Characterizing the Effects of Rapid LTE Deployment: A Data-Driven Analysis

Kareem Abdullah*, Noha Korany*, Ayman Khalafallah*, Ahmed Saeed†, Ayman Gaber‡
*Alexandria University, †Georgia Institute of Technology, ‡Vodafone Egypt

***Abstract*—The deployment of LTE, while old news to developed countries, is ongoing in developing countries with some starting LTE deployments only a year ago. LTE is essential for developing countries where broadband access is mostly available through wireless connection. The process of LTE deployment is costly and time consuming. Network operators attempt to minimize both overheads. With the current challenges facing network operators to acquire new sites, 4G deployment always depends on reusing the already existing 2G/3G macro layer. In this paper, we study the approach of rapid 4G deployment by analyzing data from a major network operator in a developing country. In particular, we look at network measurements from a single large cluster right after LTE deployment. We analyze the impact of the deployment approach on user throughput and find that LTE throughput could be improved further for the majority of LTE users in more than 50% of the studied cells by improving either interference or coverage. We find that this is mainly because the LTE system is more sensitive to interference when compared to 3G. Finally, we show a data-driven approach to detect the affected cells and mitigate the issue through physical optimization, this approach balances employed LTE transition best practices with cost efficiency and rapid deployment.**

I. Introduction

The deployment of 4G/LTE networks marked a historic transition in how people consumed digital content on their phone [1]. This is especially critical for developing countries where reliable and affordable broadband access might be limited to cellular networks [2]. This transition is somewhat old news in developed countries (e.g., US and Japan had their initial deployments in 2010 [3]). However, in developing countries this transition is fairly recent and ongoing (e.g., El Salvador [4] and Egypt [5] had their initial deployments in 2017). The deployment of 4G/LTE was marked by fierce competition between network operators trying to reach the market first. This competition can determine the benefits gained by the operator from the network upgrade [6]. That requires rapid deployment of 4G network, cost efficiency and attempting to exploit the full benefit of 4G spectrum efficiency.

With 5G deployments already underway, 4G deployment techniques remain relevant for several reasons. First, 5G deployment includes expansion of use of 4G bands, relying on similar modulation and spectrum access techniques used in 4G deployments [7]. Second, cellular operators all over the world are currently just starting to perform "3G shutdown" [8, 9]. This operation is performed, hand in hand with 5G deployment, in order to keep up with customer demand by refarming spectrum used by 3G networks to be used by 4G networks instead [10]. Finally, rapid 4G deployment techniques can be very helpful for community-operated cellular networks that service underdeveloped regions (e.g., [11]).

Standard 4G cellular network deployment implements the following steps [12, 13, 14]. First, network dimensioning takes into account user demand and density. It also accounts for the desired Quality of Service (QoS) provided by the operator. Then, network planning determines optimal radio sites locations and parameters. To achieve rapid deployment, it is typical for mobile operators to reuse the already existing 3G radio sites and infrastructure for 4G deployment. This approach capitalizes on two factors: 1) having a single Radio Access Network (RAN) that can already support both 3G and LTE networks, and 2) the overlap in supported spectrum for both networks. However, 3G and 4G technologies are significantly different. This requires several extensive optimization operations to be conducted before and after 4G activation. The optimization operations are determined based on site measurements, collected terrain information, and capacity estimation. The outcome is a set of new radio antenna parameters for eNodeBs (4G radio sites) to ensure that each site achieves its coverage without interference. This operation is time consuming and is known to be costly [15] which conflicts with any operator objectives, motivating further reduction in deployment steps. *In this paper, we assess the side effects of reducing the deployment overhead by reducing the cost of post-activation optimization.*

Post-activation optimization is needed due to the difference between physical layer protocols used in 3G and 4G networks. In particular, 3G networks tolerate overlapping coverage between multiple cells, while such conditions cause problems for 4G networks. In 3G networks, Soft Handover (SHO) handles interference gracefully by allowing a single user to receive data from multiple cells at the same time [16]. SHO is enabled by Code Division Multiple Access (CDMA). On the other hand, graceful handling of interference is not feasible in 4G/LTE networks because it relies on Orthogonal Frequency-Division Multiplexing (OFDM) which does not have the same macro-diversity as CDMA. Overlapping coverage areas in 4G networks lead to several problems including ping-pong handovers, handover failures, waste of the network resources, and degradation in signal quality [17].

Performing rigorous post-activation optimization to handle the difference between 3G and 4G is one of the costliest steps in the deployment. That is because it involves reviewing the coverage of tens of thousands of cells based on some propagation and interference models then verifying the plans

978-3-903176-17-1 / © 2019 IFIP

using rigorous drive tests. This cost can be reduced when taking into account the effects of overlapping coverage areas on 3G networks. In particular, overlapping coverage areas waste resources in 3G networks. Moreover, SHO has limitations in terms of maximum number of overlapping cells it can handle. Hence, a well-planned 3G network should have minimal overlap between coverage areas.

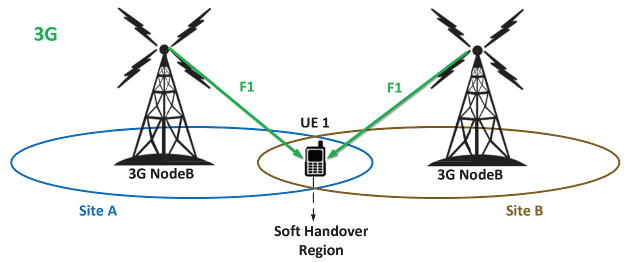
In this paper, we study the rapid deployment of 4G, testing the idea of reusing the existing 3G sites as is[1], minimizing the extensive post-activation optimization steps efforts. This is enabled by relying on a well-planned 3G network with minimal overlap in coverage areas. Our goal is to characterize the side effects of the rapid deployment. We refer to these side effects as *bad radio conditions*. Then, we determine their extent in terms of both their effect on user performance as well as their prevalence in the studied sites. Finally, we present a simple approach to handle affected site. Our aim is to reach the sweet spot between cost reduction, rapid deployment and optimum user experience. We conducted our study based on analysis of network traces and measurements from a major network operator. We collected data from a single cluster of cells covering two large cities, spanning two months in 2017 which represents the initial state of the network right after the rapid LTE deployment. The collected traces and measurements capture a snapshot of the operator's 4G deployment just before the implementation of post-activation optimization. This enables us to assess the value and side effects reducing the cost of post-activation optimization.

Our work is motivated by the appeal of reusing sites and equipment with minimal to no hardware modifications to move from 3G to 4G. This low cost operation should benefit cellular operators deploying 4G for the first time as well as performing 3G shutdown. It can also be useful for community operated cellular networks (e.g., [11]). Furthermore, our work motivates analytical models and software tools that can determine the parameters of the 4G deployment based on the 3G network the operator starts with, as well as Inter-Cell Interference Coordination (ICIC) [18] techniques that takes into account the deployment approach.

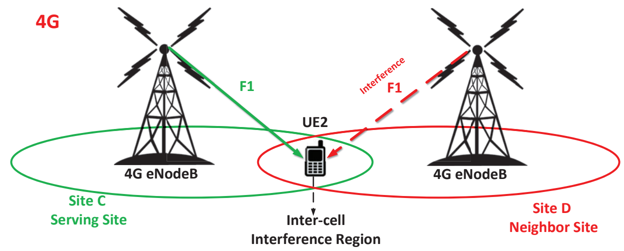
In brief, our work in this paper aims at answering the following questions about the rapid 4G deployment approach:

- *What is a good indicator of bad radio conditions?* We study correlation between different cell Key Performance Indicators (KPIs) and user throughput. We find that the most contributing factor affecting user throughput is the number of users within a cell using high-order modulation and coding scheme (§IV).
- *How prevalent are poor radio conditions, caused by coverage overlap between neighboring sites, in the operator's network?* Our goal is to help network operators to gain an insight of the extent of side effects of transitioning from 3G to 4G keeping post-activation optimization to a minimal. We use the bad radio condition indicators to identify cells with poor radio conditions. Furthermore, we validate that the main cause of the problem is overlapping coverage areas between cells by looking at sectors of the network with zero overlap (§V).
- *What is a cost-effective resolution of this problem?* We employ physical optimization, which is discussed in details in (§VI), guided by our developed approach of detecting the affected cells, to improve their performance. In particular, we employ antenna parameters optimization techniques to reduce or eliminate the effects of interference, leading to user throughput improvement by up to 114%. We believe this approach to be the best practice in rapid 4G deployment as it balances performance and cost efficiency.

[1]We omit the details of site architecture as it is part of our operators confidential information.



(a) In 3G networks, impact of overlapping coverage is mitigated through SHO.



(b) In 4G networks, overlapping coverage areas cause interference, hurting user observed performance.

Figure 1: Comparison between the effect of overlapping coverage areas in 3G and 4G networks.

II. Background

User devices, such as smartphones, tablets or modem cards connect to a radio cell over a certain radio frequency or a carrier. Each cell covers a geographic area with a directional antenna and it is common to find 3 such cells covering a full circle, approximately 120 degrees each, but there can be more or fewer cells with different coverage areas. Multiple cells covering the same direction and area can be called a sector. For coverage and capacity, there are typically multiple cells per base station, anywhere from 3 to 12, sometimes even more. There can be hundreds of thousands of cells in the network. In LTE, the site which spans all LTE sectors is called eNodeB, which is the evolution of 3G NodeBs. A Cluster is a group of radio sites covering a certain geographical area defined by the network operator.

Typical LTE deployment relies on network dimensioning in which user demand, density, along with the operator's desired

QoS provisioning are used to plan base stations placement and capacity. This step goes hand in hand with a coverage planning phase where the parameters of the antenna of each base station is determined such that all areas of interest are covered and inter-cell interference is reduced. This approach is standard and discussed in several textbooks (e.g., [12, 13, 14]). Coverage planning is one of the costliest steps in the deployment. That is because it involves reviewing the coverage of tens of thousands of cells based on some propagation and interference models then verifying the plans using rigorous drive tests. This operation is both costly and time consuming which conflicts with the goals of the operators that are eager to be the first to deploy LTE in a market that is hungry for more bandwidth. LTE Rapid deployment leverages two advantages the operator has: 1) a single RAN that operates both 3G and LTE networks allowing the same equipment to be used for either of the two technologies, and 2) re-farming of 2G and 3G spectrum for LTE operation, which is a technique by which spectrum of older technologies with diminishing demand is used by newer technologies [10]. Spectrum re-farming allowed the utilization of the same antenna configuration used in 3G to be used in LTE as well. Following this approach, the upgrade from 3G to LTE became a matter of upgrading infrastructure equipment as most of the other components, including antennas and radio modules where used as is for LTE. The main challenge of this approach is handling the difference between how 3G and LTE networks handle overlapping coverage areas.

In 3G networks, users falling in the area of coverage overlap between multiple cells don't exhibit severe problems. This is due to the macro-diversity technique that's employed by the CDMA and W-CDMA standards. This makes use of a feature called soft handover (SHO) where a cell phone is simultaneously connected to two or more cells extending its capacity, shown in Figure 1a [16]. While this approach mitigates the effects of interference from the point of view of the user, it is still problematic from the point of view of the operator. SHO comes at the expense of available resources, as one user will utilize those resources for the two, or more, base stations, while it's preferred to connect from one cell only. Hence, overlapping areas should be minimized as much as possible, though, it won't affect the customer experience in 3G as we've illustrated.

On the other hand, overlapping coverage areas in LTE is not preferable from both the user perspective and the network operator perspective. From the user perspective, the access scheme used in LTE (i.e., OFDM) does not handle interference gracefully, as shown in Figure 1b. This can lead to several problems including lower user throughput, ping-pong handovers, and handover failures. From a network perspective, overlapping coverage areas in LTE waste resources similar to the case of 3G in addition to making performance at users worse. Hence, when LTE is deployed on top of an already operational 3G network, the effect of overlapping coverage areas is amplified and becomes a main issue in the network.

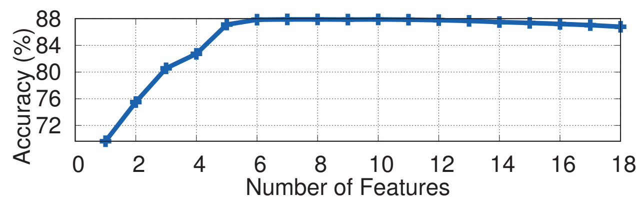


Figure 2: Throughput estimation accuracy improvement using sequential forward selection, showing that the first six features are enough to provide best estimation accuracy.

III. Data Set

We use a set of KPIs collected by a top-tier mobile operator with more than 30 million subscribers serving an entire country. All of our measurements are collected from the operator network point of view. The data we use is collected from a cluster within the operator's network that spans two major cities. We collect data spanning two months in 2017 that represent the state of the network immediately post rapid LTE deployment on that cluster. All data reported are anonymized and collected in aggregate.

The collected KPIs are measured at the eNodeBs, capturing both network KPIs and per-cell aggregate end-user KPIs. Our measurements are limited to KPIs collected and aggregated by eNodeB equipment. We follow this approach to maintain the ability to collect data at scale by avoiding tracking KPIs of individual users. This allows us to provide an approach that can be realistically used by the operator to detect cells affected by the deployment methodology.

KPI measurements are collected at 15-minute interval spanning two consecutive months. We believe that this period and monitoring granularity to be representative of normal network operations capturing hourly load variability and trends. We filter the data to focus on trends at peak hours because that is what the network provisions for. As indicated earlier, user information is collected in aggregates per cell and does not contain any personal or identifiable information about owners of devices or exact base stations info. The data is in a columnar format, with 4.2 million records, each record represents the KPIs of a certain cell over the 15-minute period.

Our goal is to characterize the impact of rapid 4G deployment. For that to happen, capitalizing on the studied cluster's dataset, we devised a supervised machine learning model to classify the cells into two groups: good and poor from the user throughput perspective. The throughput threshold used to label these cells was picked by the mobile operator reflecting the minimum throughput required to achieve a targeted customer experience and we omit the exact threshold value the operator uses. Then, multiple feature selection algorithms were applied to choose the most impacting features (KPIs) on user throughput out of the available KPIs to the operator. The best performing algorithms were the Sequential search methods [19, 20] especially, the "floating" methods [21]. Figure 2 shows the sequential floating forward selection algorithm performance for a decision tree classification algorithm [22]. It's apparent that a selected group of six features achieved the best estimation accuracy, we list them below and we will study each of them closely against the user throughput in (§IV).

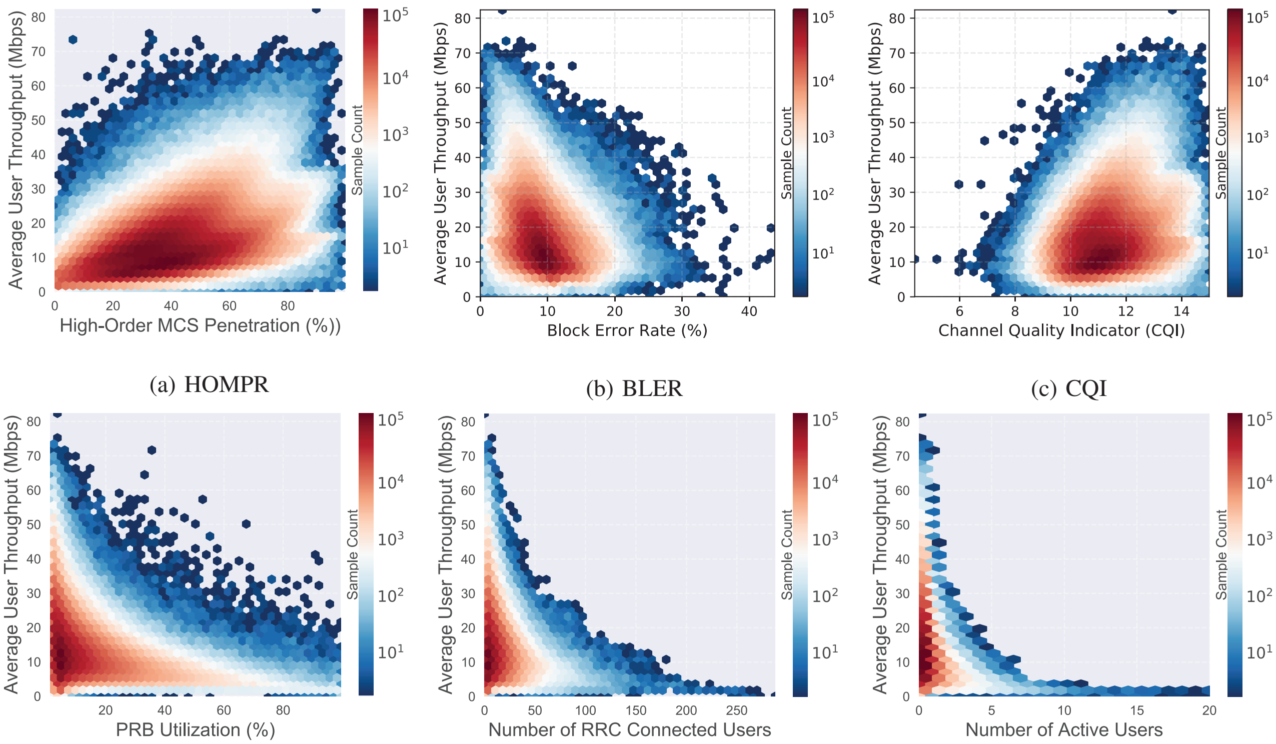
We categorize the KPIs we found into three categories: 1) Quality of Experience Indicators (QoEI), 2) Quality of Coverage Indicators (QoCI), and 3) Capacity Indicators (CI). Each of the available KPIs had different aggregation levels (e.g., average and max values per cell), and the following are the selected KPIs, all of which are averaged per cell:

- **[QoEI] User Downlink (DL) Throughput:** This KPI is the main User Quality of Experience indicator used by the operator in the radio domain [23]. The higher the user throughput, the more it is indicated that the customers served by that cell in the downlink direction are having a good experience, and vice versa. One of our goals is to find how other KPIs affect this one. During network planning and maintenance, this KPI is used as the main drive for upgrades and provisioning of resources to different cells.
- **[QoCI] High-Order MCS Penetration Rate (HOMPR):** This KPI captures the ratio of end users in a certain cell employing high-order Modulation and Coding Scheme (MCS). In particular, this KPI indicates the percentage of users using 64-QAM modulation, the highest modulation available at the current network implementation. Users are able to utilize high-order MCS when radio conditions are good. The higher the employed MCS, the higher the data rate the user can use. Cases where individual users utilize lower-order MCS are typical (e.g., user at the border of the cell or behind a signal reflector). Also, inter-cell interference has a major impact on the granted MCS, the higher the interference is, the lower the MCS order will be. Thus, when a large ratio of users of a certain cell rely on lower-order MCS, it indicates that there is a problem at the cell level.
- **[QoCI] Channel Quality Indicator (CQI):** This KPI is reported from users to the cell to indicate channel quality seen by the user. This report helps the cell scheduler determine the MCS to be used by that user. This metric is part of the LTE standard unlike signal-to-noise-and-interference-ratio (SINR). The way CQI is calculated takes into account a vendor-based definition of SINR along with the decoding capabilities of the user's device. When the CQI is high, this indicates good experienced radio conditions within the cell.
- **[QoCI] Block Error Rate (BLER):** This is the fraction of blocks delivered on the channel that fail CRC (Cyclic Redundancy Check). A high BLER percentage indicates a problem within the cell.
- **[CI] PRB Utilization:** A Physical Resource Block (PRB) is the smallest unit of resources that can be allocated to a user which is specified in terms of radio sub-carriers and time. The PRB Utilization KPI captures the ratio of the utilized physical resource blocks with respect to all the available resource blocks. This KPI reflects how overloaded a specific cell is which can be used to attribute lower user throughput to lack of resources at the cell when it is high.
- **[CI] RRC-Connected Users:** This KPI reflects the average number of users that establish a connection with the RRC (Radio Resource Control) layer. This includes all users within a cell whether they have any data to send or not. Generally, as the number of users increases, this leads to a decrease in the average user throughput within the cell due to the finite radio resources.
- **[CI] Number of Active Users:** Active users are users with data remaining in the transmit buffer of the cell at a certain transmission time interval (TTI). This KPI is the average number aggregated over the collection interval. *Note that this number does not reflect all users utilizing the DL as it only counts users with buffered data.*

Our goal is to assess the network behavior when the rapid 4G deployment mentioned earlier is used. In particular, we are interested in assessing user quality of experience of users in the 4G deployment, represented by average user throughput per cell. We are interested in understanding the relationship between network parameters and user experience. To that end, we compare the effect of Capacity Indicators (CIs) and Quality of Coverage Indicators (QoCIs) on the Quality of Experience Indicator (QoEI), which is the average user throughput in our case. Figure 3 shows the relationship between Average User Throughput and the different KPIs. We use hexagonal binning to visualize the distribution of collected 15-minute data samples from all cells. Hexagonal binning allows us to compare the distribution of data across two KPIs.

We start with QoCIs. Figure 3a shows that the higher the High-Order MCS Penetration Rate the higher the throughput users might experience. The highest concentration of samples is along the area corresponding to this relationship. Samples with high rate of High-Order MCS Penetration and low throughput can be justified by low available capacity. Invertedly, the rare occasions of having bad radio conditions and still relatively high throughput occur when high bandwidth and low user count occur. We observe a similar trend the CQI metric (Figure 3c). However, CQI exhibits less correlation with throughput than the HOMPR. Figure 3b shows the Block Error Rate (%) which naturally has negative correlation with throughput. Visually, it is clear that among all the QoCIs, the HOMPR has the highest correlation with the user throughput.

Figure 3d shows a clear relationship with PRB utilization. The higher the utilization, the lower the attained throughput as contention for resources increases. The samples indicating the high utilization and still relatively high user throughput occur when spectrum resources are utilized by a few users. CIs are less correlated with total cell population compared to QoCIs. Figures 3e and 3f both relate the number of users to user throughput. In case of the RRC-connected users, there is no clear correlation as connected users do not necessarily utilize and resources. On the other hand, the negative correlation is clear between active users and user throughput.



(a) HOMPR (b) BLER (c) CQI

(d) PRB Utilization (e) RRC-Connected Users (f) Number of Active Users

Figure 3: Hexbin scatter comparing distribution of values of different KPIs and Average User Down Link Throughput over collected data.

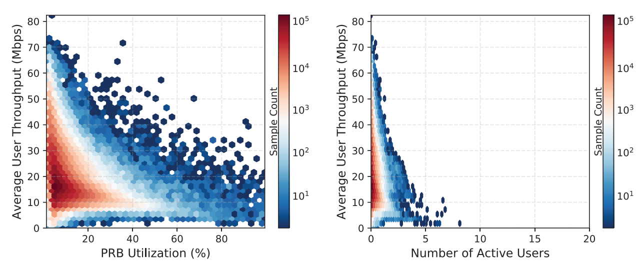
Correlation with Average User Throughput	HOMPR	CQI	BLER	PRB Utilization	RRC Connected Users	Number of Active Users
All cells	**0.454163**	0.401373	-0.316519	-0.311759	-0.120421	-0.18782
Cells in good radio conditions	0.265883	0.236271	-0.312613	**-0.370343**	-0.162819	-0.229172

Table I: Correlation between different KPIs and Average User Down Link Throughput.

IV. Relationship Between Bad Radio Conditions and Low User Throughput

The first row of Table I summarizes our finding showing the Pearson correlation coefficient between all KPIs. As clear from Figure 3a, HOMPR has the highest correlation between the other KPIs. Our hypothesis is that this correlation is an artifact of the LTE deployment approach which leads to coverage overlapping between cells. To understand our hypothesis, consider an ideally operating network. In the ideal case, user throughput should be a factor only of available capacity in a cell (i.e., low throughput should only occur when demand exceeds capacity). However, in cases of bad radio conditions, the coverage quality becomes an factor in determining user throughput. This relationship wastes capacity by having nodes operating at lower throughput than available at the cell. It also leads to poor user expience.

To validate our hypothesis, we started by looking at cells in good radio conditions. Our goal is to show that when the coverage is good, capacity is the only contributor to user throughput. In our experience, good radio conditions are associated with HOMPR higher than 40%. Figure 4 along with the second row of Table I show that under good radio conditions capacity becomes the dominant factor controlling average user throughput, which is the expected behavior. This shows that poor radio conditions lead to a change in factors affecting user throughput. This motivated further analysis as shown in the next section.



(a) PRB Utilization (b) Number of Active Users

Figure 4: Hexbin scatter comparing distribution of values different KPIs and Average User Down Link Throughput over sample with Higher-Order MCS Penetration Rate is higher than 60%.

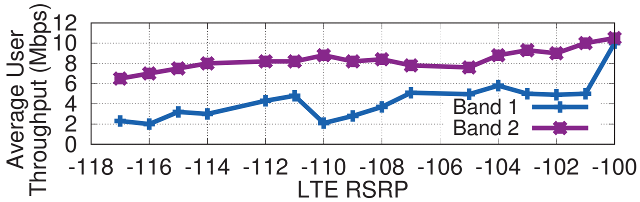


Figure 5: Comparing average user throughput for users potentially experience bad radio condition (Band 1 users) and users not experiencing bad radio conditions (Band 2 users).

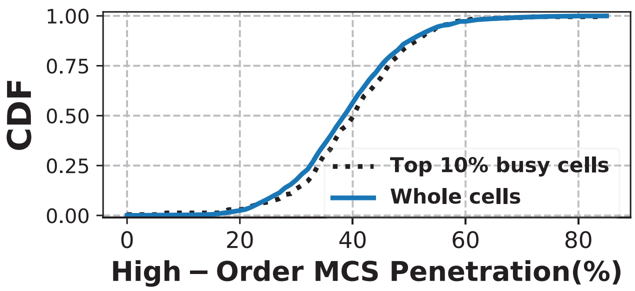


Figure 6: CDF of ratio of users operating at high modulation schemes for all studied cells and the top 10% busiest cells showing similar distribution in both cases.

V. Prevalence of Poor Radio Conditions

Our next step is to look at the extent of this issue in terms of its effect on throughput as well as the number of affected cells when rapid deployment of 4G is employed. Our goal is to determine whether bad radio conditions are likely to occur when rapid deployment techniques are employed, and if so, quantify its effect. As 3G cells are planned with minimal overlap, one might expect that an LTE network deployed on top 3G coverage planning should have little problems. However, we find that this is not always true.

We start with assessing the extent of effect of bad radio conditions on average user throughput. In particular, we study a set of cells where users can access 4G through two different frequency bands. We refer to them as Band 1 and Band 2. A load balancing algorithm ensure that, for each cell, the number of users operating at each band is roughly the same. The main difference between the two frequency bands is the cells had neighbouring cells operating in Band 1, potentially having overlap in their coverage area in that band. On the other hand, all studied cells had no neighbouring cells operating in the Band 2, ensuring having no bad radio conditions due to interference. We compare average user throughput for both scenarios where Reference Signal Received Power (RSRP) is low (i.e., areas with poor coverage). Figure 5 shows the result of the experiment. We find that average throughput is up to 3x higher for users operating in Band 2 where there are no bad radio conditions due to overlapping coverage areas. *Hence, we conclude that cell overlap has significant negative impact on user throughput.*

Next, we look at the prevalence of the problem in the studied cluster of cells. Figure 6 shows the CDF of High-Order MCS Penetration Rate (HOMPR) over all LTE cells of the studied cluster. We use HOMPR as a proxy for cells suffering from bad radio conditions as it is the KPI with the highest correlation with average user throughput. In particular, a cell with very low High-Order MCS Penetration Rate reflects that the majority of users can only decode robust lower-order modulation schemes. This is unlikely to happen due to individual users having bad radio conditions (e.g., large scale fading), rather it is more likely that a large population of cell users are facing a common source of interference.

The figure shows that 50% of the cells have a HOMPR lower than 40%, which is the level we found to make average user throughput more dependent on capacity than radio conditions. To better understand the impact of the problem, we look at the top 10% busiest cells and find that they exhibit a very similar distribution. *Hence, we conclude that bad radio conditions are quite common in cases of rapid deployment and can lead to significant deterioration in user throughput.*

VI. Resolving Poor Radio Conditions

Typically, mobile operators target a certain customer experience each year, trying to maintain an edge over other competitors. And after estimating year-over-year expected traffic increase, taking inputs from commercial teams and meeting some certain strategies, they dimension the needed capacity across all the mobile network's domains, including RAN, Transport and Core Networks. For RAN domain, the list of cells likely to be congested (i.e., hot spots) are estimated in advance, to be handled with a proper expansion according to their configurations in term of number of carriers and sectors at the site [24]. Huge investments are dedicated, in terms of spectrum, license and hardware expansions to work proactively on those cells to make sure they would accommodate the foreseen capacity needs.

Apart from that, dealing with congested cells is a normal day to day activity for the RF optimizers at the mobile operator [25]. Several approaches are implemented for the hot spot resolution. Usually, the solution involves bandwidth expansion after assuring the highest possible spectrum utilization, adding more cells/sectors or even planning new sites. These expansion procedures are very costly but this can be afforded by the operator to assure a better experience for the end user, and consequently, a higher revenue. However, these approaches to handling hot spots are not necessarily the right approaches in when LTE is deployed following the methodology discussed in Section V.

From practical observations, it was found that hot spots, following a soft LTE deployments, are mostly formed due to overlapping coverage areas between different cells. Hence, relief of this congestion can be done by re-planning coverage areas of affected cells. This approach requires: 1) detecting cells suffering from this problem, and 2) providing some guidelines for handling the interference problem.

Affected Cells Detection: We rely on High-Order MCS Penetration Rate to detect cells affected by our LTE deployment approach. In particular, cells with 40% or low High-

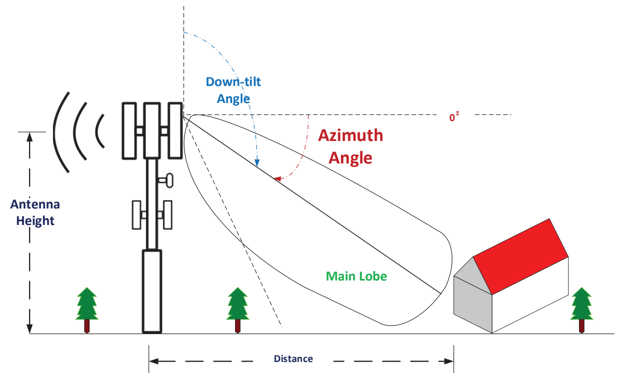
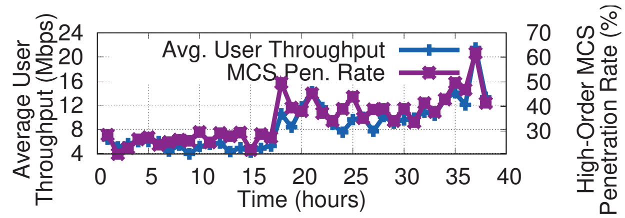
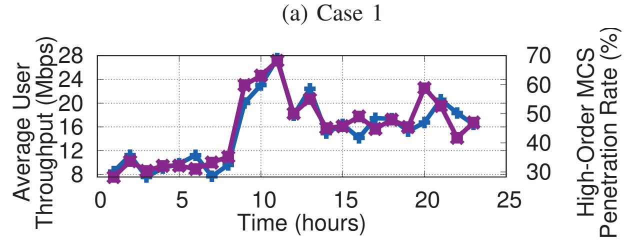


Figure 7: Illustration of cell physical parameters.



(a) Case 1



(b) Case 2

Figure 8: Two case studies showing the effect of physical optimization that improved High-Order MCS Penetration Rate on improving average user throughput by up to 114%.

Order MCS Penetration Rate are considered affected. This is consistent with our finding in Section IV where we found that for cells with High-Order MCS Penetration Rate larger than 40%, the main factors affecting user throughput are capacity factors not interference factors.

Physical Optimization: To address the problem of overlapping coverage areas, the parameters determining the coverage area of each cell are optimized using standard planning tools that simulate the coverage of affected cells. These tools identify optimal parameters for each antenna, in terms of the elevation, azimuth and height (Figure 7), to make sure the dominance of the serving cells and isolation from the neighboring cells as much as possible. And those predictions can be calibrated with performing a field drive test to capture the coverage measurements and the results are sent back for the planners to analyze and confirm their coverage plan. Finally, this parameters are applied to the affected site.

We demonstrate the effectiveness of this approach by showing the following two cases where the methodology was applied. Both cases are identified by having HOMPR below 30%. Physical optimization was applied at hour 17 and 11 for case 1 and 2, respectively. Figure 8 shows that, for both cases, once physical optimization was applied both HOMPR and average user throughput improve significantly. Average user throughput is improved by 97% and 114% on average. This is a significant impact in throughput that is achieved without provisioning any extra resources. Hence, we note that this approach allowed our operator to save costs in two ways. First, it allowed for a fast deployment of LTE where not all cells needed re-planning. Second, it allows the operator to have an approach to resolve hot spots other than the costly resource provisioning (e.g., creating more cells or licensing more spectrum) to check before going to that step.

VII. Related Work

Data-driven Cellular Network Optimization: We categorize those approaches based on where measurements are collected into the following categories:
1. Measurements taken at the network side, capturing cell KPIs: Earlier work in this direction has focused on 3G networks. We focus on recent measurements from major operators. For example, in [26] sector KPIs used to infer user QoE. This allows the network operator to detect underperforming 3G sectors. The work in [24] uses sector KPIs to forecast which cells will underperform under specific loads. Our approach falls under this category to augment earlier work. In particular, it provides a specific metric to detect underperforming cells due to 4G transition. It can be used in conjunction with other approaches to provide better cell planning and provisioning.
2. Measurements taken at the network side, capturing user traces: Several systems rely on user traces to detect [27], predict [28] and react to [29] user throughput in 4G networks.
3. Measurements taken at the user side: This approach relies on measurements that are either collected by the operator through wardriving (e.g., [30]), or crowdsourcing (e.g., [31]). This direction of work is orthogonal to ours that can alert operators to coverage gaps that cannot be detected from the operator side.

Interference cancellation and mitigation: Such techniques have been deployed with varying degrees of success for more than 20 years. They are generally categorized into three major techniques as shown in [32]: interference cancellation, interference averaging and interference avoidance techniques. Although in theory many proposed interference mitigation techniques have shown promising results, in practice the gains do not seem to materialize and better approaches need to be evaluated in realistic scenarios suitable for implementation[33]. For example, interference cancellation promises significant gains but is likely limited to the LTE UL due to processing complexity, and will require real-time exchange of information between base stations every few milliseconds to maximize the gain for an LTE system. Beam-forming technologies such as organized beam hopping show significant theoretical promise; however, as practical technologies for a deployed LTE system they are still unproven [18].

Also, LTE offers the capability to provide a flexible dynamic inter-base station approach to interference coordination through the use of inter-base station signaling capabilities using the X2 interface defined between the base stations, including the use of UL reactive overload indicators (OIs) and proactive high interference indicators (HIIs) that provide bit maps of interference conditions on a per RB basis. DL inter-cell interference coordination (ICIC) is supported through the use of DL relative narrowband transmit power (RNTP) bit maps providing a coarse power indication on a per RB basis [18]. Although this capability was already employed, further physical optimization was still required achieving extra enhancement in user throughput as we showed earlier.

VIII. Conclusion

We use a large-scale LTE data set from a major telecom operator to assess the value of a rapid LTE deployment approach that relies on transitioning from 3G to 4G, minimizing the overhead of extensive post-activation optimization steps. We also show a data-driven approach to detect underperforming cells and present the physical optimization approach to handle the detected cells. Overall, the approach balances LTE transition best practices with cost efficiency to allow the operator to provide a high performing network rapidly. We believe this approach will be beneficial for operators deploying 4G for the first time and performing 3G shutdown as well as community-operated networks servicing rural underdeveloped areas.

References

[1] I. E. Sørensen, "Content in context: The impact of mobile media on the British TV industry," *Convergence*, 2018.

[2] B. Sanou, "ICT Facts and Figures 2017," *Geneva: The International Telecommunication Union (ITU)*, 2017.

[3] J. Place, S. Keeping, S. Akayama, and A. Schlautmann, "The Business Benefits of 4G LTE," *Arthur D. Little UK*, 2012.

[4] I. PLC, "El Salvador embraces LTE, gets its first 4G networks," 2017, https://ovum.informa.com/resources/product-content/te0009-001664.

[5] D. N. Egypt, "4G service to launch on Thursday," 2017, https://dailynewsegypt.com/2017/09/24/4g-service-launch-thursday/.

[6] L. Duan, J. Huang, and J. Walrand, "Economic analysis of 4G network upgrade," in *Proc. of IEEE INFOCOM*, 2013.

[7] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, 2014.

[8] A. Weissberger, "AT&T to shut down 3G network in 2022; Verizon at end of 2019," IEEE ComSoc Tech Blog, 2019, http://techblog.comsoc.org/2019/02/22/att-to-shut-down-3g-network-in-2022-verizon-at-end-of-2019/.

[9] A. Choros, "Australian 3G Network Shutdown: Everything you need to know," Whistle Out, 2019, https://www.whistleout.com.au/MobilePhones/Guides/Australian-3G-network-shutdown-what-you-need-to-know.

[10] S. Han, Y.-C. Liang, and B.-H. Soong, "Spectrum refarming: A new paradigm of spectrum sharing for cellular networks," *IEEE transactions on communications*, 2015.

[11] S. Hasan, M. C. Barela, M. Johnson, E. Brewer, and K. Heimerl, "Scaling community cellular networks with communitycellular-manager," in *Proc. of USENIX NSDI*, 2019.

[12] A. ElNashar, M. A. El-Saidny, and M. Sherif, *Design, deployment and performance of 4G-LTE networks: A practical approach*. John Wiley & Sons, 2014.

[13] J. T. Penttinen, *The LTE-Advanced Deployment Handbook: The Planning Guidelines for the Fourth Generation Networks*. John Wiley & Sons, 2016.

[14] T. V. Krishnamurthy and R. Shetty, *4G: Deployment Strategies and Operational Implications: Managing Critical Decisions in Deployment of 4G/LTE Networks and their Effects on Network Operations and Business*. Apress, 2014.

[15] Y. Chen, L. Duan, and Q. Zhang, "Financial analysis of 4G network deployment," in *Proc. of IEEE INFOCOM*, 2015.

[16] A. J. Viterbi, A. M. Viterbi, K. S. Gilhousen, and E. Zehavi, "Soft handoff extends CDMA cell coverage and increases reverse link capacity," *IEEE Journal on selected areas in communications*, 1994.

[17] G. P. Pollini, "Trends in handover design," *IEEE Communications magazine*, 1996.

[18] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4g networks," *IEEE Communications Magazine*, 2009.

[19] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, 1997.

[20] S. Das, "Filters, wrappers and a boosting-based hybrid for feature selection," in *Proc. of ICML*, 2001.

[21] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, 1994.

[22] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, 1991.

[23] P. Casas, B. Gardlo, R. Schatz, and M. Mellia, "An Educated Guess on QoE in Operational Networks Through Large-Scale Measurements," in *Proc. of Internet-QoE Workshop*, 2016.

[24] J. Serrà, I. Leontiadis, A. Karatzoglou, and K. Papagiannaki, "Hot or not? Forecasting cellular network hot spots using sector performance indicators," in *Proc. of IEEE ICDE*, 2017.

[25] L. Song and J. Shen, *Evolved cellular network planning and optimization for UMTS and LTE*. CRC press, 2010.

[26] I. Leontiadis, J. Serrà, A. Finamore, G. Dimopoulos, and K. Papagiannaki, "The good, the bad, and the KPIs: how to combine performance metrics to better capture underperforming sectors in mobile networks," in *Proc. of IEEE ICDE*, 2017, pp. 297–308.

[27] T. Mangla, E. Zegura, M. Ammar, E. Halepovic, K.-W. Hwang, R. Jana, and M. Platania, "VideoNOC: Assessing Video QoE for Network Operators Using Passive Measurements," in *Proc. of ACM MMSys*, 2018.

[28] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of LTE: effect of network protocol and application behavior on performance," *ACM SIGCOMM Computer Communication Review*, 2013.

[29] 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?" in *Proc. of ACM HotMobile*, 2015.

[30] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G Using WiFi," in *Proc. of ACM MobiSys*, 2010.

[31] "OpenSignal," https://opensignal.com/, 2018.

[32] M. Rahman, H. Yanikomeroglu, and W. Wong, "Interference avoidance with dynamic inter-cell coordination for downlink LTE system," in *Proc. of IEEE WCNC*, 2009.

[33] D. J. Ramirez, "Interference mitigation techniques for 4G networks," Ph.D. dissertation, Supélec, 2014.