Peeking Over the Cellular Walled Gardens - A Method for Closed Network Diagnosis -

Byeongdo Hong, Shinjo Park, Hongil Kim, Dongkwan Kim, Hyunwook Hong, Hyunwoo Choi, Jean-Pierre Seifert, *Member, IEEE*, Sung-Ju Lee, *Fellow, IEEE*, and Yongdae Kim, *Senior Member, IEEE*

Abstract—A cellular network is a closed system, and each network operator has built a unique "walled garden" for their network by combining different operation policies, network configurations, and implementation optimizations. Unfortunately, some of these combinations can induce performance degradation due to misconfiguration or unnecessary procedures. To detect such degradation, a thorough understanding of even the minor details of the standards and operator-specific implementations is important. However, it is difficult to detect such problems, as the control plane is complicated by numerous procedures. This paper introduces a simple yet powerful method that diagnoses these problems by exploiting the operator-specific implementations of cellular networks. We develop a signaling collection and analysis tool that collects control plane messages from operators and finds problems through comparative analysis. The analysis process consists of three different control plane comparison procedures that can find such problems effectively. These individual procedures use a time threshold, control flow sequence, and signaling failure as the basis for comparison. To this end, we collect approximately 3.1 million control-plane messages from 13 major cellular operators worldwide. As a case study, we analyze the circuit-switched fallback technology that triggers generation crossover between third generation and long-term evolution technologies.

Index Terms—Cellular network diagnosis, performance degradation, Control plane, LTE, CSFB

1 Introduction

Cellular networks have been constantly evolving since the deployment of the analog first-generation (1G) network in 1983. We now enjoy not only wide coverage but also high speeds and low latency. Cellular networks are developed based on standards, and the 3rd Generation Partnership Project (3GPP) is the standard for all networks, from the Global System for Mobile Communications (GSM) to long-term evolution (LTE) technologies. Cellular standards, by nature, provide the operators with implementation freedom. Thus, each operator can create a "walled garden" [1], [2] in a unique manner, as each can employ different operational policies, network configurations, and implementation optimizations. However, this diverse environment can often cause unexpected problems in one network that do not arise in other networks, if operators do not configure their cellular networks carefully. It is difficult to uncover these problems through simple analysis of the standard documentation only, as these documents are extensive and do not necessarily specify all of the operational details. Furthermore, it is difficult to detect all performance problems and diagnose their exact causes using local measurements alone.

Effective diagnosis and optimization of cellular networks are ongoing challenges. Cellular operators are known to use local measurements or user feedback for performance diagnosis, and this trend is also observable in the cellular research community. Previous studies have revealed various problems in cellular networks, such as voice-data interference [3], instability in mobility management [4], problematic interactions in the control plane [5], and unreliability issues with voice over LTE (VoLTE) technology [6]. In those investigations, various stress-testing methods and local measurements were used to diagnose problems. However, significant time and effort were required, as several experiments were necessary to identify the point at which the problem occurred. In addition, diagnosing unnecessary control plane procedures to optimize a cellular network seems difficult. From the dataset used in this study (discussed below), we found that some operators constantly perform redundant control plane procedures, causing multi-second delays on every call. Existing diagnosis methods have failed to identify these problems.

In this paper, we introduce a novel diagnosis method that finds performance bugs by cross-checking cellular procedures among different operators. By employing this method in the study reported herein, we discovered performance issues that were not revealed by existing approaches. Our mechanism exploits certain details of cellular network implementation and operation. Cellular networks have complex structures and large numbers of specifications, and each operational network has proprietary implementation and operational policies. Interestingly, the individuality of

- *B. Hong, H. Kim, D. Kim, H. Hong, H. Choi, S. Lee, and Y. Kim are with KAIST, Daejeon 34141, Republic of Korea. E-mail: {byeongdo, hongilk, dkay, hyunwook.h, zemisolsol, profsj, yongdaek}@kaist.ac.kr.*
- *S. Park and J. P. Seifert are with Telekom Innovation Laboratories, Technische Universität Berlin, Berlin 10623, Germany. E-mail: {pshinjo, jpseifert}@sec.t-labs.tu-berlin.de.*

Manuscript received 25 May 2017; revised 24 Dec. 2017; accepted 28 Jan. 2018. Date of publication 12 Feb. 2018; date of current version 29 Aug. 2018. (Corresponding author: Yongdae Kim.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2018.2804913

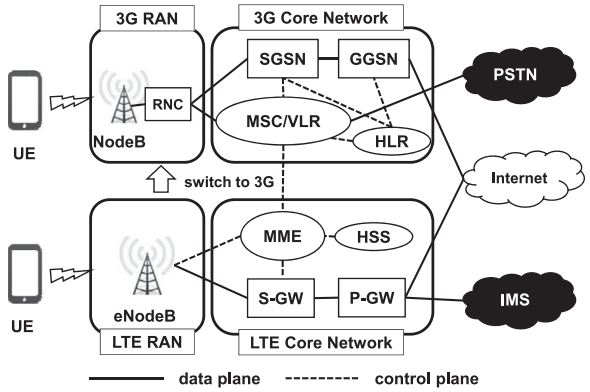


Fig. 1. Cellular network architecture.

each network can facilitate problem diagnosis and network optimization. The control flows of different operators for the same service can easily reveal differences in implementation policies and performance. In this study, the control plane procedures of different operators were compared based on this feature. To improve the comparison efficacy, we collected control plane messages from 13 major operators in seven countries.

As a case study, we investigated the control protocols for circuit-switched fallback (CSFB) technology [7], which triggers a generation crossover between third-generation (3G) and LTE technologies. CSFB technology includes most control procedures for 3G/LTE mobility management (MM), session management (SM), connection management (CM), and radio resource control (RRC). Because these complex procedures are merged, optimizing CSFB is more difficult than optimizing 3G or LTE individually. Moreover, generation crossover technology is essential for upcoming fifth-generation (5G) networks, which are expected to be commercialized in approximately 2020 [8]. At present, the access and core networks of LTE/5G are expected to be combined [9], [10]. Therefore, we chose CSFB technology for our case study; however, our method is also applicable to other technologies.

Our analysis consisted of automatic and manual analyses. For the automatic analysis, we developed a tool called the signaling collection and analysis tool (SCAT). SCAT consists of two parts: (1) *SCATm*, which archives the timings and call flows along with the error messages from the cellular network, and (2) *SCATa*, which compares the timings and call flows within and across different networks to detect problems. Using the SCAT results, we conducted manual analyses. That is, we compared and analyzed anomalous call flows using the 3GPP standards and, whenever possible, we interviewed cellular operators to share and confirm our findings.

Using our methodology, we discovered that different operators were experiencing different performance problems that had not been discovered previously. In detail, (1) three problems each existed for one unique operator, while (2) another three problems were prevalent for four or five operators; however, those operators had not detected and/or resolved those problems at the time of our study. Only one of the three problems in (1) was notable for being discovered previously [5]; in that case, the study [5] reported that the difficulty was due to *design problems* related to the 3GPP standard. However, we discovered that this problem existed in one network only! This example clearly demonstrates the need for a comparative study involving multiple networks. In that case, if other networks had been investigated, it would not have been concluded that the problem stemmed from the design of the 3GPP standards.

This paper makes the following contributions:

- It introduces a novel, simple, and effective measurement-based diagnosis method for cellular networks. Using this method, we found six problems (including five new ones) and six new causes of cellular performance degradation in this study. Three of these problems (Sections 4.1, 5.1, 5.2) are caused by misimplementation/configuration and the other three (Sections 4.2, 6.1, 6.2) are related to optimization.
- Using SCAT, we collected and analyzed approximately 3.1 million control-plane messages (17,710 calls) from 13 major cellular operators in seven countries. We plan to release both this dataset (with operator permission) and our tool for other researchers.
- Our analysis of the global dataset shows that different operators use different implementations, some of which cause various degrees of performance degradation. We discovered examples that are difficult to find through local measurements, such as cases in which more than 50 percent of the calls had 0.5 s median delays for several operators and a significant fraction of users experienced more than 1 s median delays during generation crossover.

2 Preliminaries

This section reviews the overall architectures of the 3G and fourth generation (4G) LTE networks. We then review related works.

2.1 Cellular Background

Cellular network architecture. A cellular network consists of two architectural components: a radio access network (RAN) and a core network (CN). Fig. 1 depicts the 3G and LTE network architectures. RAN refers to a wireless network connecting an item of user equipment (UE) to the CN through a base station, e.g., evolved NodeB (eNodeB) in LTE and NodeB in 3G. The CN supports cellular services such as data and voice calls by connecting to the Internet, the public switched telephone network (PSTN), or the Internet Protocol (IP) multimedia subsystem (IMS). As shown in Fig. 1, the specific components of the RAN and CN differ for each generation. For a RAN, the radio network controller (RNC) controls a group of NodeBs in 3G, while both the base station and its controller are combined into the eNodeB in LTE.

The CNs in 3G and LTE differ significantly, based on how they deliver data. The network domains for 3G are separated into the packet-switched domain for the Internet and the circuit-switched domain for voice calls. Gateways for packet-switched Internet connection consist of serving general packet radio service (GPRS) support node (SGSN) and gateway GPRS support node (GGSN). On the other hand, LTE uses only the packet-switched domain for both voice and data. The LTE gateways consist of serving gateway (S-GW) and packet data network (PDN) gateway (P-GW). For 3G, mobility

management and user authentication are handled by both the mobile switching center (MSC) and SGSN, using a visitor location register (VLR) and home location register (HLR). For LTE, a mobility management entity (MME) is used in conjunction with a home subscriber server (HSS). To support LTE-3G generation crossover, the MME is connected to the MSC and SGSN for voice and data, respectively.

Troubleshooting and optimization in cellular networks. Operators and manufacturers implement cellular networks based on standards and policies. The network service is stabilized and optimized based on a field test or troubleshooting [11], [12]. Troubleshooting involves performance of a number of tests to detect network failures, e.g., the *out-of-service* condition. This diagnostic method is relatively simple compared to the optimization method because the target problem is exposed during testing. On the other hand, service optimization is difficult, because unexposed problems must be detected. For service optimization, unnecessary procedures must be found among the many control plane procedures, and the optimal arrangement of the normal procedures must be considered. However, this task is quite difficult, because it requires an overall understanding of the relationships between the protocols and implementation policies.

Each service provider attempts to find the delay factors via time comparison, e.g., by comparing times required for call setup according to changes in various settings (e.g., the CSFB call method and CM/RRC state) related to the UE or network [13], [14]. To reduce the delay, the network timer settings or parameters are adjusted [15]. However, this optimization strategy, which changes the configuration of a single implementation, cannot detect problems such as unnecessary procedures (Section 6) or inefficient control plane procedure arrangements (Section 4.2). For performance optimization, it is necessary to compare the different implementations of the cellular network.

Circuit switched fall back. Because LTE operates via packet switching for both voice and data, cellular operators must deploy VoLTE, i.e., an implementation of voice over IP, on LTE networks. As VoLTE is still in the early stages of deployment, the 3GPP specifies CSFB, which utilizes legacy circuit-switched calls through generation crossover between 3G and LTE.

Upon receiving a CSFB call, the serving base station switches the UE to another generation such as 3G. The serving base station requests a generation crossover to the target RAN system through the MME (in LTE). Once the request is accepted, the serving base station starts to switch the UE to the target network. The UE then configures the radio control and data link for the target RAN accordingly. After the UE connects to the target 3G network, it updates the quality of service parameters and security contexts of the target network, and releases all resources from the previous network. The crossover procedure in the reverse direction (from 3G to LTE) is omitted for brevity.

2.2 Related Work

Problem diagnosis in cellular networks. Problem diagnosis in commercial cellular networks is known to be difficult. Further, problem detection via user-level analysis is especially difficult, because messages between CN components are invisible at the user end. Nonetheless, performance problems in cellular networks have been examined in a few studies. Tu et al. [3] determined a relation between voice and data, showing that a CSFB call can break LTE connectivity or degrade transmission control protocol performance. Further, Tu et al. [5] conducted a cause analysis of the *out-of-service* issues occurring during CSFB calls. Jia et al. [6] reported user experience problems for VoLTE, such as muting during a voice call, and Li et al. [4] identified issues related to instability in mobility management.

The above works focused on finding problems for specific regions or operators only (mostly in the U.S.). However, diagnosis generalization based on local measurements may yield incorrect conclusions. For example, Tu et al. [5] claimed that elimination of the 3G context can delete the LTE context (causing LTE to become unavailable) during CSFB, because of a faulty standard design. However, our measurements show that most operators (10 of the 13 examined operators) do not experience the *out-of-service* problem. Furthermore, we discovered that LTE context elimination can be caused by other factors, such as time-related misconfiguration of the MME handover or security context mapping errors (see Sections 4.1 and 5.2). We found that a problem previously claimed to be an *LTE design fault* was, in fact, triggered by *implementation or configuration errors*. Therefore, we performed a comparative study of data collected from 13 different operators. We argue that such a comparative study is necessary for problem diagnosis in cellular networks, to avoid misleading conclusions.

Signaling analysis tools for cellular networks. As activities on the mobile control plane are not directly visible to users, various baseband-specific signaling[1] analysis tools and libraries have been developed. Among them, OsmocomBB was designed for second-generation (2G) technology only, and xgoldmon [16] can monitor signaling messages in 2G/3G for the Intel baseband only. A tool developed by P1 Security [17] supports LTE but was intended for the Samsung LTE data stick only; thus, voice support is not provided. Spaar has discussed LTE monitoring in the Qualcomm baseband [18], and the SnoopSnitch app [19] and MobileInsight [20] both facilitate such monitoring. Several libraries and tools for detecting malicious activities and anomalous behaviors targeting the baseband have been developed. For example, SnoopSnitch also includes cellular information leakage detection for phones with a Qualcomm baseband, being capable of detecting silent short message services (SMSs) and other hidden activities. Darshak [21] and the Android IMSI Catcher Detector (AIMSICD) [22] can also detect malicious activities on the control plane in the Intel baseband.

Our tool, SCAT, focuses on detecting problems in cellular networks by examining call flows, statistical data, and error messages from operators. SCAT works on both the Qualcomm and Samsung basebands, which comprise 86 perc ent of the market share [23].

3 Analysis Overview

The control plane of a cellular network manages the system configurations of the access and core networks and consists of various protocols and procedures. As a service targets the UE, some control plane flows can be observed at the UE.

1. Control plane message.

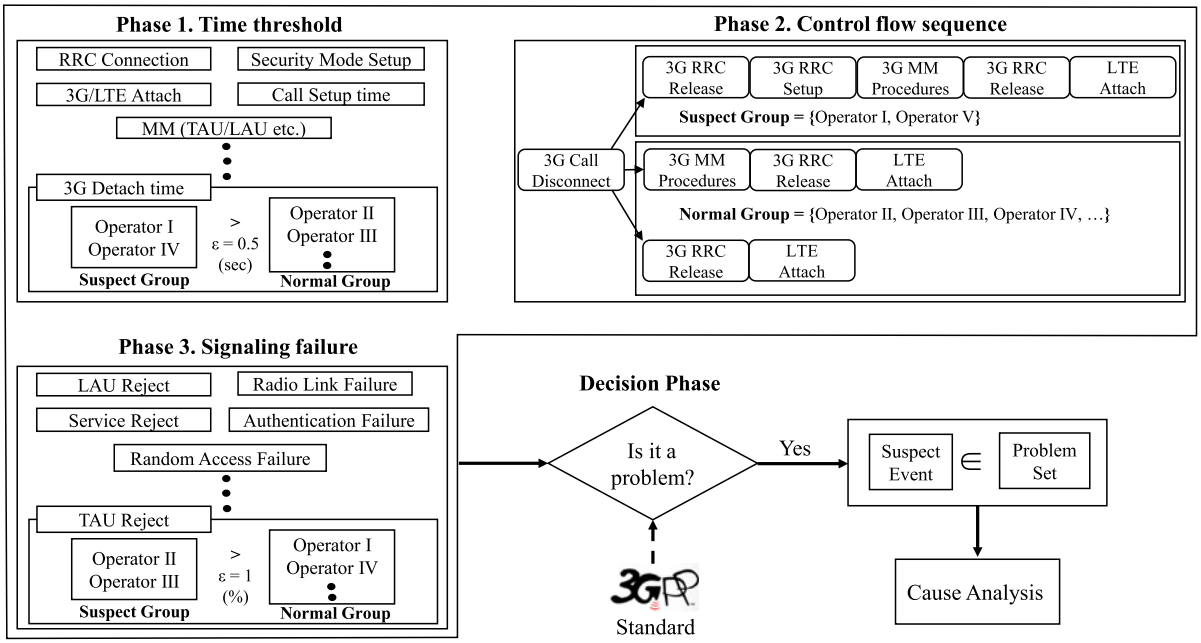


Fig. 2. Data analysis process.

In previous studies [3], [4], [5], [6], problems were diagnosed by examining experimental results for specific settings. These approaches required identification of the particular problematic situations, which required significant effort. Further, these approaches are inappropriate for discovering optimization issues degrading the user experience, as they usually cause service delays.

Cellular networks are often called "walled gardens [1], [2]," as cellular vendors and operators do not share their implementation and configuration details. While this individuality may cause problems in some networks, it motivates the use of comparative methods for problem diagnosis. In this paper, we describe how problems can be diagnosed simply and effectively using our method and the user-side control plane message dataset. As noted above, we chose CSFB as a case study for the cellular network fault diagnosis. CSFB is a complex technology that combines generation crossover (2G/3G and LTE) with many control protocols for session, connection, and mobility management. Therefore, optimization is difficult and implementation errors are likely. To implement this technology correctly and efficiently, it is necessary to consider combinations of operational policies and standard protocols, in various respects. However, the 3GPP standard is extensive, and it is difficult to understand all combinations. Although previous studies have reported a few implementation issues for CSFB [3], [5], it is highly likely that problems exist that have not yet been found, for the reasons mentioned above.

3.1 Methodology

Both automatic analysis using SCAT and manual analysis were employed. Fig. 2 shows our analysis process: all the problems discussed in this paper can be identified using this process.

SCAT: Automatic analysis tool. Because signaling messages are transparent to mobile operating systems, baseband manufacturers expose the interface to the baseband via diagnostic messages. When the baseband receives a specific logging command, it sends messages that contain diagnostic information through a specific interface, such as a kernel driver or micro universal serial bus (USB). A few diagnostic tools [4], [16], [17], [18], [19], [24], [25] already exist, but each has limitations, e.g., 2G/3G support only, unknown implementation details, no platform independence, or high cost.

To perform large-scale automatic analysis, we built SCAT, a Python program run on a laptop. SCAT consists of SCATm, which simply logs control-plane and error messages from the cellular network, and SCATa, which compares the time and call flows within and across different networks to detect anomalous events.[2] To collect signaling messages, SCATm sends a logging command to the baseband of a phone connected to a laptop and begins logging signaling messages to a database, similar to other diagnostic tools [24], [25]. Then, it automatically starts and disconnects voice calls for a pre-defined period to collect call traces. In our experiment, we configured SCATm to wait for 5–40 s and to set up a call for 10–30 s. Then, SCATa analyzed the collected data by comparing the time and call flows within/across 3G and LTE and reported anomalies.

The detailed analysis process is as follows. When the signaling messages arrive sequentially, SCATa measures all the MM, CM, SM, and RRC procedure times. To diagnose service-specific problems, the related procedures must be added. (Here, we added call setup and attachment/detachment as CSFB measurement factors.) SCATa measures the statistics[3] for each procedure and records the procedural sequence. SCATa is divided into three phases and requires data from at least two operators. The first phase detects anomalous phenomena, based on a time threshold. In this method, it is assumed that similar times are required for detailed procedures for the same signal strengths.

2. In Fig. 2, SCATa is depicted as a large box. SCATm is omitted, because it simply logs messages.

3. 10th percentile, 90th percentile, median, mean, minimum, maximum.

TABLE 1
Time/Probability Threshold of Our Method (ϵ: Value)

Time threshold		Signaling failure	
Procedure	ϵ	Procedure	ϵ
3G RRC connection	0.5 s	RRC connection reject	
LTE RRC connection	0.3 s	Attach reject	
Call setup	0.3 s	Authentication failure	
LTE attach	0.8 s	Random access failure	
3G detach	0.5 s	Service reject	1%
Routing area update	1.1 s	Security mode failure	
Location area update	1.5 s	Tracking area update reject	
Tracking area update	0.9 s	Location area update reject	
		Routing area update reject	

If the time[4] required to complete a specific procedure for one operator exceeds the sum of the time required by the other operator and the threshold value, this event is assigned to the suspect group. Table. 1 lists the threshold values of our method. The second phase finds a specific case based on the control plane procedure order. SCATa compares the control-flow sequence among the operators and selects any control flow having an order different from that for the other operators. Selected flows are (1) those having the same procedures but ordered differently and (2) those exhibiting omitted or added procedures. These cases are classified into suspect groups and examined manually. The third phase diagnoses the anomalies based on a signaling failure threshold. The control plane establishes connections through request and response messages. If a failure (or rejection) related to a connection occurs with more than a certain probability, it can be assigned to a suspect group. Table 1 lists examples of the threshold values for phases 1 and 3. The values can be adapted depending on the purpose.

Compared to previous diagnostic tools, SCAT provides platform independence, as it is written in Python. Moreover, SCAT supports both the Qualcomm and Samsung basebands. We are currently refactoring the code to release SCAT as an open-source tool.

Manual analysis. The SCAT output includes potential problems that require further manual analysis. First, a check is performed to determine whether each problematic item is listed in the 3GPP standard, which states the root causes of certain problems. If it is listed, the analysis is stopped. Otherwise, the normal and anomalous call flow are compared and the different procedures are extracted. Then, the procedures in the 3GPP standards are investigated in more detail. After filtering out problems unrelated to the 3GPP standard, the possible root causes of the remaining problems are listed. Some can be confirmed in an interview with the operator. However, because the data are collected from the end device and the CN remains a black box, some root causes cannot be confirmed.

Example analysis 1. As a sample analysis, we considered time-related misconfiguration between the RRC and non-access stratum (NAS) (see Section 4.2). First, the duration of each procedure was computed for the collected data. After analysis based on a time threshold (SCAT phase 1), we noticed that the 3G detachment times for some operators were higher than those for other operators. In addition, we recognized that the sequences of the RRC and NAS procedures were different (SCAT phase 2). Based on this observation, we analyzed the standard and found the root cause of the time-related misconfigurations of the RRC and NAS (manual analysis).

Example Analysis 2. We considered 3G redundant location updating (see Section 6.1). This problem was detected by two methods in SCAT phases 1 and 2. The time threshold (phase 1) scheme extracted seven operators with time delays exceeding the threshold for each procedure. For four (US-I, DE-I, DE-III, and FR-II) of these operators, 3G location update procedures were added after the voice call was triggered. The addition of MM procedures was differentiated from the control flow sequence through comparison with other operators and detected using the phase 2 method. This event was classified as suspicious, and its problem classification was confirmed subsequently (decision phase). That is, analysis of the standard confirmed that 3G location updating is optional, not mandatory.

3.2 Dataset

A summary of the data collected using SCAT is presented in Table 2. For this data collection, we selected seven of the top-ranked countries, with regard to LTE subscriber numbers in 2014 and 2015 [26]: one in North America (the U.S.), four in Europe (France, Germany, Spain, and the U.K.), and two in Asia (Japan and South Korea), and chose 13 operators from those countries. Our dataset consisted of 17,710 CSFB calls, including 3,056,907 control plane messages (e.g., for the RRC and NAS) collected from Nov. 2014 to Nov. 2015. These data were collected for approximately one week in each location when we attended conferences and project meetings. For operators that supported VoLTE (Japan, Korea, and the U.S.), we disabled VoLTE to use CSFB. Note that all experiments were performed in the late evening and in a stationary environment, to minimize side effects such as network overhead or other unexpected mobility problems. Throughout this paper, we denote each operator by abbreviated symbols; each symbol consists of a nation code followed by a Roman numeral and letter (e.g., JP-X, FR-Ix, DE-IIx), denoting the country, operator, and test date/region, respectively.

3.3 Comparison with Existing Processes

There exist other approaches similar to ours for discovering generic control plane problems. CNetVerifier [5] constructs a protocol model and usage scenarios in advance based on common user demand and standards related to MM, SM, RRC, etc. Following addition of cellular-specific properties to these models and scenarios, a model checker generates counterexamples that do not satisfy these properties. Scenarios based on these counterexamples are built and checked through user studies. However, this tool has several limitations compared to our approach. First, there are numerous optional procedures in the 3GPP standards. Building models for all combinations of these procedures is infeasible. Furthermore, consideration of non-existent combinations is unnecessary. Lastly, as shown in Section 6, some options are unnecessary. Therefore, model construction based on current 3GPP standards is inappropriate. In

4. Median time required for a set of specific procedures.

TABLE 2
Summary of Our Dataset

Continent	Country	Operator	Date	Place	Device	# of calls	# of signalings	Reason
North America	U.S.A.	US-Ia	Nov 2014	Arizona	Galaxy S4, G3	601	66,549	C
		US-Ib	Feb 2015	San Diego	Galaxy S4, G3	121	34,657	C
		US-Ic	Apr 2015	Atlanta	Galaxy S5, G3	746	105,440	P
		US-II	Apr 2015	Atlanta	Galaxy S5	998	119,953	P
Europe	France	FR-Ia	Dec 2014	Paris	Galaxy S4, G3	99	15,235	C
		FR-Ib	Sep 2015	Paris	Galaxy S4, Galaxy S5, G3	418	97,547	C
		FR-II	Sep 2015	Paris	Galaxy S4, Galaxy S5, G3	1,055	193,051	C
	Germany	DE-Ia	Dec 2014	Hamburg	Galaxy S4, G3	98	19,329	C
		DE-Ib	Aug 2015	Berlin	Jolla	982	130,660	L
		DE-Ic	Sep 2015	Berlin	Galaxy S5, Galaxy S6, G3, Nexus 5	2,305	966,842	L
		DE-IIa	Dec 2014	Hamburg	Galaxy S4, G3	108	13,632	C
		DE-IIb	Apr 2015	Berlin	Jolla	49	5,778	L
		DE-IIc	Aug 2015	Berlin	Jolla	497	39,607	L
		DE-IId	Sep 2015	Berlin	Galaxy S6, G3, Nexus 5	1,297	381,204	L
		DE-IIIa	Apr 2015	Berlin	Jolla	500	48,268	L
		DE-IIIb	Sep 2015	Berlin	Galaxy S4, Galaxy S6, Jolla	2,416	343,017	L
	Spain	ES-I	Jul 2015	A Coruña	Jolla	282	30,682	V
		ES-II	Jul 2015	A Coruña	Jolla	142	13,283	V
	U.K.	UK-I	Oct 2015	London	Galaxy S6, Jolla	269	41,438	P
Asia	Japan	JP-I	Apr 2015	Tokyo	Galaxy S5	337	19,898	P
	South Korea	KR-Ia	Apr 2015	Daejeon	Galaxy S4	2,713	173,008	L
		KR-Ib	Nov 2015	Daejeon	Galaxy Note 4	1,041	134,729	L
		KR-II	Apr 2015	Daejeon	Galaxy S4	636	63,100	L

(The column entitled "Reason" lists the purpose of each visit: P: Project meeting, C: Conference, L: Local, V: Vacation trip.)

addition, analysis of common user demand does not necessarily capture optimization problems.

Jia et al. devised a tool to measure the audio quality or power consumption of a voice call [6]. They examined the user experience by changing various environmental variables such as signal strength or traffic volume. Thus, they discovered problems related to user experience alone, and their approach did not necessarily capture optimization problems as well.

Unlike the above two approaches, our approach diagnoses problems through comparative study. Incidentally, it is unnecessary to consider user demands and environmental variables.

3.4 Summary of Our Results

Analysis of the data collected in this study allowed identification of six performance problems and their categorization as time-related misconfigurations (Section 4), synchronization problems (Section 5), and redundant procedures (Section 6).

The time-related misconfiguration category included two different cases. The first occurred because of timing issues due to the MME load balancing mechanism and the user tracking area update (TAU). This problem caused subscribers to experience an *out-of-service* issue for up to 11 s (Section 4.1). The second case occurred because of time-related misconfiguration between the device-to-base station and device-to-MME communications during generation crossover from 3G to LTE. This case forced subscribers to wait unnecessarily in 3G for 0.5–1.8 s (Section 4.2).

The synchronization problem also consisted of two different cases. The first occurred for one operator, when the access network broadcast incorrect frequency information from the other-generation network. As a result, subscribers first experienced the *out-of-service* state for 30 s and were then held in 3G for up to 100 s (Section 5.1). The second case occurred frequently for one operator, wherein security-related information in 3G was sent to the LTE MME, delaying subscriber attachment to LTE (Section 5.2).

Likewise, there were two redundant procedure cases, which occurred frequently for seven operators. The first involved redundant location updating that caused 1.0–6.5 s delays during switching between 3G and LTE (Section 6.1). The second case occurred because of security-related information, causing up to 0.45 s delays (Section 6.2).

In the next three sections, we discuss these categories in detail.

4 Time-Related Misconfiguration

Signaling interactions among the participating entities (including the UE) in cellular networks are complicated. Correct sequences of signaling interactions at the appropriate times are crucial for reliable services. We examined two problematic cases rooted in time-related misconfiguration that cause performance degradation.

4.1 MME Handover and TAU

MME is a key component of the LTE core network. It provides mobility management for the LTE network and supports subscriber authentication, roaming, and handover to other networks through the NAS protocol. When a subscriber attempts attachment to an LTE network, he/she

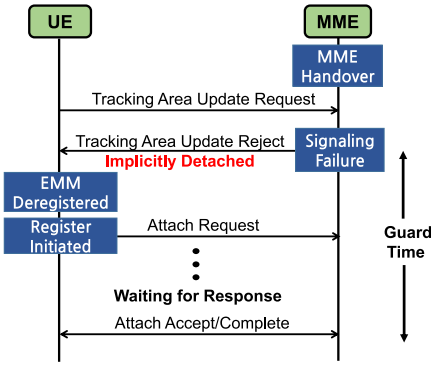


Fig. 3. Guard time with MME handover.

must be authenticated by the MME, assisted by the HSS. If the MME rejects an attach request, the subscriber cannot access the LTE service. For one operator, we found that the MME occasionally did not respond to such a request for approximately 11 s.

Question. Why does the MME in this particular network not respond within a certain time period?

Problem analysis. To guarantee MME availability, many cellular operators implement load balancing. When a subscriber moves repeatedly between 3G and LTE, he/she may be attached to a different MME during each LTE attachment. This procedure is called MME handover. To accelerate this process, most operators combine the attachment procedures and TAU. In US-II, we noted that the MME became silent for a long period of time (∼11 s) after the TAU message was rejected.

Dataset analysis. Our dataset for US-II contained 998 CSFB calls, which included 22 TAU rejection messages sent by the MME.[5] When such a message was sent, the UE connected to a different MME.[6] After receiving a TAU rejection from the MME, the UE transmitted an attach request message to the MME. However, the MME did not respond to any control plane message from the UE for 10.4–11.3 s. After this silent period, the MME delivered an identity request message to the UE. Fig. 3 illustrates this procedure.

Root cause. The standard [27] specifying the procedure between the MME and UE prioritizes signaling messages within the CN over those sent to the UE. If the UE transmits TAU and attachment request messages, MME handover may also occur. If this handover is not complete, any signaling messages from the S/P-GW are rejected. This signaling includes messages from the S/P-GW to complete the user attachment request [27]. The standards resolve this waiting period in two different ways. One method is to wait until the location update is complete. The other utilizes the "guard timer"[7] set by the P-GW, which prevents signaling messages from other entities. When the guard timer is activated, the CN resolves the delayed attach requests from the UE. If this timer has expired, the CN begins receiving signaling messages from the UE.

5. The rejection message stated "Implicitly Detached" as the cause.

6. One can check this change in the call flows by examining whether the globally unique temporary identifier is in the TAU request message, as this message includes an MME code (MMEC) that represents the MME identity in the operator.

7. Each node in the cellular network has several types of timers for efficient operation. These timers are set to wait for the next step.

TABLE 3
Attachment Time After TAU Rejection

Operator	90th percentile	Median Time	10th percentile
US-II	11,253	10,909	10,738
DE-IId	1,999	1,864	1,516
DE-IIc	1,959	1,797	1,680
ES-I	1,425	1,310	1,196

The presented operators have more than 10 TAU reject messages (units: ms).

Note that TAU rejection messages were found for some operators in our dataset. However, as shown in Table 3, the time required for network attachment after TAU rejection was short in each case, except for US-II, where 10.4–11.3 s elapsed. Hence, we concluded that the timer value was set to approximately 10 s.

Reasoning behind our analysis. It is difficult to analyze the exact behavior inside a CN. However, by comparing statistical data and the signaling message sequences for user devices, the time-related misconfiguration problem in the CN could be extracted here. Note that, if we had checked a small dataset from a single operator only, it would have been difficult to notice this problem. We informed US-II of this constant delay problem and expect that they will rectify this issue soon.

Solution. To reduce the *out-of-service* period, one may simply shorten the timer for the TAU rejection. However, this solution is not fundamental. To resolve the handover failure problem, load-balancing techniques such as S1-flex optimization (Section 7.1) may be employed, which can prevent frequent MME handovers if the serving MME has sufficient capacity for the UE. Caching old mobility contexts and forwarding request messages to the new MME without deactivating the evolved packet system (EPS) bearer context may also be a solution. However, this solution requires a change of standard.

4.2 RRC and NAS

The UE implements different protocols to communicate with the network nodes, such as the RRC [28], [29] and NAS [30]. The RRC is a control protocol between the UE and access network. It handles the connection establishment, connection release, and call paging. The NAS is an RRC upper-layer protocol handled by the MME. The NAS and RRC are separated for several reasons, such as security (to prevent eavesdropping). Consequently, the base station cannot read the NAS messages. In fact, the CN does not fully trust the access network. However, this separated structure may cause problems such as those mentioned below.

Question. Does miscommunication between the RRC and NAS affect user experience?

Problem analysis. As the NAS is an upper layer of the RRC, a mismatch between these two layers can cause problems. An example is timing mismatch; there are many time-related configurations for the NAS and RRC layers in the UE, base station, and MME. The standard sets the default timer, but the operators can utilize custom configurations. Furthermore, UE manufacturers can set some timers. In our dataset, we found one problem caused by time-related misconfiguration during generation crossover, due to the timer being set by the operator. To crossover to another

TABLE 4
Duration of Delayed 3G Detachment According to Layer Mismatch

Operator	# of Mismatch / # of LAU	Frequency (per call)	Duration (s)		
			10th	Med.	90th
DE-Ia	87/95	88.7%	0.65	0.79	1.13
DE-Ic	802/2461	34.7%	0.71	0.96	1.24
DE-IIa	9/17	8.3%	1.28	1.51	1.76
DE-IId	18/69	1.3%	0.78	1.24	1.29
FR-Ia	96/99	96.9%	0.52	0.56	0.64
FR-II	114/119	10.7%	0.52	0.58	0.65
US-Ia	42/58	6.9%	0.89	1.06	1.33
US-Ib	55/163	45.4%	0.64	0.72	0.77
US-Ic	261/304	34.9%	0.84	1.06	1.38

Here, "10th" and "90th" represent the 10th and 90th percentiles, respectively, and "Med." is median time.

generation (e.g., from 3G to LTE) after a call, the UE performs one of the following three actions: (i) immediately releasing the 3G RRC connection and attaching to the LTE; (ii) conducting the remaining NAS procedures such as location updates, releasing the 3G RRC connection, and attaching to the LTE; or (iii) immediately releasing the 3G RRC connection, but re-establishing the 3G RRC connection and conducting the remaining procedures as in case (ii). In the last case, the UE must reconnect to the RRC and conduct NAS procedures. Below, we discuss the problematic case (iii) in more detail.

Dataset analysis. In our dataset, we discovered that five of the 13 operators (US-I, DE-I, DE-II, FR-I, and FR-II) encountered the above problem (iii) during crossover from 3G to LTE (see Table 4). The UE immediately released the 3G RRC connection after a CSFB call and the re-established this connection to conduct the 3G location update. This behavior can be interpreted as follows: the UE first releases the 3G RRC, but it realizes that it must perform the NAS procedure. To complete this procedure, it then re-establishes the 3G RRC connection. However, this scenario should be handled as in case (ii), in which the 3G RRC connection is not released immediately. This mismatch delays the UE detachment from 3G for 0.56–1.51 s. This is not a small problem in terms of cellular network optimization, especially when its frequency is considered (see Table 4).

Root cause. The above problem is caused by time-related misconfiguration between the NAS and RRC layers in the UE, base stations, and CN. The RRC connection is managed by the access network, while MM procedures, such as location area updates (LAUs), are managed by the CN. In the case of a mismatch, miscommunication occurs when the access network in 3G releases the radio connection, but the UE attempts to reconnect to 3G to perform LAUs. The communication problem between *the access network*, which considers the 3G connection to be unnecessary, and *the UE* performing LAUs is the root cause.

Solution. The RNC in the access network releases the 3G RRC connection when it determines that the connection is unnecessary. In this case, the UE enters the 3G idle mode or performs a handover to return to the LTE network. If the UE maintains the 3G RRC idle mode, the above problem may be caused. To prevent this scenario, it is reasonable to direct a connection to a preferred network (here, an LTE network). The standard allows the insertion of redirection information as an extension in the *RRC Connection Release* message. In this extension field, when the RNC sets the available frequency list of the LTE networks as inter-radio access technology (RAT) information, the UE receiving the 3G *RRC Connection Release* message can leave the 3G network and attempt to attach to the LTE network [28]. The MM procedures that are not conducted in 3G can be performed in LTE, in combination with the EPS MM procedures [7]. In this case, the remaining 3G procedures do not cause additional delays, because they are simply incorporated into the LTE procedures yet to be conducted.

5 Synchronization Problem

In this section, we discuss the problems caused by misconfigurations during the CSFB for synchronization purposes, which degrade the network performance.

5.1 Misconfigured Cell Reselection

During crossover between LTE and 3G, the network provides information on which networks the UE should use. There are explicit and implicit methods for providing crossover information, and 3G and LTE use different messages for this purpose. If no explicit network crossover information is provided, implicit information is used instead. If this implicit information is incorrect or inconsistent within the network, performance problems are caused.

Question. How does incorrect generation crossover network information affect the network delay?

Problem analysis. Explicit crossover information is provided during a release of the 3G or LTE connections. In the case of 3G-to-LTE crossover, the frequency information of the LTE network is sent as evolved universal terrestrial radio access absolute radio-frequency channel number (EARFCN) values. The UE selects the LTE network based on the EARFCN values, which are not prioritized [28]. Most operators list EARFCN values of 0 or 1 during 3G-to-LTE crossover, even when they operate the LTE network on multiple frequencies.

Implicit crossover information is provided by the system information messages on the 3G and LTE networks. System information messages contain network parameters such as the mobile country and network, cell ID, and location information (e.g., current tracking area (TA) information). They are broadcast in the form of system information blocks (SIBs). System messages contain information on crossovers to other network generations (LTE SIB 6 and 3G SIB 19) [28], [29]. Unlike the EARFCNs listed in explicit crossover information, those in implicit information are prioritized, and the UE follows this priority during crossover.

We found configuration errors during 3G-to-LTE crossover for three operators, which negatively affected performance.

Dataset analysis. DE-I had configured the LTE network information on their 3G network by listing both available and unavailable networks, differentiated by priority alone; this is not a typical configuration. When combined with misinterpretation of the LTE network information on the UE side, this misconfiguration degraded the performance during 3G-to-LTE crossover. ES-I and US-I had similar misconfigurations, but those did not affect the 3G-to-LTE crossover performance.

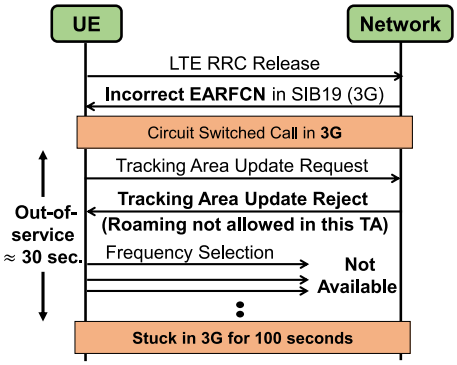


Fig. 4. Crossover procedure for receipt of incorrect EARFCN information in SIB 19 message.

Before collecting signaling messages, we scanned the network to check which EARFCN was being used in the area and whether the user could connect to the network operating on the EARFCN. If the EARFCN was not used or if the user was unable to connect to the network, we considered that EARFCN to be incorrect.

Root cause. Both network and UE misconfigurations can cause this problem. While releasing the 3G RRC connection, ES-I and US-I list multiple EARFCNs of the LTE network as the crossover information, even though they operate LTE on only one frequency in the area. In these cases, the unused EARFCNs do not significantly affect the average performance of the crossover to the LTE, as only one of the EARFCNs is actually used. If none of the listed EARFCNs are used, crossover is delayed.

DE-I did not list the LTE EARFCN during a release of the 3G RRC connection, and listed an inaccurate LTE EARFCN in 3G SIB 19. As a result, the implicit information on 3G SIB 19 was used and the performance was degraded. Fig. 4 illustrates the 3G-to-LTE crossover procedure for DE-I. DE-I merges with another operator (DE-IV[8]), allowing a domestic 3G roaming agreement between the two networks; however, this does not extend to LTE. This characteristic is reflected in SIB 19: DE-I places a higher priority on its own LTE network while listing the DE-IV LTE network as lower-priority. Simultaneously, DE-IV does the opposite. Some UEs ignore this priority and perform cell selection themselves, which causes them to be held in 3G during crossover.

When the UE on DE-I performs 3G-to-LTE crossover, the LTE network information on 3G SIB 19 is used. SIB 19 only indicates the availability of the network and does not reveal that only specific users are allowed on this network. If the UE camps on the DE-IV network, the LTE TAU fails, as LTE roaming between DE-I and DE-IV is not possible. As a result, the network informs the UE that roaming is prohibited and returns the UE to 3G (which takes 30 s, in the worst case). Until the UE performs crossover to the LTE network of DE-I itself, it remains held in 3G (for 100 s, in the worst case).

If a user is moving around their area and several overlapping TAs are available at the user position, it is possible to select the DE-IV LTE network until the UE obtains the same response from all available TAs. Consider the case of a user moving in a car. The TA area in German cities is approximately 10–30 km^2 [31], which corresponds to a radius of approximately 1.7–3.0 km. Assuming that the car travels around the city at 60 $\mathrm{km/h}$, the TA can be changed every 3.4–6 min. In our stationary experiments, we observed four nearby TAs, and the UE was held in 3G for an extended period when it performed TAU in the wrong network. Even after all available TAs were marked as forbidden, the life cycle of the forbidden TA list could be determined by many factors [30]. If the forbidden TA list was reset, the UE could perform TAU on an unavailable TA and, could again, experience performance degradation.

TABLE 5
Duration of 3G/LTE RRC Connection After SIB Messages with Correct and Incorrect Frequencies for DE-I

Event	Duration (ms)		
	10th	Med.	90th
SIB with correct freq.	1,218	1,283	1,585
SIB with incorrect freq.	1,848	2,379	3,196

The duration column represents percentiles.

In addition, we compared the durations of all the 3G and LTE RRC connections after SIB messages with correct and incorrect frequencies, as reported in Table 5. The RRC connections following misconfigured SIB messages took a median 1,096 ms longer than those following correctly configured SIB messages.

Solution. On the network side, there are two possible solutions: (i) properly configuring the cell selection preferences on an implicit crossover and (ii) explicitly specifying the LTE EARFCN when releasing the 3G RRC connection.

Similar to the configuration used in DE-I, ES-II has a domestic roaming agreement with another operator, ES-III,[9] up to 3G networks. During generation crossover in ES-II, both 3G and LTE RRC connections release messages containing the EARFCN and universal terrestrial radio access absolute radio-frequency channel numbers (UARFCN; the 3G counterpart of EARFCN) of ES-II; thus, no implicit network selection is required. Roaming is only allowed on one side (ES-II to ES-III), unlike the case of DE-I (DE-I and DE-IV users can roam into both 3G networks). As a result, cell re-selection is performed correctly. Furthermore, when there are changes in the operating frequencies of the network caused by operator policies (e.g., operator mergers and acquisitions) or regional policies (e.g., frequency spectrum auction), the operators must properly configure the system messages to reflect the current network situation.

On the UE side, following the network suggestions for implicit cell re-selection is recommended. We could not verify the exact manner in which cell selection functions, as the baseband firmware is not generally accessible to the public.

5.2 Security Context Sharing Problem

During the initial mobility procedures, UE and mobile networks establish security contexts to protect integrity and to encrypt communication. To reduce the signaling and computational loads caused by establishing new security contexts for each network generation, security contexts previously used in one network can be re-used in another

8. This operator is not featured in our dataset in Table 2.

9. This operator is not featured in our dataset in Table 2.

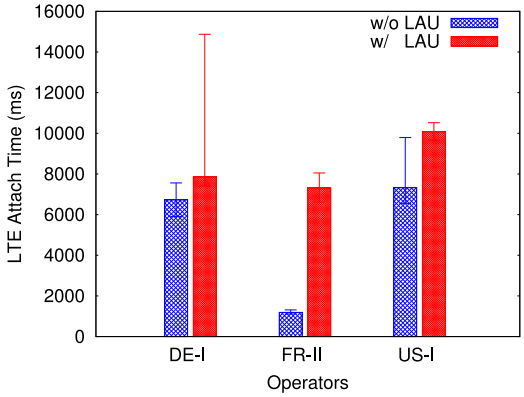


Fig. 5. Time delay (10th percentile, median, and 90th percentile) of LTE attachment procedure.

network. The 3GPP standards define this procedure as security context mapping.

Question. If security contexts are not mapped correctly, can user experience be affected during mobility management?

Problem analysis. Generation crossover from 3G to LTE can take two different directions with respect to security context: (1) use of the security context mapped from 3G (the standard specifies this context as KSI_{SGSN}) or (2) generation of a new security context (the standard specifies this context as KSI_{ASME}). The TAU request contains information about which approach the UE has chosen. When the UE requests use of KSI_{SGSN}, the LTE CN must derive the security context from its 3G counterpart [32]. Failure to derive the security context causes the UE to perform initial attachment procedures again, including establishment of a new security context.

Dataset analysis. While most of the operators examined in this study had implemented security context sharing correctly, ES-I had problems deriving the LTE security context from the 3G security context. This issue was visible for the TAU failures with "implicitly detached" as the failure cause, when KSI_{SGSN} was specified as the TAU security context.[10]

Nearly every TAU with KSI_{SGSN} as the security context failed. Specifically, TAU rejection occurred for 88 of 89, i.e., 98.8 percent TAUs. The total number of TAUs in ES-I was 261 (88/261 = 33.7 percent). Thus, the problem occurred more than once in every three phone calls. In one exceptional case, the key update procedure activated and reestablished the shared key. This high failure rate was visible for ES-I only.

After a TAU was rejected because of "implicitly detached," 10 of the 88 (11.3 percent) following attachment requests also failed. Thus, the UE was returned to 3G and the LTE service was unavailable until another attachment request was made. Other requests such as TAU also failed during this period. Between the failed TAU and subsequent successful attachment request, the user could not use the mobile network. The time between TAU failure and successful attachment was 1.24–1.52 s. The duration of the delayed TAU (1.49–1.77 s) was six to seven times (596–708 percent) that of the average TAU (0.25 s) in ES-I.

Root cause. If security context mapping from 3G to LTE is incorrectly implemented or unavailable, the LTE network can reject TAUs with mapped 3G security contexts. Because we could not find this problem in networks other than ES-I, we assume that this issue is related to the ES-I configuration.

Solution. There are two possible solutions: (i) The 3GPP standard [32] recommends generation of a new security context after 3G-to-LTE generation crossover for security reasons. As a short-term solution, generating this new security context eliminates the security context sharing problem; (ii) If the security context is mapped from 3G to LTE, the CN should check whether synchronization of the MME mapping state is required.

6 Redundant Procedures

To support generation crossover, cellular operators must implement complicated control protocols. However, because of the complexity of cellular networks, implementation of these protocols can involve unnecessary procedures such as redundant updates.

Question. Are there any redundant procedures during generation crossover that affect user performance?

6.1 Location Update

As specified in the standard [7], the UE should conduct a location update to inform the network of its current location. For example, the UE updates its location when it initially attaches to a network or moves to a new location area (LA) in 3G or a tracking area in LTE. Location updates are also run periodically, if the UE is required to report its location regularly, at predefined time intervals. However, such location updates degrade performance when used too frequently.

Problem analysis. For inter-operability of generation crossover, the standard allows operators to conduct 3G location updates in LTE[11] [7]. Therefore, once the operator conducts the location update in LTE, there is no need to do so again when the UE is in 3G during the generation crossover. However, several operators in our dataset conducted redundant location updates in 3G, even when the LA had not changed.

Note that our dataset was collected in a stationary state within one LA (i.e., there was no LA border). In this dataset, four operators (US-I, DE-I, DE-III, and FR-II) had a high probability of conducting redundant 3G location updates (71.9, 78.9, 100, and 45 percent, respectively) when they entered the 3G network from LTE, while the nine other operators did not or rarely conducted such updates (0–19.3 percent). As the operators had already obtained location information from the LTE, location updates after 3G entry were not also required. We identified two redundant updates in (i) LTE attachment and (ii) call setup.

Three operators (US-I, DE-I, and FR-II) conducted 3G LAUs after the CSFB call. Fig. 5 shows the difference between the LTE attachment times with and without LAU. The DE-I and US-I attachments were delayed for approximately 1 and 3 s, respectively, while FR-II had a 6.5 s delay. Our results indicate that the LTE attachment was delayed by the redundant LAUs. The LTE attachment time of FR-II was, surprisingly, 6.1 times longer when performing LAU.

10. Combined with an LTE attachment request message.

11. Combined attach/TAU procedure.

TABLE 6
Frequency of and Median Time to Complete 3G Authentication Procedures During Generation Crossover

Operator	Prob.	Time	Operator	Prob.	Time
US-I	8.4%	71 ms	ES-I	100%	439 ms
US-II	20.1%	157 ms	ES-II	7.5%	71 ms
FR-I	100%	163 ms	UK-I	8.6%	10 ms
FR-II	73.8%	110 ms	JP-I	1.3%	75 ms
DE-I	100%	245 ms	KR-I	1.0%	121 ms
DE-II	1.1%	271 ms	KR-II	0.0%	0 ms
DE-III	63.1%	214 ms			

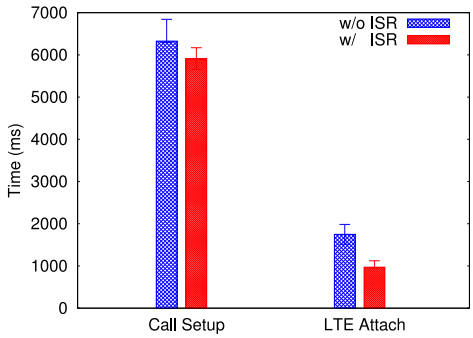


Fig. 6. Comparison of call setup and LTE attachment time (10th percentile, median, and 90th percentile) in JP-I.

Unfortunately, FR-II also exhibited a mismatch problem for 3G LAU (see Section 4.2); furthermore, an *out-of-service* condition was triggered after 3G LAU. Consequently, FR-II exhibits the largest difference in Fig. 5.

DE-III conducted 3G LAU as soon as the UE fell to 3G. Therefore, the time required for LAU was always added before a CSFB call was made. Note that, as the UE of DE-III always conducted 3G LAU, we were unable to compare cases with and without LAU directly. Therefore, the 3G LAU delay time of DE-III was estimated to be 0.41 s by measuring the time from the LAU request to the LAU acceptance.

Root cause and solution. Redundant LAU is the root cause. In our dataset, US-II, JP-I, and KR-II did not conduct any 3G location updates, and six operators had very low probabilities of conducting LAU (lower than 6 percent). FR-I initially performed 3G LAU (FR-Ia), but was later configured to omit the update procedure in CSFB calls (FR-Ib). Therefore, if the 3G-cell LA was identical to that of the updated value in LTE, the operators did not have to force the UE to update the 3G location; as the standard [7] suggests, this redundant procedure is "optional."

The CSFB standard [7] allows implementation freedom for the procedures used to return to LTE after the UE disconnects a call. Each operator implements this process in a unique manner. The UE typically enters the 3G MM idle mode after a CS call [33]. If the network releases only a signaling radio bearer, the UE can perform LAU to enter the MM active mode. An alternative means of returning to LTE is to disconnect 3G. However, UE attachment to LTE simply due to disconnection of the 3G RRC cannot be guaranteed. In this case, as mentioned above (Section 4.2), the 3G RRC can be re-established and LAU can be performed. A better (and probably the best) way to return to LTE is to insert a valid EARFCN list into the extension field of the *3G RRC Connection Release* message. It is then possible to induce the UE to receive a broadcast channel message from LTE without conducting LAU.

6.2 3G Security Context

LTE is considered to be more secure than 2G/3G. The 3GPP standard strongly recommends that the CN in LTE re-authenticate the UE utilizing the authentication and key agreement (AKA) during generation crossover from 2G/3G to LTE [32]. Thus, operators update the UE security contexts during generation crossover from 2G/3G to LTE. However, the standard does not consider the opposite case (LTE to 2G/3G generation crossover), because remapping of the security context can be internally processed in the CN. The standard does not recommend re-authentication of the UE either [32]; this is set as an optional procedure [7], with security-related procedures depending on the operator's implementation. In addition, the standard allows the MSC to modify the authentication frequency to accelerate the CSFB procedures [30].

Problem analysis. Table 6 shows the frequency of 3G authentication and its duration during CSFB generation crossover. During generation crossover from LTE to 3G, seven operators (US-I, DE-II, ES-II, UK-I, JP-I, KR-I, and KR-II) conducted 3G authentication with up to 8.6 percent probability, three operators (US-II, FR-II and DE-III) performed it very frequently (20.1, 73.8 and 63.1 percent, respectively), and three other operators always performed this authentication (100 percent). Note that the 3G authentication procedure is not time-consuming. Even with this additional procedure, UK-I (with a low probability of 8.6 percent) spent only 10 ms on 3G authentication. As the worst case, ES-I always authenticated the UE (100 percent) during generation crossover. Further, the time for 3G authentication (439 ms) was significantly longer than for the other cases, being a large penalty for subscribers. The root cause of this large time difference is unclear.

Root cause and solution. This security procedure blindly follows the security policies of the operators. One might think that this is a tradeoff between time and security. However, the security context of LTE is already mapped to 3G. The MME transfers the cipher key (CK) and integrity key (IK) with KSI to SGSN, i.e., the 3G network obtains the security context from LTE. In this case, 3G authentication is unnecessary. As a solution, operators can skip this redundant procedure.

6.3 Suggested Solution for Redundant Procedures

3GPP [30] suggests a solution called idle mode signaling reduction (ISR) for redundant signaling messages during generation crossover. In our dataset, only JP-I utilizes the ISR to reduce signaling messages. The same standard requires the UE to support ISR, but implementation of the ISR on the CN is optional.

If the ISR is activated, the UE can maintain resources for sessions in both generations, unlike in typical generation crossover, during which all resources in the source network [12] are released. Many procedures for generation crossover are eliminated in this case, such as the TAU/LAU and AKA. Fig. 6 shows a comparison of the user experiences

12. When generation crossover from LTE to 3G occurs, the LTE and 3G networks are called the source and target networks, respectively.

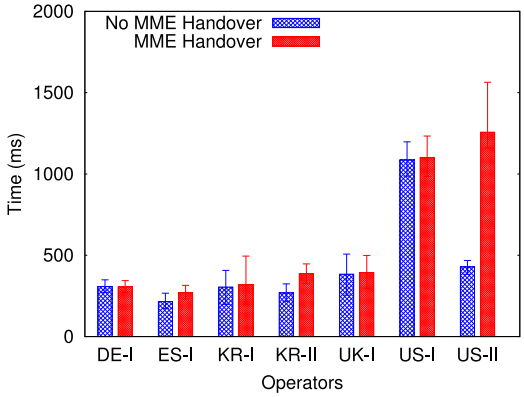


Fig. 7. LTE attachment time among operators with MME pooling, with and without MME handover (10th percentile, median, and 90th percentile).

with and without signaling reduction in JP-I. When ISR was adopted, the median call setup time is decreased by 0.47 s (from 6.37 to 5.90 s) and the median LTE attachment time is decreased by 0.77 s (from 1.74 to 0.97 s). These decreases occur because ISR eliminates additional location updates and security context establishments, which are performed by the other operators. The graph shows that these procedures impact user performance. Thus, it is clear that ISR improves the user experience. However, one interesting question remains: *although the user experience improves, does ISR increase the network overhead?* We leave this question for future work. Another interesting question is the cost of ISR deployment. One operator we interviewed decided not to adopt ISR because of the tradeoff between cost and user experience.

7 Discussion

In this section, we discuss other minor performance issues as well as the lessons learned and limitations identified through this study of operational cellular networks.

7.1 Other Performance Issues

S1-Flex Implementation Irregularities. To provide availability and manage resources efficiently in MMEs, cellular operators utilize an MME load-balancing technology called S1-flex. Before adoption of S1-flex, a single MME controls a group of base stations. If this MME becomes unavailable, the group of base stations for which the MME is responsible cannot provide user connections, although other MMEs can provide proper assistance. S1-flex can solve this problem by allocating multiple MMEs to multiple base stations.

When an operator utilizes S1-flex, the UE may communicate with different MMEs[13] while re-attaching to the network, even if the UE is connected to the same base station. All the operators in our dataset utilize S1-flex, and seven among them frequently change the serving MME when the UE re-attaches to the network of the operator.

Fig. 7 presents the LTE attachment times for operators who do and do not offer serving MME changes. Six among the seven operators show negligible time differences between the cases in which the serving MME was maintained and changed. However, US-II exhibits a large time difference (the median delay for the former case is 1,256 ms and that for the latter is 428 ms). This difference could be caused by misconfiguration by the operator or implementation flaws in S1-flex. As the detailed S1-flex procedures were not seen by the UE, we were unable to analyze the exact cause of the performance degradation.

Further Optimization on Inter-RAT Handover. Even for a single generation, 3GPP standards are continuously evolving through different "releases." Problems found in any given cellular measurement or analysis may not exist for other operators deploying technology using different releases. Tu et al. [5] considered only two US operators, and among the CSFB methods defined in the standard [34], only the simplest, Release 8 ("R8" hereafter) redirection without system information, was mentioned. We found operators using alternative CSFB methods defined in the standard, i.e., the R8 PS handover and another redirection method defined in Release 9.

R8 redirection without system information is the simplest means of crossover from LTE to 3G, as this method specifies 3G channel information only when the LTE RRC connection is released. As a result, the UE must find the 3G network based on the channel information, and perform the basic mobility procedures. Because the R8 redirection-based CSFB does not require mobility management between multiple networks inside the CN, this method is most commonly used among the operators in our dataset (in eight out of 13 cases).

Compared to R8 redirection, R8 PS handover with data radio bearers (DRBs) ("R8 PSH" hereafter), and Release 9 redirection with system information ("R9" hereafter) accelerate LTE-to-3G crossover. Specifically, R8 PSH accelerates the procedures between the LTE RRC connection release and part of the 3G RRC connection setup, and R9 reduces the time spent on 3G cell searching by delivering nearby 3G cell information when releasing an LTE RRC connection. Readers may refer to the 3GPP standard [34] for details of the procedure and performance differences among operators.

To deliver partial 3G network information via an LTE network, both R8 PSH- and R9-based CSFB require additional mobility management procedures, which are not necessary for R8 redirection-based CSFB. Thus, a relatively small number of operators use the advanced method: DE-III and UK-I use R8 PSH, and DE-II, US-II, and FR-I use R9 (five of 13). For DE-III in this study, R8 PSH had better performance to 3G compared to R8 redirection-based CSFB, by ~800 ms on average.

7.2 Operator Interviews

We found six performance problems (five of which were novel), with root causes that were not discovered in previous studies, as summarized in Table 7. Most causes were indisputable, as we analyzed anomalous procedures based on standards. To confirm each cause, we contacted and interviewed four cellular operators.

For the Section 4.1 case, the operator refused confirmation, stating that the configuration related to mobility management is considered a trade secret. One of the operators confirmed the problem in Section 4.2 and considered how

13. In some cases, UEs always connect to the same MME, although their operators utilize S1-flex in their own particular configurations.

TABLE 7
Summary of Problems

Problem	When?	Delayed Procedure	Observation	Method			OP
				1	2	3	
Time-related misconfiguration between TAU request and MME handover (Section 4.1)	When conducting TAU (2.2% per TAU, once in 9 min)	LTE attach, TAU	Out-of-service 10.4–11.3 s			✓	1
Time-related misconfiguration between RRC and NAS (Section 4.2)	After the CSFB call ends (96.9% per call, in worst case)	3G detach	Delay 0.56–1.51 s	✓	✓		5
Attachment with incorrect frequency channel (Section 5.1)	When changing TA (once in 3.4–6 min if traveling at 60 km/h)	LTE attach, TAU RRC connection	Out-of-service 30 s Stuck in 3G 100 s RRC delay 1.1 s	✓			1
Synchronization error of 3G security context (Section 5.2)	When connecting to LTE (33.7% per LTE attach)	LTE attach, TAU	Delay 1.2–1.5 s	✓		✓	1
Redundant location update in 3G (Section 6.1)	When connecting to 3G by CSFB (every call in worst case)	LTE attach, 3G detach Call setup	Attach/Detach delay 1.0–6.5 s Call setup delay 0.4 s	✓	✓		4
Redundant security context update in 3G (Section 6.2)	When connecting to 3G by CSFB (every call in worst case)	Call setup, 3G attach	Delay 0.4 s		✓		5

(Method 1: Time threshold, 2: Control flow sequence, 3: Signaling failure; OP: # of operators).

to mitigate the problem using RRC and NAS timer settings. Because Section 5.1 is obvious, we skipped that case. We found the root causes of the problem in Section 5.2 and reported them to the operator. The causes in Section 6.1 and Section 6.2 were indisputable based on the standards, and we reported them to two operators. One had already recognized the problem and confirmed our assessment, but the other was unaware of this issues. From the interviews, we realized that operators can often overlook certain details of standards or misunderstand them.

7.3 Limitations and Possible Extensions

In this work, we focused on data obtained from the UE. The control plane messages inside the CN were invisible to us and, therefore, were not considered. In addition, because of economic difficulties regarding simultaneous testing of many different mobile devices with different LTE subscriptions, our tests were conducted using a small number of UEs at the same location.

Our case study was limited to CSFB. We performed experiments with prepaid subscriber identity module (SIM) cards for most operators, who offer CSFB and 3G CS calling as voice call technology to prepaid SIM users. Because our experiments required a various comparative data, the latest services (e.g., VoLTE, Internet of Things (IoT)) were excluded. Furthermore, as VoLTE is supported for residents only, large-scale VoLTE was infeasible at the time of our experiment. When LTE technology supporting IoT (e.g., narrowband IoT; NB-IoT) is standardized, signaling messages for such LTE can also be tested. Further, we could not investigate the impact of different equipment manufacturers, as the feasibility of fingerprinting LTE core equipment is currently unknown. Our approach seems more suitable for unicast than broadcast messages, because our problem diagnosis method depends on the time measurement or failure probability of the request/response. In the case of a broadcast channel, the network usually transmits information in one direction and does not wait for a response from the device. Nevertheless, as there are many different ways of constructing broadcast systems in cellular networks, we may obtain meaningful results simply by comparing other systems.

We collected signaling messages when the signal strength was relatively strong. While variations of these conditions, such as testing with myriads of smartphones or with weak signal strengths, could serve as the basis for yet another interesting paper, we leave these tests as future work for economic reasons. In addition, generation crossover can occur when the signal strength in one generation becomes weak. This crossover requires movement through different areas, which makes data collection quite limited. Therefore, we did not consider this case either. Nevertheless, we identified six major problems and their causes, which had not been discussed previously.

7.4 Automation Challenges

Two parts of our methodology involve manual analysis: (1) threshold value determination, and (2) root cause analysis. For root cause analysis, manual analysis is unavoidable, as messages within CNs are not visible and the 3GPP specifications do not explain implementation details and operational policies. However, the other parts of our analysis could be automated.

One approach is to simplify complex control plane messages to render them comparable, e.g., by introducing a state machine to represent control-plane procedures with timing information. As the 3GPP standards define states, building a simple state machine is feasible. However, for such a state machine to be useful for automatic analysis, several challenges remain. (1) While current state information is defined separately in the MM, SM, and RRC, their interaction must be represented. (2) Even after combining these separate states, the resulting state machine must include sufficient information on aspects such as timing, along with detailed information on each signaling message (e.g., error messages) for analysis. (3) Comparison of large graphs is required, which is known to be difficult, as the entire state machine including this information would become very large.

Solving challenge (2) in particular seems difficult, because two identical control plane procedures could have

different meanings. For example, without checking the MME code, it is impossible to know whether MME handover has occurred. In other words, all the different information included in a signaling message must be included in the state machine representation. If these three challenges are resolved, it might be possible to compare two or more state machines and extract differences automatically. However, manual root cause analysis may still be necessary.

7.5 Lessons Learned

As described in Section 2, while operational cellular networks have been analyzed in several studies, recent measurement studies have focused on small numbers of operators, especially in the U.S. Measurement studies neglecting operator characteristics may be insufficient for analyzing the problems of operational cellular networks, however, because the implementations and configurations of each service may differ among operators.

Operators may have powerful self-diagnostic tools to analyze their networks. However, as our results show, many operators still experience various kinds of problems, as described in Table 7. This finding does not necessarily mean that our methodology can cover a superset of problems, but it does mean that other diagnosis methods are also needed.

In this work, instead of relying on a few cellular operators, we conducted a comparative study based on a large dataset obtained from 13 operators in seven countries. The results revealed six different problems in CSFB, which was taken as a case study. Hence, we learned that a comparative measurement study is a simple yet effective mechanism for analyzing problems in operational cellular networks. However, we believe that other existing problems are currently unknown, because of the limitations described in Section 7.3.

8 Concluding Remarks and Future Work

We presented a novel diagnosis method that finds performance bugs by cross-checking cellular procedures among different operators. To evaluate our method for CSFB as a case study, we collected 17,710 calls and 3,056,907 control-plane messages from 13 major LTE operators in seven different countries around the world. Using our simple and effective analysis methodology of comparing the call flows and times between the operators, we discovered six major issues, five of which were not discussed in previous studies. We also provided in-depth analyses of the root causes of these problems. We found that different operators employed different implementations yielding various degrees of performance degradation. Because of the diversity among operators, we argue that hasty generalization in cellular network research may be hazardous. To prevent such errors, we strongly recommend examining traffic from multiple operators over multiple regions.

Future topics of investigation include automating parts of the analysis procedure and expanding the analysis to data services, VoLTE [6], [35], and 5G. We also plan to release our dataset (with operator approval) and our analysis tool, which can be run on any platform to detect and diagnose cellular service performance problems, as open-source code. We believe that this tool can significantly simplify the cellular network diagnosis process and reduce troubleshooting time costs.

Acknowledgments

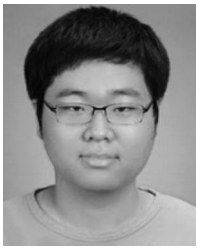
This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2015-0-00403) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

[1] A. Ahmad, *Wireless and Mobile Data Networks*. Hoboken, NJ, USA: John Wiley & Sons, 2005.
[2] M. Riegel, A. Chindapol, and D. Kroeselberg, *Deploying Mobile WiMAX*. Hoboken, NJ, USA: John Wiley & Sons, 2009.
[3] G.-H. Tu, C. Peng, H. Wang, C.-Y. Li, and S. Lu, "How voice calls affect data in operational LTE networks," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 87–98.
[4] Y. Li, J. Xu, C. Peng, and S. Lu, "A first look at unstable mobility management in cellular networks," in *Proc. 17th Int. Workshop Mobile Comput. Syst. Appl.*, 2016, pp. 15–20.
[5] G.-H. Tu, Y. Li, C. Peng, C.-Y. Li, H. Wang, and S. Lu, "Control-plane protocol interactions in cellular networks," in *Proc. ACM Conf. SIGCOMM*, 2014, pp. 223–234.
[6] Y. J. Jia et al., "Performance characterization and call reliability diagnosis support for voice over LTE," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 452–463.
[7] TS 23.272, Circuit Switched (CS) Fallback in Evolved Packet System (EPS); Stage 2, Jul. 2015.
[8] GSM Association, "The 5G era: Age of boundless connectivity and intelligent automation," (2017). [Online]. Available: https://www.gsmaintelligence.com/research/?file=0efdd9e7b6eb1c4ad9aa5d4c0c971e62&download
[9] Ericsson, "4G/5G RAN architecture: How a split can make the difference," (2016). [Online], Available: https://www.ericsson.com/en/ericsson-technology-review/archive/2016/4g5g-ran-architecture-how-a-split-can-make-the-difference
[10] KT 5th Generation Radio Access; Overall Description; (Release 1), (Aug. 2016). [Online]. Available: http://file.kt.com/kthome/business/kt5g/5G_300_v1.2.pdf
[11] SK Telecom, "SK telecom completes field trial for its 5G system," (Apr. 2016). [Online]. Available: https://http://sktelecom.com/en/press/detail.do?idx=1161
[12] Cisco, "Troubleshooting TechNotes," Sep. 2015. [Online]. Available: https://http://www.cisco.com/c/en/us/support/wireless/mme-mobility-management-entity/products-tech-notes-list.html
[13] M. Anehill, M. Larsson, G. Strömberg, and E. Parsons, "Validating voice over LTE end-to-end," *Ericsson Rev.*, vol. 1, pp. 4–10, 2012.
[14] Qualcomm, "CSFB performance," (Oct. 2012). [Online]. Available: https://www.qualcomm.com/media/documents/files/4g-world-2012-csfb.pdf
[15] J. E. V. Bautista, S. Sawhney, M. Shukair, I. Singh, V. K. Govindaraju, and S. Sarkar, "Performance of CS fallback from LTE to UMTS," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 136–143, Sep. 2013.
[16] T. Engel, "xgoldmon," (2013). [Online]. Available: https://github.com/2b-as/xgoldmon
[17] P1 Security, "LTE_monitor_c2xx," (2013). [Online]. Available: https://github.com/P1sec/LTE_monitor_c2xx
[18] D. Spaar, "Tracing LTE on the phone," (2013). [Online]. Available: http://www.mirider.com/weblog/2013/08/index.html
[19] SRLabs, "SnoopSnitch," (2014). [Online]. Available: https://opensource.srlabs.de/projects/snoopsnitch
[20] Y. Li, C. Peng, Z. Yuan, J. Li, H. Deng, and T. Wang, "Mobileinsight: Extracting and analyzing cellular network information on smartphones," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 202–215.
[21] R. Borgaonkar and S. Udar, *Understanding IMSI Privacy*. Las Vegas, NV, USA: Black Hat, 2014.
[22] SecUpwN, "Android IMSI-catcher detector," (2012). [Online]. Available: https://github.com/CellularPrivacy/Android-IMSI-Catcher-Detector/
[23] Forbes, "Qualcomm retains lion's share of LTE baseband market; further gains expected in 2016," (Feb. 2016). [Online]. Available: https://www.forbes.com/sites/greatspeculations/2016/02/24/qualcomm-retains-lions-share-of-lte-baseband-market-further-gains-expected-in-2016/

[24] Qualcomm, "Qxdm," (2012). [Online]. Available: https://www.qualcomm.com/documents/qxdm-professional-qualcomm-extensible/diagnostic-monitor
[25] Innowireless, "Optis-s," [Online]. Available: http://www.innowireless.co.kr/eng/sub.asp?localNum=2&pageNum=1&subNum=1&subNum2=1
[26] Statista, "Ranking of countries/territories by LTE mobile subscribers in 2014," (2015). [Online]. Available: https://www.statista.com/statistics/309599/lte-mobile-subscribers-by-country/
[27] 3GPP TS 23.401, General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, Oct. 2015.
[28] 3GPP TS 25.331, RRC; Protocol specification, Oct. 2015.
[29] 3GPP TS 36.331, Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification, Oct. 2015.
[30] 3GPP TS 24.301, Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3, Oct. 2015.
[31] A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, and J.-P. Seifert, "Practical attacks against privacy and availability in 4G/LTE mobile communication systems," in *The Network and Distributed System Security Symposium*. Reston, VA, USA: Internet Society, 2016.
[32] 3GPP TS 33.401, System Architecture Evolution (SAE); Security architecture, Oct. 2015.
[33] H. Kaaranen, *UMTS Networks: Architecture, Mobility and Services*. Hoboken, NJ, USA: John Wiley & Sons, 2005.
[34] 3GPP TS 36.300, E-UTRA and E-UTRAN; Overall description; Stage 2, Sep. 2015.
[35] H. Kim, et al., "Breaking and fixing VoLTE: Exploiting hidden data channels and mis-implementations," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Security*, 2015, pp. 328–339.

Byeongdo Hong received the MS degree from the Department of Mathematical Sciences, Seoul National University, in 2012. He is working toward the PhD degree in the Graduate School of Information Security at KAIST. His current research interests include several topics related to cellular networks, such as problem diagnosis, performance degradation, security, and privacy.

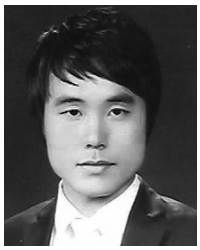


Shinjo Park He received his BS degree from KAIST in August 2012, and his MS degree from the Graduate School of Information Security at KAIST in August 2014. He is working toward the doctoral degree with the Technical University of Berlin. He is interests include various aspects of cellular security, including cellular control plane, cellular network entities, and basebands.

Hongil Kim received the BS and MS degrees from the Department of Electrical Engineering at KAIST, in 2013 and 2015, respectively. He is working toward the PhD degree in the School of Electrical Engineering at KAIST, under the supervision of Prof. Yongdae Kim. His main research interests include security analysis of mobile communications, and the development of security-enhanced mobile technologies.

Dongkwan Kim received the BS and MS degrees in Electrical Engineering and Computer Science from KAIST, in 2014 and 2016, respectively. He is working toward the PhD degree in the School of Electrical Engineering at KAIST. He is interests include various fields of security: software, embedded devices, cellular networks, and sensing/actuation systems. He has participated in several hacking CTFs (DEFCON, Codegate, PlaidCTF, Whitehat Contest, HDCON) as a member of KAIST GoN, and now as a member of KaisHack.

Hyunwook Hong received the BS degree in computer science from KAIST, in February 2009, and the MS degree from the Graduate School of Information Security at KAIST, in August 2013. He was awarded a PhD by the Graduate School of Information Security at KAIST in August 2017. He is interest include the security of cellular networks.



Hyunwoo Choi received the BS and MS degrees from the Department of Information Security, Soonchunhyang University, in 2009 and 2011, respectively. He was awarded a PhD by the Graduate School of Information Security at KAIST, in 2017. His research interests include system security issues for low-level attacks and mobile computing systems.

Jean-Pierre Seifert received the degree in computer science and mathematics from the Johann Wolfgang Goethe University Frankfurt am Main, where he was awarded a PhD in 2000, under the guidance of Prof. Dr. C. Schnorr. Afterward, he gained intensive practical experience in research and development of hardware security at Infineon, Munich, and Intel. At Intel (2004–2006), he was responsible for the design and integration of new CPU security instructions for microprocessors to be integrated in all Intel microprocessors. From 2007-2008, he developed the world's first commercial secure cell-phone for Samsung Electronics. Since 2008, he has been a professor heading the Security in Telecommunications Group at Technische Universität Berlin and Deutsche Telekom Laboratories. In 2002, he received the "Inventor of the Year" award from Infineon, and in 2005, he received two Intel Achievement Awards. Approximately 40 patents have been granted to him in the field of computer security. He is a member of the IEEE.

Sung-Ju Lee received the PhD degree in computer science from the University of California, Los Angeles, in 2000, and spent 15 years in the industry in Silicon Valley before joining KAIST. He is associate professor and KAIST Endowed chair professor at KAIST. His research interests include computer networks, mobile computing, network security, and HCI. He is the winner of the HP CEO Innovation Award, the Best Paper Award at IEEE ICDCS 2015, and the Test-of-Time Paper Award at ACM WINTECH 2016. He is a fellow of the IEEE.

Yongdae Kim received the PhD degree from the Computer Science Department, University of Southern California. He is a professor of the School of Electrical Engineering and an Affiliate professor of GSIS at KAIST. Between 2002 and 2012, he was an associate/assistant professor of the Department of Computer Science and Engineering, University of Minnesota—Twin Cities. Between 2013 and 2016, he served as a KAIST chair professor. He received the NSF career award and McKnight Land-Grant Professorship Award from the University of Minnesota in 2005. Currently, he is serving as a steering committee member of NDSS and an associate editor of ACM TOPS. His current research interests include security issues for various systems such as cyber physical systems, cellular networks, P2P systems, and embedded systems. He is a senior member of the IEEE.