Characterizing and Optimizing Cellular Network Performance During Crowded Events

M. Zubair Shafiq, Lusheng Ji, *Senior Member, IEEE*, Alex X. Liu, Jeffrey Pang, Shobha Venkataraman, and Jia Wang

***Abstract*— During crowded events, cellular networks face voice and data traffic volumes that are often orders of magnitude higher than what they face during routine days. Despite the use of portable base stations for temporarily increasing communication capacity and free Wi-Fi access points for offloading Internet traffic from cellular base stations, crowded events still present significant challenges for cellular network operators looking to reduce dropped call events and improve Internet speeds. For an effective cellular network design, management, and optimization, it is crucial to understand how cellular network performance degrades during crowded events, what causes this degradation, and how practical mitigation schemes would perform in real-life crowded events. This paper makes a first step toward this end by characterizing the operational performance of a tier-1 cellular network in the U.S. during two high-profile crowded events in 2012. We illustrate how the changes in population distribution, user behavior, and application workload during crowded events result in significant voice and data performance degradation, including more than two orders of magnitude increase in connection failures. Our findings suggest two mechanisms that can improve performance without resorting to costly infrastructure changes: radio resource allocation tuning and opportunistic connection sharing. Using trace-driven simulations, we show that more aggressive release of radio resources via 1–2 s shorter radio resource control timeouts as compared with routine days helps to achieve better tradeoff between wasted radio resources, energy consumption, and delay during crowded events, and opportunistic connection sharing can reduce connection failures by 95% when employed by a small number of devices in each cell sector.**

***Index Terms*— Cellular network, crowded events, performance.**

Manuscript received March 25, 2014; revised October 18, 2014; accepted December 17, 2014; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Chong. Date of publication April 20, 2016; date of current version June 14, 2016. This work was supported in part by the National Science Foundation under Grant CNS-1318563 and Grant CNS-1524698, the National Natural Science Foundation of China under Grant 61472184 and Grant 61321491, the Jiangsu Future Internet Program under Grant BY2013095-4-08, and the Jiangsu High-Level Innovation and Entrepreneurship (Shuangchuang) Program. The preliminary version of this paper, titled "A First Look at Cellular Machine-to-Machine Traffic—Large Scale Measurement and Characterization," was published in the Proceedings of the ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), Pittsburgh, PA, USA, June 2013. *(Corresponding author: Alex X. Liu.)*

M. Z. Shafiq was with Michigan State University, East Lansing, MI 48824 USA. He is now with the Department of Computer Science, The University of Iowa, Iowa City, IA 52242 USA (e-mail: zubair-shafiq@uiowa.edu).

L. Ji, J. Pang, S. Venkataraman, and J. Wang are with AT&T Laboratories—Research, Bedminster, NJ, USA (e-mail: lji@research.att.com; jeffpang@research.att.com; shvenk@research.att.com; jiawang@research.att.com).

A. X. Liu is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: alexliu@cse.msu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2016.2533612

I. INTRODUCTION

CROWDED events, such as football games, public demonstrations, and political protests, put an extremely high demand for communication capacity on cellular networks around the duration of the events [1]. Cellular networks are facing unprecedent challenges in dealing with such spiky demand. First, cellular network utilization has already been rapidly approaching its full capacity throughout the world due to the increasing prevalence of cellular devices such as smartphones, tablets, and Machine-to-Machine (M2M) devices. Even in the United States, cellular network usage is at an all-time high even under normal operating conditions and projections show that traffic volume will further increase by 26 times by 2015 as compared to 2010 [2], [3]. Second, the spiky demand caused by crowded events is often extremely high because there maybe a large number (often tens of thousands) of users gathered in a small region (such as a football stadium) that is covered by only a small number of cell towers. Even worse, people tend to use their cellular devices more than usual during the events to either talk with their friends or access the Internet (such as uploading a photo to Facebook or a video clip to YouTube during a football game). Third, it is critical for cellular networks to cope with such high demand during crowded events because poor performance will affect a large number of people and cause widespread user dissatisfaction. Although cellular network operators have deployed remediation solutions, such as portable base stations called Cells on Wheels (COWs) for temporarily increasing communication capacity and free Wi-Fi access points for offloading Internet traffic from cellular base stations, crowded events still remain a major challenge for cellular network operators.

To the best of our knowledge, this paper presents the first thorough investigation of cellular network performance during crowded events. Based on the real-world voice and data traces that we collected from a tier-1 cellular network in the United States during two high-profile crowded events in 2012, we aim to answer the following three key questions.

How does cellular network performance degrade during crowded events as compared to routine days? To answer this question, we characterize cellular network performance during both the pre- and post-connection phases (i.e., before or after radio access bearer assignment), which helps us to understand user experience before and after acquiring radio resources. For pre-connection phase, we find that pre-connection failures dramatically increase during the crowded events by 100-5000 times as compared to their average on routine days. These failures occur because when too many users attempt to acquire radio resources at the same time, they exhaust the limited

bandwidth of the signaling channel resulting in connection timeouts and failures. We find that this resource exhaustion occurs not only at the event venue, but also as far as 10 miles around the event as users arrive and depart. Moreover, some failures, such as dropped and blocked voice calls, are most likely to occur in bursts just before, after, and during event intermissions. For post-connection phase, we find that voice network performance in terms of dropped and blocked calls degrades during crowded events by 7-30 times, and data network performance in terms of packet loss ratio and round trip time (RTT) degrades during crowded events by 1.5-7 times, compared to their average on routine days.

What causes the performance degradation? To answer this question, we analyze user traffic patterns in terms of both aggregate network load and user-level session characteristics. For aggregate network load, we find that uplink traffic volume increases by 4-8 times, and both downlink traffic volume and the number of users increase by 3 times, during the crowded events as compared to their average on routine days. We conclude that the large number of users trying to access radio resources at the same time is a major cause of the observed excessive pre-connection failures. For user-level session characteristics, we find that the average byte volume per session decreases by 0.5 times during the events even though the average session length increases. Our investigation in Section IV-B suggests that this change in workload is due to a change in application usage during these events, such as the increased use of online social networks. We conclude that lower byte volume per session, despite an increase in average session length, is a major cause of the waste of radio resources in the post-connection phase.

How would practical mitigation schemes perform in real-life? To answer this question, we investigate two practical mitigation schemes that do not require making significant changes to the cellular infrastructure: radio resource allocation tuning and opportunistic connection sharing. Radio resource allocation tuning addresses the issue of inefficient radio resource allocation in the post-connection phase by adjusting cellular network resource allocation parameters. Cellular networks allocate resources to each user using a Radio Resource Control (RRC) state machine, which is synchronously maintained by the network and devices. Different states of the RRC state machine correspond to different amount of radio resources allocated by the network and energy consumption by cellular devices. Since a large number of users contend for limited radio resources during crowded events, we show that more aggressive release of radio resources via 1-2 seconds shorter RRC timeouts helps to achieve a better tradeoff between wasted radio resources, energy consumption, and delay during crowded events. Note that cellular network operators often know the time and location of large crowded events beforehand; thus, it is practical for them to adjust cellular network parameters before events and restore them after events. In practice, network operators can automatically tune RRC timeouts based on network load measurements.

Opportunistic connection sharing addresses increased pre-connection failures (due to poor RF quality, interference, or radio resource exhaustion, etc.) by aggregating traffic from multiple devices into a single cellular connection. That is, by having some devices share their cellular connection with nearby devices over their Wi-Fi or Bluetooth interface (*i.e.*, "tethering"), opportunistic connection sharing should reduce the number of overall cellular connection requests, thereby reducing request congestion and connection failures. Using trace-driven simulations, we show that connection sharing can reduce connection failures by more than 95% when employed by a small number of devices in each cell sector. Although much work has been done on opportunistic connection sharing to address issues such as mobility, energy use, and incentives [4]–[6], no prior work has demonstrated the significant benefit that such connection sharing can achieve based on real-life cellular network data.

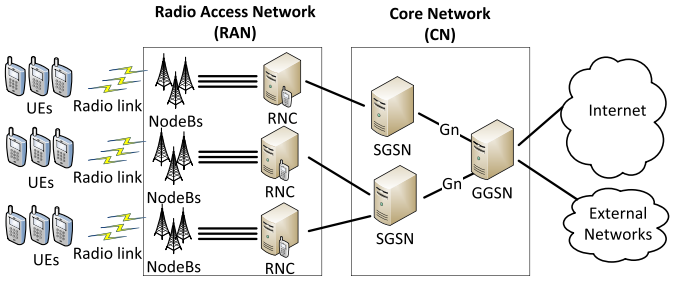


Fig. 1. Cellular network architecture.

The rest of this paper is organized as follows. In Section II, we present details of the data collection process. Section III presents the characterization of performance issues during the crowded events and Section IV presents various aspects of user traffic patterns to study the underlying causes of performance issues. We conduct trace-driven simulations to evaluate radio network parameter tuning and opportunistic connection sharing in Section V. Section VI reviews related work and the paper is concluded in Section VII.

II. Data Set

A. Background

A typical 3GPP Universal Mobile Telecommunications System (UMTS) cellular network, shown in Figure 1, consists of two components: Radio Access Network (RAN) and Core Network (CN). RAN consists of NodeBs and Radio Network Controllers (RNCs). Each NodeB has multiple antennas, where each antenna corresponds to a different cell sector. CN consists of Serving GPRS Support Nodes (SGSNs) facing the user and Gateway GPRS Support Nodes (GGSNs) facing the Internet and other external networks. A user via user equipment (UE) connects to one or more cell sectors in the RAN. The traffic generated by a UE is sent to the corresponding NodeB by cell sectors. Each RNC controls and exchanges traffic with multiple NodeBs, each of which serves many users in its coverage area. RNCs manage control signaling such as Radio Access Bearer (RAB) assignments, transmission scheduling, and handovers. RNCs send traffic from NodeBs to SGSNs, which then send it to GGSNs. GGSNs eventually send traffic to external networks, such as the Internet.

RAN dynamically allocates resources to a UE. Specifically, every UE negotiates allocation of radio resources with the RAN based on a wide range of factors, such as available radio resources and signal strength [7]. Every UE follows the RRC protocol for dynamic acquisition and dropping of radio resources. The RRC state machine of each user is

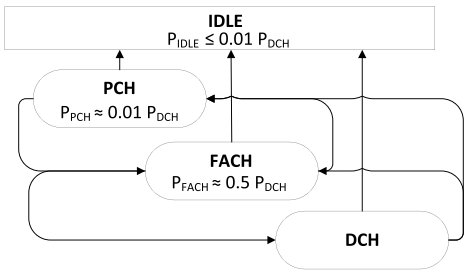


Fig. 2. RRC protocol state transitions.

synchronously maintained by the UE and network. Different states in the RRC state machine correspond to different amounts of allocated radio resources by the network and energy consumption by UEs. In this paper, we are only concerned about UE energy consumption at the radio interface. Thus, we do not account for other types of UE energy consumption, e.g., computation, screen-on, etc. Figure 2 shows all RRC protocol state transitions. According to RRC protocol, a UE transitions to Dedicated Channel (DCH) state or Forward Access Channel (FACH) state for uplink or downlink data transfer. RAN assigns a dedicated or shared channel for DCH and FACH states, respectively. If a UE does not have any data to transfer, it transitions to Paging Channel (PCH) state before the transition to IDLE state. Generally, state promotions are controlled by data buffer size thresholds and state demotions are controlled by inactivity timeouts. Furthermore, the energy consumption by a UE is roughly inversely proportional to the amount of allocated radio resources. The energy consumption by a UE is maximum in DCH state, which is about halved when it transitions to FACH state, and is reduced to less than $1/100$th in PCH and IDLE states [8]. Moreover, given stable signal strength, a UE's energy consumption from its radio interface is generally stable regardless of type of transmissions (e.g., uplink vs. downlink) or throughput.

B. Data Collection

The data set used in this study contains anonymized logs collected from RAN and CN of a tier-1 UMTS cellular network in the United States serving over 100 million customers. The cellular network employs High Speed Packet Access (HSPA) technology. Our data set consists of two separate collections, each covering a metropolitan area during a high-profile event in 2012. The collections include information from hundreds of thousands of users and thousands of cell locations over multiple days including the event days. The first event, referred to as Event A hereafter, is an outdoor sporting event that consists of two segments of activities separated by an intermission. The second event, referred to as Event B hereafter, is an indoor professional conference event that consists of multiple segments of activities separated by intermissions of varying lengths. In terms of publicly available attendance statistics, event A is roughly twice the size of event B. Both events attracted on the order of tens of thousands of users. At peak usage during event A, more than fifty thousand users were connected to the cellular network. The activity segments in both events are illustrated by gray bars in all timeseries figures presented in this paper. Furthermore, it is noteworthy that free Wi-Fi service was provided to all users during both of the events to offload as much cellular network traffic as possible. The free Wi-Fi service was intended to offload data traffic and cannot offload voice calls. However, we do not have measurements on the network traffic that was offloaded to these Wi-Fi services; thus, we acknowledge that our results may be biased by this offloading. For instance, users are likely to prefer Wi-Fi over cellular connectivity for high bandwidth activity (e.g., audio/video streaming) due to data caps. Therefore, aggregate traffic volume statistics may be dampened and certain application types may be over- or under-represented in our data set.

The anonymized logs collected at an RNC in RAN contain throughput and RRC protocol request/response information. Using RRC requests from UEs and responses from the RNC, the RAB status of all UEs can be monitored. The anonymized logs collected from the CN contain TCP header information of IP flows carried in PDP context tunnels. They are collected from the Gn interfaces between SGSNs and GGSNs in the core network. They contain timestamp, per-flow traffic volume, content publisher, RTT computed during TCP handshake [9], and estimated packet loss ratio for each TCP flow aggregated in 5 minute bins. All device and user identifiers (*e.g.*, IMSI, IMEI) are anonymized to protect privacy without affecting the usefulness of our analysis. The data set does not permit the reversal of the anonymization or re-identification of users. We note that logs collected at RNCs encompass both voice and data traffic, whereas logs collected from the CN contain only data traffic information. The voice traffic volume is almost an order of magnitude less than the data traffic volume. This trend is inline with recent usage reports and surveys [10].

Before we delve into the performance and workload characterization, we analyze the aggregate movement of user population during the event and routine days for both events. In Figure 3, we plot the cumulative distribution functions (CDFs) of the number of users connected to the cellular network as a function of distance from venue. Comparing CDF plots for the event and routine days for both events, we observe that CDF curves rise faster for the event day as compared to the routine day. This observation highlights that user population is located closer to the venues on the event day. For example, during 9 pm-12 am interval for event A, 40% and 70% of the user population is located within 1 mile radius from the venue on the routine and event day, respectively. Comparing different time intervals, we observe that users' proximity to the venue gradually increases on the event days. For example, the fraction of user population located within 1 mile radius from the venue for event A is 50%, 60%, and 70% for 9 am-12 pm, 3 pm-6 pm, and 9 pm-12 am, respectively. These findings indicate that cellular infrastructure likely faces higher load and performance issues at the cell sectors that are situated closer the venue on the event days.

Next, we characterize performance issues during the aforementioned two high-profile events in Section III. To study the underlying causes of the identified performance issues, we then correlate network performance with various aspects of user traffic patterns in Section IV. It is noteworthy that the while our study uses logs from a UMTS cellular network (with HSPA), the analysis and findings about traffic patterns are generalizable to networks that use other cellular technolo-

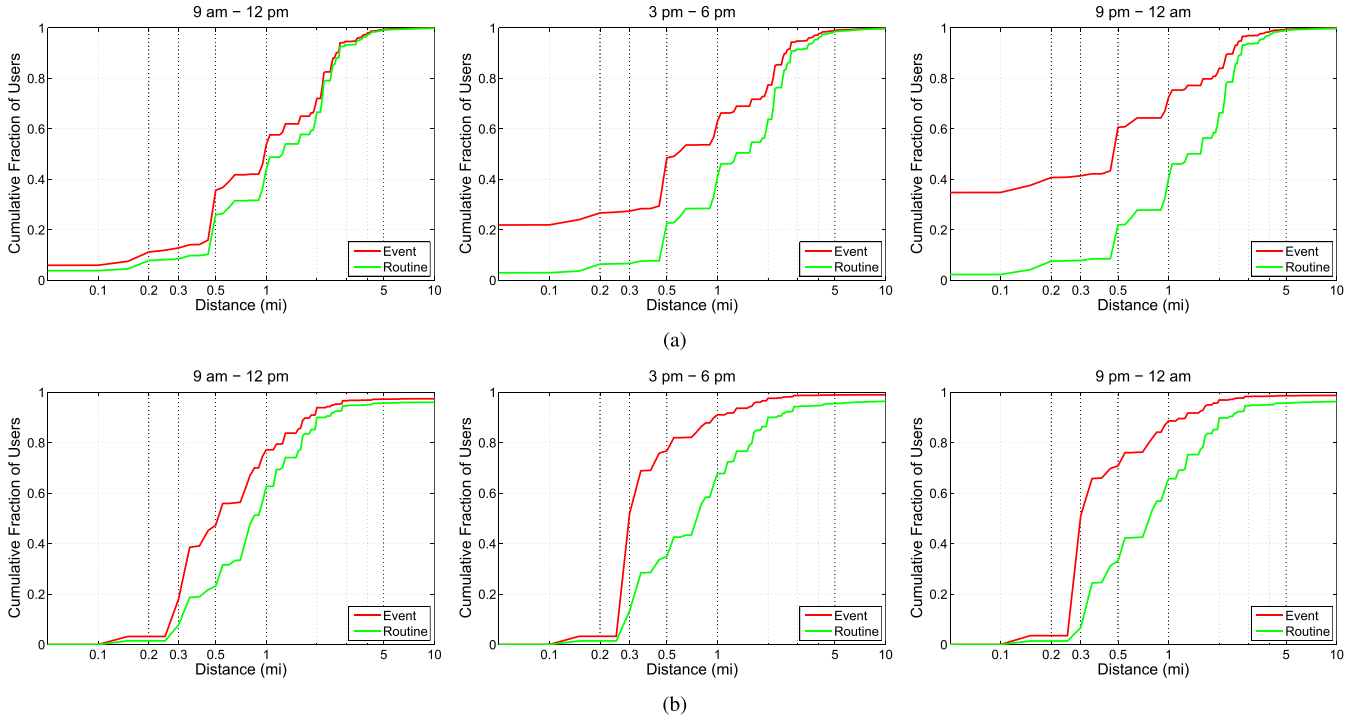


Fig. 3. CDF of number of users as a function of distance. (a) Event A. (b) Event B.

gies (e.g., LTE, WiMAX). Although the network parameters and performance metric are specific to the technology, our analysis and findings (e.g., reliability, inactivity timers) qualitatively hold for types of cellular technologies. Throughout, we present results of the event day in relation to a routine day for baseline comparison. We normalize the actual measurement values by their mean values on the routine day (unless stated otherwise); our results thus effectively represent how the event differs from routine conditions. We omit absolute numbers from some non-normalized plots due to proprietary reasons.

III. Characterizing Performance Issues

Generally speaking, a user's experience about network performance can be divided into two phases. The *pre-connection phase* is characterized by the UE attempting to establish a connection with the cellular network, or in other words establishing a RAB. In this phase, the user waits for connection establishment, while not being able to exchange traffic at this time. The *post-connection phase* starts after a RAB is assigned. In this phase, user experience is related to more traditional voice call performance metrics such as call drop and block rate or end-to-end TCP performance metrics, such as delay and packet loss. Below, we separately discuss both pre- and post-connection network performance experienced by users during both events.

A. Pre-Connection Network Performance

Users may experience difficulty in establishing RABs in the pre-connection phase due to a wide variety of reasons. Every time a request to allocate more radio resources is denied by the RNC, a RRC failure and its underlying reason is logged by our measurement apparatus. In our analysis, we study the logs of various types of RRC failures that are collected at the RNC. Each type of RRC failure corresponds to a specific problem in the cellular network operation. The 3 most common types of failures observed in our data set are the following.

1) *Radio link setup failures* occur when a user's request to setup a radio link is not served due to poor RF channel quality, which is often caused by increased interference.
2) *Radio link addition failures* occur when a user's request to add a radio link to an existing radio connection for soft handovers is denied.
3) *Too many serving cell users* indicates blocking for new users which results when all available RABs are occupied by existing users. For example, our data shows that a cell sector simultaneously supports around 32 users with DCH assignments (for downlink HSPA). If a lot more users try to connect and request DCH assignment, their RAB requests will be blocked.

Figure 4 plots the timeseries of the most common types of RRC failures on the event and routine days for both events. We observe that RRC failures increase sharply on the event days, whereas they are negligible (and steady) on the routine days for both events. For both events, RRC failures start occurring around noon and generally reach their peak either just before or during the event. Specifically, *radio link addition failures* peak at more than $700\times$ their average on the routine day for event A and *too many serving cell users* peak at more than $5000\times$ their average on the routine day for event B.

The nature of RRC failures for both events indicates that their potential root cause is high network load and congestion due to a large number of competing users at cell sector level. Therefore, we next analyze RRC failures at cell sector level before, during, and after the events as a function of distance from the venue. Figure 5 shows the scatter plots between the distance of cell sectors from the venue (in miles) and the ratio of the number of RRC failures on the event day to that on the routine day. The horizontal dashed line at $y = 1$ is a reference

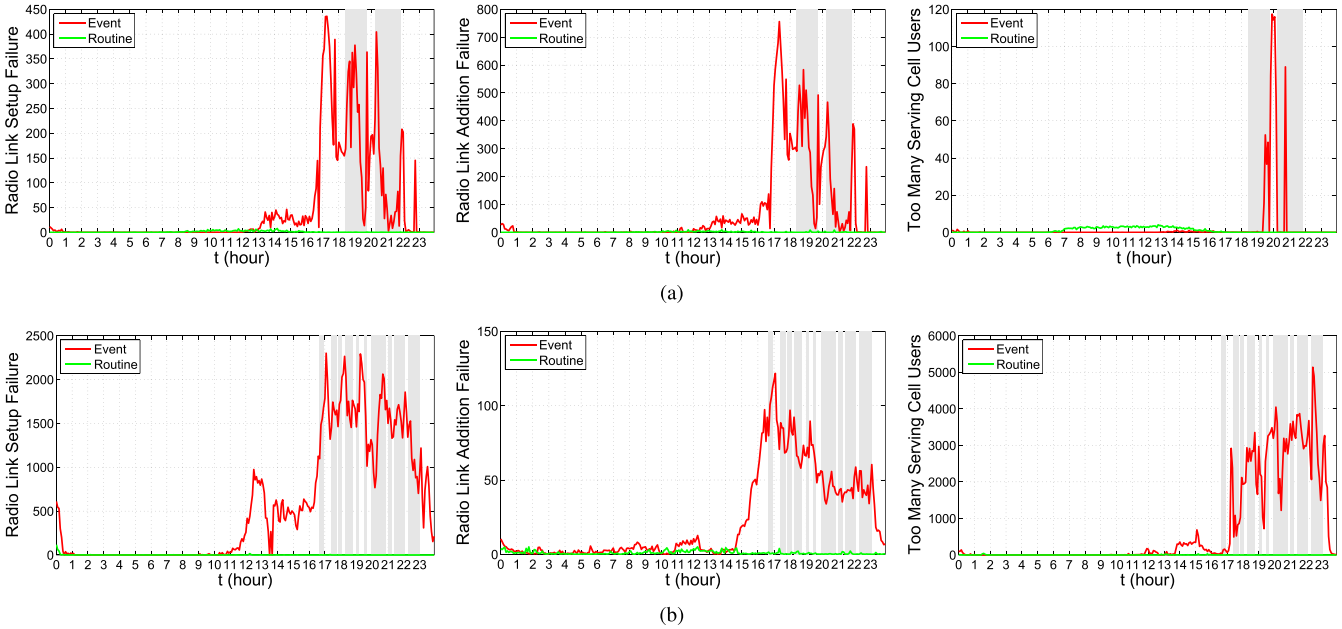


Fig. 4. (Normalized) Timeseries of common types of RRC failures. (a) Event A. (b) Event B.

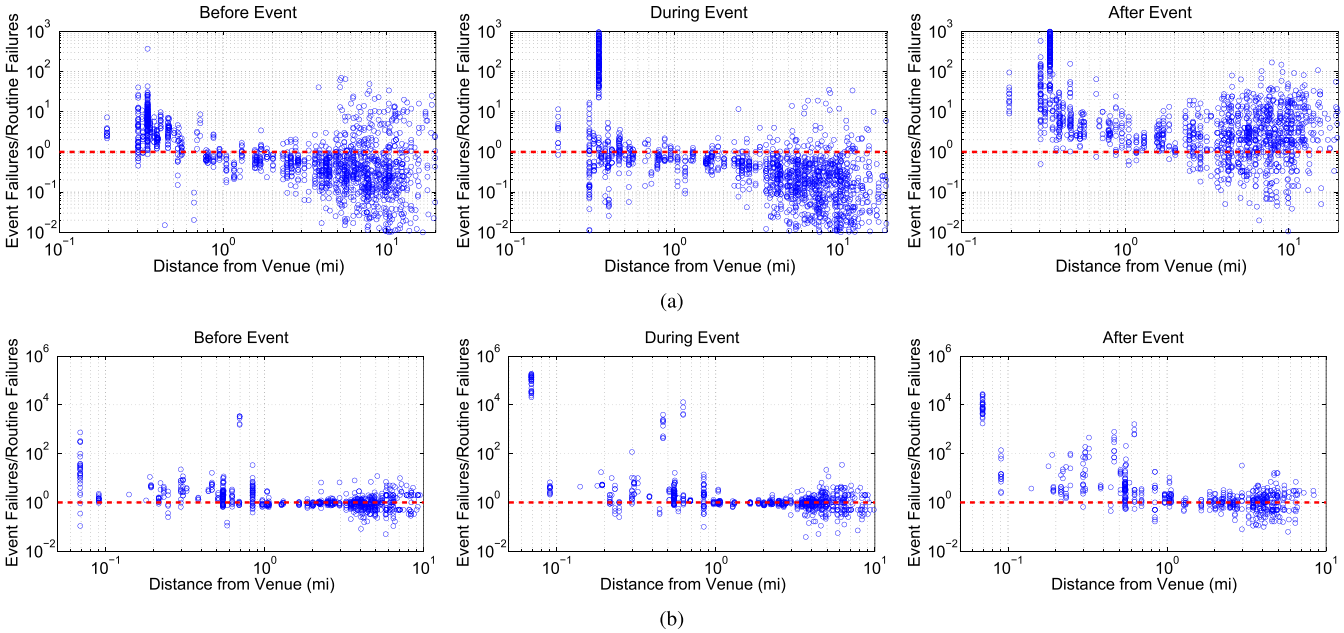


Fig. 5. RRC failure ratios plotted as a function of distance to the venue and for time intervals before, during, and after the event. (a) Event A. (b) Event B.

for the data points where RRC failures on the event and routine days are equal. So the data points above the reference line represent cell sectors that have more RRC failures on the event day than the routine day. Likewise, the data points below the reference line represent cell sectors that have less RRC failures on the event day than the routine day. Both x- and y-axes are converted to logarithmic scale for the sake of clarity. Note that there are many cell sectors equidistant from the venue, especially those cell sectors that are close to the venue. These cell sectors are mounted at the same cell tower but face different directions, and have different tilt angles and frequencies.

Overall, we observe that cell sectors closer to the venue have 2-3 orders of magnitude more RRC failures on the event day than the routine day. The RRC failure ratios progressively decrease as the distance of cell sectors to the venue increases. For both events, we observe interesting dynamics across the scatter plots for time intervals before, during, and after the event. For event A, we observe that the failure ratios generally increase by 2-3 orders of magnitude for cell sectors less than half a mile from the venue throughout the event day. In contrast, for the cell sectors that are far from the venue, their failure ratios drop during the event and jump by 1-2 orders of magnitude after the event finishes. The aforementioned observations can be linked to the sporting nature of event A, where people swarm the venue before and during the event, creating a void in surrounding areas. The post-event jump in the failure ratio is likely correlated with most people leaving

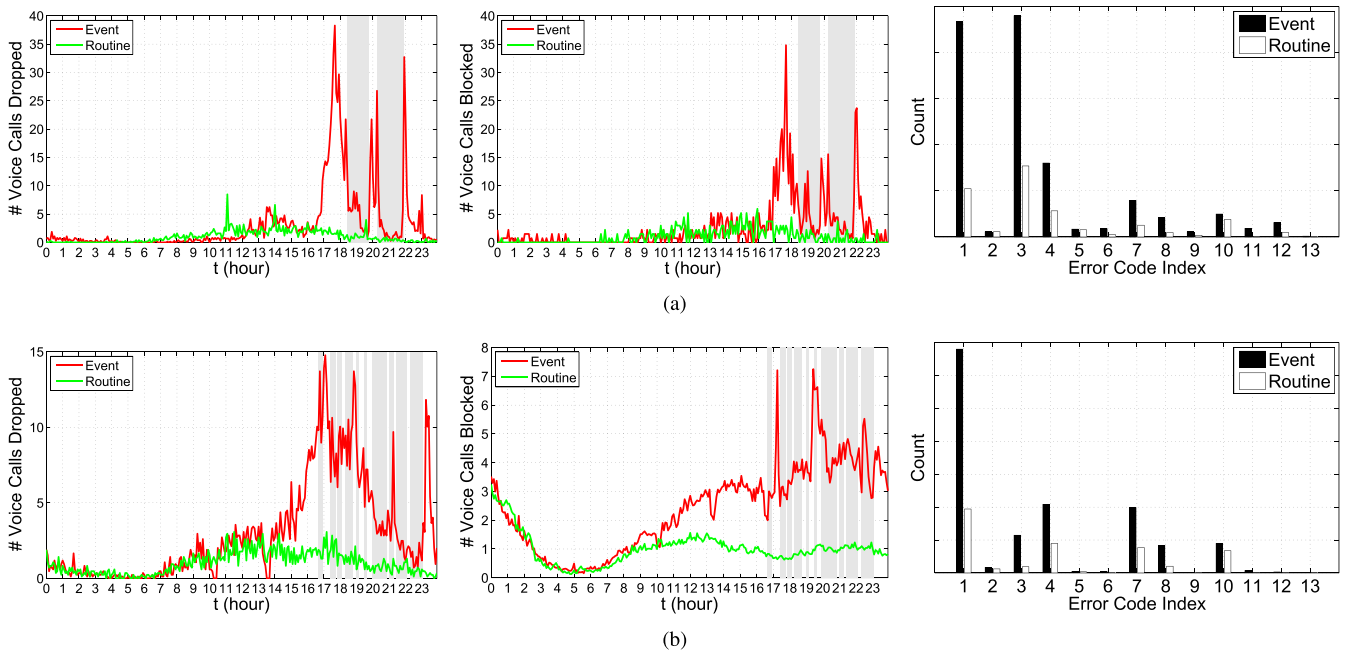


Fig. 6. (Normalized) Voice performance measurements. (a) Event A. (b) Event B.

the venue and using their devices to share their experience with others via voice calls or social network posts (we show later in this section that the observed user activity supports this hypothesis). We observe similar trends for event B as well; however, the post-event jump in the failure ratio is clearly visible only for cells within 1 mile of the venue. For these reasons, while characterizing user network traffic in the next section, we focus our attention on the cell sectors that are within 1 mile radius of the venues for both events.

Summary: Pre-connection failures (especially those pertaining radio link addition and indicating too many serving cell users) peak by a factor of 700 (for event A) and 5000 (for event B) relative to their average on the routine days. These failures increase by 2-3 orders of magnitude in cell sectors very close to the venues before and during the events, but only increase in cell sectors further away after the venues.

B. Post-Connection Network Performance

As discussed in Section II, during the RAB setup phase, the RNC verifies that the needed radio resource for the request actually exists before it assigns a RAB. In other words, if a device has successfully acquired a RAB for communication, its performance should theoretically remain acceptable per operator's configuration even if the overall network demand level exceeds network capacity. This is because excessive demand requests will get blocked off by the RNC from acquiring any RAB. However, network conditions can quickly change even for UEs that have already acquired a RAB because of factors such as interference, mobility, *etc.* Such dynamic network conditions can force UEs to request a change in current RAB status, initiating a series of RRC failures which could in turn result in degraded voice and data performance. Below, we separately discuss voice and data performance.

1) Voice Performance: To quantify voice performance, we study voice call drop and block rates for both events in Figure 6. Similar to our observations about pre-connection network performance, the number of voice call drops and blocks increase substantially on the event days as compared to the routine days for both events. Specifically, we observe peak increases of more than 30× and 7× relative to their average on the routine days for events A and B, respectively. It is noteworthy that voice call drop and block rates peak just before the start of the events, during the intermissions, and at the end of the events. This observation is consistent with our expectation that users are less likely to make voice calls during event activities and more likely to make voice calls either before the start of events, after the end of events, or during intermissions between event activities. To further investigate the root causes of voice call blocks and drops, we also plot the histogram of their error codes in Figure 6. The error code descriptions in Table I indicate that the two most common categories of error codes for both events are related to radio connection supervision and soft handovers, which in turn point to interference and mobility as the root cause. The number of "unspecified" errors also increases substantially; however, our measurement infrastructure does not provide any information to investigate them further.

2) Data Performance: To quantify data performance, we study two key end-to-end TCP performance metrics: packet loss ratio and RTT for both events.[1] Packet loss ratio quantifies network reliability. We only have packet loss ratio measurements for TCP flows, which constitute approximately 95% of all flows in our data set. RTT quantifies network delay and is defined as the duration of time taken by a packet to reach the server from the UE plus the duration of time taken by a packet to reach the UE from the server. It is important to note that RTT measurements are biased by differences in the paths between different UEs and the external servers they communicate with. Similar to packet loss ratio measurements, we only

[1]Because end-to-end TCP performance also involves additional parameters such as back-haul bandwidth and even remote server load, we leave a more detailed investigation of TCP performance to future work.

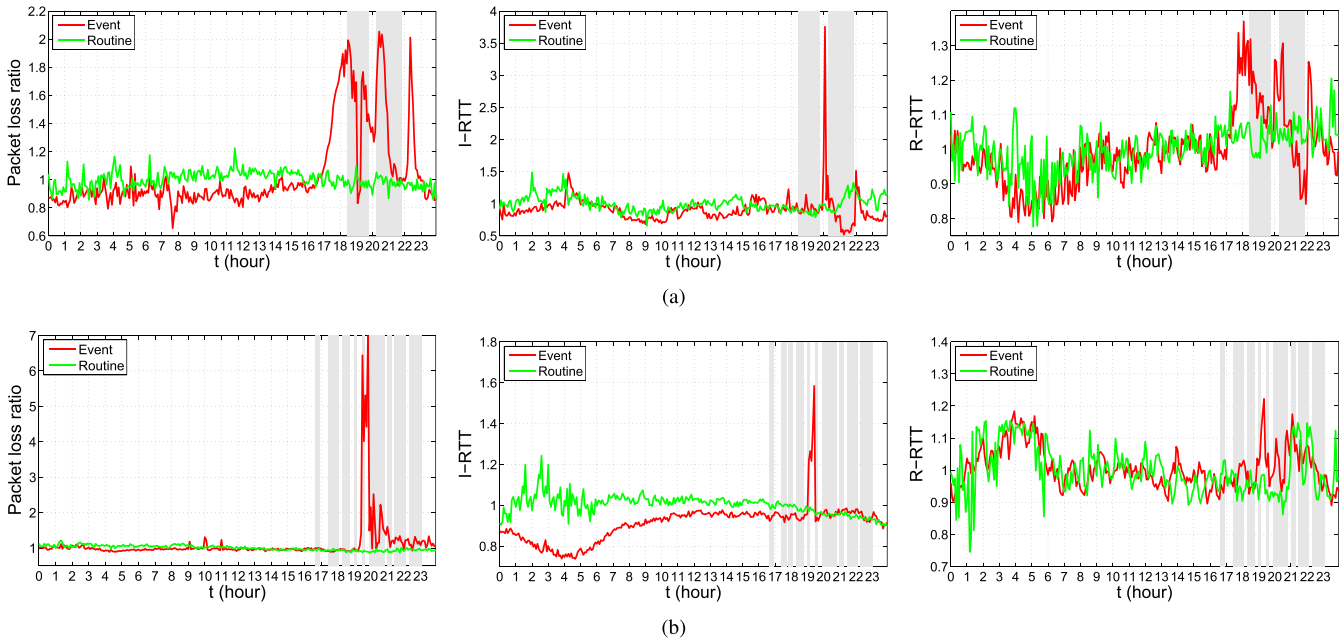


Fig. 7. (Normalized) Data performance measurements. (a) Event A. (b) Event B.

TABLE I
DESCRIPTION OF VOICE CALL ERROR CODES

Index	Category	Description
1	Unspecified	All cases which do not map to the ones described below
2	Radio Connection Supervision	Radio Link Control (RLC) unrecoverable
3	Radio Connection Supervision	Maximum number of RLC retransmissions
4	Radio Connection Supervision	Expiry of timer
5	Radio Connection Supervision	Radio link failure indication
6	Operations & Management	Cell lock indication
7	Soft Handover	No active set addition update
8	Soft Handover	No active set deletion update
9	Soft Handover	No active set replacement update
10	Soft Handover	Cell not in the neighbor set
11	Soft Handover	High speed-downlink shared channel cell change failure
12	Inter-Frequency Handover	Inter-frequency handover failure
13	Channel Switching	Transition to DCH state not completed

have RTT measurements for TCP flows. RTT measurements for TCP flows are estimated by SYN, SYN-ACK, and ACK packets in the TCP handshake. In a cellular network, RTT essentially consists of two components: radio network RTT and Internet RTT. Radio network RTT (R-RTT) is the time duration between the SYN-ACK packet from server passing the Gn interface and the ACK packet from the UE passing the Gn interface. Internet RTT (I-RTT) is the time duration between the SYN packet from the UE passing the Gn interface and the SYN-ACK packet from the server passing the Gn interface. Thus, RTT = R-RTT + I-RTT.

Figure 7 shows the timeseries plots of packet loss ratio, Internet RTT, and radio network RTT for both events. Packet loss ratio peaks at 2× and 7× relative to its average on the routine days for events A and B, respectively. Median packet loss ratio on both event days was around 2%, which is inline with measurements reported in prior literature (e.g., [11], [12]). We observe different trends for radio network RTT and Internet RTT for both events. There is only a minor increase in radio network RTT on the event days. Median radio network RTT on both event days was around 20 milliseconds. Internet RTT increases during the intermissions for both events; however, this increase indicates congestion at remote servers caused by increased event-driven traffic. Median Internet RTT on both event days was around 200 milliseconds. RTT measurements are also inline with measurements reported in prior literature (e.g., [11], [12]). Overall, data performance results indicate that users experience data connection performance issues to varying extents during the two events.

Summary: Post-connection performance degradation is observed for both voice and data network during the events. Specifically, voice call failures (dropped calls and call blocks) increase by a factor of as much as 30 (for event A) and 7 (for event B). Moreover, packet loss ratio increases by a factor of 2 (for event A) and 7 (for event B); while the RTT increases by a factor of 3.5 (for event A) and 1.5 (for event B). While these indicate a degradation in performance experienced by users already connected to the network, this is substantially smaller than the pre-connection failures discussed in Section III-A. Overall, pre- and post-connection network performance results highlight that limited radio resources are the major bottleneck during crowded events.

IV. UNDERSTANDING PERFORMANCE ISSUES

Next, we characterize user network traffic to identify patterns that correlate with the observed pre- and post-connection

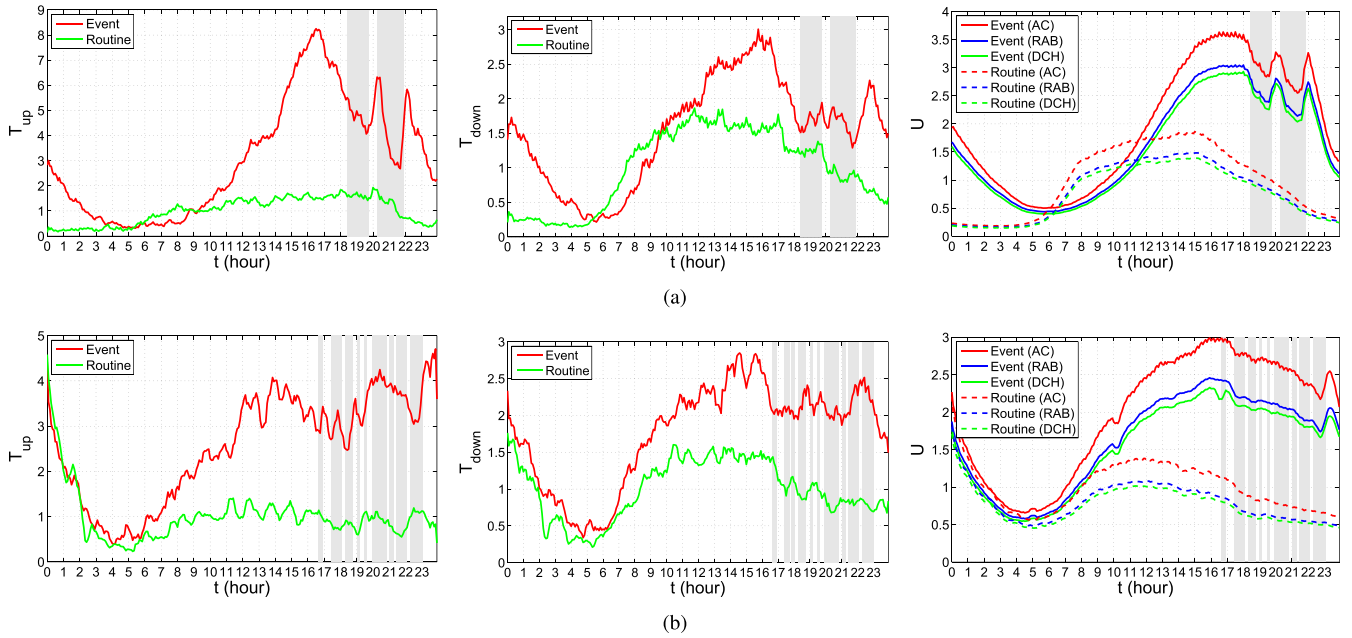


Fig. 8. (Normalized) Network load measurements. (a) Event A. (b) Event B.

performance degradation during the events. Using the insights obtained from this characterization, we aim to identify network optimization opportunities that can potentially improve end-user experience in crowded locations. We characterize network traffic in terms of both aggregate network load and user-level session characteristics.

A. Aggregate Network Load

We quantify aggregate network load in terms of the following two metrics: data traffic throughput and user counters. Throughput or bit-rate is sampled for all UEs at the RNC every couple of seconds. Based on the direction of traffic, we can split the throughput into *uplink throughput* (T_{up}) and *downlink throughput* (T_{down}). Figure 8 plots the timeseries of uplink and downlink throughput on the event and routine days for both events. For the routine days, both uplink and downlink throughput peak around the noon time and decline steadily afterwards, reaching the bottom during late night and early morning. We observe a different trend for uplink throughput on the event days. For instance, the peak uplink throughput on the event day is more than 8x and 4x the average throughput on the routine day for events A and B, respectively. We also observe that the uplink throughput peaks and event activities are approximately aligned. For instance, uplink throughput sharply increases at the start and end of the second segment for event A. Similar, though less pronounced, patterns are also observable for event B. In contrast to the uplink throughput, increases in the downlink throughput timeseries are steadier for both events.

To further analyze traffic volume characteristics, we plot the traffic flow count histograms for top content publishers in Figure 9. We focus on flows rather than bytes to avoid bias towards high volume applications, such as video streaming. We observe that flow counts of social networking content publishers more than double on the event day as compared

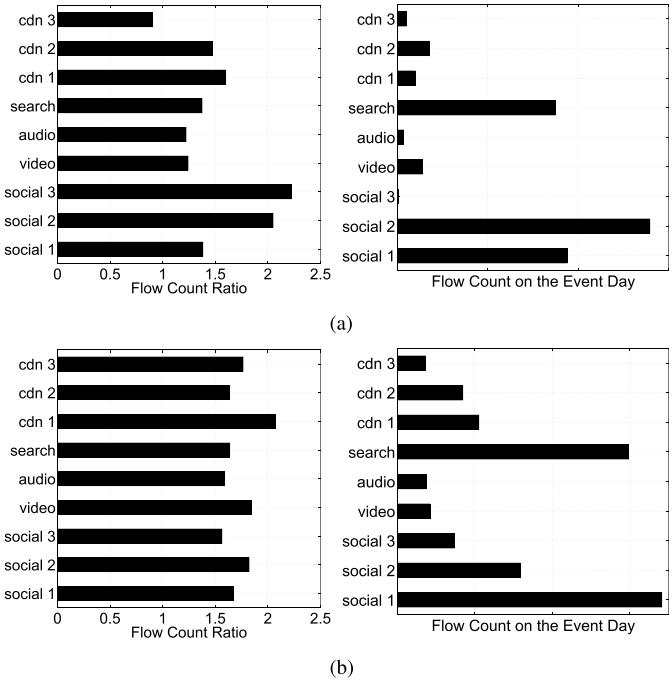


Fig. 9. Flow count histograms for top content publishers in our data set. (a) Event A. (b) Event B.

to the routine day for event A. Likewise, social networking content accounts for most flows on the event day for event B. Our further investigation (not shown here) revealed that social networking content is at least 2× more upstream heavy as compared to other content types, which explains the increase in uplink throughput during both events.

We also analyze user counters for the event and routine days for both events. Users are classified into the following overlapping categories based on their RRC states: *admission control* (AC), *radio access bearer* (RAB), and *dedicated*

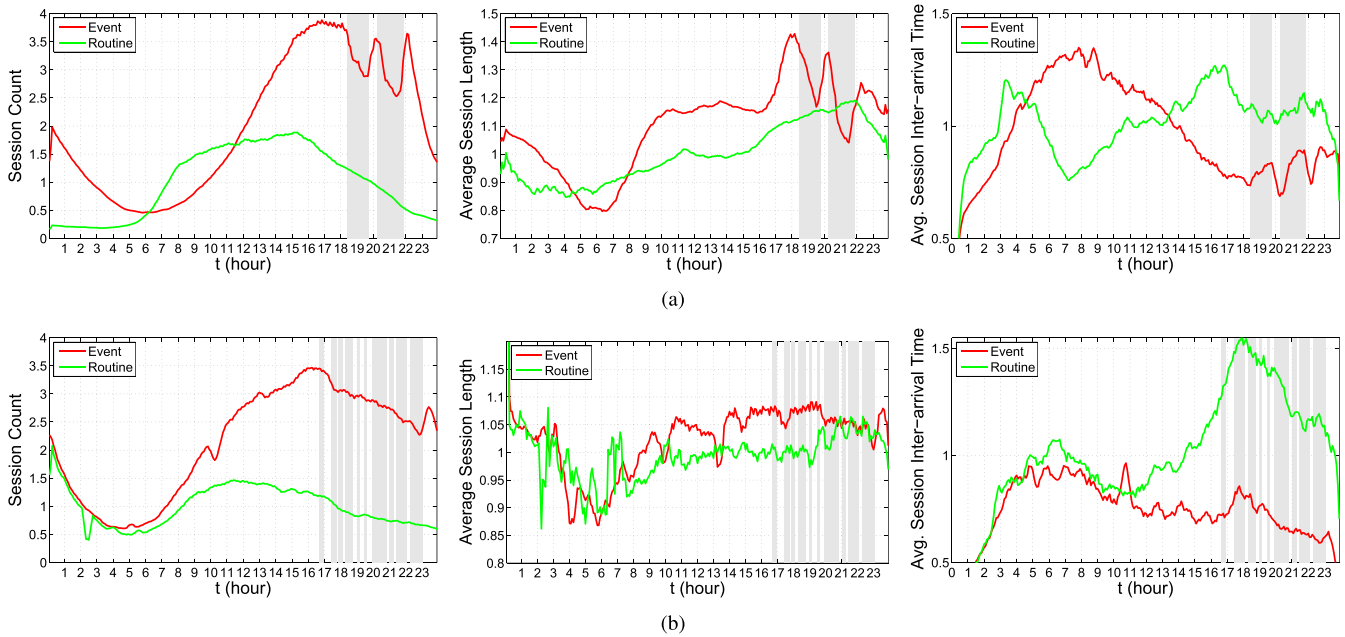


Fig. 10. (Normalized) Session Count, Average Length, Average Inter-arrival Time. (a) Event A. (b) Event B.

channel (DCH). AC category includes the users who have completed the admission control procedure. RAB category includes the users who have been assigned a RAB after admission control. Such users are typically in either FACH or DCH state. Finally, DCH category only includes the users who are in DCH state. Let U denote the number of users, also let U_{AC}, U_{RAB}, and U_{DCH} denote the number of users in the aforementioned categories. As a general rule, $U_{AC} \geq U_{RAB} \geq U_{DCH}$. Figure 8 plots the timeseries of number of users in AC, RAB, and DCH categories. These timeseries show a trend similar to the throughput measurements. All user counters have higher values on the events days as compared to the respective routine days for both events. Specifically, the number of users with admission control peaks at more than $3\times$ during the events as compared to its average on the routine days.

Summary: Both aggregate uplink and downlink throughput increase during the event days; uplink throughput increases by a factor of as much as 8 and 4 (for events A and B respectively), while downlink throughput increases by a factor of 3 (for both events). Moreover, there is a substantial increase in the traffic volume of social networking content during the events, which is relatively more upstream heavy. Likewise, number of users with admission control increase by a factor of 3 for both event days. Overall, our aggregate network load characterization shows that increased user activity during the events, specifically in terms of uplink throughput and user counters, is correlated with increased pre-connection failures. To reduce the impact of increased network load during crowded events, we will investigate the effectiveness of opportunistic connection sharing in Section V-B.

B. User-Level Sessions

We now analyze characteristics of user-level traffic sessions for both events. A session consists of consecutive time intervals with uplink or downlink byte transfer and its end is marked by an inactivity timeout of τ seconds. The results presented in this section are computed for $\tau = 5$ seconds. Changing the value of τ does not qualitatively affect the analysis results. Figure 10 shows the timeseries of session count, average session length, and average session inter-arrival time for both events. Session count follows a similar trend to the earlier aggregate network load metrics – at peak, there is more than $3.5\times$ increase relative to the average on the routine days for both events. Furthermore, we observe an increase in average session length on the event days as compared to the routine days, *e.g.*, there is more than $1.4\times$ increase for event A. On the contrary, average session inter-arrival time decreases sharply on the event days as compared to the routine days – this indicates that users are initiating sessions much more frequently during the events. To further investigate the nature of changing session patterns, we plot the timeseries of average downlink bytes per session (B_{down}), average uplink bytes per session (B_{up}), and the average ratio of downlink bytes to uplink bytes per session in Figure 11. We observe that average downlink bytes per session sharply decreases up to 0.5x during the event days; whereas, average uplink bytes per session exhibits a mixed trend. The ratio (B_{down}/B_{up}) also sharply decreases during the events, which is due to the increased traffic volume of upstream-heavy social networking content.

Summary: User sessions are on average longer during both events (by a factor of as much as 1.4) – as well as more numerous and more frequently initiated. However, users exchange only as much as half the bytes per session on average. This change in workload is due to a change in the application usage during these events, such as greater proportion of social networking flows observed earlier. These trends point to potential waste of radio resources by UEs, which can be mitigated by tuning radio network parameters. Towards this end, we will investigate the effectiveness of varying RRC timeouts in Section V-A.

V. Evaluating Mitigation Schemes

In this section, we evaluate two proposals to mitigate cellular network performance degradation during crowded events.

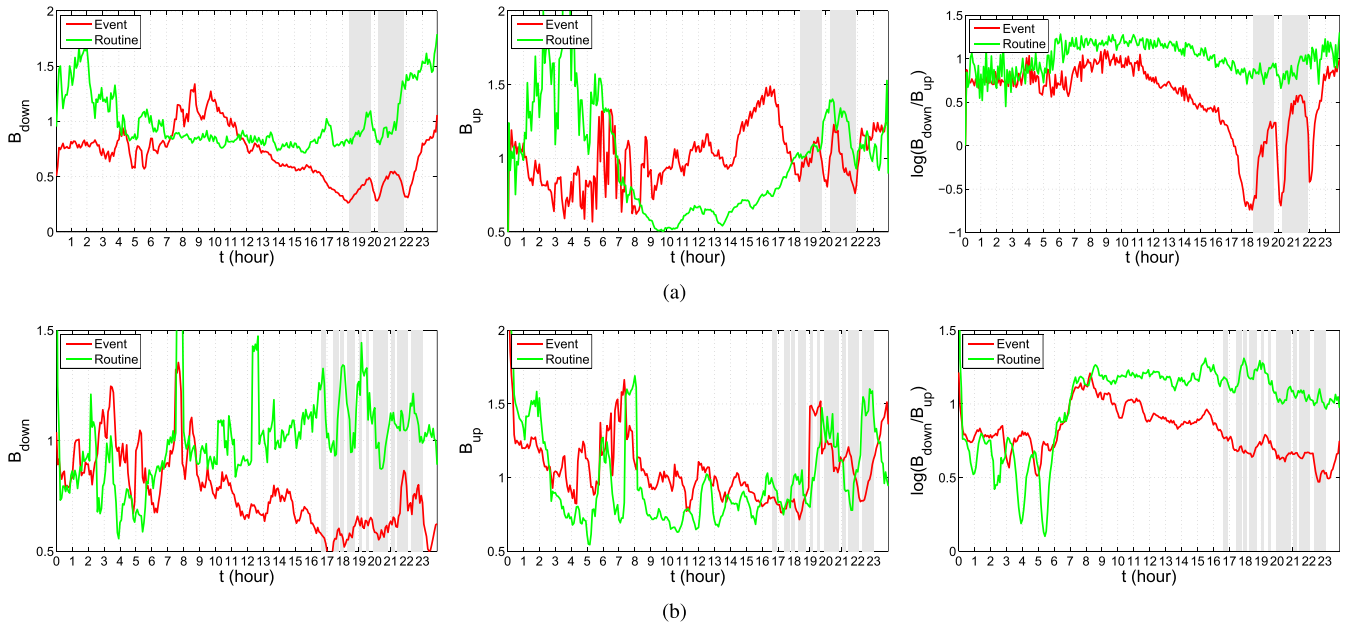


Fig. 11. (Normalized) Per-Session Downlink Bytes (B_{down}), Uplink Bytes (B_{up}), Ratio ($B_{down}/(B_{up}$). (a) Event A. (b) Event B.

A. Radio Network Parameter Tuning

We first investigate whether tuning radio network parameters can result in more efficient radio resource usage during crowded events. As mentioned in Section II, UEs acquire and release radio resources by transitioning to different RRC states. A UE is promoted to a higher energy state depending on buffer occupancy and it is demoted to a lower energy state depending on timeouts. Here, we study how RRC timeouts can be tuned for more efficient radio resource utilization, without explicit feedback from individual UEs. Recall from Figures 10 and 11 that average bytes per session decreases during the events, despite the increase in average session length. This observation highlights potential waste of radio resources and UE energy consumption in crowded locations. Therefore, a natural suggestion would be to reduce RRC timeouts to mitigate the radio resource wastage. However, reducing RRC timeouts can result in more frequent state promotions, which can introduce additional signaling traffic and state promotion delays resulting in degraded user experience [8], [13]. Hence, there is a tradeoff between performance and resource efficiency.

1) Simulation Setup: We conduct trace-driven simulations to quantitatively study the tradeoffs involved in changing RRC timeouts. We simulate the RRC state machine of every user using the RNC logs while focusing on the DCH state, which has the highest allocated radio resources and energy consumption among all RRC states. Specifically, we study DCH→FACH RRC timeout parameter, which is denoted by α hereafter. As mentioned earlier, changing RRC timeouts introduces tradeoffs among radio resource wastage, user experience, and UE energy consumption. We use the following three performance metrics to quantify these factors. (1) The DCH state idle occupation time, denoted by $T_{\text{DCH-IDLE}}$, quantifies the radio resources wasted by UEs in DCH state. (2) The promotion delay quantifies the additional delay caused when UEs transition to DCH state from FACH state. (3) The power consumption quantifies the total energy consumed by UEs during DCH state occupation and in FACH to DCH transitions. We use the following simulation parameters in our experiments (inferred by Qian *et al.* in [8]): (1) FACH→DCH promotion radio power = 700mW, (2) DCH state power = 800mW, (3) FACH→DCH promotion delay = 2 sec, and (4) RLC buffer threshold = 500 bytes. To conclude, RRC state machines are simulated using our data traces and the performance metrics are simulated using the parameter values from prior literature. Specifically, we simulate RRC state machines with varying DCH→FACH RRC timeout value. For example, RRC state machine would quickly transition from DCH state to FACH state for smaller timeouts. As a result, users' DCH state idle occupation time and energy consumption will reduce while promotion delays will increase.

2) Results and Discussions: Similar to the evaluation of opportunistic connection sharing, we evaluate radio network parameter tuning for a subset of cell sectors that are within 1 mile radius of the venues. We conduct trace-driven simulations of individual users' RRC state machines on this subset for both event and routine days. Figure 12 shows the timeseries plots of the aforementioned three performance metrics for varying α values. We observe that the DCH state idle occupation time and UE energy consumption increase for larger α values. On the other hand, promotion delay decreases for larger α values. These observations indicate that decreasing the RRC timeout values reduces the waste of scarce DCH channels and UE energy consumption. However, this benefit is achieved at the cost of increased promotion delay that may degrade user experience, especially for applications that are not delay-tolerant.

To systematically study the tradeoffs between these performance metrics on the event days and compare them to routine days, we plot them as a function of α. Figure 13 plots the max-normalized average of the performance metrics as a function

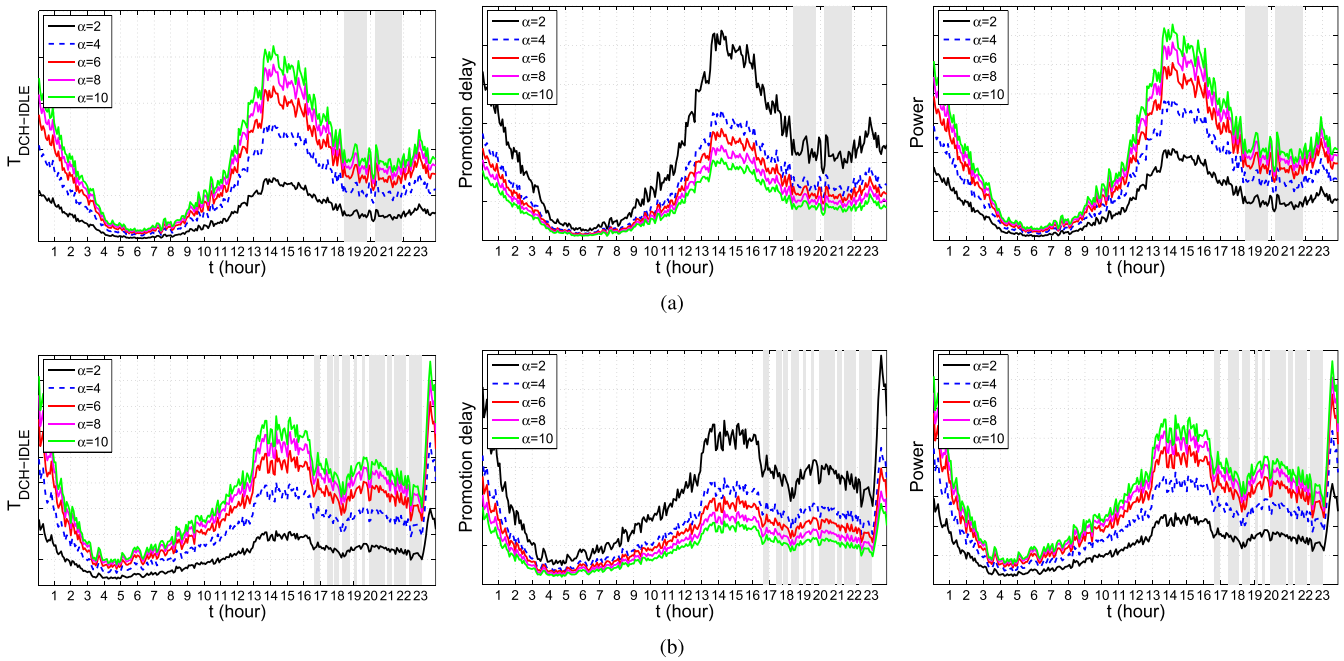


Fig. 12. Experimental results for radio network parameter tuning. (a) Event A. (b) Event B.

of α for the event and routine days. In theory, we want to select a value of α which simultaneously minimizes the values of all performance metrics. In this case, the crossover points (highlighted by black circles in Figure 13) and their corresponding α values represent suitable performance tradeoff. We find that these crossover points shift to smaller α values – by 1-2 seconds – on the event days as compared to the routine days. It is noteworthy that the type of normalization does not qualitatively affect the analysis results. In practice, α is typically set to achieve a target delay or resource overhead. In this case, as observable from Figure 13, we can tune α to smaller values to achieve the same targets and achieve strictly better performance during crowded events.

B. Opportunistic Connection Sharing

We now evaluate a simple opportunistic connection sharing scheme to reduce the network load at individual cell sectors for eradicating RRC failures observed in Section III-A. The basic idea is that users can share their connection to NodeBs with other users to reduce the overall network load in terms of occupied radio channels. In this scheme, a selected set of UEs act as Wi-Fi hotspots for other UEs in their vicinity. Therefore, other UEs, instead of wastefully establishing separate connections, can connect to NodeBs via the UEs acting as Wi-Fi hotspots. Using this approach, we aim to reduce the number of UEs that are directly connected to NodeBs to free up channels, although the overall throughput carried by the network remains the same.

1) Simulation Setup: To evaluate the potential benefit of the opportunistic connection sharing scheme, we conduct cell sector level trace-driven simulations. We assume that users are static within 1 minute time bins. This is a reasonable assumption for crowded events in stadiums, auditoriums, and conference rooms. We do not have fine-grained location information of users in our data set; therefore, we have to simulate

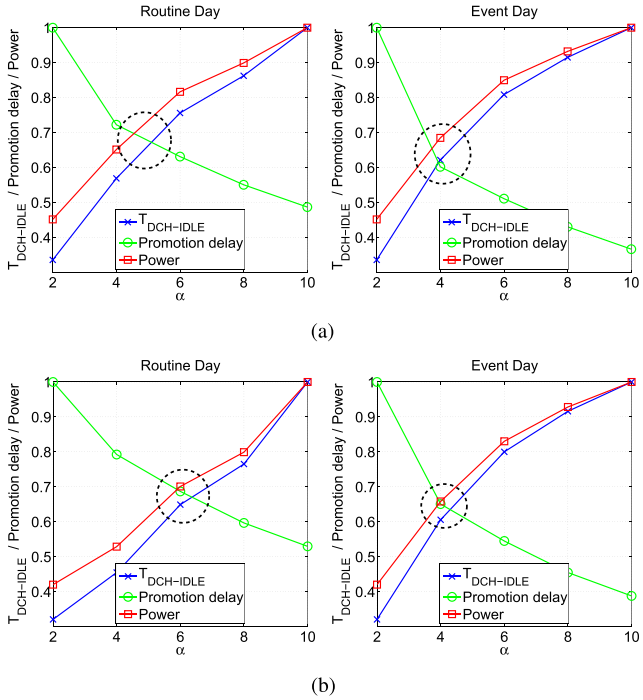


Fig. 13. Tradeoff between performance metrics for varying RRC timeout (α) values. Y-axis is max-normalized for each metric. α values corresponding to black circles achieve better performance tradeoff. (a) Event A. (b) Event B.

the locations of users. In this paper, we aim to generate the locations of users in a grid-like scenario – similar to how people are typically seated in stadiums and conferences. Towards this end, we use Complete Spatial Randomness (CSR) point generation model with hard-core inhibition [14]. CSR with hard-core inhibition does not allow neighbors within a predefined radius around the randomly generated points, resulting in a grid-like setting. The points in the realizations denote the locations of users in our simulations. In our simulations,

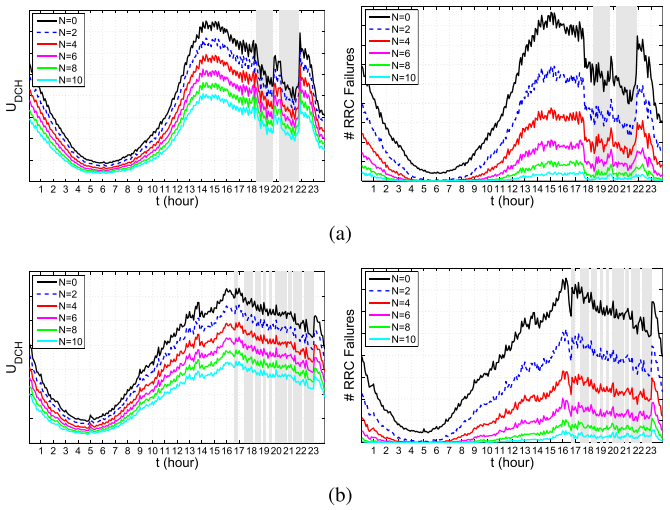


Fig. 14. Experimental results for opportunistic connection sharing. (a) Event A. (b) Event B.

Wi-Fi hotspots are randomly selected among UEs because we do not have access to other relevant information, such as battery life and signal strength, that may be used to optimize this selection. Once a UE connects to a Wi-Fi hotspot, it is disconnected after 1 minute of inactivity. The locations of inactive users are updated using the above CSR model. In our simulations, the cell sectors are set to have $2,250,000$ ft^2 coverage area, the inhibition radius is set to 2 ft, the hotspot Wi-Fi range is simulated as $\mathcal{N} \sim (200 \text{ ft}, 20 \text{ ft})$, and the upper limit on the number of simultaneous connections for each Wi-Fi hotspot is set to 5. The cell sector coverage area is in typical range for crowded urban locations, the inhibition radius is set to be reasonably large, and the Wi-Fi range and the maximum number of simultaneous connections are conservatively set. The range of a Wi-Fi hotspot, once randomly selected, is kept fixed. We minimize adjacent-channel interference by selecting only a few Wi-Fi hotspots per cell sector that can choose non-overlapping Wi-Fi channels in 2.4 GHz and 5 GHz bands. Moreover, we neglect co-channel interference because the number of simultaneous connections for each Wi-Fi hotspot is limited to 5. We assess the benefit of the connection sharing scheme in terms of the following metrics: the number of users in DCH state (U_{DCH}) and the number of RRC failures. The performance metrics (number of users in DCH state and RRC failures) are simulated using RNC logs. Specifically, we exclude a user from DCH state occupation and corresponding RRC failures if she/he is in the range and served by a nearby Wi-Fi hotspot in our simulation instead of cellular connectivity.

2) Results and Discussions: Since we are primarily interested in deploying this scheme in congested locations, we focus our evaluations on a subset of cell sectors in our data set that are within 1 mile radius of the venues. We evaluate the opportunistic connection sharing scheme using trace-driven simulations on this subset on the event days. The results plotted in Figure 14 are the average of 1000 independent simulation runs. We plot the timeseries of the number of occupied DCH channels (U_{DCH}) for varying number of Wi-Fi hotspots per cell sector (denoted by N). As expected, we observe that U_{DCH} values become smaller for larger values of N, freeing up DCH channels that are now available for UEs unable to transition to the DCH state due to RRC failures. We also plot the number of RRC failures for varying values of N in Figure 14. Again, as expected, we observe that RRC failures decrease for increasing values of N. Consequently, based on instantaneous load conditions, the cellular network can dynamically vary the required number of users acting as Wi-Fi hotspots to minimize RRC failures. We note that this connection sharing scheme successfully eradicates more than 95% RRC failures for both events when $N = 10$. This substantial reduction in the number of RRC failures in congested cell sectors will likely result in improved performance for users.

3) Practical Issues: Below, we discuss some practical issues of opportunistic connection sharing.

a) Wi-Fi hotspot selection: The selection of Wi-Fi hotspots can be mediated by the cellular network based on a variety of factors, such as battery life and signal strength. UEs acting as Wi-Fi hotspots may experience high energy drain and may run out of battery power. To address this issue, the role of Wi-Fi hotspot can be periodically rotated among the user pool by the cellular network. The cellular network should prefer UEs with better signal strength because UEs consume significantly more energy and suffer reduced effective bit rate when the signal strength is poor [15]. On the other hand, the UEs that are unable to get RAB assignments can discover Wi-Fi hotspots in their range using the standard Wi-Fi discovery methods. In case of multiple options, UEs should prefer hotspots with better signal strength.

b) Initial connection delay: After a device connects to a Wi-Fi hotspot, similar to RRC protocol, it disconnects after a pre-defined inactivity timer expires. However, the value of this timer should be set much higher than the corresponding RRC timers so that the device does not have to incur initial delay, which is up to several seconds, for every data transfer. In our simulations, the inactivity timer was set to be 1 minute.

c) Out of range: A device has to request RAB assignment when it moves out of a hotspot's Wi-Fi range. If it is unable to get a RAB due to congestion then the RNC can dynamically assign more Wi-Fi hotspots in the cell sector to provide connectivity to more users.

d) Radio technologies: Opportunistic connection sharing is only usable when a majority of devices in the cellular network have built-in Wi-Fi capability. In our simulations, we assume that all devices have Wi-Fi capability. In case Wi-Fi is not available, other technologies such as Bluetooth can also be used. Bluetooth has lower power consumption, smaller radio range, and supports less data rate as compared to Wi-Fi. Consequently, it can be used as a low power alternative for small transmissions such as tweets.

e) Wi-Fi-cellular handovers: Working extensions to the Wi-Fi standard already address the issue of smooth handovers between Wi-Fi and cellular networks, including 3GPP Access Network Discovery and Selection Function (ANDSF), Hotspot 2.0 initiative [16], and other techniques [17].

f) Voice traffic offloading: In this opportunistic connection sharing scheme, voice traffic can be tunneled via the Wi-Fi connection using the well-known Voice over Wi-Fi solutions, such as Wi-Fi certified Voice-Enterprise [18].

g) Incentives: Cellular network operators may provide billing based incentives to users for participating in this opportunistic connection sharing scheme.

C. Limitations

Below, we briefly mention two limitations of our trace-driven simulation evaluations. First, our simulation based evaluations cannot account for changes in traffic workload resulting from different network conditions due to our proposed mitigation schemes. Second, they also cannot account for low-level dependencies between performance metrics and network load. For example, some types of RRC failures are impacted by interference, which in turn is a function of network load. Addressing these limitations requires experiments on operational cellular networks, which are beyond the scope of this work. However, despite these limitations, we believe that the sheer magnitude of the improvements observed in our simulations indicates that the mitigation schemes discussed in this paper would accrue some benefit in practice.

VI. Related Work

Cellular Performance Characterization: Cellular performance characterization have recently received much attention by the research community. For example, small-scale studies have characterized application performance [11], [19] and fairness [7]. Large-scale studies have characterized throughput and airtime [20], smartphone traffic [21], M2M device traffic [22], smartphone app traffic [23], [24], and heavy users [25]. In contrast to these studies, we believe that we are the first to analyze cellular performance changes specifically during crowded events. Erman and Ramakrishnan studied traffic characteristics of LTE cellular network of a large network operator during the 2013 Super Bowl [26]. Similar to our findings, they observed a spike in LTE traffic volume during the event. While they did not evaluate opportunistic connection sharing or radio network parameter tuning, they studied the feasibility of trading-off delay for delay-tolerant applications to mitigate peak congestion. Their results showed that 5 minute delay tolerance can achieve 25% reduction in peak usage.

Radio Network Parameter Tuning: Prior work on radio network parameter tuning study the impact of RRC timers on network performance and smartphone energy consumption. Most prior work is based on user-end measurements performed using a few cellular devices. For instance, Liu *et al.* characterized performance in a 1xEV-DO network using measurements obtained from two laptops equipped with Sierra Wireless data cards [27]. Balasubramanian *et al.* proposed a UE based approach, called TailEnder, to alter traffic patterns based on the prior knowledge of RRC state machine [28]. Some studies are based on theoretical analysis and simulation. For instance, Liers *et al.* proposed a scheme to adaptively tune RRC timeout parameters based on the demand and load situation, and validated it using simulations [29]. Yeh *et al.* proposed a scheme to tune RRC timeout parameters using analytical models based on available radio resources, energy consumption, quality of service, and processing overheads of the radio access network [30]. Athivarapu *et al.* proposed a client-side solution to mine program execution for optimizing radio state transition parameters [31]. Qian *et al.* conducted trace-driven RRC state machine simulations using network-end measurements to investigate the optimality of RRC timeout parameters [8]. Furthermore, they proposed a client-side, application-aware tail optimization protocol to simultaneously optimize radio and energy resources [32]. Client-side solutions have access to fine-grained information and are profiled for individual users, thus they may achieve better performance. However, their implementation requires OS modification at clients and cooperation from cellular carriers, which may be prohibitive for wide-scale adoption. In contrast to client-side solutions (e.g., [8], [28], [31], [32]), we focus on network-end tuning of RRC timeouts without any cooperation from UEs.

Opportunistic Connection Sharing: We build on existing work on opportunistic traffic offloading [4], [5]. Luo *et al.* proposed a unified architecture, where mobile clients use both 3G cellular link and Wi-Fi based peer-to-peer links for routing packets via peer-to-peer links to the appropriate destinations [4]. Han *et al.* proposed content-specific opportunistic communication scheme to offload cellular traffic via Wi-Fi or Bluetooth [5]. However, neither of these proposals were evaluated using real-world traces, and both approaches require architectural changes to network protocols and hardware. Our work complements these proposals by showing that their simplest and most practical instantiation — a simple one-hop connection sharing scheme that does not require architectural changes — can be very effective in real-life crowded events. To the best of our knowledge, this paper is the first to evaluate practical connection sharing techniques on real-world traces.

VII. Conclusion

This paper presents the first performance characterization of an operational cellular network during crowded events. We make three key contributions in this study based on the real-world voice and data traces that we collected from a tier-1 cellular network in the United States during two high-profile crowded events in 2012. First, we measured how cellular network performance degrades during crowded events as compared to routine days. Second, we analyzed what causes the observed performance degradation. Third, we evaluated how practical mitigation schemes for the observed performance degradation would perform in real-life crowded events using trace-driven simulations. Our findings from this study are crucial for cellular design, management, and optimization during crowded events. The measurement and modeling techniques developed in this paper are potentially useful for other applications as well [33]–[36].

References

[1] (Jun. 2012). *Actix Press Release*. [Online]. Available: http://www.actix.com/sites/www.actix.com/files/Actix_Hotspots_Study_Findings.pdf.

[2] S. Lawson, *Wireless Networks Are Near Capacity*, accessed on Jul. 2011. [Online]. Available: http://www.pcworld.com/businesscenter/article/235964/survey_wireless_networks_are_near_capacity.html.

[3] "Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015," Cisco, San Jose, CA, USA, White Paper, Feb. 2011.

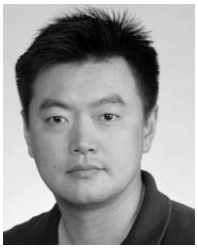
[4] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: A unified cellular and ad-hoc network architecture," in *Proc. ACM MobiCom*, 2003, pp. 353–367.

[5] B. Han *et al.*, "Cellular traffic offloading through opportunistic communications: A case study," in *Proc. 5th ACM Workshop Challenged Netw.*, 2010, pp. 31–38.

[6] *Architecture Enhancements for Non-3GPP Accesses*, accessed on Sep. 2012. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23402.htm.

[7] V. Aggarwal, R. Jana, K. K. Ramakrishnan, J. Pang, and N. K. Shankaranarayanan, "Characterizing fairness for 3G wireless networks," in *Proc. 18th IEEE Workshop LANMAN*, Oct. 2011, pp. 1–6.
[8] F. Qian *et al.*, "Characterizing radio resource allocation for 3G networks," in *Proc. 10th ACM IMC*, 2010, pp. 137–150.
[9] H. Jiang and C. Dovrolis, "Passive estimation of TCP round-trip times," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 3, pp. 75–88, 2002.
[10] (Jun. 2013). *Ericsson Mobility Report*. [Online]. Available: http://www.ericsson.com/mobility-report.
[11] J. Huang *et al.*, "Anatomizing application performance differences on smartphones," in *Proc. ACM MobiSys*, 2010, pp. 165–178.
[12] J. Sommers and P. Barford, "Cell vs. WiFi: On the performance of metro area mobile connections," in *Proc. ACM IMC*, 2012, pp. 301–314.
[13] P. P. C. Lee, T. Bu, and T. Woo, "On the detection of signaling DoS attacks on 3G wireless networks," in *Proc. IEEE INFOCOM*, May 2007, pp. 1289–1297.
[14] R. S. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis With R*. New York, NY, USA: Springer-Verlag, 2008.
[15] A. Schulman *et al.*, "Bartendr: A practical approach to energy-aware cellular data scheduling," in *Proc. ACM MobiCom*, 2010, pp. 85–96.
[16] B. Orlandi and F. Scahill, "Wi-Fi roaming—Building on ANDSF and Hotspot2.0," Alcatel-Lucent, Boulogne-Billancourt, France, Tech. Rep., 2012.
[17] *Offload Service*, accessed on Apr. 2016. [Online]. Available: http://www.devicescape.com/offload-service.
[18] "Wi-Fi CERTIFIED Voice-Enterprise: Delivering Wi-Fi voice to the enterprise," Wi-Fi Alliance, White Paper, May 2012.
[19] M. P. Wittie, B. Stone-Gross, K. C. Almeroth, and E. M. Belding, "MIST: Cellular data network measurement for mobile applications," in *Proc. IEEE BROADNETS*, Sep. 2007, pp. 743–751.
[20] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 882–890.
[21] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling Internet traffic dynamics of cellular devices," in *Proc. ACM SIGMETRICS*, 2011, pp. 305–316.
[22] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "A first look at cellular machine-to-machine traffic: Large scale measurement and characterization," in *Proc. ACM SIGMETRICS/PERFORMANCE*, 2012, pp. 65–76.
[23] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3G cellular data network," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1341–1349.
[24] Q. Xu *et al.*, "Identifying diverse usage behaviors of smartphone apps," in *Proc. ACM IMC*, 2011, pp. 329–344.
[25] A. Botta, A. Pescape, G. Ventre, E. Biersack, and S. Rugel, "Performance footprints of heavy-users in 3G networks via empirical measurement," in *Proc. 8th Int. Symp. Modeling Optim. Mobile, Ad Hoc Wireless Netw. (WiOpt)*, May/Jun. 2010, pp. 330–335.
[26] J. Erman and K. K. Ramakrishnan, "Understanding the super-sized traffic of the super bowl," in *Proc. ACM IMC*, 2013, pp. 353–360.
[27] X. Liu, A. Sridharan, S. Machiraju, M. Seshadri, and H. Zang, "Experiences in a 3G network: Interplay between the wireless channel and applications," in *Proc. ACM MobiCom*, 2008, pp. 211–222.
[28] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: A measurement study and implications for network applications," in *Proc. ACM IMC*, 2009, pp. 280–293.
[29] F. Liers and A. Mitschele-Thiel, "UMTS data capacity improvements employing dynamic RRC timeouts," in *Proc. IEEE 16th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2005, pp. 2186–2190.
[30] J.-H. Yeh, J.-C. Chen, and C.-C. Lee, "Comparative analysis of energy-saving techniques in 3GPP and 3GPP2 systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 432–438, Jan. 2009.
[31] P. K. Athivarapu *et al.*, "RadioJockey: Mining program execution to optimize cellular radio usage," in *Proc. ACM MobiCom*, 2012, pp. 101–112.
[32] F. Qian *et al.*, "TOP: Tail optimization protocol for cellular radio resource allocation," in *Proc. 18th IEEE ICNP*, Oct. 2010, pp. 285–294.
[33] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.
[34] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Inf. Sci.*, vol. 295, no. 1, pp. 395–406, 2015.
[35] J. Qiao *et al.*, "Enabling device-to-device communications in millimeter-wave 5G cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 209–215, Jan. 2015.
[36] W. Du, W. Jia, G. Wang, and W. Lu, "Analysis of channel allocation scheme for wireless cellular networks," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 4, nos. 3–4, pp. 201–209, 2009.

M. Zubair Shafiq received the Ph.D. degree from Michigan State University in 2014. He is currently an Assistant Professor of Computer Science with The University of Iowa. His research interests include networking, security, and privacy. He received the 2012 IEEE ICNP Best Paper Award and the 2013 Fitch-Beach Outstanding Graduate Research Award at Michigan State University.



Lusheng Ji received the Ph.D. degree in computer science from the University of Maryland, College Park, in 2001. He is currently a Principal Technical Staff Member with AT&T Laboratories–Research, Bedminster, NJ, USA. His research interests include wireless networking, mobile computing, wireless sensor networks, and networking security. He is a Senior Member of the IEEE.

Alex X. Liu received the Ph.D. degree in computer science from The University of Texas at Austin in 2006. His research interests focus on networking and security. He received the IEEE and IFIP William C. Carter Award in 2004, the National Science Foundation CAREER Award in 2009, and the Michigan State University Withrow Distinguished Scholar Award in 2011. He also received best paper awards from ICNP-2012, SRDS-2012, and LISA-2010. He is an Associate Editor of the IEEE/ACM TRANSACTIONS ON NETWORKING, an Editor of the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, and an Area Editor of *Computer Communications*.

Jeffrey Pang received the Ph.D. degree in computer science from Carnegie Mellon University in 2009. He is currently a Researcher with AT&T Laboratories–Research. He builds systems to measure and optimize cellular networks. His research interests include networking, mobile systems, distributed systems, and privacy.

Shobha Venkataraman received the Ph.D. degree in computer science from Carnegie Mellon University in 2008. She is currently a Researcher with AT&T Laboratories–Research. Her main research interests span computer security, networking, and systems problems using techniques from machine learning, data mining, and algorithm design.

Jia Wang received the Ph.D. degree in computer science from Cornell University in 2001. She has been with AT&T Laboratories–Research since 2001, where she is currently a Principal Technical Staff Member with the Network Measurement and Engineering Research Department. Her research interests focus on network measurement and management, network security, performance analysis and troubleshooting, IPTV, social networks, and cellular networks.