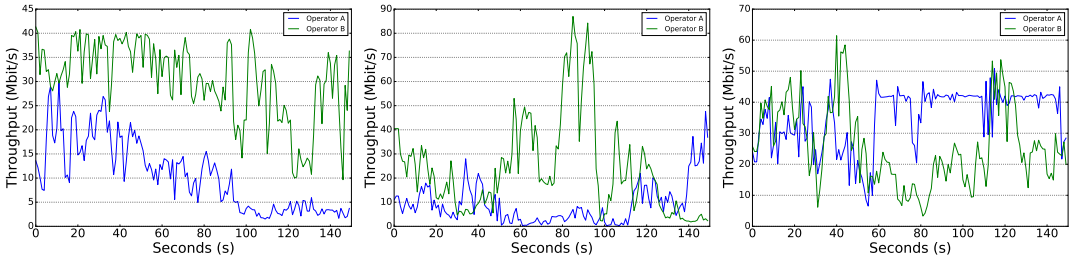
Bus Bus traces using public transport around Cork city. We gathered traces during weekdays and at the weekends to capture different congestion patterns.

Car Car traces over the city and suburban routes. This subcategory of our dataset contains the most traces.

Train While our goal is to collect 4G traces, a majority of the train traces are a mixture of 3G and 4G for both operators, due to the availability of 4G within major urban areas only.

We now provide a more detailed overview of the Throughput, Channel and Context information provided in our dataset:

[4]https://opencellid.org/



(a) Static **(b) Bus** **(c) Car**

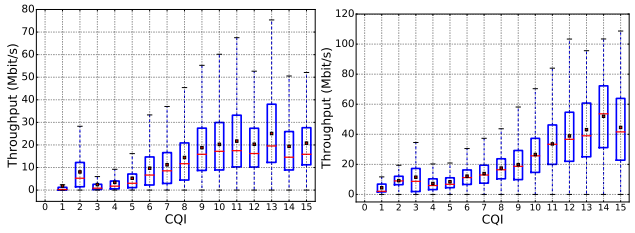
Figure 1: Time-series of application throughput for different mobility patterns and mobile operators

Table 2: Average and Variation Range of Application Throughput (Mbit/s) across different mobility patterns and mobile operators

	Mobility Patterns									
Operator	**Static**		**Pedestrian**		**Bus**		**Car**		**Train**	
	Avg.	*Var. Range*	*Avg.*	*Var. Range*	*Avg.*	*Var. Range*	*Avg.*	*Var. Range*	*Avg.*	*Var. Range*
A	5.3	(0.9, 9.3)	9.9	(0.4, 28.0)	8.0	(0.08, 20.3)	11.4	(0.92, 27.9)	4.7	(0, 11.3)
B	42.6	(21.3, 77.2)	18.2	(5.6, 34.2)	13.5	(2.0, 29.1)	22.3	(3.2, 49.1)	6.6	(0.3, 16.5)
Num. Traces	15		31		16		53		20	
Trace Dur. (m)	254		560		180		1265		650	

Throughput Figure 1 illustrates a time-series of application throughput for both network operators across different mobility pattern setups (we show randomly selected competing traces). Furthermore, Table 2 depicts average application throughput and variation including the number of traces and total trace duration across all traces for different mobility pattern categories and two mobility operators. By definition, variation range is a percentile-wise measure of variation. Let's define R as application throughput during time interval the $(t, t+1)$. Then we can define variation range as the interval $[R^L, R^H]$, where R^L represents a 10^{th} percentile of R, and analogously R^H a 90^{th} percentile of R [9]. This range defines boundaries where 80% of measured throughput lies. From the values shown in Table 2, operator B has a significantly higher average than operator A for all mobility pattern cases. There could be different reasons for this observation including better coverage, and operator's internal traffic policy (e.g., traffic limitation and shaping). Looking at each case individually, there are different changes in average value and variation range depending on the operator itself, e.g., for A, a static case has significantly lower average than the pedestrian case. A rationale for this result could be in coverage discrepancy for indoor and outdoor scenarios. We note that experiments run indoor have a weaker signal in 90% of cases.

Channel Measured throughput is a combination of the eNodeB environment (load, scheduler policy), wireless channel characteristics and mobile device receiver capabilities. Additional information about the channel environment in addition to throughput values can increase accuracy and granularity, paving a way to more accurate prediction. In Figure 2, we analyse this relationship and show boxplot of CQI against application throughput. Boxplot shows the range of throughput values for each CQI separately. Overall, we can observe an increasing trend in throughput proportional to CQI. However, the range of throughput values oscillates significantly for each CQI. Furthermore, for operator A, the average throughput of CQI equals 14 is lower than the throughput for CQI 15. A similar observation holds for operator B as well. Finally, this result is strengthened even more with the calculation of correlation between throughput and CQI, yielding a relatively low correlation coefficient of 0.6 and 0.38, for operator A and B, respectively. However, this correlation is even lower for other cases; in particular for the static case where the correlation coefficient equals 0.35. CQI is calculated on the mobile device (based on wireless channel condition) and represents the maximum rate the device can receive with low error. However, actual rate (number of allocated resources blocks per frame) is assigned by eNodeB (scheduler). Many factors can influence eNodeB decision, including the number of users, other users throughput demand, their CQIs values, etc.



(a) Operator A **(b) Operator B**

Figure 2: Boxplot of CQI vs application throughput for both network operators (car)

Context Our dataset provides additional context information such as device's GPS positions and velocity. Figure 3 shows a randomised selected train route from our dataset. We provide estimated GPS coordinates of serving eNodeBs and distance between them using Haversine formula.

Additionally, Table 3 shows average and variation range of device velocity across different mobility patterns and network operators. Intuitively, speed increases as we move from static (not shown) to

Table 3: Average and Variation Range of device velocity (kph) across different mobility patterns and mobile operators

	Mobility Patterns							
Operator	**Pedestrian**		**Bus**		**Car**		**Train**	
	Avg.	*Var. Range*	*Avg.*	*Var. Range*	*Avg.*	*Var. Range*	*Avg.*	*Var. Range*
A	2.4	(0.0, 4.0)	17.2	(0.0, 34.0)	23.7	(0.0, 54.0)	60.6	(0.0, 109.4)
B	1.5	(0.0, 3.0)	10.7	(0.0, 30.0)	35.1	(0.0, 56.0)	53.9	(0.0, 114.0)

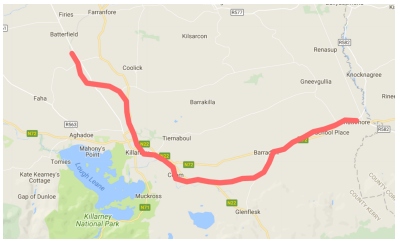
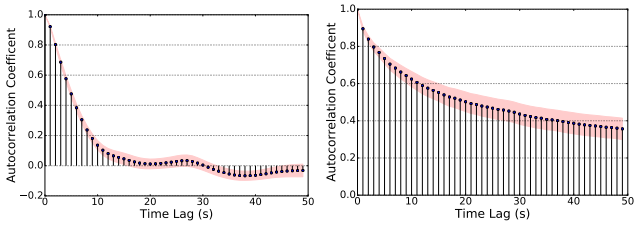


Figure 3: GPS coordinate for the train mobility pattern



(a) Simulation **(b) Real case (car)**

Figure 4: The autocorrelation coefficient of throughput

pedestrian and finally train scenario. A similar observation holds for variation range as well. Velocity values are alike for both network operators as the same phones/patterns were used for both operators.

Caveats This production dataset contains a considerable amount of information. However, there are a number of limitations. Firstly, our sampling granularity is only one second. This limitation is due to G-NetTrack and the Google channel API. Even with direct access to the API, granularity does not significantly increase [27]. Secondly, not all records have all values. The most prominent example represents RSSI, which doesn't get logged for every sample. Similarly, for geo-locations of eNodeB, we use opencell.org database. Unfortunately, this database doesn't contain GPS coordinates for all eNodeBs. One approach we use to deal with missing data is imputation methods [14].

Synthetic Dataset To supplement our real-time 4G production network dataset, we also provide a synthetic dataset generated from a large-scale 4G ns-3 simulation. As pointed out, our production dataset has medium sample granularity and only contains information gathered at the client. As an alternative, we provide simulation traces that have high granularity (250ms) and complement the simulation traces with network-side measurements. These additional pieces of information can only be collected at the network/operator, which in practice, is a ponderous task. We provide additional information such as competing metrics for users connected to the same cell, leveraging eNodeB scheduling principles, and ultimately provide a means of large-scale evaluation of key performance indicators in multi-cell mobility scenarios.

Due to space limitation, we provide a brief overview of our synthetic dataset, with full details of trace output, simulation testbed, and associated code and setup located at our website[5]. In our synthetic dataset, we utilise a large-scale 4G ns-3 simulation that includes 100 users randomly scattered across a seven-cell cluster. Every user has a constant moving speed (80 kph) and uses Gauss-Markov mobility model for movement emulation. Half of the devices are downloading at 32 Mbit/s rate, and the other half are uploading at 2 Mbit/s rate. We use UDP instead of TCP as the transport protocol. A motivation for this decision is a removal of any adaptation mechanisms from the client.

[5]http://www.cs.ucc.ie/misl/research/datasets/ivid_4g_lte_dataset/

Interestingly, if we compare autocorrelation coefficient of throughput for different time lags between simulation and the real case we conclude that simulation throughput exhibits more randomness, as depicted in Figure 4. This result is intuitive as simulation uses pseudo-random generators so one can expect this result.

5 POSSIBLE USE-CASES

In this section, we outline some of the possible use-cases for the dataset. We start with HAS algorithms, where our dataset enables the comparison of different algorithm strategies depending on the information they require for optimisation of chunk selection. Most algorithms calculate on throughput samples only, with some of them requiring finer granularity than chunk duration. However, going beyond throughput requirement, new strategies mandate channel and context information, allowing them to make more accurate throughput prediction. The proliferation of Commercial Virtual Reality (VR) technology is increasing download demands and is a distinct candidate for evaluation using our dataset. Although VR typically uses progressive download, it is expected that VR will switch to HAS mechanism in the near future [15]. This switch will result in the need for designing new adaptation algorithms suitable for VR specific needs (adapting the quality level of tiles).

Another use-case would be handover analysis and prediction. Handover procedure is crucial in cellular networks as it allows continuous connection across different eNodeBs. There are various mechanisms and approaches for handover prediction [1, 7, 13]. To benefit these approaches, our dataset contains information about handover events and also information about GPS position of current cell and device, channel metrics for serving and the neighbouring cell. Finally, generating new bandwidth traces based on the existing traces is a very interesting and demanding challenge, as multi-dimensional statistical analysis is needed over all available KPIs. For this task, one approach could be leveraging machine learning techniques. As a result, a large number of realistic traces would be generated and thus relieving researchers of manually collecting vast amounts of network traces, which can be a very tedious task.

6 CONCLUSION

In this paper, we present both production and synthetic 4G trace dataset, with low bandwidth throughput sampling granularity, and invaluable client-side cellular channel and context information, from a diverse set of routes across two mobile operators (production) and a large range of clients in a multi-cell cluster (synthetic). The throughput values of both datasets permit detailed analysis with respect to oscillation in the transmission medium, while the channel and context metrics of the production dataset far exceed the original goal of the dataset with respect to HAS evaluation for throughput prediction.

We provide a high-level overview of the dataset which provides insight into different mobility patterns across both mobile operators, with respect to application throughput, average and variation in bandwidth, and channel and context metrics. We also illustrate a number of possible use cases for the dataset. To the best of our knowledge, this is the first publicly available dataset that contains throughput, channel and context information for real-time analysis of a production 4G network.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Science Foundation Ireland (SFI) under Research Grant 13/IA/1892. We also thank Noel Bourke and Yusuf Sani for their invaluable help with the dataset collection.

REFERENCES

[1] I. M. Bălan, B. Sas, T. Jansen, I. Moerman, K. Spaey, and P. Demeester. 2011. An enhanced weighted performance-based handover parameter optimization algorithm for LTE networks. *EURASIP Journal on Wireless Communications and Networking* 2011, 1 (17 Sep 2011), 98. https://doi.org/10.1186/1687-1499-2011-98

[2] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero. 2011. An Open Source Product-oriented LTE Network Simulator Based on Ns-3. In *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '11)*. 293–298.

[3] A. Bokani, M. Hassan, S. S. Kanhere, J. Yao, and G. Zhong. 2016. Comprehensive Mobile Bandwidth Traces from Vehicular Networks. In *Proceedings of the 7th International Conference on Multimedia Systems (MMSys '16)*. Article 44, 6 pages.

[4] L. De Cicco, V. Caldaralo, V. Palmisano, and S. Mascolo. 2013. ELASTIC: A Client-Side Controller for Dynamic Adaptive Streaming over HTTP (DASH). In *2013 20th International Packet Video Workshop*. 1–8. https://doi.org/10.1109/PV.2013.6691442

[5] Cisco. 2017. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021. (2017). https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[6] B. A. Forouzan. 2002. *TCP/IP Protocol Suite* (2 ed.). McGraw-Hill, Inc., New York, NY, USA.

[7] H. Ge, X. Wen, W. Zheng, Z. Lu, and B. Wang. 2009. A History-Based Handover Prediction for LTE Systems. In *2009 International Symposium on Computer Network and Multimedia Technology*. 1–4. https://doi.org/10.1109/CNMT.2009.5374706

[8] T. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson. 2014. A Buffer-based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service. In *Proceedings of the 2014 ACM Conference on SIGCOMM (SIGCOMM '14)*. 187–198.

[9] M. Jain and C. Dovrolis. 2005. End-to-end Estimation of the Available Bandwidth Variation Range. In *Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '05)*. 265–276.

[10] J. Jiang, V. Sekar, and H. Zhang. 2014. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming With Festive. *IEEE/ACM Transactions on Networking* 22, 1 (Feb 2014), 326–340. https://doi.org/10.1109/TNET.2013.2291681

[11] L. Li, K. Xu, D. Wang, C. Peng, Q. Xiao, and R. Mijumbi. 2015. A measurement study on TCP behaviors in HSPA+ networks on high-speed rails. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. 2731–2739. https://doi.org/10.1109/INFOCOM.2015.7218665

[12] Y. Li, C. Peng, Z. Yuan, J. Li, H. Deng, and T. Wang. 2016. Mobileinsight: Extracting and Analyzing Cellular Network Information on Smartphones. In *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*. 202–215.

[13] W. Luo, X. Fang, M. Cheng, and X. Zhou. 2011. An optimized handover trigger scheme in LTE systems for high-speed railway. In *Proceedings of the Fifth International Workshop on Signal Design and Its Applications in Communications*. 193–196. https://doi.org/10.1109/IWSDA.2011.6159423

[14] R. Mazumder, T. Hastie, and R. Tibshirani. 2010. Spectral Regularization Algorithms for Learning Large Incomplete Matrices. *J. Mach. Learn. Res.* 11 (Aug. 2010), 2287–2322.

[15] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan. 2016. Optimizing 360 Video Delivery over Cellular Networks. In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges (ATC '16)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/2980055.2980056

[16] D. Raca, A. H. Zahran, C. J. Sreenan, R. K. Sinha, E. Halepovic, R. Jana, and V. Gopalakrishnan. 2017. Back to the Future: Throughput Prediction For Cellular Networks Using Radio KPIs. In *Proceedings of the 4th ACM Workshop on Hot Topics in Wireless (HotWireless '17)*. 37–41.

[17] H. Riiser, P. Vigmostad, C. Griwodz, and title = Commute Path Bandwidth Traces from 3G Networks: Analysis and Applications booktitle = Proceedings of the 4th ACM Multimedia Systems Conference series = MMSys '13 year = 2013 isbn = 978-1-4503-1894-5 location = Oslo, Norway pages = 114–118 numpages = 5 http://doi.acm.org/10.1145/2483977.2483991 doi = 10.1145/2483977.2483991 acmid = 2483991 publisher = ACM address = New York, NY, USA keywords = 3G, adaptive streaming, bandwidth traces, bitrate adaption, fluctuating bandwidth, mobile internet, wireless Halvorsen, P. [n. d.].

[18] Y. Sani, A. Mauthe, and C. Edwards. 2017. Adaptive Bitrate Selection: A Survey. *IEEE Communications Surveys Tutorials* 19, 4 (Fourthquarter 2017), 2985–3014. https://doi.org/10.1109/COMST.2017.2725241

[19] K. Spiteri, R. Urgaonkar, and R. K. Sitaraman. 2016. BOLA: Near-optimal bitrate adaptation for online videos. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2016.7524428

[20] Statista. 2016. Number of smartphone users worldwide from 2014 to 2020. (2016). https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[21] T. Stockhammer. 2011. Dynamic Adaptive Streaming over HTTP –: Standards and Design Principles. In *Proceedings of the Second Annual ACM Conference on Multimedia Systems (MMSys '11)*. 133–144.

[22] D. Stynes, K. N. Brown, and C. J. Sreenan. 2016. A probabilistic approach to user mobility prediction for wireless services. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*. 120–125. https://doi.org/10.1109/IWCMC.2016.7577044

[23] J. van der Hooft, S. Petrangeli, T. Wauters, R. Huysegems, P. R. Alface, T. Bostoen, and F. De Turck. 2016. HTTP/2-Based Adaptive Streaming of HEVC Video Over 4G/LTE Networks. *IEEE Communications Letters* 20, 11 (2016), 2177–2180.

[24] H-L. Wang, S-J. Kao, C-Y. Hsiao, and F-M. Chang. 2014. A moving direction prediction-assisted handover scheme in LTE networks. *EURASIP Journal on Wireless Communications and Networking* 2014, 1 (15 Nov 2014), 190. https://doi.org/10.1186/1687-1499-2014-190

[25] Q. Xiao, K. Xu, D. Wang, L. Li, and Y. Zhong. 2014. TCP Performance over Mobile Networks in High-Speed Mobility Scenarios. In *2014 IEEE 22nd International Conference on Network Protocols*. 281–286. https://doi.org/10.1109/ICNP.2014.49

[26] X. Xie, X. Zhang, S. Kumar, and L. E. Li. 2015. piStream: Physical Layer Informed Adaptive Video Streaming over LTE. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15)*. 413–425.

[27] C. Yue, R. Jin, K. Suh, Y. Qin, B. Wang, and W. Wei. 2017. LinkForecast: Cellular Link Bandwidth Prediction in LTE Networks. *IEEE Transactions on Mobile Computing* PP, 99 (2017), 1–1. https://doi.org/10.1109/TMC.2017.2756937

[28] A. H. Zahran, D. Raca, and C. Sreenan. 2018. ARBITER+: Adaptive Rate-Based InTElligent HTTP StReaming Algorithm for Mobile Networks. *IEEE Transactions on Mobile Computing* (2018), 1–1. https://doi.org/10.1109/TMC.2018.2825384

[29] X. K. Zou, J. Erman, V. Gopalakrishnan, E. Halepovic, R. Jana, X. Jin, J. Rexford, and R. K. Sinha. 2015. Can Accurate Predictions Improve Video Streaming in Cellular Networks?. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*. 57–62.