Demystify Undesired Handoff in Cellular Networks

Chunyi Peng
Department of Computer Science Engineering
The Ohio State University
Columbus, OH 43210
Email: chunyi@cse.ohio-state.edu

Yuanjie Li
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
Email: yuanjie.li@cs.ucla.edu

***Abstract*—Handoff is a critical mechanism in cellular networks. When the mobile device moves out of the coverage of the serving cell (*i.e.*, base station), a handoff is performed to switch its serving cell to another and thus to ensure seamless network access. To provide nice user experience, it is desirable to select the preferred cell (*e.g.*, 4G rather than 3G/2G in most cases) among multiple candidate cells which all are around and able to serve the device if needed. In this paper, we examine the property of *desired reachability* in the current design and practice of handoff. We show that handoff is designated to be configurable in order to accommodate diverse requirements by users and operators. However, handoff misconfigurations exist and they make the device stuck in an undesired target cell (*e.g.*, 2G when 4G available). We model the distributed mobility management as an *iterative* process and use a formal analysis to classify the causes. We further design a software tool to detect handoff misbehaviors and run it over operational networks. We validate the identified issues on two major US mobile carrier networks.**

***Index Terms*—Cellular Network, Handoff, Mobility Management, Desired convergence, Reachability**

I. Introduction

Mobility support is widely regarded as a fundamental utility service to the evolving Internet. To support billions of mobile-ready devices (including smartphones, tablets, wearables, Internet of Things, *etc.*), cellular networks play a pivotal role in offering "anytime, anywhere" mobility support in reality. The key lies in its micro-mobility management scheme, which determines the serving cell (also known as base station[1]) and migrates the mobile device from the currently serving cell to the next neighboring one when *necessary*. This procedure is also called as *handoff*.

Handoff is designed to meet versatile demands from mobile carriers and users. They include, but not limited to, sustaining pervasive network availability, providing high-speed data service, offering seamless voice/data support, balancing traffic loads between cells. Moreover, coexistence of heterogeneous technologies (*e.g.*, 3G, 4G LTE, LTE-advanced, small cells) further results in diverse handoff configurations. As a matter of fact, 3GPP standards defines a variety of handoff mechanisms with distinct logic and tunable parameters [4]–[8], [10]–[13]. In some cases, carriers have freedom to determine their own handoff decision logic and parameters to use.

Given such flexibility, a question arises. Will handoff configurations at different cells conflict with each other? If so, what are their negative impacts in reality? This work is stimulated by our recent studies on handoff stability [24], [25]. We have disclosed that mobility management (MM) misconfigurations do exist among different cells so that the handoff process may never converge in some cases. Instead, it oscillates among multiple cells in a persistent loop and incurs excessive resource waste and sharp performance degradation or even failures. In this work, we move forward to another structural properties: reachability (desired convergence). Reachability states the handoff eventually settles down at a choice (converges) and at a nice choice (*e.g.*, selecting 4G rather than 2G/3G when all available). By "Nice", we mean that the decision conforms to user and/or operator preferences and will elaborate it later in each instance.

Our efforts cover from theory to practice. We start from a handoff model and then conduct a formal analysis to derive the conditions for undesired reachability. We further design an in-device software tool and carry out real experiments over two top-tier US carrier networks to validate the existence of such misbehaviors and assess their impacts. Our study shows that undesired handoffs do occur in our real life. The device stays in 2G when 4G available, or even becomes out of service (can't connect to 4G) when it moves from femtocells (user-deployed small cells) to 4G. We also uncover that the handoff to 2G takes over the one to 3G due to device-network misconfigurations on MM. To the best of our knowledge, this work is the first effort to examine (un)reachability due to mobility management misconfigurations.

The rest of the paper is organized as follows. §II reviews the background of handoff configurations and related work. §III and §IV describes our analytical efforts and empirical findings. §V discusses the remaining issues and fix solutions; §VI concludes this paper.

II. Background and Related Work

The 3G/4G network is the largest wireless infrastructure deployed to date. Each cell tower serves one geographic area called a cell, denoting the coverage of radio access to devices in proximity. At a given location, a device is usually covered by multiple, possibly overlapping cells.

Handoff process. Given a mobile device and its current serving cell, a handoff is to determine whether to switch the

[1] Each base station may manage multiple cells (antennas), each of which covers a geographical area. In this paper, we use cells and base stations interchangeably, for a slight abuse of notations.

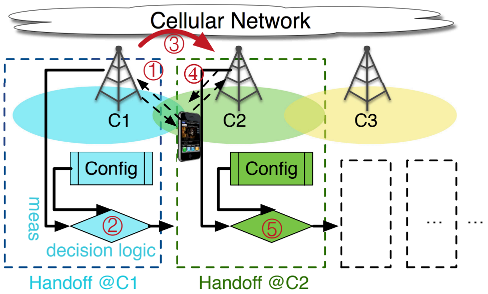


Figure 1: Distributed handoff process with each atomic handoff executed at the serving cell.

current cell and which to select among multiple candidates. This process is illustrated in Figure 1. Each decision is made locally at a cell or by the mobile device. The operation has three components: the local decision logic (rules), tunable configuration parameters and runtime measurements. The decision logic takes both pre-configured parameters and runtime measurements as inputs (①), and determines the next appropriate cell (②). Once the decision is made, it executes the handoff procedure (③) and migrates the device to the chosen next cell. Once the previous handoff procedure completes, the device switches to a new cell; New handoff procedures can be invoked and new serving cells will be further selected and switched to (④ and ⑤) as long as the handoff criteria is met. This way, through a sequence of handoff events, the mobile device retains its radio access to the cellular network no matter where it goes or stays.

In essence, the handoff process is distributed in nature. There is no central point which collects all the information and makes a global decision. Instead, each decision is made locally and iteratively until it settles down at one certain cell.

Handoff types. There are two types of handoffs in 3G/4G networks. (1) *Idle-state handoff*: it is performed by the mobile device, when the device is at the idle state (without ongoing voice/data traffic) and has no active connection to the serving cell. This is to make the device ready for network access at any time. (2) *Active-state handoff*: it is initiated by the serving cell, when the device is actively served by the current cell for its ongoing data traffic through the established radio connection.

Handoff serves as generic mobility support to satisfy versatile (sometimes conflicting) demands such as selecting the best radio quality, boosting high-speed access, sustaining seamless data/voice support, load balancing, to name a few. As a result, 3GPP standards regulate a variety of procedures related to MM to fulfill different purposes. Table I lists the main procedures. They include *initial attach, cell (re)selection, active handoff, voice support via CSFB (Circuit Switch Fallback) and SRVCC (Single Radio Voice Call Continuity), offloading, load balancing* (*e.g.*, via self-organizing networks). Each works with certain radio access technology (RAT, say, 4G/3G/2G), and/or various service types (say, active data/voice/both or idle).

Specifically, the *initial attach* and *cell-(re)selection* procedures are used to look for a serving cell or another better cell when the device has no active association with the serving one (idle). They are performed regardless of whether mobility is involved or not. The decision is based on the measured radio quality from different cells, the cell preference and radio evaluation criteria preconfigured by the device or reconfigured by the associating cell. The used parameters for the idle-state handoff have been standardized in [13]. The *active handoff* procedure regulates the cell switch with ongoing traffic, and its primary goal is to ensure seamless services. It exhibits many forms, including inter-RAT handoff (*e.g.*, 4G↔3G) and intra-RAT handoff (*e.g.*, within 4G), soft handoff (with simultaneous connectivities to multiple cells) and hard handoff (disconnect-and-connect). Moreover, several handoff procedures are designed for different goals. For instance, 4G LTE leverages 3G/2G systems to carry voice through CSFB and SRVCC, thus invoking 4G↔3G/2G handoffs, whereas the normal handoff often triggers the switch to 4G because 4G is likely faster. Some carriers encourage offloading to small cells or user-deployed femtocells, or traffic redirection to different cells for load balancing or carrier-specific optimizations. Compared with the idle-state handoff, the active-state handoff decision logic as well as the configuration parameters, are not standardized and carriers have freedom to customize them.

Procedure	Standard	RAT	Service
Initial attach	23.401 [6]	all	idle
Cell (re)selection	25.304 [8],36.304 [13]	all	idle
Active handoff	23.009 [5]	all	active
CSFB and SRVCC	23.272 [7],23.216 [4]	4G	active(voice)
Femtocell offloading	25.367 [11]	3G,4G	active & idle
WLAN offloading	23.261 [10]	3G,4G	active & idle
Load balancing	32.500 [12]	all	active

Table I: Main MM procedures in 3GPP standards.

Related work. Mobility support over cellular networks has been a long-lasting research topic. Extensive early efforts have been devoted to different forms of optimization, including VoIP support [20], radio link failure reduction [16], [21], [27], and handoff algorithm enhancement [17], [26], [29]. In recent years, most studies focus on mobility support for new needs such as traffic offloading [15], [19], [31], cognitive radio cellular networks [22], femtocells over LTE-advanced network [35] and unified mobility support for 5G [36]. In addition, data service performance under handoff and its optimization has been actively studied in the literature (*e.g.*, [23], [33]).

However, the performance of handoff itself in operational cellular networks has been largely overlooked, especially those rooted in the fundamental conflicts in mobility management (say, inconsistent decision logics and configurations). We take the first step to examine the impacts of MM misconfigurations in recent studies [24], [25]. We uncover that the current MM configuration might be inconsistent and thus the handoff process might never converge under the invariant environment.

Roadmap. In this work, we look into a different problem. Rather than whether it converges, we explore how well the convergence performs (assuming it converges). We are particularly interested in whether the handoff process settles down at a desired target cell. Given certain network conditions, a target cell is usually designated as the one that yields best performance by the operator or the user. Failing to converge to this target typically leads to worse performance. This is called as the desired reachability problem. To address it, we start from a formal analysis and derive the conditions for handoff unreachability (§III). We then validate the existence of such potential misbehaviors and assess their impacts in real cellular networks (§IV).

III. Analysis on Desired Reachability

Desired reachability specifies the quality of handoff convergence. In this section, we first model the handoff process and then use analysis to derive the causes for unreachability.

A. The Handoff Model

Our handoff model generally follows a discrete-event style. Each handoff is abstracted as an atomic transition from the serving cell to the next target. The whole process is modeled as an iterative one that consists of multiple (at least one) cascading handoff(s).

An atomic model. Each atomic handoff in current 3G/4G networks is *configurable*. Three components work in concert to make a handoff decision: the *decision logic*, the *tunable configuration parameters* and the *runtime observations (i.e., measurements)*. The decision logic takes both parameters and observations as inputs, and selects the next cell. Tunable parameters specify what kinds of metrics are of interest to the device and the operator. Runtime observations collect latest measurements, thus capturing dynamic network conditions. We next elaborate on three components for idle-state and active-state handoffs.

○ *Decision logic.* This is the algorithm to choose the target cell. The decision logic likely varies in both types. For instance, the device might prefer a cell with strongest signal strength while idle, whereas it chooses a 4G LTE cell with reasonable signal strength (say, >-100dBm) when active. The idle-state handoff logic is standardized in 3GPP specifications [8], [13]. Its exact form will be described in Figure 2. In contrast, the active-state handoff logic is customizable which gives carriers freedom to develop proprietary handoff algorithms for their sake.

○ *Configurable parameters.* They are used by the decision logic. For idle-state handoff, two types of parameters are used: the cell preference and the radio assessment thresholds. Table II summarizes the parameter notations, which are abstracted from actual configurations in operational networks. The active-state handoff allows to customize its parameter set.

○ *Runtime observations.* They are usually on the dynamic radio quality measured at the device, and serve as inputs to the handoff execution. The device collects and transfers such observations to the decision logic. The idle-state handoff accepts cell radio quality assessments as inputs, while the active-state one can use both the radio quality values and customizable observations (*e.g.*, cell loads). In practice, these observation metrics are typically pre-processed before handoff decisions are made. For example, the received signal strengths used in the handoff have been averaged to filter out noises and transients [8], [13]. To stay focused, we assume the observations remain unchanged during each handoff decision iteration.

Symbol	Description
Symbols for the abstract model	
$s \xrightarrow{\Omega_s} t$	One iteration with s as the serving cell, t as the target
C, c	C: List of available cells, c: one candidate cell, $c \in C$
Ω_s	the decision logic executed when s is serving
G_s	List of all configuration parameters when s is serving
O_s	List of runtime observations when s is serving
Parameters for configurations and observations	
γ_c	Received signal strength of cell c
$P_{s,c}$	Preference of cell c at cell s
Θ_s^{serv}	Threshold of γ_s when s is serving
$\Theta_{s,c}$	Threshold of γ_c when s is serving
$\Theta_{s,c}^{low}$	Threshold of γ_c when s is serving and $P_{s,c} < P_{s,s}$
$\Theta_{s,c}^{eq}$	Threshold of γ_c when s is serving and $P_{s,s} = P_{s,c}$
$\Theta_{s,c}^{high}$	Threshold of γ_c when s is serving and $P_{s,c} > P_{s,s}$

Table II: Notations.

We now model each atomic handoff execution as follows.

$$\text{Atomic handoff:} \qquad t = \Omega_s(G_s, O_s), t \in C_s, \tag{1}$$

where s is the serving cell, and t is the target cell selected from candidate cells C_s (often represented as C regardless of the serving cell). Given the serving cell s, Ω_s, G_s and O_s denote the handoff decision logic, tunable parameters and runtime observations, respectively. If the serving cell does not exist (*e.g.*, the devices just powers on), we have $s = \emptyset$ as a special case and the decision is initially made by the device.

Idle-state handoff. We start with the idle-state handoff which is fully regulated by 3GPP standards [3], [9]. This offers a basic and generic form which serves as the most important decision criteria for both idle-state and active-state handoffs.

Figure 2 shows the standardized decision logic Ω_s for the idle-state handoff. The decision logic chooses the target cell through pairwise comparison (the serving cell versus each candidate). The runtime observations are the received signal strength values each from one candidate cell (γ_c), measured by the user device. For each candidate cell c, the serving cell s defines two types of configurable parameters: the preference level ($P_{s,c}$) concerning a candidate cell c and a series of signal strength thresholds ($\Theta_s^{serv}, \Theta_{s,c}^{low}, \Theta_{s,c}^{eq}, \Theta_{s,c}^{high}$) that help Ω_s to make a decision. Note that both types of parameters are needed. Radio signal strength is directly related to wireless transmission performance, as well as the cell type (3G, 4G, macro-cells, or femtocells). The cell preference reflects the precedence of cell types from the perspective of the carrier or the user or both. It supplies a flexible mechanism for the device/network to adjust the priorities.

Specifically, each cell is evaluated with its pre-configured preference and runtime received signal strength. A target cell is chosen when one of the following criteria is satisfied:

```
Idle-state handoff
Input:   serving cell $s$, neighboring cell list $C$, radio measurements $O_s = \{\gamma\}$ tunable parameters and $G_s = \{P_{s,c}, \Theta_s^{serv}, \Theta_{s,c}^{low}, \Theta_{s,c}^{eq}, \Theta_{s,c}^{high} | c \in C\}$
Output:  target cell $t$
Step1:   initialize candidate cell list $L \leftarrow [\ ]$
Step2:   pairwise cell comparison
         for each cell $c \in C$,
             $L$.append($c$), only if one below rule is satisfied
             (1) when $P_{s,c} > Ps, s$, $\gamma_c > \Theta_{s,c}^{high}$
             (2) when $P_{s,c} = Ps, s$, $\gamma_c > \gamma_s + \Theta_{s,c}^{eq}$
             (3) when $P_{s,c} < Ps, s$, $\gamma_s < \Theta_s^{serv}$ and $\gamma_c > \Theta_{s,c}^{low}$
Step3:   target cell decision
         $t = \begin{cases} s & \text{if } L \text{ is empty} \\ c & \text{if } c = \arg\max_{c \in L} P_{s,c} \text{ (using } \gamma_c \text{ if a tie)} \end{cases}$
```

Figure 2: Idle-state handoff decision logic.

1) it is more preferred than the serving cell, and its signal strength is higher than a threshold;
2) it is equally preferred to the serving cell, and its signal strength is offset higher than the serving cell's;
3) it is less preferred than the serving cell, but the serving cell's signal strength is lower than a threshold, while the target cell's signal strength is higher than another threshold.

If more than one cell outperforms the serving cell, the one with the highest preference could be chosen. If a tie exists, the signal strength is used to break the tie.

Active-state handoff. We now extend the idle-state handoff model to the active-state one. It follows the same form $s \xrightarrow{\Omega_s(G_s,O_s)} t$ with various Ω_s, G_s and O_s in the active-state handoff context.

The main difference is that the active-state handoff allows the operator to customize its decision logic and use some network-side configurations and measurements which are not accessible on the device side. Take load balancing as an example. It may be designed to handoff from the serving cell to another when (1) the current one is overloaded and the neighboring one not, and (2) the neighboring cell offers satisfactory radio quality (say, signal strength larger than one threshold). The mobile device has no access to the first criterion and it only has partial information to infer the handoff decision logic.

Consider most carriers are reluctant to provide public access to network-side (usually proprietary) handoff information. In this work, we focus on the study from the device perspective. Namely, our model is used to infer possible MM misbehaviors primarily based on the limited information available on the device side. As a result, we divide the active-state handoff model into the observable part (on the radio access) and the unobservable one (on the network-side). The observable one uses the radio criteria based on measured signal strength and network preferences, which are similar to the idle-state handoff criteria. The unobservable one models the network-side decision logic. So we have the active-state handoff modeled as

$$t = \Omega_s(G_s, O_s), \text{ iff } \begin{cases} t = \Omega_s^{(radio)}(G_s, O_s) \\ t = \Omega_s^{(network)}(G_s, O_s) \end{cases}. \quad (2)$$

The $\Omega_s^{(radio)}(G_s, O_s)$ takes the same form as the idle-state handoff. For example, we observe that each candidate cell has to meet the radio quality requirement (here, >-106dBm) for load balancing [24].

Note that the radio criteria only partially determine the handoff result. Namely, they serve as the necessary but *not sufficient* conditions in the active-state handoffs whereas they are the necessary and sufficient conditions in the idle-state one.

Distributed handoff process. Finally, we put them together and model the whole handoff process. It is represented as an *iterative* one each with a transition from the serving cell to the target one. At each iteration, the target cell is determined by the current handoff decision logic, with the tunable parameters and runtime observations as its inputs. It can be performed by the current serving cell or the user device during the active or idle state. For each iteration, there are two possible outcomes. (1) If $t \neq s$, the serving cell switches to t at the next iteration and the handoff process continues. (2) Otherwise, if $t = s$, the handoff process stops unless the environment (through observations) varies and triggers another handoff procedure. In short, the handoff process can be denoted by the following sequence of serving cells.

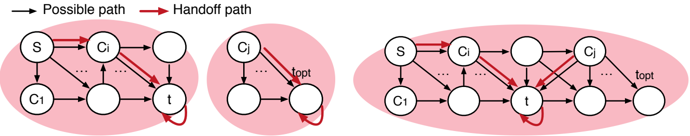
$$s \xrightarrow{\Omega_s} c_1 \xrightarrow{\Omega_{c_1}} c_2 \xrightarrow{\Omega_{c_2}} \cdots c_i \xrightarrow{\Omega_{c_i}} \cdots \rightarrow t, \quad c_i, t \in C \quad (3)$$

We assume that the handoff process converges to a target cell t. The non-convergence problem has been investigated in [24], [25]. The desired reachability is violated when the convergence may not settle down at the desired target cell. Given certain network conditions, a target cell is usually designated as the one that yields best performance by the operator or the user. Let t_{opt} be the desired target from all the candidate cells. It satisfies that $t_{opt} = \arg\max_{c \in C} \Phi(c)$, where $\Phi(c)$ represents the performance metric of our interests. This represents a globally optimal choice regardless whether it is feasible through the distributed, iterative handoff process.

Desired reachability states that (1) the handoff process converges to a target cell t and (2) $t = t_{opt}$. Therefore, undesired reachability implies that

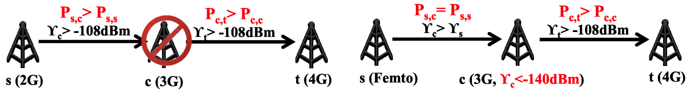
$$\begin{cases} s \xrightarrow{\Omega_s} \cdots c_i \xrightarrow{\Omega_{c_i}} \cdots \boxed{\mathbf{t} \xrightarrow{\Omega_t} \mathbf{t}}, & c_x, t \in C, \\ t \neq t_{opt}, & t_{opt} = \arg\max_{c \in C} \Phi(c) \end{cases}. \quad (4)$$

One thing noting worth is that our modeling settings strive to be as simple as possible, if not overly simplistic in some cases, while still capturing the essence and neglecting secondary details. In particular, this model take no account into the timing issue (how long the handoff takes) and the handoff cost (how much radio and network resource consumption). We assume that each handoff always succeeds once the decision is made. It turns out that these factors will not change the structural property on reachability (only the damage of unreachability).



(a) convergence split (b) Premature convergence

Figure 3: Two categories of undesired convergency.



(a) Relay cell unaccessible (b) Weak relay cell

Figure 4: Two instances of convergence split due to missing configurations.

In reality, there are few or even only one iteration(s) in Equ. (4). The best handoff is expected to directly switch to and settle down at the desirable cell in one iteration It indeed holds true in most cases but our study also discloses certain misbehaviors.

B. Analysis: Classification of Undesired Reachability

In principle, there are two classes of undesired convergence, as illustrated in Figure 3.

○ *Convergence split.* In the first category, the convergence depends on the initial serving cell. The sequence of handoffs for the given device does converge but settles down at a cell other than the desired target because there is no path from s to t_{opt} (Figure 3a). Let us use a directed graph to represent all possible handoff transitions. The problem here is that the initial cell and the target cell exist in two isolated graphs so that t_{opt} is unreachable no matter how the handoff take places.

○ *Premature convergence.* In the second category, the convergence is independent of the initial serving cell. Theoretically, there exists a path from s to t_{opt} (Figure 3b); However, the actual process for the given device is either unable to reach the desired target or stops early before it reaches the target.

Both fail to achieve the expected goal which should be avoided. We further deduce their root causes. It turns out that they are caused by misconfigurations and inappropriate device-network coordinations. In other words, they are rooted in the fundamental conflicts or implementation glitches, regardless of dynamic network environments.

We further uncover three concrete categories, concerning the quality of convergence.

○ *C1: Unaccessible intermediate cells due to missing configurations.* In this case, the handoff process prematurely stops before reaching the target cell because of missing configurations. Basically, it is identified through checking whether the initial cell and the target one lie in two isolated directed graphs (independent sets).

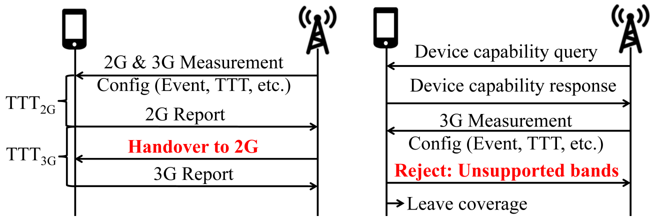
Figure 4a shows an example validated in the real trace. The device initially stays in an area with only 2G coverage, but later moves into a new spot with both 2G and 4G coverage. However, the device does not move to 4G as expected. Despite strong radio coverage from 4G, the device gets stuck in 2G. This problem has been repeatedly reported by users [18], [28], but its root cause is not disclosed.

Our trace analysis shows that, the 2G cell does not configure a local handoff rule to 4G, but only has a handoff rule to 3G. However, in no presence of a 3G cell, the 2G cell cannot hand over the device to the 4G cell. Therefore, the root cause is that the 2G cell lacks proper handoff configurations for the 4G cell. The issue arises in practice possibly because 2G has been phasing out and the operator mainly focuses on deploying new 3G or 4G cells. When new cells are in operation, old 2G cells do not have the configuration update. The intermediate 3G cell in the 2G cell configuration can be inaccessible for various reasons. The user device has radio compatibility issue to access the 3G cell (*e.g.*, it only supports certain 3G technology such as TD-SCDMA, but not others), or the device's signal strength to the 3G cell is too weak.

We observe another similar issue caused by missing configurations but among Femtocell, 3G and 4G cells. The device is trapped in the current cell since it does not have any configuration that is capable of reaching the target. In Figure 4b, the device becomes out of service once moving outside the 3G femtocell coverage, despite the existence of a 4G cell. The root cause is that, the 3G femtocell has no configuration rule to the 4G cell, but only has the rule to a 3G public cell. When a 3G cell is not accessible (here, 3G is extremely weak), the migration to 4G (via the intermediate 3G cell) is infeasible. The device is thus stuck at out-of-service in this case. This femtocell deployment indeed follows the common guideline, which suggests the femtocell to be deployed with weak macrocell coverage [34]. Unfortunately, here such guideline would still trigger this problematic instance.

It reveals a practical challenge that mobile networks are facing. Not all the cells have a direct path to any other cells and the reachability from s to e has to depend on the intermediate cell (here, 3G). However, the existence of intermediate cells are not guaranteed. The unpleasant consequence is that the big investment on advanced technology (here, 4G) goes futile due to 2G's configuration glitch. The blame can be that 2G or 3G Femtocells lack proper configurations to 4G. However, it is not without rational. There was no 4G when 2G was deployed and the 2G infrastructure is likely not updated to date due to heavy cost (possibly retire soon). Femtocells may be configured so under the premise that 3G has been largely deployed. With versatile access technologies and rich options (different frequency bands and small cells), it is not guaranteed that each cell has a direct path to all possible cells. Mobile networks should be painstaking on their decision procedure or rigorous on their infrastructure deployment or both.

○ *C2: Blocked decision by others.* This category belongs to the first class where the desired target cell is ideally reachable. However, the convergence process to the target cell may also halt when it is disrupted by another candidate cell. It implies that the problem lies in the order in making the decision. The undesired cell is chosen first and thus blocks the chance to selecting the desired one. Basically, it is identified



(a) Blocked decision (b) Problematic coordination between device and network

Figure 5: Two instances of premature convergency.

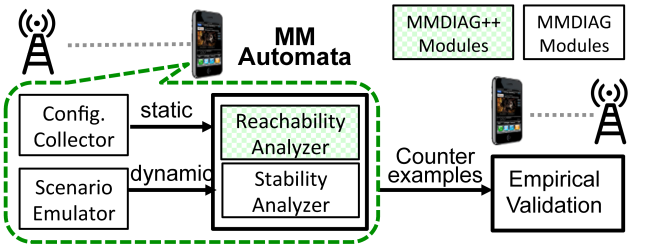


Figure 6: The `MMDIAG++` architecture. The earlier version of `MMDIAG` is developed in [25].

through a reachability analysis over the directed graph. Given the initial cell, decision logic Ω_s, parameter configurations G_s and runtime measurements O_s, we replay the handoff procedure and obtain the time order of each result. It might be problematic once the undesirable one happens first.

Figure 5a shows such a real-world scenario. The user device is at the active state and about to leave its 4G serving cell (here, 4G). The new location has both 2G and 3G cells, but these cells cannot reach each other. To initiate the handoff decision, the serving cell asks the device to measure and report signal strengths from both 2G and 3G cells. For each candidate cell, the 4G serving cell configures the device with (1) the report criteria; (2) the measurement duration `TTT` (TimeToTrigger) to ensure stable measurements. The problem arises when both 2G and 3G signal strengths are good. If the serving cell uses the first-come-first-serve (FCFS) strategy and the device reports 2G first, the serving cell may immediately hand over the device to 2G, without waiting the device to finish its 3G measurement. Given the good radio quality from the 2G cell, handoff to 2G is activated. A premature convergence to 2G occurs, thus ruling out the desired handoff to the 3G cell.

The root cause lies in improper coordinations between the network and the device. The network acts as the master to control the device (the slave) to conduct measurements for the handoff. However, its FCFS response to the device reports does not work well with the device which has freedom to conduct its measurements of candidate cells in any order. In this case, both the user and the network have their valid reasons. The serving cell wants to expedite the handoff decision to minimize the handover latency, whereas the device decides its own order for measurements since it does not know the decision logic at the serving cell. However, it turns out that both get penalized.

◦ *C3: Trapped due to problematic, device-network coordination.* We also uncover that premature convergence can be caused by problematic coordinations between the network and the device.

Figure 5b shows a real scenario. The 3G cell supports multiple frequency bands, but the device supports only one of them (a common case since many phone models cannot support all). Without taking into account the device's capability, the serving cell requests the user device to monitor all 3G frequency bands. Upon this request, the device rejects this command, even though it can still access some bands. No measurements would be conducted by the device thereafter. The serving cell could not initiate any handoff without measurement reports. If the user also leaves the current serving cell, the device loses its network access.

IV. Empirical Study on Desired Reachability

In this section, we present our tool to detect undesired reachability and empirical assessment in two top-tier US carrier networks using this tool.

A. `MMDIAG++`: In-Device Automatic Detection Tool

With above analytical findings, we next design and implement `MMDIAG++`, an in-device diagnosis tool to detect and validate undesired reachability in handoff. This tool is built on top of `MMDIAG`, which was previously developed for instability detection [25]. Given the configurations from cells at a location, our tool reports handoff configuration conflicts that may incur undesired reachability and uncovers their root causes.

We take the device-based approach, since the carriers are reluctant to provide public access to their mobility management configurations and runtime information for handoff decisions. Our approach is deemed a viable solution, because we can leverage the signaling exchanges to bypass this major constraint. The underlying premise is that, the serving cell has to send their main parameters and decision logics to the device. Its effectiveness has been validated in our previous work [25].

Figure 6 plots the architecture of `MMDIAG++`. Following the design of its predecessor `MMDIAG`, it is still divided into two phases: detection and validation. The core of the detection phase is an MM automata which models the MM decision logic based on the 3GPP standards (elaborated in §III-A). We feed this model with real configurations collected directly from the device and indirectly from the serving cell, as well as dynamic environment settings created for various scenarios. `MMDIAG++` then run model checking to first ensure the handoff convergency (via *stability analyzer*) and then compare it with the desired target (via *reachability analyzer*). Once undesired convergence is found, we move to the second phase for device-based validation. For each counterexample, we set up the corresponding experimental scenario and conduct measurements in operational networks for validation.

`MMDIAG++` reuses four `MMDIAG` modules (*configuration collector, scenario emulator, stability analyzer* and *validation*) and devises one new module (*reachability analysis*) and upgrades the tool for in-device use. We briefly introduce how common modules work (details in [25]) and elaborate on new components.

- *Configuration collector* retrieve parameters from the signaling messages exchanged between the serving cell and the device. We log signaling messages through MobileInsight [1], an in-device cellular signaling collector developed by us. This acts like QXDM [2] and XCAL [30], proprietary software used by professionals to record message exchanges over the air.
- *Scenario emulator* is based on the MM automata. In particular, we create runtime scenario parameters (*e.g.*, radio signal strength and traffic loads) and feed them into the MM model. We enumerate all the options when the number is limited and sample them if unlimited.
- *Stability analyzer* is to check whether the handoff converges. With handoff configurations and scenario observations as input, it enumerate the possible handoff transitions and examines the convergence rules.
- *Reachability analyzer* is built on top of the stability analyzer. Its core role is to compare the converged cell and other candidates and infer whether two problematic scenarios (convergence split and premature convergency) might occur. If so, it outputs the counterexamples.
- *Empirical validation* is to construct test scenarios, run experiments, collect real traces, and confirm whether the identified problems appear, given the hints from the counterexample.

MMDIAG++ pushes detection online. Compared with MMDIAG, all the modules are developed in the device side so that it can facilitate measurement and diagnosis in the wild.

B. Experiments Over Operational Carrier Networks

We run the designed tool to validate undesired convergence in two top-tier US carrier networks (denoted by OP-I and OP-II). We run experiments in two metropolitan cities: Los Angeles in the west coast and Columbus in the midwest.

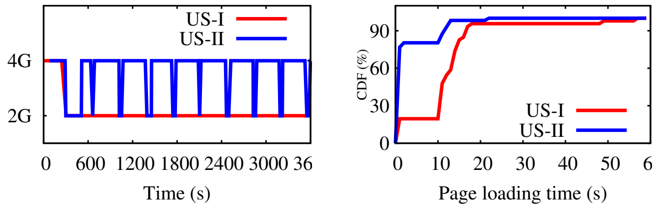
We conduct both outdoor and indoor experiments. The outdoor experiment covers 63 different locations over 240 km^2 in the west coast and 260 km^2 in the east coast. We also collect information on indoor experiments at 50 spots in two 8-floor office buildings and one apartment. In this indoor setting, we mainly collect the radio quality observations at various spots, since most cells, as well as their configurations, are similar across locations. We deploy four 3G Femtocells in office and at home for indoor tests. We use four phone models: Samsung Galaxy S4, S5 and Note 3, and LG Optimus G. The results are similar for all phone models.

We collect all cells' active and idle-state handoff decision profiles, as well as their measured radio quality assessments. This is used to feed MMDIAG++ and test if their handoff decisions may violate the reachability conditions. Once a violation is identified, we perform more tests under this scenario to quantify the impacts.

Table III summarizes the outdoor test settings. The cell distribution at different outdoor locations confirms that today's deployment is quite dense and hybrid. At most locations, there are about 8–16 cells. On average, there are about 11 cells in OP-I and 10 cells in OP-II. The number of unique cells, excluding those observed at multiple locations, are 275 (4G: 120, 3G: 97, 2G: 58) in OP-I and 222 (4G: 92, 3G: 66, 2G: 64) in OP-II. It confirms that 4G cells have smaller coverage and denser deployment whereas the 2G coverage is much larger. The indoor setting has similar cell density as the outdoor one. The results in OP-II are similar and thus omitted.

	Avg. cell#/spot		Unique cell#	
	OP-I	**OP-II**	**OP-I**	**OP-II**
#4G	2.6	2.1	120	92
#3G	3.4	2.4	97	66
#2G	5.4	5.6	58	64
#All	11.4	10.1	275	222

Table III: Statistics of outdoor cell deployment.



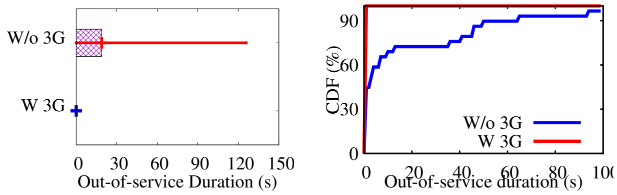
(a) Log of serving cells (b) Webpage loading time

Figure 7: Log and performance in the missing-configuration case (C1) where the phone gets stuck in 2G when 4G is available.

We observe all four instances in reality through this tool and validate the effectiveness of MMDIAG++.

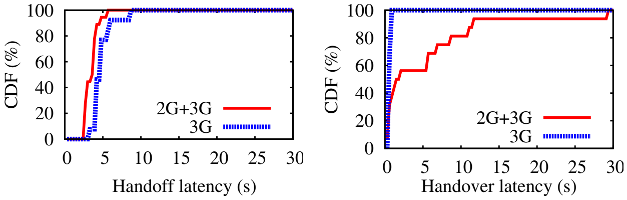
○ *Fail to reach 4G from 2G (C1).* Due to missing configuration in 2G cells, the device may not reach 4G in some areas with weak/no 3G coverage. We examine how likely the problem happens in reality. Among 63 locations we tested, none of the 2G cells have the idle (and active) state handoff rules to 4G in OP-I. In OP-II, all 2G cells are observed to have idle-state handoff rules to 4G, but no active-state handoff rules. We discover that 2G is deployed in all locations in both carriers. But in OP-I there exist 5 out of 63 locations with 2G and 4G, yet with 3G's signal strength less than -105dBm.

It hurts user experience since 2G is slower than 4G. We run the webpage browsing test for 20 times. we use Firefox to fetch the webpage (www.cnn.com) every 1min. Figure 7a shows the cell the device is associated with in a 1-hour test. In OP-I, once the first call is made, the phone gets stuck in 2G afterwards. In OP-II, the phone can switch back to 4G after the voice call. The minimal switch time is 30s, and the maximum switch time is 253s. Figure 7b shows the page loading time in two carriers. In OP-I, except before the first call is made, the user device's page loading suffers from 2G's low data rate. The average loading time is 15.4s. In OP-II, the average loading time is 3.7s. Depending on whether in active state or not, the phone in OP-II may still suffer from low-rate 2G temporarily. 2G slows down by 35.8x on average (*i.e.*, 15.4s for 2G, and 0.4s for 4G).



(a) Histogram (b) CDF

Figure 8: Duration of out-of-service time in case the device moves from the femtocell coverage to a 4G one (C1).



(a) OP-I (b) OP-II

Figure 9: Active-state handoff latency in OP-I and OP-II in the 2G-blocking-3G case (C2) .

○ *"Out of service" when moving to 4G (C1).* We observe this problem when a phone is about to leave the femtocell and moves to an area with 4G. We find that all four femtocells have no direct handoff rule to 4G. This problem thus happens once the femtocell is deployed in areas with no or weak 3G. We observe that 5 of 63 areas have 2G and 4G without 3G. We quantify the impact through a comparison experiment with/without 3G. We deploy a femtocell at two indoor places: one without 3G coverage, while the other with 3G signal strength in (-80dBm, -90dBm). We place the phone at the coverage boundary of the Femtocell, and record the switching time from the femtocell to 4G. Figure 8 shows the result. With 3G, the device works well; without public 3G, the phone may be out of service up to 125.8s (25 seconds on average). This is because the device has to scan all frequency bands to find 4G after the device loses its femtocell access. The handoff fails.

○ *3G blocked by 2G (C2).* We observe that the handoff selects 2G rather than 3G in both carriers, even though both 2G and 3G show satisfactory signal strength based on serving cell's measurement criteria. Our outdoor tests show that, there are 60 out of 63 locations (95.2%) in OP-I and 100% locations in OP-II satisfy this condition. In OP-I, its active-state handoff decision is always responsive to the first message. However, when both 2G and 3G cells satisfy the measurement report criteria, all the tested phones choose to report 2G first. So the phone hands over to 2G with 100% probability even when 3G is available. In OP-II, the handoff decision may not be always responsive to the first measurement report. In our indoor test, the probability of handoff to 2G is 5.7%, whereas the probability to 3G is 94.3%.

We note that, OP-II does pay the cost of large handoff latency to alleviate 2G/3G blockage. Figure 9 shows the handoff latency in OP-I and OP-II at the same condition with 2G+3G and 3G only (by manually disabling 2G on the device). The handoff in OP-II is delayed for about 1-12 seconds due to waiting for the 3G report. In the worst case, it is up to 30 seconds. The long latency arises when the 3G signal strength is not satisfactory, so the user device sends 2G reports only. Note that such long latency is not necessary. Based on serving cell's configuration, it takes up to 1.28s to complete the measurements of both 2G and 3G. Without receiving a 3G report after 1.28s, the serving cell knows that the 3G signal is weak and may stop waiting. Even worse, this delay may lead to service failure. We run voice calls (since data service in 2G is too bad) and find that the call drop ratio is 10.8% when 2G and 3G are enabled in OP-II. In contrast, no call would be dropped if only 3G is enabled.

○ *"Out of service" when moving to 3G (C3).* We find that the problem also occurs in the setting of Figure 5b, when the device moves to a 3G area. This is because when the device moves out of a femtocell coverage to another area, the serving cell asks the device to monitor all 3G frequency bands but it is rejected by the phone, which fails to support all bands. Once the device moves away, no handoff would be triggered and the device will be 100% out of service. In our test, all phones are observed to have this issue.

V. Discussion

We now elaborate on several issues not fully covered in this work so far, and describe our recommended fixes.

Practical factors. In our modeling and analysis, we assume ideal handoff execution and invariant observations during each handoff iteration. Several practical factors are simplified for ease of the analysis. For example, transient fluctuations such as time-varying radio signal strength values are not considered (though they has been widely explored in literature, *e.g.*, [32]). Other practical issues are also largely ignored, including the handoff timing and overhead, handoff failures, the roaming speed, measurement inaccuracy, and implementation issues (e.g., we did observe that certain phone model may not follow the command from the serving cell), to name a few.

Desired convergence. We realize that it is challenging to determine the desired target cell in all scenarios. In this work, we select the target simply based on common wisdom, *e.g.*, 4G>3G>2G unless the preferred cell has weak radio quality. In principle, it depends on many factors including the cell type, radio quality, ongoing traffic, *etc.*. Other efforts may facilitate the proper choice, yet largely independent of our work.

Other properties. In addition to desired convergency, other structural properties such as convergence speed, robustness, and availability, are worth exploring. They are not considered in this paper and will be investigated in the future work.

To address the identified configuration issues, we recommend some fixes on the device side and on the network side.

Fix on the device. It is probably easier for the user to apply quick fixes on his/her device. The phone is not only the device that interacts with the serving cell and all available candidates, but also the entity that performs handoffs and suffers from undesired convergence. The user thus has incentives to apply the fix.

The user device can act as an implicit controller for three functions. First, it runs self checking. It thus verifies whether the handoff configuration for each cell satisfies the desired reachability condition in §III-B. If not, the device may elect to not honor such configurations from the cell, thus avoiding undesired convergency. Second, it can record the available and desired choices in the recent past. When the serving cell is not the desired one, it probes more on its own (thus not being restricted by the instructions from the serving cell). Third, the device can leverage crowd-sourcing to retrieve problematic areas and suggested serving cells reported by others. These functions can be implemented as part of the functions on the chipset. The downside of this solution is to raise computation overhead at the device side. More computation and communication is required from top to down. Another limitation of the device-side fix is that, without assistance from the network side, the phone may not have complete information (*e.g.*, active-state handoff decision) or cannot control the network actions (*e.g.*, which report(s) to respond and the order). It raises another possible downside that uncoordinated behaviors between the phone and the network may impede network optimization in some cases (*e.g.*, the phone rejects to obey the decision made by the serving cell for load balancing).

Network-side approach. We also recommend two fixes to the network. First, the network deploys a centralized controller, which collects and coordinates the handoff decision functions and configurations among cells. This is a long-term solution which is aligned with 5G trends [14]. Second, the network corrects common misconfigurations identified in our work. For example, it should add one handoff rule to 4G at those colocated 2G cells. It also needs to remove those inconsistent preference settings over femtocells at 3G and 4G cells, both of which should prefer to femtocells or have equal preference.

VI. Conclusion

Mobility management is a key utility function offered by 3G/4G cellular networks. Like all operational networks, mobile carriers allow for flexible handoff configurations to realize versatile handoff policies. However, this management-plane aspect on mobility has been largely overlooked by past research efforts. This work, following our previous efforts, continues to make a study of mobility management configurations toward high-quality handoff convergency. Our study discloses that mobile devices may fail to reach the desired serving cell (*e.g.*, 2G when 4G/3G available or temporally out of service). In the broader context, our study moves beyond the current focus on both data and control planes. Management plane of 3G/4G networks (likely also the upcoming 5G) is still a wide-open research area and deserves more attention.

References

[1] Mobileinsight project. http://metro.cs.ucla.edu/mobile_insight.

[2] QUALCOMM eXtensible Diagnostic Monitor. http://www.qualcomm.com/media/documents/tags/qxdm.

[3] 3GPP. TS25.331: Radio Resource Control (RRC), 2006.

[4] 3GPP. TS 23.216: Single Radio Voice Call Continuity (SRVCC), 2011.

[5] 3GPP. TS23.009: Handover Procedures, 2011.

[6] 3GPP. TS23.401: GPRS Enhancements for E-UTRAN Access, 2011.

[7] 3GPP. TS23.272: Circuit Switched (CS) fallback in Evolved Packet System (EPS), 2012.

[8] 3GPP. TS25.304: User Equipment (UE) Procedures in Idle Mode and Procedures for Cell Reselection in Connected Mode, 2012.

[9] 3GPP. TS36.331: E-UTRA; Radio Resource Control (RRC), 2012.

[10] 3GPP. TS23.261: IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2, 2014.

[11] 3GPP. TS25.367: Mobility procedures for Home Node B, 2014.

[12] 3GPP. TS32.500: Self-Organizing Networks (SON); Concepts and requirements, 2014.

[13] 3GPP. TS36.304: E-UTRA; User Equipment Procedures in Idle Mode, 2015.

[14] N. Alliance. NGMN 5G White Paper, 2015.

[15] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting mobile 3g using wifi. In *ACM MobiSys*, June 2010.

[16] C. Brunner, A. Garavaglia, M. Mittal, M. Narang, and J. V. Bautista. Inter-system Handover Parameter Optimization. In *VTC Fall*, 2006.

[17] M. Z. Chowdhury, W. Ryu, E. Rhee, and Y. M. Jang. Handover between Macrocell and Femtocell for UMTS Based Networks. In *IEEE ICACT*, 2009.

[18] A. S. Communities. iPhone 5 Gets Stuck on EDGE Network. *https://discussions.apple.com/thread/5113660*.

[19] W. Dong, S. Rallapalli, R. Jana, L. Qiu, K. Ramakrishnan, L. Razoumov, Y. Zhang, and T. W. Cho. ideal: Incentivized dynamic cellular offloading via auctions. *TON*, 22(4):1271–1284, 2014.

[20] H. Fathi, R. Prasad, and S. Chakraborty. Mobility management for voip in 3g systems: evaluation of low-latency handoff schemes. *Wireless Communications, IEEE*, 12(2):96–104, 2005.

[21] D. Flore, C. Brunner, F. Grilli, and V. Vanghi. Cell Reselection Parameter Optimization in UMTS. In *Wireless Communication Systems*, 2005.

[22] W.-Y. Lee and I. F. Akyildiz. Spectrum-aware mobility management in cognitive radio cellular networks. *Mobile Computing, IEEE Transactions on*, 11(4):529–542, 2012.

[23] L. Li, K. Xu, D. Wang, C. Peng, Q. Xiao, and R. Mijumbi. A Measurement Study on TCP Behaviors in HSPA+ Networks on High-speed Rails. In *INFOCOM*, April 2015.

[24] Y. Li, H. Deng, J. Li, C. Peng, and S. Lu. Instability in distributed mobility management: Revisiting configuration management in 3g/4g mobile networks. In *ACM SIGMETRICS*, 2016.

[25] Y. Li, J. Xu, C. Peng, and S. Lu. A First Look at Unstable Mobility Management in Cellular Networks. In *HotMobile*, Feb 2016.

[26] M. Liu, Z. Li, X. Guo, and E. Dutkiewicz. Performance Analysis and Optimization of Handoff Algorithms in Heterogeneous Wireless Networks. *IEEE Transactions on Mobile Computing*, 7(7):846–857, July 2008.

[27] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan. Coordinating Handover Parameter Optimization and Load Balancing in LTE Self-Optimizing Networks. In *VTC Spring*. IEEE, 2011.

[28] MacRumors. Stuck in Edges. *http://tinyurl.com/zzy2h7u*.

[29] J. McNair and F. Zhu. Vertical handoffs in fourth-generation multinetwork environments. *Wireless Communications, IEEE*, 11(3):8–15, 2004.

[30] Mediatek. Xcal-mobile. http://www.accuver.com.

[31] C. Paasch, G. Detal, F. Duchene, C. Raiciu, and O. Bonaventure. Exploring mobile/wifi handover with multipath tcp. In *Proceedings of ACM SIGCOMM Workshop on Cellular Networks (CellNet)*, 2012.

[32] G. P. Pollini. Trends in Handover Design. *IEEE Communications Magazine*, 34(3):82–90, 1996.

[33] F. P. Tso, J. Teng, W. Jia, and D. Xuan. Mobility: A Double-Edged Sword for HSPA Networks: A Large-Scale Test on Hong Kong Mobile HSPA Networks. *IEEE Transactions on Parallel and Distributed Systems*, 23(10):1895–1907, 2012.

[34] Wikipedia. Femtocell. http://en.wikipedia.org/wiki/Femtocell.

[35] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis. Mobility management for femtocells in lte-advanced: key aspects and survey of handover decision algorithms. *Communications Surveys & Tutorials, IEEE*, 16(1):64–91, 2014.

[36] V. Yazıcı, U. C. Kozat, and M. Oguz Sunay. A new control plane for 5g network architecture with a case study on unified handoff, mobility, and routing management. *Communications Magazine, IEEE*, 52(11):76–85, 2014.