Analyzing Throughput and Stability in Cellular Networks

Ermias Andargie Walelgne*, Jukka Manner* Vaibhav Bajpai† and Jörg Ott†
*Aalto University, Finland , †Technical University of Munich, Germany
(ermias.walelgne | jukka.manner)@aalto.fi, (bajpaiv | ott)@in.tum.de

***Abstract*—The throughput of a cellular network depends on a number of factors such as radio technology, limitations of device hardware (e.g., chipsets, antennae), physical layer effects (interference, fading, etc.), node density and demand, user mobility, and the infrastructure of Mobile Network Operators (MNO). Therefore, understanding and identifying the key factors of cellular network performance that affect end-users experience is a challenging task. We use a dataset collected using `netradar`, a platform that measures cellular network performance crowd-sourced from mobile user devices. Using this dataset we develop a methodology (a classifier using a machine learning approach) for understanding cellular network performance. We examine key characteristics of cellular networks related to throughput from the perspective of mobile user activity, MNO, smartphone models, link stability, location and time of day. We perform a network-wide correlation and statistical analysis to obtain a basic understanding of the influence of individual factors. We use a machine learning approach to identify the important features and to understand the relationship between different ones. These features are then used to build a model to classify the stability of cellular network based on the data reception characteristics of the user. We show that it is possible to classify reasons for network instability using minimal cellular network metrics with up to 90% of accuracy.**

I. INTRODUCTION

According to the International Telecommunication Union (ITU) [1] (2015) report, there are more than nine billion mobile cellular telephone subscriptions in the world. Moreover, the International Data Corporation (IDC) prediction shows that, in 2017 smartphones will have a share of 70% of the total market of all device types, exceeding desktop and tablet computers [2]. There is also quite a variety of radio technologies (including LTE, HSPA, UMTS and HSPA+) with various range of bandwidth performance.

With continuous technology evolution and demand increase, predicting and understanding cellular network performance is challenging for several reasons. Performance depends on various factors: Hardware and operating system (OS) platforms, radio technology, the wireless link characteristics and the resulting bandwidth [3]; carrier networks and running applications [4]; user behavior and mobility [5]; and time of the day and location [6] all may influence the quality and performance of the cellular network. In addition, cellular network subscribers may be tied to bandwidth caps [7], [8].

Consequently, a prediction mechanism is desirable that (a) is based on comprehensive measurements able to capture the broad set of features while (b) not requiring large data volumes for measuring each individual data point. (b) is also important because it minimes the time each measurement takes (so that we obtain more exact point measurements if a user is moving) and it reduces the measurement overhead and memory footprint for storing each data record. The breadth of (a) is crucial to allow identifying the determinant factors of cellular networks. This may benefit both users and operators: Consumers may use such information as a guidance for choosing their mobile carrier networks and mobile phones, whereas mobile network operators can utilize such analysis to optimize their network to improve the end user experience.

The essence of this paper is to investigate factors affecting the throughput and stability of cellular networks based on a longitudinal dataset collected using the `netradar` [9] measurement platform. We start by analyzing different factors affecting the performance of TCP download and upload data rates in cellular networks. We then identify important features to classify the stability of a given cellular network based on the receive rate of a user. To classify network stability and predict bit rates, we apply a `randomForest` [10] machine learning algorithm tested with bootstrap and bagging. Our results show that it is possible to classify the stability of a given network using 'minimal information' (such as a device model, radio technology, signal strength, and battery level), which can be easily collected from mobile devices operating systems. The main **contributions** of this paper are:

1) We—expectedly—confirm that the time of a day and the location affect the throughput: in cities, the throughput drops during peak hours as congestion increases. We also observe that switches happens from legacy radio technologies to more advanced ones during the course of a day (see section IV).
2) We show that it is possible to predict the probability sudden drops in bit rate in a cellular network with 90% accuracy and 78% of kappa value, just relying on easily accessible information such as device model, location, network technology type, time of the day and latency (see section V).
3) We trained a model that can predict the average TCP download speed based on the first 5 seconds of median bit rate value of throughput (see section V-E).

In the remainder of this paper, we first review related work in section II and describe the dataset used in this paper in section III. We then present the analysis regarding cellular network thoughput as a function of various factors in section

Network Operator	Total # Of Measurements	LTE	HSPA/	HSPA+/	HSDPA	UMTS	Others
Elisa	373K	45.75%	8.49 %	14.70%	1.12%	25.43%	5%
DNA	235K	63.30%	7.52%	8.17%	10.69%	0.9%	9.42%
TeliaSonera	140K	34.74%	1.04%	14.27%	1.29%	33.60%	15.06%

TABLE I: Measurement Distribution per radio Technology for each MNO in Finland

IV and the machine learning model for stability prediction in section section V. We conclude this paper in section VI with a brief summary and review the lessons learned.

II. Related Work

Chen *et al.* [11] applied an ensemble learning technique called RuleFit [12] to identify important factors that affect the performance of 3G (UMTS) networks. Huang *et al.* [4] presented factors affecting the application performance of a cellular network for various smartphones. Nikravesh *et al.* [13] studied longitudinal patterns of cellular network performance and showed that there is a performance difference across factors such as network operators and radio technologies. Dräxler *et al.* [14] proposed a stochastic model used to anticipate the future data rates of a mobile device, which was then used to adjust the streaming video quality and time. Zou *et al.* [15] showed that the battery power of smartphones could be one factor impacting data throughput in cellular networks. Tuan *et al.* [16] also showed that network type, transfer packet size, and link quality impact smartphones' battery consumption. A study by Zou *et al.* [17] showed that it is possible to get accurate predictions of throughput (up to 98% accuracy) for short time periods by observing network performance on a stationary client device. Dziugas *et al.* [18] studied the characteristics of packet loss on UMTS network based on an active measurement data and showed that most of the packet loss occurred during state transitions in Radio Resource Control (RRC).

Most previous studies on cellular network performance and throughput prediction have been either limited to a specific set of application types, including P2P and video streaming [19], [20], or focused on a single network technology [18], [21]. Moreover, there is a gap in addressing the network performance related to the stability of the cellular link. Especially, little attention has been given towards predicting the stability of cellular connectivity, i.e., the probability of having a persistent data transmission in a given time interval. This paper close this gap by first studying the characteristics of individual network features to get the baseline understanding of the cellular network and then building a network stability classification and prediction model using minimal information, thus minimize the cost of data intensive measurements.

III. Data Set

We use the data set collected using `netradar` [9], a measurement platform for crowdsoursing information about cellular network performance from mobile devices. The tool measures and collects information including throughput (using TCP), signal quality and strength, radio technology type, round-trip time (using UDP towards Amazon Cloud instances and Aalto University servers), time and location, information about the base station and network operators (e.g., Cell Id, Mobile Network Code (MNC), and Mobile Country Code (MCC)) and information related to the mobile devices (hardware, Operating Systems (OS)). Each measurement records a series of records throughout a measurement session, not just a single value.

In the `netradar` dataset, we focused on studying the characteristics of the network when the user is not moving during a ten second measurement session (for details see [22]). We focus on stationary nodes to minimize the variability that might arise from mobility when we analyse link stability in section V. To do this, we eliminate all cellular data measurements which are not from stationary nodes. A node is considered stationary if the user's moving speed—as recorded for each measurement session—is zero. As the accuracy of user movement speed improves over time ((A)GPS is often only turned on only when a measurement starts), we took the last record to determine whether a node is moving or not.

We further limit the measurement data to the three Finnish mobile network operators covering one year (2015). Thereby, we can minimize the potential impact of diverse network infrastructures across countries. The resulting data set has a total of about 750K measurements with 13+K unique installation IDs (i.e., individual user devices). Table I summarizes the data set along with the distribution of the measured data points per network type and mobile network operator.

IV. Network Throughput

A. Device Model

Device manufactures continuously improve their mobile devices, changing hardware ingredients and form factor. Newer device models may bring a network performance improvements to the end user. The difference in chipset design, memory, processor (# cores, clock rate), graphics chips, screen resolution, etc. could influence the performance of applications running on the user device.

Usually, it is assumed that newly released devices come with better performance than the old ones. We use a device's year of release to confirm this hypothesis for throughput. We extracted the top device models for which we collected enough measurement samples using the `netradar` platform for each radio technology. Firstly, we examine how much the release of the new device model improved the performance of TCP downlink speed. Figure 1 shows the global TCP download speed trends for all device models colored by their year of release and aggregated per radio technology type used during the measurement. The figure shows no obvious direct

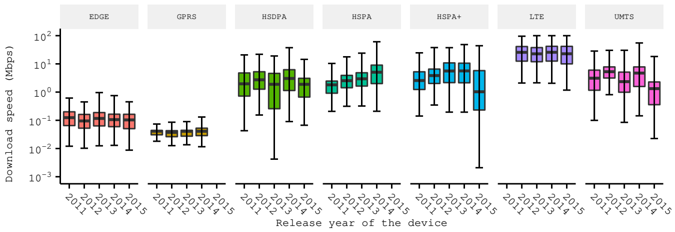

Fig. 1: TCP download speed of different device models per network technology. Except for the HSPA network, the release year of the device models do not correlate to TCP download speed.

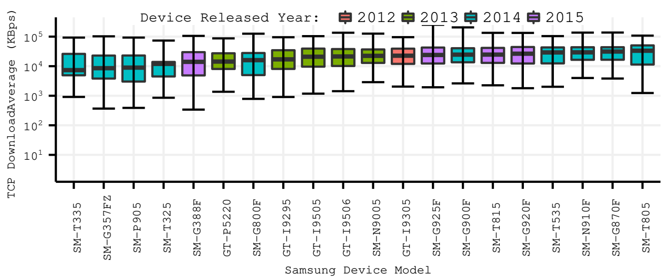

Fig. 2: TCP download speed of different Samsung device models (for LTE networks) sorted by median value. The colors show the respective release year of each device. We observe that even for the same brand it is not always the newest device model whose TCP download performance is best.

relationship between the devices year of release and TCP download speed for most of the radio technologies, except for HSPA. Due to the restrictions to read the radio technology type in iOS devices, we do not have enough samples for those and hence do not include iOS devices.

To further investigate the above (which we saw for all device types around the world), we look into individual device models based on their brand name. Figure 2 shows TCP download speeds for top 20 Samsung device models (based on the number of measurements collected). There are ~2K measurements per device model on LTE networks. We observe that the release year of the model has no direct impact on TCP downlink speed; even for the same brand and radio technology (Samsung and LTE, respectively, in this case). The figure shows that some old devices have better downlink speed than the most recent ones. For example, model `SM-G388F`, released in 2015, has a median downlink speed of about 14 MBps, while the older model `GT-I9305`, released in 2012, recorded a median downlink speed of 22 MBps. We observe similar network performance variation for other device brands, too. This indicates that irrespective of the radio technology, the type of devices and its specification remains equally important for the overall improvement of network performance.

We also notice that there is a high range of variation between different device brands accessing the same radio technology. For example, we observe that Sony's device model LT25i, which was released in 2012, scored a high download speed

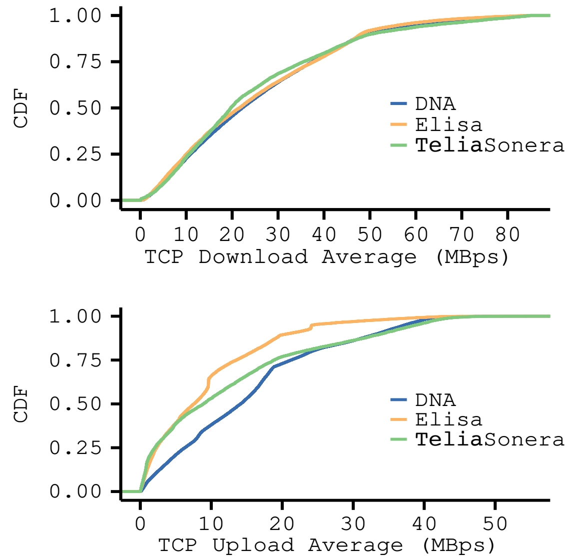

Fig. 3: Mean TCP throughput for LTE networks of 3 MNO downlink (top) and uplink (bottom) speed. There is clear difference between upload and download distributions (>1K users in each MNO).

with the median 21.69 MBps. This is much higher than the Samsung model of SM-T325, release in 2014; with median 13.41 MBps download speed for the same LTE network type.

B. Mobile Network Operator

This subsection investigates network performance variation based on three MNOs in Finland: Elisa, DNA, and TeliaSonera. We choose these MNOs because we have large enough measurements for most radio technologies as per in table I.

Figure 3 shows the CDF of TCP download and upload rates for LTE. We observe that the LTE network of DNA outperforms the other two MNOs in upload speed. There is also a considerable difference in download speed among MNOs; e.g., the median speed of DNA is 13% and 8% faster than TeliaSonera and Elisa, respectively. We applied the Anova [23] procedure to observe TCP uploading and downloading variance in each operator per radio technology: we filter equal sample size from each carrier for every radio technology. The Anova result shows that the overall p-value is very small, which allows us to reject the null hypothesis. For example, we observe that all carriers do not have identical mean uploading speed for LTE. The p-value of LTE uploading speed shows that the mobile network operator is of significant importance for network performance variation. To observe the uploading speed difference across operators, we further analyze the data using a pairwise t-test. The result shows that the difference in means is not significant between Elisa and TeliaSonera (p-value = 1), but both are significantly different from DNA's uploading speed (p-values of 1.6e-05 and 1.8e-05, respectively). Hence, we can conclude that the mean uploading speed of LTE is significantly different for DNA. Another multiple comparison method called Turkey's confirmed that the

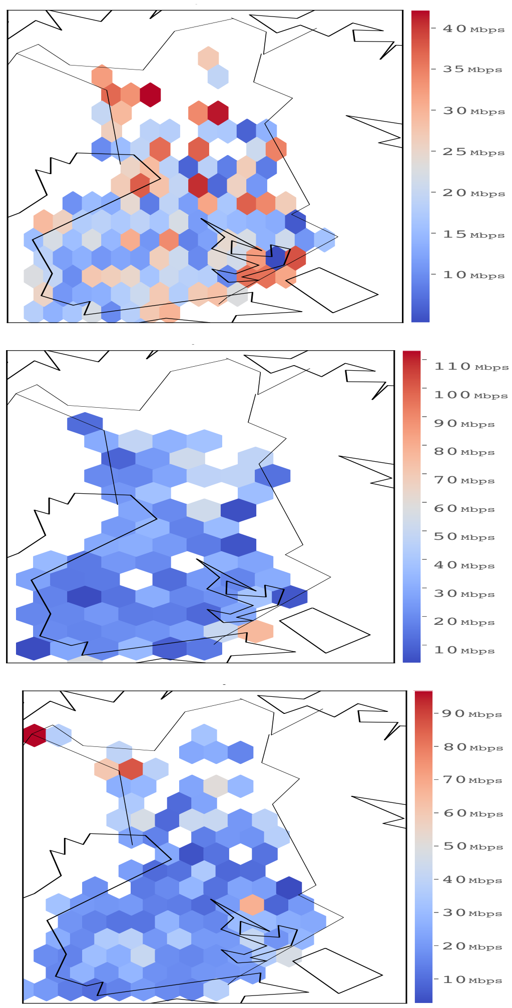

Fig. 4: Mean TCP throughput distribution by area in Finland for LTE networks of three MNO: Elisa (top), DNA (middle), TeliaSonera (bottom). The comparison across MNOs shows a large variation in throughput per locations. Note the different scales on the three plots.

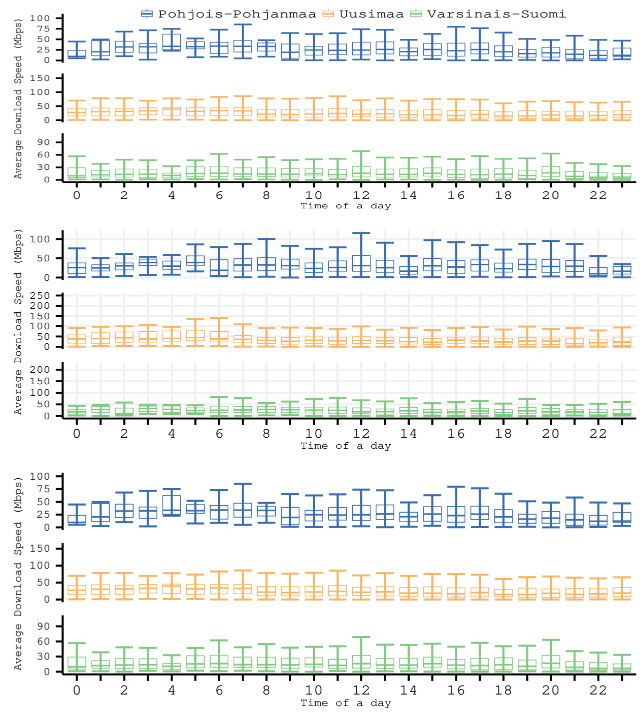

Fig. 5: TCP download distribution of LTE network in three major states for three mobile operators. Elisa (top), DNA (middle), TeliaSonera (bottom). Note the scale difference on the y-axis.

Elisa-DNA and TeliaSonera-DNA differences are significant (p-values of 0.0000156 and 0.0000181, respectively), while the TeliaSonera-Elisa difference is not (p-value = 0.9994570).

Figure 4 shows the geographical distribution of throughput values for three MNO's in Finland. It shows that there is a high variation in throughput by MNO for different locations in LTE networks. In the Uusimaa region, DNA has the highest median throughput of 28.31 Mbps, compared to Elisa and TeliaSonera that have 21.28 and 19.6 Mbps, respectively. In another area such as Pirkanmaa (Tampere), Elisa and TeliaSonera show 12.5% and 8,4 % median throughput improvement over DNA, respectively. We observe such variations in other areas, too. These observations suggest that MNO may support some locations with better technology and infrastructure based on their subscriber base.

C. Subscribers Location

We next investigate the impact of location and time of day on cellular network performance. For this analysis, we restrict the data set to measurements collected for three major provinces in Finland: *Uusimaa*, *Pirkanmaa*, and *Pohjois-Pohjanmaa*, which cover Southwest Finland (Helsinki, Espoo, and Vantaa), Tampere area and northern Finland, respectively. Each province has more than 94K, 45K, and 24K LTE based measurements, respectively. Figure 5 shows that the median download speed varies in all areas for the TeliaSonera network irrespective of the time of the day. Users that are living in the metropolitan area, i.e., Uusimaa state, and are subscribed to DNA or Elisa get a better download speed than users living in the other two areas. This is likely due to better infrastructure provisioning: higher base station density (within short distance of the users) combined with sufficient (core) network capacity.

DNA exhibits more stable throughput throughout the day than the other two operators in all of the areas. The download speed decreases starting from 10:00 in the morning in both TeliaSonera and Elisa network in all the three states. The difference between daytime and nighttime for the Elisa network is about 9% in the Uusimaa area. We explore a possible reason for this in the next section.

D. Radio Technology Changes and Time of Day

When studying the effect of radio technology switches in the measurement data, we observe that most devices that start with a network type of UMTS, HSPA, or HSDPA shifted to HSPA+ or LTE after few seconds of measurement. For instance, for the device model SM-T335, we observe that 74.36% and 23.5% of the measurements start with UMTS and HSPA+ network,

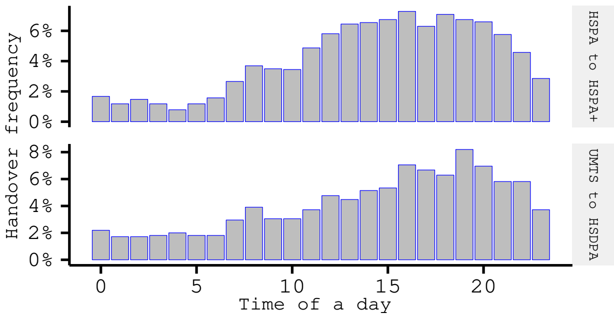

Fig. 6: Frequency of radio technology switches over time of day, for the TeliaSonera network (similar to other MNOs). The occurrence of switches (from legacy to more advanced technology e.g UMTS to HSDPA) increases during peak hours.

respectively. After the duration of few milliseconds (between 300 and 500 ms) more than 90% of UMTS measurements start using HSPA+ until the end of the measurement session. We observe a similar trend in most of the measurements and we further explore if this is related to the time of day. For this, we focus on the data of the specific MNO - TeliaSonera.

We observe that most HSPA+ network measurements start appearing at 10:00 in the morning and increase almost monotonically until 20:00 in the afternoon, at which point they start descreasing in number and switch to other technologies such as UMTS. On the other hand, the use of UMTS seems to follow the opposite pattern of HSPA+ as shown in Figure 6. The reason for such network behavior would be best explained if we knew the network operator's algorithms for resource management. However, looking at the trend, we can at least speculate that the ISP network runs some scheduling algorithm based on time of the day and number of users in a specific area. This could be also related to resource management based on the presence of users so that operators can save resources such as energy by switching off some base stations.

V. Network Stability

Stability of a given network link can be defined as the probability of having persistent connections over a given interval of time. We define the stability of a network in terms of the period of time t during which a given measurement session yields receive bit rates of zero in a row. We call this zero bit rate duration a *sudden dropout*. The longer the *sudden dropout* duration is, the more unstable the network is. The `netradar` measurement server tests the download speed by sending a random data over TCP for 10 seconds. During the measurement session, both the client and the server record the number of bytes transferred every 50 ms.

Different real-time and multimedia applications such as VoIP, online gaming, video-on-demand (VoD), among others, are very sensitive to bandwidth and latency variation, and often entail strict requirements on throughput. For instance, [24] states that for proper voice comprehension in VoIP calls the end-to-end delay should be within 150-300 ms. Another previous study [25] on web application response time show

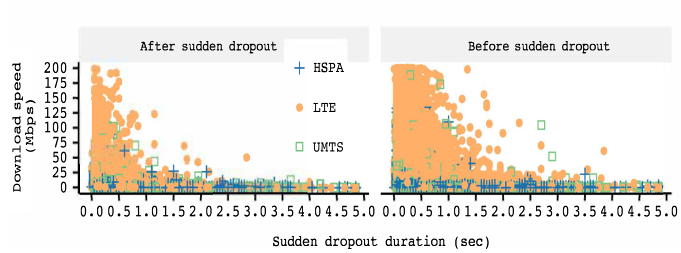

Fig. 7: TCP maximum download rate observed before and after a sudden dropout *of a certain duration (≥ 200 ms) per radio technology.*

that users feel the delay when the response time is more than 100 ms. *Sudden dropouts* obviously affect instant throughput but will also affect latency (if packets are queued in the network during a dropout) or lead to losses (if they aren't)—the latter of which may cause repair actions at the transport layer, which again turns into delay. In the `Netradar` dataset, we have observed network fluctuations and instability that could potentially affect user experience (e.g delay in response time) when using delay-senstive applications.

We take our observation from `Netradar` measurement data and results from previous work as a motivation and study how many of `Netradar` measurements in our dataset exhibit zero bit rate transmission for at least 200 ms. We choose 200 ms heuristically also considering suggestions from previous works work. Since the number of bytes downloaded is recorded at every 50 ms, a zero bit rate transmission for 200 ms means that at least 4 samples were missed or not received by the client. We understand that 200 ms might be relatively small to cause a major degradations for many applications. Later on, we will analyze and determine which *sudden dropout* duration will have a major impact on throughput.

We observe that, within the 10 second data transmission measurements, about 30% of cellular measurements suddenly dropped to zero bit rates that lasted for multiple recording intervals (in this case for at least 200 milliseconds and more) and sometimes for several seconds in a row. As mentioned earlier, one important question is how to determine the sudden dropout duration that would have a measurable impact on bit rate. Although the direct effect of sudden dropout duration on the user experience would vary based on the context and the type of application the user is running, we can still apply a heuristic approach to determine the value of dropout duration that has a maximum impact on bit rate transmission.

We observe the effect of the *sudden dropout* duration for different dropouts on our measurement data. For each session containing a dropout, we consider the maximum download bit rate recorded before and after the dropout. Figure 7 shows a scatter plot illustrating the effect of sudden dropout duration against the maximum TCP bit rate value before the dropout happens and after the dropout is over for different radio technologies. Note that, to avoid the variability due to a possible change in radio technology during the measurement session, we only consider measurements that used the same radio technology type, at least at the beginning and at the end

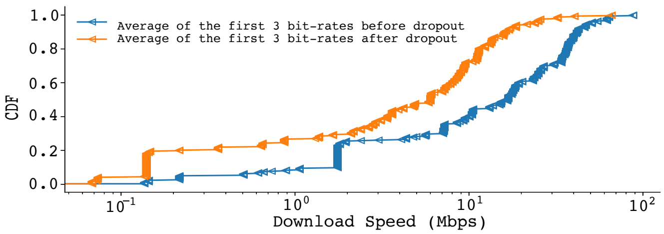

Fig. 8: Average of the first 3 bit rates samples before and after a sudden dropout.

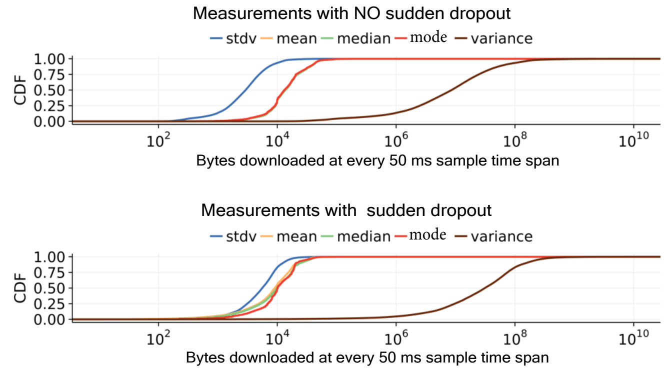

Fig. 9: Bytes downloaded during each 50 ms sample times pan.

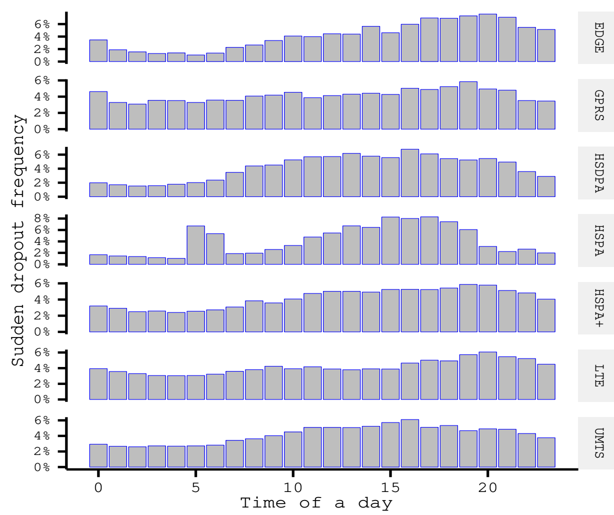

Fig. 10: Sudden dropout *occurrences per network technology over time of a day.*

of the measurement sessions.

During our analysis, we skip the first 5 seconds of each measurement session so that the TCP slow start phase is over [22] and the measured throughput becomes stabilizes. Then, we run an algorithm to detect such sudden dropout. We observed that a sudden sudden dropout duration that stayed for at least 200 ms does have an impact on download bit rate. As shown in figure 7, this impact is visible especially after the dropout period is over (right side of the figure).

Figure 8 shows the average of the first 3 bit rate samples before and after the *sudden dropout*. The average throughput of the three samples received after *sudden dropout* is consistently lower than before the *sudden dropout*. This shows that the effect of *sudden dropout* is reflected even after the *sudden dropout* (zero bit rate) is over, which may be a result of TCP recovery mechanisms.

A. Received Data per Recording Interval

We now study the distribution (the variation) of the number of bytes received every 50 ms (on average) within the 10 seconds of the measurement session. For this, we first classified the data into two groups, `dropped` and `non-dropped`. The former refer to measurement sessions that experience a sudden dropout for more than 200 ms, the latter refer to sessions without this phenomenon. Then, we compute the number of bytes received every 50 ms. Again, to avoid the effect of TCP slow start, only the last 5 seconds of the measurement are considered. Figure 9 shows that the mean, mode and median of `dropped` measurements diverge with a relatively higher standard deviation than the `non-dropped` measurements. This indicates that more network inconsistency and jitter is present in the `dropped` measurements than in `non-dropped` ones, which could increase the likelihood of application disruptions.

B. Sudden Dropout by Network Technology

We observe that *sudden dropouts* happen mostly in UMTS and HSPA networks especially during daytime from 11:00 to 19:00 hours as shown in figure 10. The distributions of *sudden dropout* per geographical area are almost similar except for a few island areas that experience more sudden dropouts out than metropolitan region. We observe that the Southwest Finland (Varsinais-Suomi) area and metropolitan (Uusimaa) areas have a median of 0.2 s and 0.15 s *sudden dropout* duration with LTE networks, respectively. Note that both areas have almost the same number of measurements.

C. Switching Radio Technology

We study the effect of changing radio technology in terms of download bit rate for those measurements that showed a *sudden dropout*. We classify the dataset into two: one covers those measurement sessions during which the mobiles change their radio technology at least once and those during which the radio technology remains unchanged. We then compare the bit rates measured immediately after the *sudden dropout* is over for both cases as well as the maximum receive bit rate after the sudden dropout.

Figure 11 shows the bit rate immediately after the sudden dropout is over as well as the maximum bit rate observed during the rest of the sessions after a the sudden dropout with respect to switching the access technology. It shows that for those that do not change the radio access technology, there is a more variability between the immediate bit rate and the maximum bit rate, whereas those that do change

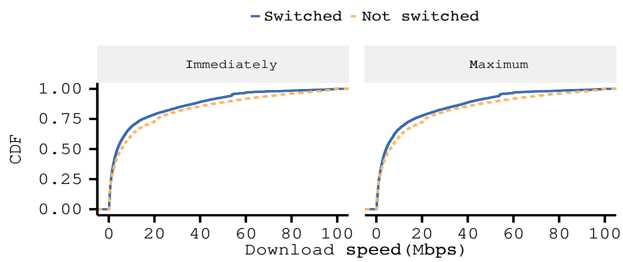

Fig. 11: Immediate and maximum bitrate after radio technology switch. Distribution of bytes downloaded during each 50 ms sample time span after sudden dropout *session is over. Those not changing radio technology have a marginal advantage.*

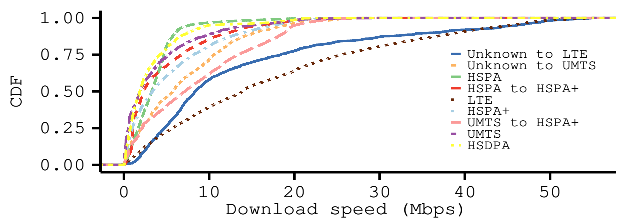

Fig. 12: Impact of radio technology switches for download speed.

their radio technology are more stable. We also observe that those that remain using the same network technology have a chance to get maxim bit rate after the dropout than those that do change their access technology. When we compare the immediate bit rate value after the *sudden dropout* session is over, those that do not change their radio technology have got the slowest download speed than those that change their access technology. As shown in figure 12, we find that while keeping the cell ID unaltered, a switch in radio technology from 3G to 4G causes a significant variation in TCP download speed. Especially a radio technology change from UMTS to LTE or HSPA or HSPA+ reduces the download speed. Even for small *sudden dropout* duration, if the measurement was starting with UMTS radio technology, it has a high probability of changing the radio technology into some other technologies.

D. Feature Selection and Classification Model

As we discussed in previous sections, cellular network performance depends on a number of factors. Most of these factors—or features—are (at least partly) related with one another. The relations among these features may not be of simple linear nature. Therefore, determining the major cause for the effect of a *sudden dropout* is non-trivial. In order to address these challenges, we use machine learning to identify the important network features that indicate (upcoming) network interruption. The machine learning approach is also useful to understand the interdependence if various network features. These identified network features can then be used to build a classification algorithm.

At first, we seek to predict the probability of network stability through classification. We pre-processed the dataset by classifying it into two categories based on the duration

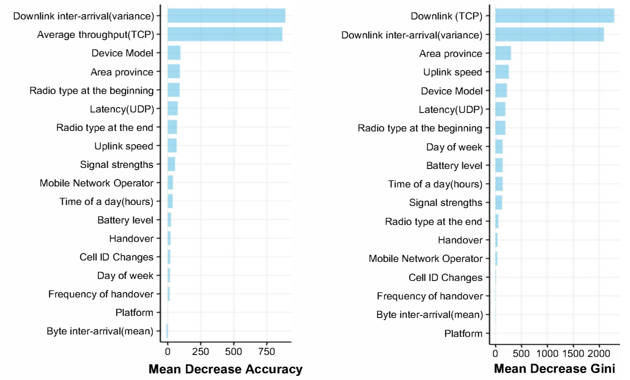

Fig. 13: Importance of network features for the classification model when using throughput.

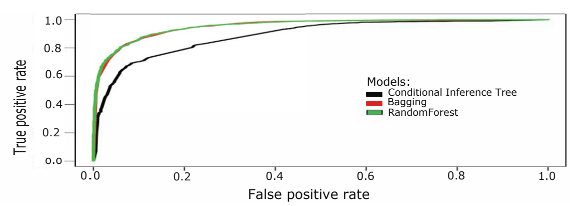

Fig. 14: True positive and false positive rates of the three classification models; random forest shows the bestROC curve.

of sudden bit rate dropout. The two categories are classified similar to the approach we used in V-A. Measurements which experienced a continuous zero bit rate transmission for at least 200 ms or more in a row are classified as `dropped`, the others as `non-dropped`.

Considering our dataset (which is a mix of both numerical and categorical variables) we make different tests for selecting a training model. We evaluate three prediction models namely, bagging based classification [26], random forest, and conditional inference tree [27] using different features (i.e TCP downlink speed, latency, radio technology, time of the day, location, operator network, signal strength, base station information, OS platform, device information and battery level). We also consider the technology switch happening during the measurement session. For this, we take the radio technology that was accessed at the beginning and at the end of the measurement into account. Finally, we consider how many times does the radio technology change during the test.

We trained the model with and without using TCP throughput values. The two types of feature importance measures are shown in figures 13 (when throughput is used to train the model) and 15 (without using the throughput). The *meanDecreaseAccuracy* tests show how the classification model performs without each variable. The results show that TCP download and upload rate average, latency, signal strength starting network technology, mobile network operator, and device model type were among the top important predictive variables for classification. *MeanDecreaseGini* measures the mean decrease in node impurity and a high score indicates that predictive variables were important. The variable features depicted in figure 13 show the overall features (variables)

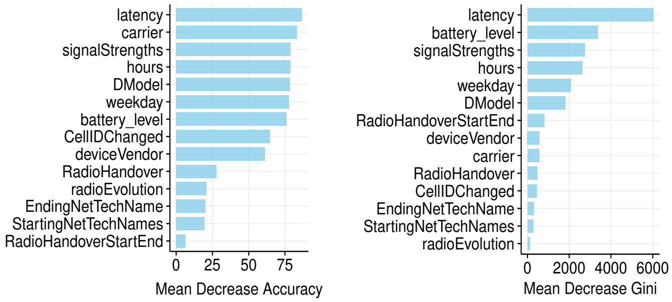

Fig. 15: Importance of network features used for the classification model without using TCP throughput.

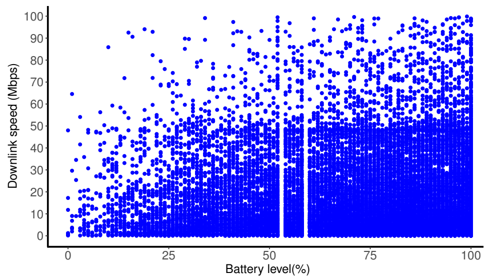

Fig. 16: Battery level vs. throughput.

and their importance. However, since throughput has a direct influence on the sudden dropout duration, using throughput as a predictor variable would create a bias. Hence, we trained our model without using the throughput features. The importance of features without using throughput is shown in figure 15. Note that all the results reported in the paper are training result without using the throughput.

We split the data into a training dataset (75%) and a validation (testing) data set (25%) using [28]. After experimenting with the three different learning algorithms, we find that random forest with bootstrap aggregation and bagging produces better prediction with accuracy of 90% and error rate of 10.2% on the testing dataset. Figure 14 shows the ROC curve of `randomForest`, `Bagging`, and `Conditional Inference Tree` with accuracy of 90%, 87%, and 77%, respectively. These classification models are trained using the aforementioned network features. The random forest implementation offers some flexibility for improving the classification accuracy by combining different sets of classifiers into ensembles. We trained random forest using tree size of 5K and by randomly sampling 9 variables at each split.

E. Predicting the Average Throughput

An accurate prediction of bit rates is important for different applications. For instance, it can be applied to anticipate the future data rate when using video streaming and thus adapt the streaming quality, accordingly. A previous study based on bit rate estimations also demonstrated that an accurate prediction can improve video performance in cellular networks [17].

Specifically, we want to study how to predict the overall mean bit rate only using the first five seconds of TCP bit rate measurement. We applied the caret [28] package of R to train the model. After trying different models including tree-based, GBM and bootstrapping, we found that random forest is (again) the best predictor model. We tested the training model with different variable numbers sampled as candidates at each split point of the random forest tree. It turns out that `mtry` of 67 is the best predictor model for the mean bit rate with Root Mean Square Error (RMSE) of 0.003 Mbps and 98% R-squared value. The model is cross-validated with both testing and validating datasets. This is important because we manage to predict the user throughput, using only the first five seconds download rates (which would be easily available in practice right after startup). We also observe that a relation exists between the amount of battery level in the smartphone and the average download rate for an LTE network as shown in figure 16. This implies that the battery level contributes a considerable value during both predicting the bit rate and classifying the stability of a cellular network, which is confirmed by figure 15 showing the feature importance ranking.

VI. Conclusion

We presented an analysis of end-to-end throughput of cellular networks using one year of measurements from mobile devices. We show that the throughput observed by a cellular network user depends on various factors such as device model, location and time of the day where metropolitan areas during peak hours showed more drops throughput. By examining the data received by the client every 50 ms, we find that about 30% of the total measurements experience sudden drops to zero bit-rate for at least 200 ms. TCP is designed to deliver data reliably and its performance is sensitive to losses and latency and jitter, so that a *sudden dropout* could create noticeable performance degradations and also real-time applications using other transports would be affected. We classified TCP throughput stability based on *sudden dropouts* only using device model, location, network technology type, time of the day and latency. This allows inferring network stability using the "minimal information" (e.g without measuring download and upload rates), which in turn reduces measurement overhead. We also trained a model that can predict the average TCP throughput based on the throughput of the first five seconds (mostly TCP slow start phase). Such models are useful to anticipate (near) future performance and to adjust application demands (e.g., streaming quality) accordingly. As part of our future work, we are working on extending our predictive approach to build adaptive video streaming that look-ahead off time and predict the throughput based on geospatial crowdsource measurement of `netradar` dataset.

Acknowledgements

This work is partially funded by the FP7 Marie Curie Initial Training Network (ITN) METRICS (grant no. 607728) and the Horizon 2020 project SSICLOPS (grant no. 664866).

References

[1] ITU. (2017) Global ICT developments ICT Facts and Figures The world in 2015 . [Online]. Available: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

[2] IDC, "IDC idc 50th anniversary transformation everywhere," 2017. [Online]. Available: https://www.idc.com/IDC_50anniversary/download/IDC50thAnniversary.pdf

[3] M. C. Chan and R. Ramjee, "TCP/IP Performance over 3G Wireless Links with Rate and Delay Variation," *Wireless Networks*, vol. 11, no. 1-2, pp. 81–97, 2005. [Online]. Available: https://doi.org/10.1007/s11276-004-4748-7

[4] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl, "Anatomizing application performance differences on smartphones," in *MobiSys*, 2010, pp. 165–178. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814452

[5] E. Halepovic and C. L. Williamson, "Characterizing and modeling user mobility in a cellular data network," in *Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, PE-WASUN*, 2005, pp. 71–78. [Online]. Available: http://doi.acm.org/10.1145/1089803.1089969

[6] A. Markopoulos, P. Pissaris, S. A. Kyriazakos, and E. D. Sykas, "Cellular Network Performance Analysis: Handoff Algorithms Based on Mobile Location and Area Information," *Wireless Personal Communications*, vol. 30, no. 2-4, pp. 97–117, 2004. [Online]. Available: https://doi.org/10.1023/B:WIRE.0000049393.01548.30

[7] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 91–99, 2012. [Online]. Available: https://doi.org/10.1109/MCOM.2012.6353688

[8] C. Joe-Wong, S. Ha, S. Sen, and M. Chiang, "Do Mobile Data Plans Affect Usage? Results from a Pricing Trial with ISP Customers," ser. Passive and Active Measurement PAM, 2015, pp. 96–108. [Online]. Available: https://doi.org/10.1007/978-3-319-15509-8_8

[9] "Netradar measurement platform," http://www.netradar.org/.

[10] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf

[11] Y. Chen, N. G. Duffield, P. Haffner, W. Hsu, G. Jacobson, Y. Jin, S. Sen, S. Venkataraman, and Z. Zhang, "Understanding the complexity of 3G UMTS network performance," ser. IFIP Networking Conference, 2013, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/document/6663506/

[12] J. H. Friedman and B. E. Popescu, "Predictive learning via rule ensembles," *The Annals of Applied Statistics*, pp. 916–954, 2008. [Online]. Available: http://www.jstor.org/stable/30245114

[13] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile Network Performance from User Devices: A Longitudinal, Multidimensional Analysis," ser. PAM, 2014, pp. 12–22. [Online]. Available: https://doi.org/10.1007/978-3-319-04918-2_2

[14] M. Dräxler, J. Blobel, and H. Karl, "Anticipatory Download Scheduling in Wireless Video Streaming with Uncertain Data Rate Prediction," ser. IFIP Wireless and Mobile Networking Conference, iWMNC, 2015, pp. 136–143. [Online]. Available: https://doi.org/10.1109/WMNC.2015.21

[15] R. Mittal, A. Kansal, and R. Chandra, "Empowering Developers to Estimate App Energy Consumption," ser. Mobicom, 2012, pp. 317–328. [Online]. Available: http://doi.acm.org/10.1145/2348543.2348583

[16] T. A. Dao, I. Singh, H. V. Madhyastha, S. V. Krishnamurthy, G. Cao, and P. Mohapatra, "TIDE: A User-Centric Tool for Identifying Energy Hungry Applications on Smartphones," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1459–1474, 2017. [Online]. Available: https://doi.org/10.1109/TNET.2016.2639061

[17] X. K. Zou, J. Erman, V. Gopalakrishnan, E. Halepovic, R. Jana, X. Jin, J. Rexford, and R. K. Sinha, "Can Accurate Predictions Improve Video Streaming in Cellular Networks?" ser. HotMobile, 2015, pp. 57–62. [Online]. Available: http://doi.acm.org/10.1145/2699343.2699359

[18] D. Baltrunas, A. Elmokashfi, and A. Kvalbein, "Dissecting packet loss in mobile broadband networks from the edge," ser. INFOCOM 2015, 2015, pp. 388–396. [Online]. Available: https://doi.org/10.1109/INFOCOM.2015.7218404

[19] S. Liu, W. Jiang, and J. Li, "Architecture and Performance Evaluation for P2P Application in 3G Mobile Cellular Systems wwwaw," ser. WiCom, Sept 2007, pp. 914–917. [Online]. Available: http://ieeexplore.ieee.org/document/4340010/

[20] Q. Zhang, W. Zhu, and Y. Zhang, "Network-adaptive scalable video streaming over 3G wireless network," ser. Proceedings of the 2001 International Conference on Image Processing, ICIP, 2001, pp. 579–582. [Online]. Available: https://doi.org/10.1109/ICIP.2001.958184

[21] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," ser. MobiSys, 2012, pp. 225–238. [Online]. Available: http://doi.acm.org/10.1145/2307636.2307658

[22] S. Sonntag, J. Manner, and L. Schulte, "Netradar - Measuring the wireless world," ser. 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, WiOpt 2013, 2013, pp. 29–34. [Online]. Available: http://ieeexplore.ieee.org/document/6576402/

[23] A. Agresti and M. Kateri, "Categorical data analysis," ser. International Encyclopedia of Statistical Science, 2011, pp. 206–208. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_161

[24] M. A. El-Gendy, A. Bose, and K. G. Shin, "Evolution of the Internet QoS and support for soft real-time applications," *Proceedings of the IEEE*, vol. 91, no. 7, pp. 1086–1104, 2003. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/1215688/

[25] J. NIELSEN. (1993, January) Response Times: The 3 Important Limits. [Online]. Available: https://www.nngroup.com/articles/powers-of-10-time-scales-in-ux/

[26] A. Peters and T. Hothorn, "ipred: Improved predictors," 2017, r package version 0.9-6. [Online]. Available: https://CRAN.R-project.org/package=ipred

[27] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical statistics*, vol. 15, no. 3, pp. 651–674, 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/106186006X133933

[28] M. Kuhn, "Building Predictive Models in R Using the caret Package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: https://www.jstatsoft.org/article/view/v028i05