Mobility Support in Cellular Networks: A Measurement Study on Its Configurations and Implications

Haotian Deng, Chunyi Peng, Ans Fida, Jiayi Meng, Charlie Hu
Computer Science, Purdue University

ABSTRACT

In this paper, we conduct the first global-scale measurement study on how 30 mobile operators worldwide manage mobility support in their carrier networks. Using a novel, device-centric tool MMLab, we are able to crawl runtime configurations without the assistance from operators. Using handoff configurations from 32,000+ cells and $> 18,700$ handoff instances, we uncover how policy-based handoffs work in practice. We further study how the configuration parameters affect the handoff performance and user data access.

1 Introduction

The Internet is going mobile. To date, billions of mobile users are accessing the Internet as they move. For their Internet access, the cellular network plays a pivotal role, since it is the only global-scale infrastructure with ubiquitous mobility support. Central to its mobility management is the handoff scheme [1]. In a nutshell, handoff switches the mobile device's serving cell tower (called cell hereafter) from one to another as the user roams. It is a mechanism that may greatly affect user experience. For instance, data services are disrupted, if an expected handoff happens too late (e.g., no handoff yet while the old cell's radio quality is too weak); Data throughput drops when a handoff makes a bad choice (e.g., 2G cell instead of high-speed 4G cell).

Despite its importance, there is little study on real-world handoff practices. Current efforts focus on how handoffs affect TCP and applications [2–4] and how to improve their performance [5–9]. Instead, we study *how* and *why* handoffs are performed over operational networks, as well as their implications on data access. This is the focus of this work.

We have identified three challenges. First, practical handoffs take the *policy-based* approach. Each handoff takes into account many factors, including cell priorities, radio link quality, list of events of interest, eligible candidate cells, etc.. It runs multiple asynchronous procedures (say, measurement, reporting, decision and execution, Figure 1). Each procedure has its own configuration parameters, while following the common mechanism (logic) standardized in the specifications (3GPP [1, 10–14]). For example, certain measurement is triggered when the serving cell's radio signal strength r_S is smaller than certain threshold, $r_S \leq \Theta_{r_S}$. Through configuring distinct values for Θ_{r_S}, operators manage handoffs at each cell and different locations. Therefore, tunable configurations play an important role on policy-based handoffs. Second, no large-scale handoff traces are publicly available. Given the policy-based practice, operators are reluctant to reveal their data sets due to privacy concerns. Moreover, it is nontrivial for the operators to collect and archive handoff operations, given that handoffs are executed at each cell for each mobile device in the distributed manner over geo-distributed areas. Third, handoff configurations take many parameters and are distributed at all cells for a mobile carrier network. The standard specifications [1, 12–15] describe 66 parameters for a single 4G cell and 91 parameters for 3G/2G RATs (see Table 4 for an illustration).

To address the above challenges, we take a device-centric, rather than infrastructure-centric, approach to measurement study on handoffs. We thus design MMLab, a software tool that runs at the mobile device without operators' assistance. In a nutshell, MMLab takes the device-centric approach to crawling handoff configurations from operational networks. It leverages the fact that handoff configurations are broadcast by the serving cell and reach each nearby mobile device. It consequently extracts all configuration parameters from the signaling messages received at the mobile device, thus enabling real-world handoff configuration collection via smartphones only.

Using MMLab, we and our volunteers have collected handoff configuration traces and handoff instances[1] from global mobile carriers across three continents. Our data set D1 contains $> 18,700$ handoff instances, while our data set D2 covers handoff configurations from 32,000+ cells over 30 carriers in North America, Europe and Asia.

Based on the traces, we conduct an in-depth study on cellular mobility support. We examine why and how a handoff is triggered at a cell (*reason and procedure*), rather than which cell is eventually chosen (*consequence*). We look into *persistent* and *structural* factors that determine a handoff procedure, instead of *transient* factors like time-varying radio channel quality and network states. We provide answers to the following two questions: (Q1) What are these config-

[1]Our codes and datasets will be released to the public.

	Summary	Details	Ref.
Q1	Real-world handoff configurations are extremely complex and diverse and allow micro-level mobility management in the wild.	1a. Handoffs are policy-based and their configuration space is high-dimensional.	Tab. 2
		1b. Parameter values exhibit rich diversity (distinct distribution and dispersion). They are affected by many factors like carriers, frequency, location etc..	Fig. 5,10, 13 –22
		1c. Most configurations seem reasonable but not all are well justified.	Fig. 18
Q2	Configurations affect radio and performance as expected (by design), but not all the impacts are intuitively 'positive'.	2a. The serving cell's radio quality in handoffs changes as configured.	Fig.6, 9, 10
		2b. Data performance impacts match with expected consequences by reasoning.	Fig. 8, 7
		2c. Improving radio quality is not the sole key to better data performance. In most cases, timing of handoffs are more crucial.	Fig. 8, 9
		2d. Current configurations for both active-state and idle-state handoffs are 'questionable' in terms of performance and efficiency.	Fig. 8, 11

Table 1: Summary of our main findings.

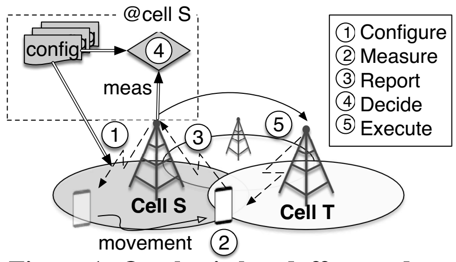


Figure 1: One basic handoff procedure.

urations in the wild? (Q2) What impacts do these policy-based configurations have on handoff performance? How do they affect data access for mobile users?

Specifically, we first use dataset `D1` to characterize small-scale handoff configurations (Q1) and empirically assess their performance impacts on handoff configurations (Q2). We further use dataset `D2` to conduct a large-scale study and show how they are configured and explore why. Table 1 summarizes our main findings to be elaborated later. Our measurement study yields many interesting results which have not been anticipated or reported. Specifically, handoff configurations may not select the cell with best radio signals, and they are extremely diverse. Operators use a few popular choices to decide their policy practice. While some configurations sound rational, a few others seem to be problematic. They may degrade handoff quality or cause problematic handoffs; In many real cases, we observe that timing, rather than better radio signal quality, is more critical to mobility performance. We discuss their implications for operators and mobile users, and identify new research opportunities inspired by our study.

2 Policy-Based Handoff Primer

Cellular networks deploy many overlapping cells across geographic areas. At a given location, a mobile device is served by one cell but covered by multiple cells in proximity. The cells may use distinct radio access technologies (RATs) of 2G, 3G and 4G. Each cell further operates over a given frequency channel (see [16] for a complete list of channels).

A handoff switches the serving cell from one to another. Depending on the used frequencies, handoff can happen over: *intra-freq* (on the same RAT and frequency channel), *inter-freq* (on the same RAT but different channels), and *inter-RAT* (on different RATs).

Handoff is generally classified into two categories: idle-state handoff and active-state handoff, depending on whether the device is at the idle/active state without/with user traffic. The idle-state handoff is performed by the device. It selects an appropriate cell for *future* access. The active-state handoff is initiated by the network. The serving cell migrates the device to another target cell to retain radio access.

2.1 Handoff Procedures

Figure 1 depicts a basic handoff procedure. It typically consists of five/four steps: *configuration*, *measurement*, *reporting* (only for active-state handoff), *decision* and *execution*. Initially, the device is served by cell S; It receives handoff configuration parameters broadcast by S (①) and learns the criteria to trigger, decide and perform a handoff, including whether to invoke measurement, what and when to measure, whether/when/what to report, how to determine the next target cell, to name a few. The subsequent steps (②-④) will be invoked when the criteria configured by S's handoff parameters are satisfied at runtime.

Active-state and idle-state handoffs differ at step ③. In an active-state handoff, the device reports its measurement results (obtained at ②) to S when the reporting criterion is met (e.g., one candidate cell's radio strength is offset stronger than S's). S then decides whether to switch to a new cell and which cell to go (④). In an idle-state handoff, step ③ is skipped. The device makes a decision locally using the decision criteria pre-configured by the serving cell. Eventually, the cell switches from S to T under network-device cooperation (⑤). Once this round completes, the device is served by T and is ready to repeat the above procedure.

2.2 Policy-Based Configurations

The cellular network uses policy-based handoff. Each handoff takes into account many factors, including cell priorities, radio link quality, list of events of interest, eligible candidate cells, etc.. According to the standard specifications [1, 12–15], our measurement study covers 66 parameters for a single 4G LTE cell and 91 parameters for four 3G/2G RATs (Tab. 4). Due to space limit, we use only 4G LTE to illustrate main parameters (Tab. 2) and their use.

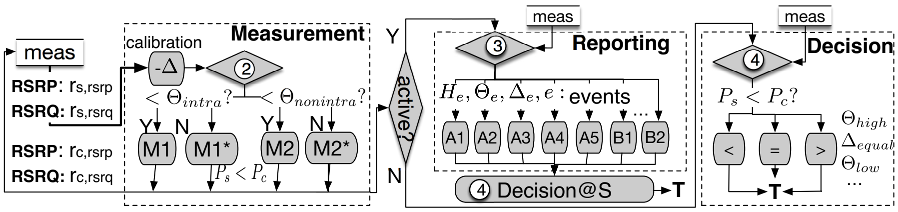


Figure 2: Flow of handoff steps in a decision tree.

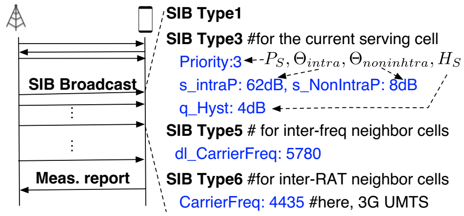


Figure 3: An example trace via MMLab.

Category	Parameter	Remark	Used for	Message
Cell	P_S	Priority of the serving cell, ranging from 0-7 with 7 being the most preferred	measurement, decision	SIB 3
priority	P_C	Priority of candidate cells in neighborhood, associated with its frequency channel, i.e., P_{freq}	measurement, decision	SIB 5,6,7,8
	Θ_{intra}	Threshold of radio strength level for intra-freq measurement ($\Theta_{intra,rsrp}$, $\Theta_{intra,rsrq}$)	measurement	SIB 3
	$Q_{nonintra}$	Threshold of radio strength level for non intra-freq measurement ($\Theta_{nonintra,rsrp}$, $\Theta_{nonintra,rsrq}$)	measurement	SIB 3
	Δ_{min}	minimum required signaling strength for handoff ($\Delta_{min,rsrp}$, $\Delta_{min,rsrq}$)	calibration	SIB 1,3,5,6,7,8
Radio	H_e, Θ_e, Δ_e	Hysteresis, threshold(s) and offset(s) used for the reporting event e (A1–A5, B1–B2)	reporting	event A1-A5,B1-B2
evaluation	H_s	Hysteresis value added to the serving cell's radio strength	decision	SIB 3
	$\Theta^{(c)}_{higher}$	Threshold of radio evaluation for a higher-priority candidate cell	decision	SIB 5,6,7,8
	$\Theta^{(c)}_{lower}$, $\Theta^{(s)}_{lower}$	Thresholds for a lower-priority candidate cell and the higher-priority serving one	decision	SIB 3,5,6,7,8
	Δ_{equal}	Offset of radio comparison for equal-priority cells, $\Delta_{s,n}$, Δ_{freq}, Δ_{cell}	decision	measurement object
	$T_{reselect}$	Time required to fulfil switching condition	measurement	SIB3,5,7
Timer	$T_{reportTrigger}$	Time to trigger when measurement report triggering criterion is always fulfilled	reporting	event A1-A5,B1-B2
	$T_{reportInterval}$	Interval for sending measurement report	reporting	event A1-A5,B1-B2
	$T_{decision}$	Time to trigger when the radio evaluation criterion is always fulfilled	decision	event A1-A5,B1-B2
	$Freq_{interest}$	List of the frequency channels of interests	measurement	SIB 5,6,7,8
Misc	$List_{forbid}$	List of forbidden candidate cells (due to access control)	measurement	SIB4
	$meas_{bandwidth}$	maximum bandwidth allowed for performing measurement	measurement	SIB5

Table 2: Main configuration parameters standardized for handoff at 4G LTE cells.

Figure 2 depicts the handoff flow in a decision tree[2].

• **Measurement.** It is unnecessary to measure all candidate cells at all times. LTE runs two types of measurements: (M1) intra-freq and (M2) non intra-freq (aka, inter-freq and inter-RAT) [12]. Both support event-based and periodic modes. If

$$r_S \le \Theta_{intra}(\Theta_{nonintra}),\ (\text{actually, } r_S - \Delta_{min} \le \Theta) \quad (1)$$

intra-freq (non intra-freq) measurement is triggered, otherwise the measurement only for those higher priority cells ($\{C|P_C > P_S\}$) is performed periodically (every $T_{higherMeas}$).

It measures the received radio signal quality using multiple metrics like RSRP (reference signal received power) and RSRQ (reference signal received power quality) for 4G LTE. They take different values: RSRP (-140dBm, -44dBm), RSRQ (-19.5dB, -3dB), and thus use distinct configuration parameters. Without loss of generality, we use RSRP hereafter unless specified. Calibration is used to compensate different transmission power and ensure fair radio comparison. It converts the actual measurement $\dot{r}_S$ into a level of radio quality $r_S = \dot{r}_S(\text{actual}) - \Delta_{min}$, where Δ_{min} is another preconfigured parameter.

• **Reporting.** It uses a set of reporting events to determine *whether*, *what* and *when* a user device should report its measurement results to the serving cell to assist an active-state handoff. LTE defines ten events (A1-A6, B1,B2, C1,C2) [12], but our measurement study shows that not all the events are used (A1-A5, B1 and B2 observed). Each event targets one specific condition and has its own configuration set (thresholds Θ_e, hypothesis H_e and offsets Δ_e).

A1 and A2 indicate that the serving cell's radio strength r_S is better or worse than a threshold; A3 and A4 indicate that one neighboring (candidate) cell is better than the serving one plus an offset (A3) or a threshold (A4); A5 indicates that the neighboring one is larger than certain threshold while the serving one weaker than another threshold; Event B1 and B2 indicate the existence of a decent inter-RAT neighboring cell (B1: larger than one threshold, B2: larger than another threshold and the serving one weaker than certain threshold).

We use A3 to illustrate the event form:

$$\begin{cases} \textit{A3. Reporting condition:} & r_c > r_s + \Delta_{A3} + H_{A3} \\ \textit{A3. Stopping condition:} & r_c < r_s + \Delta_{A3} - H_{A3} \end{cases} \quad (2)$$

Δ_{A3} is a positive offset and indicates a stronger candidate cell. H_{A3} is a hysteresis to adapt the start and stop conditions. It is expected to be positive.

• **Decision.** We consider the idle-state handoff decision. It is made by comparing radio strengths of the serving and candidate cells, given their cell priorities. It considers three cases: higher-priority, equal-priority and lower-priority. The ranking of a candidate cell is higher ($rank_c > rank_s$) when one of the following criteria is satisfied,

$$\begin{cases} (1)\text{ if } P_c > P_s, & r_c > \Theta^{(c)}_{higher} \\ (2)\text{ if } P_c = P_s, & r_c > r_s + \Delta_{equal} \\ (3)\text{ if } P_c < P_s, & r_c > \Theta^{(c)}_{lower}, r_s < \Theta^{(s)}_{lower} \end{cases} \quad (3)$$

The decision is made until the above requirements have been fulfilled for $T_{decision}$ to avoid frequent handoff caused by measurement dynamics. Δ_{equal} (>0 expected) implies the favor towards the serving cell. Other rules counts on the threshold settings to customize the criteria at a higher or lower priority. The active-state handoff may use similar rules,

[2]Execution also uses some parameters such as timers and maximum retry count. It is less critical and is omitted in this study.

along with proprietary non-radio criteria [17]. So for an active-state handoff, radio evaluation is treated as necessary but not sufficient condition as [17] does. Our measurement study shows that an active-state handoff is determined by the last reporting event (see §4).

3 Measurement Methodology

To conduct handoff configurations study at scale, we have designed a new software tool `MMLab` and conduct measurements based on it.

3.1 `MMLab` Tool Design

We have thus designed `MMLab`, a software tool that runs on Android phones to facilitate our handoff study. The structure of `MMLab` is illustrated in Figure 4.

`MMLab` is a device-centric measurement tool. which crawls configuration data and runs performance assessment on smartphones without assistance from operators. This overcomes the long-lasting research barrier of closed cellular networks, where operators do not publicize their infrastructure-side data traces.

`MMLab` is built atop `MobileInsight` [18], an open-source tool to collect cellular network signaling messages on rooted Android smartphones. It extracts all handoff configurations from signaling messages exchanged between the device and the cell. This is feasible by leveraging the fact that handoff configuration parameters are broadcast by the serving cell and received by local phone devices. Note that, `MobileInsight` in its current release does not support all needed messages. `MMLab` thus has to parse new, handoff-specific messages. It further customizes message collection tailored to 4G/3G/2G handoff configurations only (see Table 2 for main parameters and messages over 4G LTE).

Figure 3 shows an example log observed in AT&T. The last message `measurement report` indicates an active-state handoff, and its tunable handoff parameters are broadcast in System Information Blocks (SIBs) (in blue).

To make our measurements further scale out to more carriers, we develop a crowdsourced measurement infrastructure managed by `MMLab` servers. To this end, `MMLab` relies on global, participating volunteers from different countries and regions to collect configuration traces from operators worldwide. We thus distribute a mobile app version of `MMLab` to volunteers to make this participation as easy as possible. Specifically, it performs three tasks: handoff configuration collection (i.e., Type-I measurement), configuration characterization and analysis (Q1), and performance assessment/-validation experimentation (Type-II measurements, Q2). Upon new app updates, the app communicates with the `MMLab` server and loads the experimentation patches on the fly without new installation and compiling.

To make our data collection more efficient, we enable proactive cell switching for the serving cell. `MMLab` changes its preferred network type (e.g., LTE only, UMTS/CDMA only, and GSM) and even its frequency band to automate the switching of the serving cell. `MMLab` is thus able to collect handoff configurations from multiple cells at a given location. Note that, this operation intervenes the actual, default handoff procedure, and is disabled in Type-II measurements. Type-I measurement (data collection) consumes little resource (at least no extra data), and is open to participating volunteers worldwide. Type-II experiments may access data services, thus incurring considerable data usage (like `Speedtest`). They are constrained to those controlled devices (from ourselves and our partners with their explicit consents).

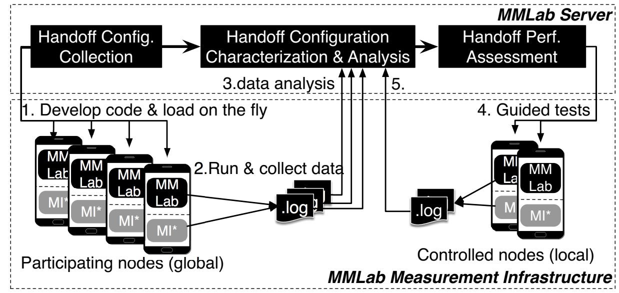


Figure 4: `MMLab` Overview.

3.2 Measurements

Using `MMLab`, we conduct two types of measurements: configuration collection only (Type-I, Q1), and configuration plus performance assessment (Type-II, Q2).

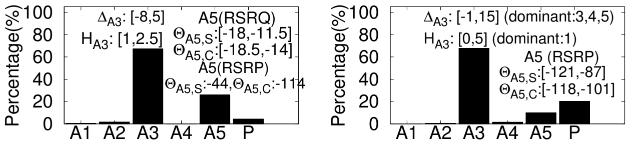
To handle configuration complexity in analysis and performance assessment, we let Type-I and Type-II help each other. Specifically, we run Type-II experiments at a small scale and gain insights on important configurations. We use these hints to guide through the large-scale configuration characterization and analysis. We also exploit results and findings in the configuration study to run Type-II experiments. For example, we run experiments around certain cells or routes with configurations of interest, to assess their impacts. Additional settings and processing information will be presented per experiment set.

4 First Look at Handoff Configurations and Performance Implications

We describe our Type-II measurement results to look into real-world handoff configurations (Q1) and quantify their impacts on user performance and handoff quality (Q2).

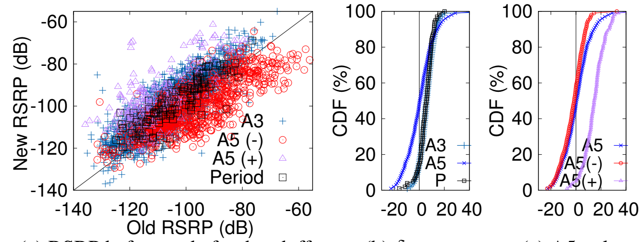
Experimental settings and dataset `D1`. We assess active-state and idle-state handoffs in three US cities (Chicago, IL; Indianapolis, IN; Lafayette, IN) and highways in between.

For active-state handoffs, we run designated data services while driving locally (<50 km/h) and on highways (90–120 km/h) Every run tests one of three data services: continuous speedtest via [19], constant-rate iPerf (5kbps and 1Mbps), and ping (Google) every five seconds. We run four-week



(a) Percentage of events (AT&T) (b) Percentage of events (T-Mobile)

Figure 5: Reporting event configurations observed in active-state handoff instances (dataset D2).



(a) RSRP before and after handoffs (b) δ_{RSRP} (c) A5 only

Figure 6: RSRP changes in active handoffs (AT&T).

experiments intermediately in the period of Jan-April, 2018, over 16 rooted Android phones (three models: Pixel 2/XL and Nexus 6P) in all four top US carriers: AT&T, T-Mobile, Verizon and Sprint. But the speed-test and constant-rate iPerf are primarily in AT&T and T-Mobile only. We use `tcpdump` to log data packets and `MI*` to collect cellular signaling messages that convey handoff configuration parameters. We study $4G \rightarrow 4G$ active-state handoffs only and collect 14,510 instances (around 8,000 Km in total).

For idle-state handoffs, we run the driving tests in the same area without running any data services. Moreover, we turn off background data as much as we can. We observe that some handoffs are still active-state due to the background traffic. We ignore them and consider $4G \rightarrow 4G$ idle-state handoffs only and have collected 4,263 instances.

4.1 Active-state handoff

We study three issues: (1) What are decisive configurations in active-state handoffs and how they look like? (2) How do they affect radio quality before and after a handoff? (3) How do they affect data performance on the go?

• **Configurations in reporting.** We observe that handoffs are triggered by different reporting events with distinct configuration values. We observe that all the handoffs (99.6%) have multiple reporting events (e.g., one or multiple A2/A5/P events) and end with one of the following events: A3, A5 and periodic reporting of neighboring cells' radio quality (P, configured by the carrier). We gauge that the last event is decisive because all the handoffs happen immediately (within 80-230 ms) once the last measurement report is sent to the serving cell. It is not hard to understand that other events are not enough to invoke a handoff. Specifically, event A2 (the current serving cell is weaker than one threshold) should not trigger a handoff unless there is a strong candidate cell (A3, A5, P); Periodic reporting and other events can be triggered when the reporting criteria is satisfied but the reported radio quality is not sufficient to make a handoff decision until the last one comes.

Fig. 5 plots the distribution of the decisive events, along with the range of their main parameter values in AT&T and T-Mobile. We clearly see uneven usage of distinct events and carrier-specific configurations. This conclusion is also applicable to other carriers (see §5.3). Operators may use different radio metrics. AT&T uses RSRP for A3 and RSRP and RSRQ almost equally for A5, whereas T-Mobile uses RSRP in most cases. Although RSRP and RSRQ are conceptually interchangeable, there is no 1:1 mapping between them. As such, we observe more A5 options in AT&T. Specifically, we observe that AT&T primarily uses A3 (67.4%) and A5 (26.1%) while P (4.4%) and A2 (1.7%) are occasionally observed. In T-Mobile, most active-state handoffs are caused by A3 (67.7%), P(20.2%) and A5 (10.0%). In both carriers, A1 and A4 are rarely observed in both carriers ($< 0.5\%$) and other events like A6, B1, B2, C1 and C2, are never observed. Moreover, parameter values are quite different; for example, Δ_{A3}, the offset value in event A3, ranges in [-1dB, -15dB] in T-Mobile, but [0dB, 5dB] in AT&T (dominated by 3dB). Parameters for A5 are more dispersed. We present detailed results in our larger-scale study (§5, e.g., Fig. 14 for AT&T).

Implications: Two policies dominate the practice by mobile operators. A3 is the most popular one for its simplificity. It uses a *relative* comparison to mandate that the new cell is better (usually, $\Delta_{A3} > 0$) than the serving one.

A5 is the most flexible one. It can make the same comparison as A3 (e.g., offset = the gap in two thresholds); Note that, it differs from A3 because A5 has additional requirement on the absolute radio quality (the serving one < $\Theta_{A5,S}$ and the candidate one > $\Theta_{A5,C}$). A5 can substitute other events with particular parameter settings, such as A2 (when $\Theta_{A5,C}$ = -140dB, worst RSRP) and A4 (when $\Theta_{A5,S}$ = -44dB, best RSRP). In fact, we do observe such configurations in A5. This explains why other events are rarely observed.

• **Handoffs for better radio?** We find that, *not all handoffs go to a cell with stronger radio signals, and this choice depends on handoff configurations.*

Fig. 6a shows RSRPs before and after handoffs under three decisive reporting events in AT&T (similar for other carriers). For comparisons, we also plot the cumulative distribution functions (CDFs) for the RSRP changes (δ_{RSRP} = $RSRP_{new} - RSRP_{old}$) in Figure 6b.

We see that, for A5, only 52% of handoffs get better in terms of RSRP (62% for RSRQ). In contrast, A3 and periodic reporting largely ensure a better radio quality: 87% of handoffs have δ>0 and the ratio goes up to 94% given that 3dB measurement dynamics is common.

This is because A5 reports two *independent* conditions: the serving cell is weaker than one threshold ($\Theta_{A5,S}$) and the candidate cell is stronger than another one ($\Theta_{A5,C}$). Given two parameter configurations, it is not ensured that the new

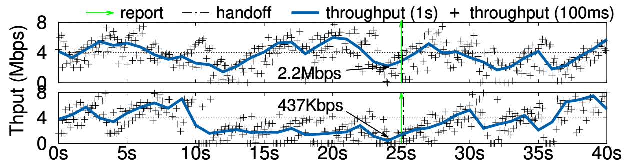
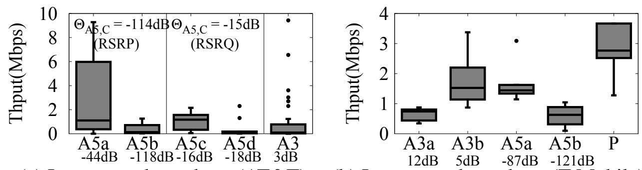
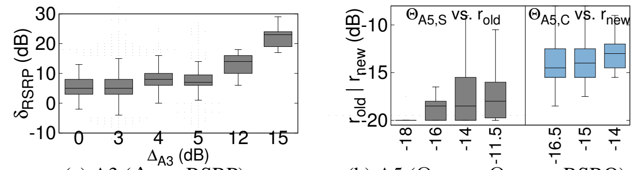


Figure 7: Throughput of two handoff examples using distinct event A3 offsets Δ_{A3}: 5dB (top) and 12dB (bottom).



(a) Impact on throughput (AT&T) (b) Impact on throughput (T-Mobile)

Figure 8: Impacts of reporting event configurations.



(a) A3 (Δ_{A3}, RSRP) (b) A5 ($\Theta_{A5,S}$, $\Theta_{A5,C}$, RSRQ)

Figure 9: Radio impacts of configurations in A3 and A5.

cell is stronger.

We observe that $\Theta_{A5,S,rsrq} > \Theta_{A5,C,rsrq}$ (e.g., -11.5dB vs. -14dB). In RSRP cases, the dominant setting is $\Theta_{A5,S,rsrp}$ = -44dB (no requirement) and $\Theta_{A5,C,rsrp} = -114$dB. That is, such A5 events do not take into account the serving cell's radio strength using RSRP in AT&T. They are responsible for the cells after handoffs with weaker radio coverage. While this finding differs from our expectations, it well matches the consequences of such configurations. We further divide δ_{RSRP} in A5 into positive (+) ($\Theta_{A5,C,rsrq} > \Theta_{A5,S,rsrq}$) and negative (-) cases; Fig.6c confirms that weaker radio is caused by negative configurations.

• Expected data performance impacts and "questionable" configurations. We show that, data performance during handoffs are also affected by such configurations.

We first present two handoff examples, which are both triggered by A3 but with different offset values: Δ_{A3} = 5dB (top) and 12dB (bottom). We align both routes with the `Measurement Report` message (t = 25s) and handoffs are performed right away after the reporting (within 180 ms). Fig. 7 shows the average throughput in two time bins (1s and 100ms) while we run continuous speedtest in T-Mobile. We see that data throughput decreases down to 2.2 Mbps (top) and 437Kbps (bottom) before handoffs. Performance is much worse in the bottom case because Δ_{A3} is 12dB, much higher than 5dB (top), which invokes the handoff very late after data throughput has already severely fell down. The handoff occurs only when one candidate cell must be much stronger than the serving one. The minimum throughput before handoffs declines by 80.1% (5× gap).

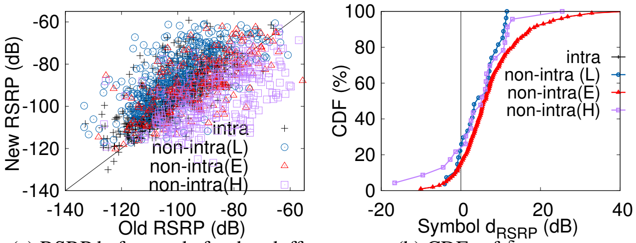
We use the minimum throughput before handoffs (reporting) to assess performance impacts of reporting configurations. Fig. 8 compares performance under representative configurations in AT&T and T-Mobile. It shows that data performance impacts match with the anticipated consequences of such configurations. In T-Mobile, A3a (12 dB) and A5b (-121dB) tend to defer or prevent handoffs to new cells, compared to A3b (5 dB) and A5a (-87dB). A5 considers the serving cell's threshold $\Theta_{A5,S}$ (RSRP: -87dB and -121dB) only. Consequently, they result in lower throughput and worse handoff quality. This is consistent with observations in AT&T. A5a ($\Theta_{S,RSRP}$: -44dB) outperforms A5b (-118dB) given the same $\Theta_{C,RSRP}$ (-114dB). It is similar in the A5c/A5d cases which use RSRQ; But the gap is much smaller as two thresholds are quite close. Such performance impacts can be somehow derived from their impacts on radio quality. Fig. 9 shows the box-plots of three pairwise relations: Δ_{A3} versus δ_{RSRP}, $\Theta_{A5,S}$ versus r_{old} and $\Theta_{A5,C}$ versus r_{new}. We choose these three pairs by the purposes of those parameters. We can see that handoffs are performed as configured.

Implications: We discover that, radio quality enhancement may not be the key to better handoff performance. "Better" configurations should invoke the handoffs in time, well before the performance degrades or is about to degrade.

Recall, when $\Theta_{A5,S}$= -44dB (RSRP), A5 performs best but stronger radio quality is not guaranteed. This choice relaxes the requirement on the serving cell and creates a larger chance to obtain measurement reports earlier, thus making early handoff possible. Compared with other strict configuration like $\Theta_{A5,S}$= -118dB (RSRP), it indeed ensures a handoff only when the current one is really poor (depending on the value of $\Theta_{A5,S}$) and avoids some handoffs (e.g., where the serving cell is stronger than -118dB with a neighboring one is even better (e.g., >-100dB)).

This illustrates two different policies for handoff management. The former is more performance driven while the latter also takes into account handoff overhead and seeks to reduces handoff frequency. It is hard to argue which one is better. As the cellular network infrastructure has been evolving with long-lasting deployment and upgrades (radio coverage likely enhanced and overhead for frequent handoffs not like a big concern), it may be the time for us to update handoff policies and their configurations.

We also note that, it is hard to compare data performance under different configurations (e.g., A3 and A5 in AT&T). A larger variance is observed in A3, because A3 only regulates the relative enhancement but the actual serving one may have large variations. This raises a question on which configuration contributes to better handoff performance. If the answer varies at cells (e.g., depending on nearby radio coverage), we seek to learn whether any mechanism or algorithm handles it at runtime (e.g., reconfiguration). Unfortunately, our following study (§5) seems to reach a negative conclusion.



(a) RSRP before and after handoffs (b) CDFs of δ_{RSRP}

Figure 10: RSRP changes in idle-state handoffs.

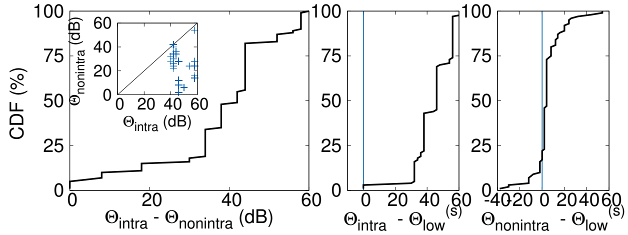


Figure 11: CDFs of representative radio thresholds used for measurement and idle-state handoff decision.

4.2 Idle-state Handoff

There is no user traffic during idle-state handoffs. We consider two issues: (1) What do these configurations look like and what are their impacts on radio quality? (2) Do measurement and decision perform efficiently?

Expected radio "enhancement" after handoffs. Fig. 11 examines the RSRP change before and after an idle-state handoff in four US carriers. The results are consistent across different carriers. We consider intra-freq and non-intra-freq cases. In non-intra freq cases, we further consider subcases of various cell priorities We observe that 4G cells use multiple priorities from 2 to 7, elaborated in §5. Unsurprisingly, handoffs are well controlled by these configurations. Almost all the handoffs (except higher-priority non-intra freq handoffs) go to stronger cells. We observe that most configuration follow the common expectations: $\Delta_{equal} > 0$ determines that it will choose a stronger cell when both have equal priorities; $\Theta^{(c)}_{lower} > \Theta^{(s)}_{lower}$ implies that the chosen cell is better the previously serving one, when the new cell has a lower priority. Only in the higher-priority cases, handoffs occur as long as the candidate cell is larger than an *absolute* value $\Theta^{(c)}_{higher}$, regardless of the serving one. It is possible that it switches to a weaker cell (20% observed).

Implications: handoffs use priority + radio evaluation to tune their policies on radio evaluation. Higher-priority cells may be preferred for better performance even with lower radio quality (4G vs 3G/2G) or non-performance reasons (e.g., operators favor some newly deployed cells and acquired bands, see the instance observed in §5.4.1).

Measurement efficiency (necessity)? We examine whether measurements run efficiently (whether all measurements are necessary). We consider two issues.

(1) *Will intra-freq measurements be always preferred over non infra-freq measurements?* This is intuitively the case, since intra-freq measurements take less time and non intra-freq ones must measure other frequency bands with larger overhead. Therefore, Θ_{intra} should be no smaller than $\Theta_{nonintra}$ so that $r_S < \Theta_{nonintra} \geq \Theta_{intra}$ always holds true. We plot the CDF of $\Theta_{intra} - \Theta_{nonintra}$ in Fig. 11 (left), along with all value pairs observed. Clearly, it holds true ($\geq$ 0) in these tested areas. We also note that, it becomes zero ($\Theta_{intra} = \Theta_{nonintra}$) in 5% cases. This means that, both measurements use the same criterion and will be invoked at the same time. However, our larger-scale study shows that, it is not always true. In other areas, non-intra freq measurements are performed more often than intra-freq ones.

(2) *Will all measurements be closely associated with the subsequent handoff decision?* We infer the occurrence of measurements by examining whether $r_S - \Delta_{min} < \Theta_{intra}$ (or $\Theta_{nonintra}$) (see Eq. (1)). We find that **in many cases, measurements are always invoked but handoffs likely not**. This happens when Θ_{intra} (or $\Theta_{nonintra}$) is extremely large and but the radio evaluation threshold in decision is small. We use one common instance for illustration: Θ_{intra} = 62dB, $\Theta_{nonintra}$ = 28dB, Δ_{min} = -122dB, $\Theta^{(s)}_{low}$ = 6dB and Δ_{equal} = 4dB. For simplicity, we consider the serving cell has the highest priority and thus idle-state handoffs occur when r_S < -122 + 6 = -116 dB (lower-priority) or $r_C > r_S$ + 4dB (equal-priority), refer to Eq. (3). However, intra-freq measurements are triggered when r_S < -122 + 62 = -60 dB. It is almost true anywhere. It implies that measurements are performed at all time even when the device is static at one place or under good radio coverage. Intuitively, these measurements are unnecessary. Fig. 11 (middle and right) plots the gaps between the measurement and decision thresholds observed. Clearly, $\Theta_{intra} - \Theta^{(s)}_{low}$ is pretty large (>30dB in 95% cases). Such big gap implies that intra-freq measurements performed when the serving cell is strong are much less necessary because handoffs only happen when the serving cell is quite weak (e.g., a small value for $\Theta^{(s)}_{low}$). We also observe another interesting finding of $\Theta_{nonintra} < \Theta^{(s)}_{low}$. It implies that non-intra freq measurements may not run in time ($\Theta_{nonintra}$ is too small) to assist handoffs.

Implications: Diverse choices may cause inconsistent or conflicting configurations and hurt handoff quality. While each single parameter configuration is sensible, the combination of multiple ones might be problematic. There exists negative compound effect that we do not anticipate, when these configurations are not well coordinated.

5 Configurations at Larger-Scale

In this section we characterize how various operators configure handoff parameters (Q1) in a larger-scale study. Given insights gained in §4, we discuss their implications.

Dataset `D2`. We run type-I measurements in several US cities sporadically in two periods: Oct 2016 - April 2017 and Aug 2017- May 2018. We also distribute `MMLab` to

Country/Region		Carriers
USA (US)	4	**A**(T&T), **T**(-mobile), **V**(erizon), **S**(print)
China (CN)	3	**C**(hina)**M**(obile), **C**(hina)**U**(nicom), **C**(hina)**T**(elecom)
Korea (KR)	2	**K**(orea)**T**(elecom), **SK**(Telecom)
Singapore(SG)	3	**ST**(arhub), **SI**(ngTel), **MO**(bileone)
Hongkong (HK)	2	**TH**(ree), **C**(hinamobile)**H**(ongKong)
Taiwan(TW)	2	**C**(hung)**W**(haTelecom), **T**(aiwan)**C**(ellular)
Norway(NO)	1	**N**(et)**C**(om)
Others	13	e.g., Orange (France), DeutscheTelekom (Germany), VOdafone (Spain), MoviStar (Mexico), $\cdots$

Table 3: Main carriers and their acronyms (in bold).

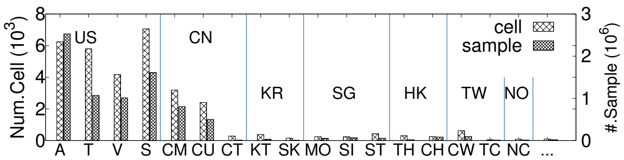


Figure 12: Number of cells and samples per carrier.

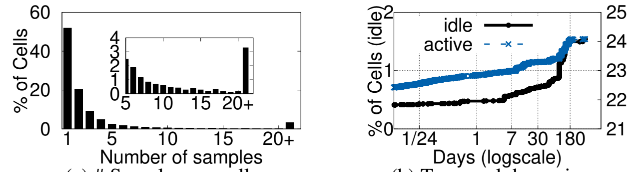
35+ volunteers (college students, professors and researchers, aged from 21 to 50) across the US and around the world to intermittently collect configuration data between Nov 2017 and April 2018. This dataset covers 7,996,149 configuration samples from 32,033 unique cells locating in 30 carriers over 15 countries and regions (main carriers shown in Tab. 3). Fig. 12 shows the breakdown per carrier. Most data is collected in USA, China, several countries/regions in Asia and a small number of cells (<100) are in France, Germany, Spain and Mexico, etc.. Note that the number of cells is relatively small in small regions like Singapore, Hongkong,Taiwan and Korea (Seoul only). We count the number of unique cells. We retreat each parameter observed as one sample. The number of samples is much larger because some cells have been collected at multiple rounds and each round observes a set of configuration parameters. Our dataset covers all five RATs and 4G LTE is dominant (Tab. 4). LTE contributes to 72% cells. This is because LTE is the latest and the most widely used technology today; 3G and 2G support two family standards. The UMTS/GSM family is more popular and EVDO/CDMA1x are only observed in Verizon, Sprint and China Telecom. We first study LTE and then investigate its differences with other RATs.

5.1 Temporal Dynamics

We first examine whether (and how) handoff configurations change over time. This is also critical to our dataset quality and data cleaning. Note that most data is collected by volunteers beyond our control and data collection does not run at all time. As such, not all the updates if existing, are captured in this dataset; Therefore, actual temporal dynamics may be underestimated. However, our following analysis shows that configurations do not vary over time frequently, and thus one-time collection is sufficient.

	4G LTE	3G UMTS	GSM	3G EVDO	CDMA1x
#. parameter	66	64	9	14	4
cell-level (%)	72%	14%	5%	5%	4%

Table 4: Breakdown per RAT.



(a) # Samples per cell (b) Temporal dynamics

Figure 13: Temporal dynamics in configurations.

We first confirm that we have enough samples to study temporal dynamics. Fig. 13a shows the number of samples across all the cells, for given serving-cell configuration parameters (in SIB3), and almost half of the cells (48.1%) have multiple samples. This indicates that at least 48.1% of cells > 1 samples for certain configuration parameter and our dataset suffices to examine temporal dynamics. We find that temporal dynamics varies with idle-state and active-state handoff configuration parameters. We plot the percentage of LTE cells with distinct samples observed over time in Fig. 13b (two y-axes). If the cell is observed with multiple samples for the same parameter in one round, it will be counted into the t=0 case. We see that **configuration updates over time are really rare; Those idle-state handoff parameters are updated less frequently.** Both idle-state and active-state do not vary too much over time (idle: 0.4% to 1.6%, active: 21.2% to 24.1%, up to 2 years, mostly in 6 months). Active-state handoffs are updated more frequently.

Implications: Given low temporal dynamics, our data collection even with only one-time observation is enough. In our following study, we consider unique samples, so as not to tip distributions in favor of cells with many same samples.

5.2 Configurations in One US Carrier

We first use one US carrier (AT&T) to characterize handoff configurations in reality and then extend to other carriers in §5.3. We find that configurations are quite complex and diverse in all carrier networks. We characterize such complexity and diversity in terms of three measures: the number of unique values, the distribution and the dispersion over the value range.

Fig. 14 plots the distribution of eight representative parameters selected from Table 2. This is no surprise that they are consistent with our findings in §4 (three cities only). We have three observations. *First, there are multiple distinct values for most parameters*, except the hysteresis for the serving cell's radio evaluation H_s $(4dB)$. On the extreme end, some parameters such as $\Theta^{(s)}_{lower}$, $\Theta_{nonintra}$ and $\Theta_{A5,S}$ have around 20+ options. *Second, their distributions vary a lot as well.* Some have a skewed distribution with one or few dominant values (e.g., Δ_{min}, the measurement calibration threshold mainly set as -122dB); Others have a relatively even distribution across most values (e.g., the priority of the serving cell P_s as 0-7 for LTE cells). This indicates that AT&T does not treat all 4G LTE cells equally with finer-grained priority settings. However, it may induce inconsistent priority settings and problematic handoffs disclosed in

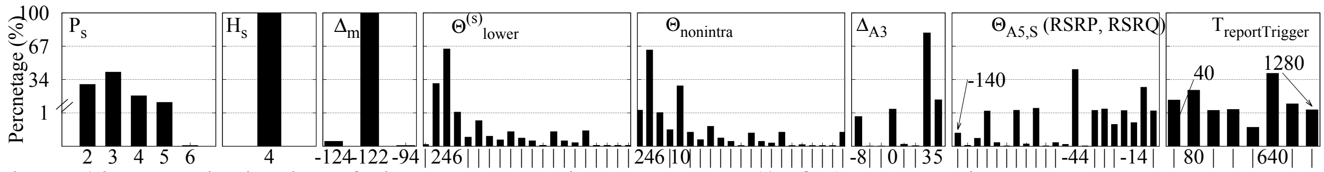
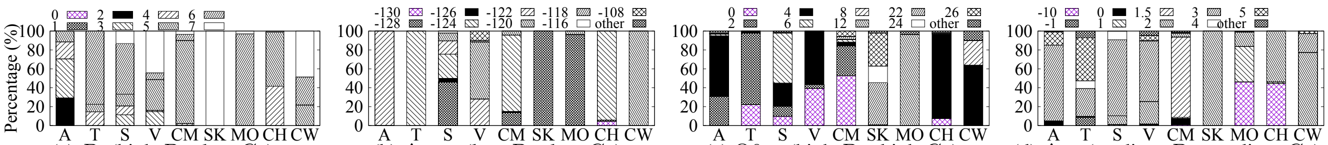


Figure 14: The distribution of eight representative parameters (AT&T). The y-axis uses two scales. The one of [0%, 1%] has been amplified for better readability. '|' represents skipped values without enough space. Left to right with its index in Fig. 16, Simpson index and coefficient of variation: P_s(index:31, 0.69,0.3), H_s(index:1,0,0), Δ_{min} (index:9,0.003,0.003), $\Theta^{(s)}_{lower}$(index:22,0.49,0.81), $\Theta_{nonintra}$(index:23,0.52,0.82),Δ_{A3}(index:20,0.33,0.34), $\Theta_{A5,S}$ (index:32,0.72,0.69), $T_{reportTrigger}$(index:35,0.78,0.84).



(a) P_s (high D + low C_v) (b) Δ_{min} (low D + low C_v) (c) Θ^s_{low} (high D + high C_v) (d) Δ_{A3}(medium D + medium C_v)

Figure 15: Illustrative distributions of four parameters observed with different diversity in AT&T in all chosen carriers.

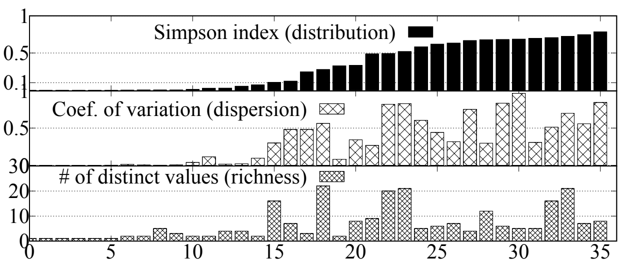


Figure 16: Diversity measures of LTE handoff parameters in AT&T.

our recent study [17]. Fo Fig. **??**. *Third, rich diversity does not only exist in their distribution but also in their value range.* Some parameters disperse in a broad range of values (e.g., [-140dB, -8dB] for $\Theta_{A5,S}$ (both RSRP and RSRQ supported), and [40ms, 1280ms] for the $T_{reportTrigger}$ timer). Such wide dispersion implies that those parameters probably affect handoff quality more as validated in §4.

Diversity metrics. To quantify such diversity, we apply two popular metrics: Simpson index of diversity [20] and coefficient of variation [21]. Simpson index is to quantify the diversity in distribution. It is better than the naive measure of the number of unique values (richness) because it takes into account the relative abundance of each value. Coefficient of variation is a well-defined, statistical measure to quantify the diversity in the value range. This complements Simspon index for measuring relative variability. They are given by

$$D = 1 - \sum_{i=1}^{m} (n_i)^2/N^2, \quad C_v = \frac{\sqrt{Var[X]}}{E[X]} \tag{4}$$

Where m is the number of unique values, n_i is the count of a single value x_i, and N is the total counts of all values $N = \sum_{i=1}^{m} n_i$. $E[X]$ and $Var[X]$ are the expectation and variance of the data X ($X_j, j = 1 \cdots N$). Simpson index ranges from $[0, 1]$ and a lower value indicates less diversity. A lower coefficient in C_v indicates lesser dispersion in value.

Fig. 16 shows the diversity measures of all handoff configuration parameters observed in AT&T, sorted in the increasing order of Simpson Index. We only observe a subset of configuration parameters because AT&T does not support 3G EVDO and 2G CDMA technologies (some parameters not applicable); Some events are not observed (say, B1, B2, A6) or rarely observed (say A1, A4). Those parameters are omitted as well. We confirm that each configuration parameter has its unique diversity pattern. The only exception is those parameters with no/low diversity (index $\leq$ 16 or 8). In fact, the first 8 parameters are single valued and No.9-16 are dominated by a single value. We find that these parameters do not exhibit rich diversity because they are primarily used for calibration or are associated with other varying parameters (e.g., Event A3 uses both an offset and a hysteresis; The hysteresis remains fixed as the offset varies). This way, carriers are still armed with sufficient power for fine-grained handoff management. Among those parameters with distinct values, diversity is multi-faceted with consistent or divergent patterns among their distribution, dispersion and richness. For instance, $\Theta_{A5,S}$ (index:32), $\Theta^{(s)}_{lower}$ (index:22) and $\Theta_{nonintra}$ (index:23), are consistently diverse, but the serving priority P_s (index: 31) is diverse in the distribution but not in dispersion and richness. In contrast, $\Theta^{(c)}_{lower}$ (index:15) and $\Theta^{(c)}_{Higher}$ (index:18) have high richness and dispersion but medium (lower) distribution diversity because one or two values are dominant in use.

Implications: operators have power to realize fine-grained handoff managements with diverse configurations.

5.3 From One to Many Carriers

We extend the above study to all other carriers. Unsurprisingly, rich diversity is observed in all other carriers. Due to space limit, we mainly present interesting results on carrier-specific diversity. We consider all four US carriers and other

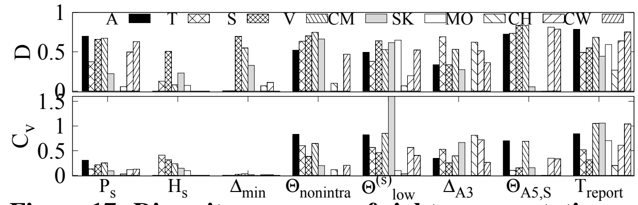


Figure 17: Diversity measures of eight representative parameters across various carriers.

representative carriers each from China (China Mobile), Korea (SK Telecom), Singapore (MobileOne), Hong Kong (China Mobile Hong Kong) and Taiwan (Taiwan Cellular). The conclusions are applicable to other carriers. We select four representative parameters with different-level diversity observed in AT&T to exemplify their distributions in those carrier networks in Fig. 15. We show diversity measures of the same eight parameters across the chosen carriers in Fig. 17.

We clearly see that each parameter configuration is carrier specific. This gives several implications. *First, parameters are likely configured by carriers, not by telecom equipment vendors (default values not in use). Second, we observe that diversity across multiple parameters is consistent for certain carriers.* For example, SK Telecom (Korea) exhibits lowest diversity for almost all the parameters. All four representative parameters (priority, radio evaluation thresholds/offsets) are single-valued; In contrast, all other carriers except Mobileone (Singapore) use highly diverse configurations for all the parameters. This implies that carriers adopt distinct (likely proprietary) configurations and policies. There might be no single answer given different goals of interests (performance, operational cost, robustness, etc.). But it might be a concern without thorough investigation on whether the current one is a winner and how far away, if not.

Implications: Carrier-specific configurations raise an interesting question, which configuration (policy) runs better? There might be no single answer given different goals of interests (performance, operational cost, robustness, etc.). It might be a concern if handoff configurations are not well managed and verified before their use.

5.4 Understanding Handoff Configurations

We next delve into a closer look at *why* they are configured so. We attempt to unveil what attributes contribute to current configuration diversity and how. We consider three factors: cell frequency, RAT, and location. We choose them because intuitively, operators may customize their policies per cell for finest-grained management (low temporal dynamics validated in §5.1). These three factors decide the cell type (what the cell is) and location (where the cell is), which are the most important cell properties visible to us.

5.4.1 Frequency

We first select P_S and P_C, the priorities of the serving and candidate cells for frequency dependence analysis. Intuitively, they should be frequency-dependent. Fig. 18 plots their breakdown per frequency channel in AT&T. All the carriers with multiple values (except SK and MO with low diversity) observe similar frequency-dependent patterns (omitted without enough space). AT&T uses 24 distinct channels, and the operating frequency for the serving cell primarily over the channels numbered as 850, 1975, 2000, 5110, 5780 and 9820, which matches with its 4G band usage [22]. The channel number is called EARFCN (LTE Absolute Radio Frequency Channel Number), and their mappings to frequency spectrum bands are regulated by [23] and can be found online, e.g., via [24].

We see that each frequency channel is mostly associated with one single/dominant value and the use of multiple frequency channels is the primary contributor to current priority diversity (exceptions explained later). There are several interesting findings. *First, AT&T uses a lower priority (here, 2) for LTE-exclusive bands* (called main bands [22], bands 12 and 17 around 700MHz), including 5110/5145 (band 12) and 5780 (band 17); Channel 1975 (band 4, AWS-1) is an exception. A higher priority (5 or 4) is mainly assigned to the 9820 channel (band 30, 2300 MHz WCS), which is recently acquired to provide additional bandwidth. Such priority setting implies that AT&T prefers the additional bands to the LTE-exclusive (main) bands and tends to use 3G-coexisting bands for LTE as much as possible. This sounds a good upgrade strategy which facilities and accelerates wide adoption of a new RAT. *Second, some frequencies use multiple values which is prone to conflicts.* For example, AT&T assigns two or three values over the channels of 1975, 2000, 2425, 5870 and 9820. This is observed at 6.3% of AT&T cells in our measurement study. Such multiple-value priority settings are also observed in other carriers. However, such inconsistent priority settings might make troubles. Consider a case where two cells believe the other has a higher priority. It is prone to a handoff loop, which was reported by our prior work [17]. Our large-scale study shows that this problem exists in many carriers and unfortunately, it is not that rare as we anticipated before. *Third, our study troubleshoot the problematic practice.* . We notice that updating priorities with new channel preference can be problematic in practice. [25] reports one recent user complaint that AT&T breaks the 4G service for the phones that do not support band 30 (here, channel 9820). But no technical cause has been reported. Now, we know why. AT&T sets the highest priority to band 30 and thus the handoff decision step likely chooses the cell over it, even when 4G LTE cells over different channels available (otherwise, the phone could not work well before). Given that not all the phones support band 30 (e.g., a Verizon iPhone 6S+ [25]), they are unable to switch to the target cell all along and thus the 4G service is disrupted.

We further quantify such frequency dependence, using a generic measure of parameter θ's dependence on a factor F.

$$\zeta_{\mathcal{M},\theta|F} = E\left[|(\mathcal{M}(\theta|F = F_j) - \mathcal{M}(\theta)|\right] \qquad (5)$$

where $\mathcal{M}(\theta)$ is the θ's diversity measure (here, D or C_v).

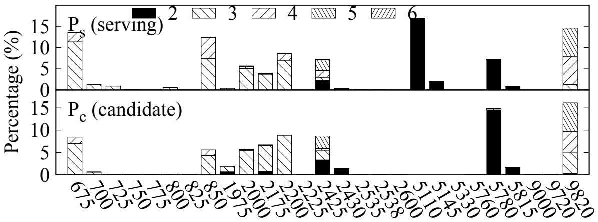


Figure 18: The breakdown of the serving (top) and candidate (bottom) cell priorities over frequency (AT&T).

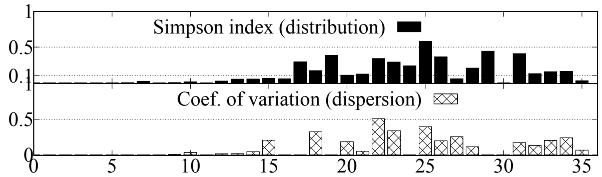


Figure 19: Measures of frequency dependence: $\zeta_{\mathcal{D},\theta|freq}$ (top) and $\zeta_{\mathcal{C}_v,\theta|freq}$ (bottom) across all the parameters in the same order of Fig. 16 (AT&T).

We compare it with the expectation of the conditional ones $\{\mathcal{M}(\theta|F = F_j)\}$. We plot $\zeta_{\mathcal{D},\theta|freq}$ and $\zeta_{\mathcal{C}_v,\theta|freq}$ for all the parameters observed in AT&T in Fig. 19. We indeed observe that frequency dependence per parameter is also carrier-specific and do not show these results due to space limit. However, it holds true to all the carriers that not all highly diverse parameters (here, No.$\geq$ 17) are frequency-dependent. Interestingly, we find that some reporting events are frequent-dependent like A2 (index: 32) and A5 (index: 33 and 34) but some not, such as A1 (index: 21), A3 (index: 21). This helps to infer the carrier's handoff policies. Here, we can see that there is a universal standard for a good cell (A2) and relative comparison (A3) but the standard for a poor cell (A2) and the absolute value setting (A5) are frequent-dependent. We also observe that some other parameters like $T_{reportTrigger}$ (index 35) and hysteresis (index: 27) are frequent-indepenent, which matches with their use.

5.4.2 Location

We quantify the impacts of location at the macro-level (city) and micro-level (proximity). We aim to answer two questions: (1) Do operators customize their configurations in cities? (2) Will diversity disappear (or greatly decline) among nearby cells?

City-level. Here, we study US cities only. We divide our dataset based on the cities where the configurations are collected and we present the results for top-5 cities (total number of cells in four US carrier): C1(Chicago: 4671), C2 (LA: 2982), C3 (Indianapolis: 2348), C4 (Columbus: 1268), C5 (Lafayette: 745). We choose P_S and normalize its distribution in each city. Fig. 20 plots the results. We observe that carriers may configure cells at different geographical locations slightly differently. In C1 (Chicago), their configurations obviously differ from those in other cities. This is understood. Operators usually divide their network domain

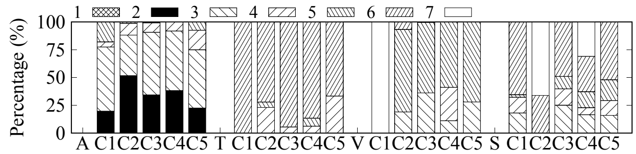


Figure 20: City-level priority distributions in five cities.

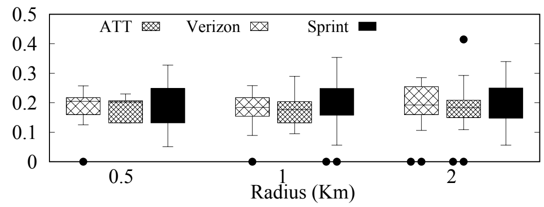


Figure 21: Spatial diversity for P_s under various Radii in Indianapolis (C3).

(one nation) into multiple market areas and they may run incremental deployment and configurations over time. The bands used may differ as well. We also check other parameters and observe location-dependent diversity.

Proximity. We further consider those cells in close proximity. Handoff is distributed. After it switches to a new serving cell, the configurations associated with the new cell take effects. So a handoff is affected by configurations at co-located cells. As our data collection is dependent upon user movement, we observe that the cells covered in our dataset are sparse except those cells collected by us. So we use a subset of dataset `D2` which is collected in a more controlled manner by us. In particular, we drive along the main roads separated every 500m –1Km and cover the whole city to get a more dense collection. We have done so in C3, C4 and C5, partially in C1 and C2. We apply Eq. (5) to define a measure of spatial diversity as $\zeta_{\mathcal{M},\theta|R}$, where R is the radius of one neighborhood, $\mathcal{M}$ is the diversity metric and θ is the parameter to study. For any cell c, we obtain the cluster of cells located in a circle of radius R km and obtain $\zeta_{\mathcal{M},\theta|R}[c]$. To illustrate its spatial diversity, Fig. 21 shows the boxplot of $\zeta_{\mathcal{M},\theta|R}[c]$ for all the cells in C3. We select various radii to gauge the change in configurations. We only show the results for AT&T, Sprint and Verizon. We observe that carriers indeed use varying values for cells located closely to each other. This indicates that even in a very small geographical area ($r < 0.5$), carriers prefer to fine tune different parameters. However, this is not the case for all the carriers. In T-Mobile, we observe that spatial diversity in close proximity is extremely small (almost zero). That is, spatial diversity does exist across small geographical areas but is also carrier dependent.

Implications: Lower dispersion is observed in a smaller range (like a city or a neighborhood). Location-dependency is likely caused by real-world deployment (the network deployment and upgrade not at the same using the same equipments). It can be also attributed to the carrier's configuration

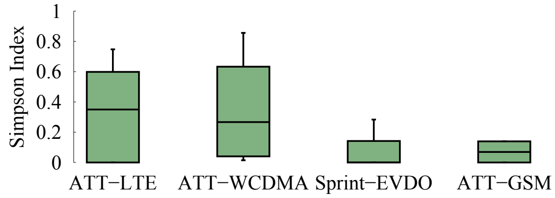


Figure 22: Boxplots of diversity metrics of all parameters used by different RATs.

over geographic area.

5.5 Evolution of RATs

We finally study the configuration patterns under other RATs and learn how they have evolved. Because different RATs use different sets of parameters, it is hard to compare each parameter across RATs. We thus calculate the diversity metric (here, Simpson index) for all the parameters and show their boxplots in Fig. 22. We see that handoff configurations are becoming more and more diverse along the RAT evolution. In particular, LTE heavily inherits from UMTS and thus they have a large number of parameters in common. CDMA2000 and CDMA1x are used by Sprint and Verizon and configured differently from LTE. They use a smaller number of handoff parameters. Most of the parameters are observed to have a single dominant value and relatively static configurations. Similarly, GSM is also observed to have an almost static configuration scheme. Their average diversity of their parameters is significantly smaller than those of LTE and WCDMA, indicating single dominant values. Thus, the evolution of RATs over time has also made cell handover procedure more convoluted and complicated where numerous more parameters with varying and diverse configurations are used.

Implications: Increasing diversity may continue in the coming 5G, especially with hybrid and more radio access options. Our study likely helps understand mobility support in 5G as well.

6 Implications and Discussions

We now discuss the implications and potential actions for operators, end users, and the research community. Our study further opens problems that warrant future efforts.

Suggestions for operators. Operators should verify the correctness and validate the expected properties of their configurations. They should also reassess their used configuration algorithms for active-state handoffs. Our study has confirmed that, certain configurations are problematic and yield nontrivial performance penalty.

We make four suggestions for operators to check their handoff configurations. First, they should look into A3 and A5 events for active-state handoffs. Some negative offset values are observed in A3, and A5 allows for more than 20 options. They may prevent or delay handoffs, thus impeding performance on the go. Second, operators should check measurement and decision parameters with nontrivial gaps for idle-state handoff configurations. They imply either premature measurements or overdue handoff decisions. Consequently, they may unnecessarily drain the battery at the mobile device. Third, operators should look into their priority settings carefully. Our study have identified several real-world instances of losing 4G access, despite the availability of local 4G cells, due to improper priority settings by operators. In fact, when performing infrastructure upgrades or reconfiguring parameters (more than priorities), operators should consider not only their impacts on the updated cells, but also those that are not upgraded but still affected (e.g., cells in proximity). Fourth, operators should take configurations into account when troubleshooting the user compliant tickets on poor performance and failures upon mobility, Improper configurations should share the blame for some cases.

Mobile user at the device side. We believe that mobile users can benefit from our study along two dimensions.

Our key finding is that, given the observable configurations, it is feasible to *predict* handoffs at runtime at the mobile device. Using our tool, the mobile user can readily collect runtime configuration parameters, and use them plus realtime measurements to forecast whether and how a handoff will occur in the near future. Moreover, such predictions can be highly accurate, given the common handoff policies being used. Such accurate predictions supply reliable heuristics at runtime to optimize TCP and application performances over cellular networks.

Mobile users can further detect improper configurations using the information collected from handoff configurations. They can leverage their device-side capabilities to eliminate or alleviate their negative impact (e.g., reducing unnecessary measurements, triggering timely handoffs, and relaxing strict requirements on radio quality).

Research community. Our study so far helps to under how operators manage handoffs and design their policies. It also partially explains why operators design their handoffs in the current forms. However, a number of research issues remain for further efforts.

• *Automated tool for configuration verification* Given the sheer scale of cells and configuration settings, we thus believe an automated solution to configuration verification is the viable approach. Our study sheds lights on how to design such automated tool. We believe that, such a tool is feasible, if we leverage runtime configurations collected from the device, the formal models for handoffs specified by the 3GPP standards, the verification techniques borrowed from programming language and AI communities, and the learning algorithm to be adapted from the machine learning and AI communities. Moreover, given such configuration checks, we can further conduct cross-layer study that spans the low-level cellular protocol stack to the higher-level TCP/IP suite.

• *What are the goals for operators to achieve in their policy-based handoffs?* Policy-based practice is not for performance only. As we have learned from the Internet BGP case, policy design is mainly shaped by nontechnical issues. However, a big difference exists for the handoff case. Handoffs are mainly for a single carrier network, without crossing administration domains. They are invoked for diverse (even conflicting) goals such as selecting the best radio quality, boosting high-speed access, sustaining seamless data/voice support, balancing loads, lowering operational cost, etc.. Note that not all configurations are exposed to us. The observed configurations are mainly on radio evaluation tuned by cell priorities. Our study shows that, it might be viable to first examine their policies on radio quality, and to then extend to non-radio components. Through relaxing their radio requirements (see the A5 examples in §4.1) and comparing with other configurations, we can possibly infer how handoff policies vary, reason the expected impacts of such changes and learn the handoff policies.

• *Will handoff configurations realize the policies and goals as expected?* From configurations, we learn that a handoff will take place under what conditions and go to which cell, as well as its required time and overhead. As shown in our measurements §4.2, we may infer their impact on intra-freq and non intra-freq measurements when the threshold reduces from 62dB to 42dB. From performance assessments, we can associate these configurations with the perceived configurations. Through large-scale learning, we can quantify their performance impacts and examine whether they induce unnecessary performance penalties.

• *Will handoff configurations introduce unexpected troubles?* Handoffs are distributed operations in nature. While single-cell configurations are well justified, there is no guarantee for no conflicts among multiple cells. Our prior studies [17,26] have shown that, misconfigurations may compromise the structural properties of stability and reachability. This study reveals rich diversity, which is prone to configuration conflicts and misconfigurations.

• *Implications to 5G* The identified results are likely applicable to 5G. 5G will continue its policy-based handoff management. Following the RAT evolution, the upcoming 5G will employ even more configuration options, with the adoption of new radio technology and support for extremely high speed (e.g., at aircrafts).

7 Related Work

There is no work that characterizes real-world handoff configurations and investigates their implications, except our preliminary studies [17, 26, 27]. Our prior results have disclosed the problems of handoff instability [17, 27] and unreachability [26] due to misconfigurations or conflicting configurations. Those studies conduct theoretical analysis (reasoning) to prove the existence and their conditions of unstable handoffs or unreachable cells, and use a small dataset of two US carriers in two cities to validate the possibilities in real-world and assess their performance impacts. As a matter of fact, this measurement study is motivated by them but it focuses on characterizing configurations at a much larger global-scale (30 carriers, 31,000+ cells). Moreover, we assess the impacts of 'normal' configurations, rather than problematic configurations only, on handoff quality and user performance. We next compare most relevant work in two categories, which may also benefit from our measurement study.

Performance in mobility. Cellular data performance on the go has been measured, analyzed and enhanced in the literature [2–9]. They look into TCP and application layers only. They show data performance indeed declines due to handoffs [2–4] and propose enhanced TCP solutions for cellular networks (e.g., recent advances in [5–8] and a survey in [9]). Our study takes a new perspective into the underlying handoff process directly and investigates performance impacts induced by diverse configurations. It can offer explicit handoff information, not only on its consequence, but also the reason and procedure. This potentially help higher TCP/application layers deal with handoffs in a better way.

Handoff parameter re-design. Several studies aim to enhance mobility support in cellular networks through handoff parameter optimization [28–30]. [28] proposes to locate optimal event parameters for 2G-3G handoffs; [29, 30] advocate to integrate user objectives into handoff decision strategies and introduce new parameters. However, they use a simplistic model which aims to optimize one objective function, but do not take into account real handoff mechanisms (distributed handoffs, policies and configurations). Our work can help handoff configuration optimization is solved in a more realistic setting.

8 Conclusion

In this work, we conduct a sizable measurement study on policy-based handoff configurations from 30 mobile carriers in the US and globally. To this end, we design a new device-centric tool `MMLab`, which collects the runtime handoff configurations without the assistance from operators. We analyze practical handoff configurations used in operational networks. This helps us to understand how handoffs are performed in reality. Our study further sheds lights on how the operators, the mobile user at the device side, and the research community can move forward to refine the handoff management in 4G/5G systems, which are the only large-scale networks with ubiquitous mobility support.

9 References

[1] 3GPP. TS23.009: Handover Procedures, 2011.

[2] Keon Jang, Mongnam Han, Soohyun Cho, Hyung-Keun Ryu, Jaehwa Lee, Yeongseok Lee, and Sue B Moon. 3g and 3.5 g wireless network performance measured from moving cars and high-speed trains. In *Proceedings of the 1st ACM*

workshop on Mobile internet through cellular networks, pages 19–24. ACM, 2009.
[3] Ruben Merz, Daniel Wenger, Damiano Scanferla, and Stefan Mauron. Performance of lte in a high-velocity environment: A measurement study. In *Proceedings of the 4th workshop on All things cellular: operations, applications, & challenges*, pages 47–52. ACM, 2014.
[4] Li Li, Ke Xu, Dan Wang, Chunyi Peng, Qingyang Xiao, and Rashid Mijumbi. A Measurement Study on TCP Behaviors in HSPA+ Networks on High-speed Rails. In *INFOCOM*, April 2015.
[5] Yung-Chih Chen, Yeon-sup Lim, Richard J Gibbens, Erich M Nahum, Ramin Khalili, and Don Towsley. A measurement-based study of multipath tcp performance over wireless networks. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 455–468. ACM, 2013.
[6] Vasilios A Siris, Maria Anagnostopoulou, and Dimitris Dimopoulos. Improving mobile video streaming with mobility prediction and prefetching in integrated cellular-wifi networks. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 699–704. Springer, 2013.
[7] Ashkan Nikravesh, Yihua Guo, Feng Qian, Z Morley Mao, and Subhabrata Sen. An in-depth understanding of multipath tcp on mobile devices: Measurement and system design. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 189–201. ACM, 2016.
[8] Prateesh Goyal, Mohammad Alizadeh, and Hari Balakrishnan. Rethinking congestion control for cellular networks. In *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, HotNets-XVI, pages 29–35, 2017.
[9] Hala Elaarag. Improving tcp performance over mobile networks. *ACM Computing Surveys (CSUR)*, 34(3):357–374, 2002.
[10] 3GPP. TS36.331: Radio Resource Control (RRC), 2012.
[11] 3GPP. TS25.331: Radio Resource Control (RRC); Protocol specification, 2006.
[12] 3GPP. TS36.133: E-UTRA; Requirements for support of radio resource management, 2017.
[13] 3GPP. TS25.304: User Equipment (UE) Procedures in Idle Mode and Procedures for Cell Reselection in Connected Mode, 2012.
[14] 3GPP. TS36.304: User Equipment procedures in idle mode, 2013.
[15] 3GPP. TS25.133: Requirements for support of radio resource management (FDD), 2017.
[16] Wikipedia. Cellular frequencies. https://en.wikipedia.org/wiki/Cellular_frequencies.
[17] Yuanjie Li, Haotian Deng, Jiayao Li, Chunyi Peng, and Songwu Lu. Instability in distributed mobility management: Revisiting configuration management in 3g/4g mobile networks. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pages 261–272. ACM, 2016.
[18] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. Mobileinsight: Extracting and analyzing cellular network information on smartphones. In *ACM MobiCom*, 2016.
[19] Continuous speed test. http://backup1.startrinity.com/InternetQuality/ContinuousBandwidthTester.aspx, 2018.
[20] M. O. Hill. Diversity and evenness: A unifying notation and its consequences. In *Ecology*, pages 427–432. Ecological Society of America, 1973.
[21] A. Skrondal B. S. Everitt. The cambridge dictionary of statistics. Cambridge University Press, 2011.
[22] What are at&t's 4g lte bands in 2018? https://forums.att.com/t5/Phone-Device-Upgrades/What-are-AT-amp-T-s-4G-LTE-bands-in-2018/td-p/5359265, 2018.
[23] 3GPP. TS36.101: E-UTRA; User Equipment (UE) radio transmission and reception, 2017.
[24] Frequency calculator. http://niviuk.free.fr/lte_band.php, 2018.
[25] The move to band 30 causing issues with other bands? https://forums.att.com/t5/Apple/The-move-to-Band-30-causing-issues-with-other-bands/td-p/5267142, 2017.
[26] Chunyi Peng and Yuanjie Li. Demystify undesired handoff in cellular networks. In *Computer Communication and Networks (ICCCN), 2016 25th International Conference on*, pages 1–9. IEEE, 2016.
[27] Yuanjie Li, Jiaqi Xu, Chunyi Peng, and Songwu Lu. A First Look at Unstable Mobility Management in Cellular Networks. In *HotMobile*, Feb 2016.
[28] Christopher Brunner, Andrea Garavaglia, Mukesh Mittal, Mohit Narang, and Jose Vargas Bautista. Inter-system Handover Parameter Optimization. In *VTC Fall*, 2006.
[29] Dhanaraj Cheelu, M Rajasekhara Babu, P Venkatakrishna, and XZ Gao. User preferences and expert opinions based vertical handoff decision strategy with the inclusion of cost parameter for 4g networks. *International Journal of Autonomous and Adaptive Communications Systems*, 10(3):261–278, 2017.
[30] Li Qiang, Jie Li, and Corinne Touati. A user centered multi-objective handoff scheme for hybrid 5g environments. *IEEE Transactions on Emerging Topics in Computing*, 5(3):380–390, 2017.