Insecurity of Voice Solution VoLTE in LTE Mobile Networks

Chi-Yu Li, Guan-Hua Tu
University of California, Los Angeles
{lichiyu, ghtu}@cs.ucla.edu

Chunyi Peng
The Ohio State University
chunyi@cse.ohio-state.edu

Zengwen Yuan, Yuanjie Li
University of California, Los Angeles
{zyuan, yuanjie.li}@cs.ucla.edu

Songwu Lu
University of California, Los Angeles
slu@cs.ucla.edu

Xinbing Wang
Shanghai Jiao Tong University, China
xwang8@sjtu.edu.cn

ABSTRACT

VoLTE (Voice-over-LTE) is the designated voice solution to the LTE mobile network, and its worldwide deployment is underway. It reshapes call services from the traditional circuit-switched telecom telephony to the packet-switched Internet VoIP. In this work, we conduct the first study on VoLTE security before its full rollout. We discover several vulnerabilities in both its control-plane and data-plane functions, which can be exploited to disrupt both data and voice in operational networks. In particular, we find that the adversary can easily *gain free data access*, *shut down continuing data access*, or *subdue an ongoing call*, *etc.*. We validate these proof-of-concept attacks using commodity smartphones (rooted and unrooted) in two Tier-1 US mobile carriers. Our analysis reveals that, the problems stem from both the device and the network. The device OS and chipset fail to prohibit non-VoLTE apps from accessing and injecting packets into VoLTE control and data planes. The network infrastructure also lacks proper access control and runtime check.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless Communication*

Keywords

Cellular Networks; LTE; VoLTE; attack; defense

1. INTRODUCTION

Voice is a simple utility service, yet vital to both mobile operators and phone users. It has been a killer application to mobile networks for decades. As the infrastructure upgrades to Long Term Evolution (LTE), the fourth-generation (4G) mobile technology, voice service is also going through its fast evolution. This solution to the 4G network is called VoLTE (Voice over LTE).

CCS'15, October 12–16, 2015, Denver, Colorado, USA.

DOI: http://dx.doi.org/10.1145/XXX.XXXXXXX.

In a nutshell, VoLTE is a Voice-over-IP (VoIP) scheme for the *packet-switched (PS)-only, all-IP* based LTE network [5]. It abandons the circuit-switched (CS), legacy call solution to 2G/3G networks. The design appears straightforward. It carries voice messages in IP packets on the data plane, no longer through the dedicated circuit. To facilitate the voice communication, each VoLTE call also maintains a separate signaling session on the control plane. This is akin to VoIP over the Internet. However, VoLTE adopts cellular-specific techniques to ensure carrier-grade quality. It leverages high-priority, quality-of-service (QoS) offered by the LTE network for both sessions.

Consequently, VoLTE offers clear benefits over its legacy 2G/3G call service, including improved quality (*e.g.*, crystal-clear calls via high-fidelity codecs), more options (*e.g.*, video calling, voicemail and conferencing), and better interoperability (*e.g.*, over mobile networks, WiFi and the wired Internet). It is thus mandatory and designated as the ultimate call solution [5]. As VoLTE shifts its design paradigm from CS to PS, we are interested in whether such substantial changes would possibly imperil the LTE network, as well as mobile users.

In this work, we examine whether VoLTE exposes new and unexpected threats. Our study stems from a simple rule of thumb in that any major change is probably a source for insecurity. With the nontrivial changes from CS to PS in its core technology, VoLTE may interfere with other system components, thereby inducing new loopholes. Technology-wise, we suspect that identical, PS-based operations for voice and data may open the door to run data over VoLTE. Moreover, as IP forwarding is easily accessible by the mobile OS, VoLTE extends the rigid CS access within the device chipset (hardware), to the more open PS access in software (OS and mobile apps). This likely invalidates the existing protection mechanisms and security defenses for traditional voice. Last, different from normal data services, VoLTE has higher priority in resource allocation to ensure better quality. It may inadvertently act as the side-channel to leak critical information.

Our study confirms all above suspicions. VoLTE may disrupt both data and voice. First, the VoLTE control session can be abused to carry PS data packets, beyond voice signaling messages. More threateningly, with no billing upon VoLTE control packets, this exploit leads to *free data access*. Such free service is available to both Mobile-to-Internet and Mobile-to-Mobile communications. Furthermore, the control-plane exploit empowers higher-priority, yet undeserved data access. This shuts down progressing data sessions (data DoS), or imposes data overcharges on the victim. Second, VoLTE is prone to new voice DoS attacks. An unprivileged

Category	Attack	Victim	Description and Threat	Vulnerability
Data (§3)	Free data	Operator	Adversary device gains free data access to the Internet or another mobile device.	V1: Lack of the control-plane access control (§3.1) V2: Imprudent forwarding in the network (§3.1) V3: Abusing no billing of VoLTE signaling traffic (§3.2)
	Overbilling	Individual	Adversary injects spams to impose excessive data bill on the victim.	
	Preemptive data	Operator, Individual	Adversary device gains undeserved higher-priority data access.	V1: Lack of the control-plane access control (§3.1)
	Data DoS	Individual	Adversary shuts down the ongoing data access on the victim phone.	V4: Abusing highest-priority allocated to VoLTE control plane (§3.3)
Voice (§4)	Muted voice (DoS)	Individual	Adversary mutes an ongoing VoLTE call on the victim.	V5: Insufficient data-plane access control (§4.1) V6: Side-channel leakage of data-plane information (§4.1)
	Enhanced muted voice	Individual	Adversary mutes the voice faster.	V5: Insufficient data-plane access control (§4.1) V7: Leakage from improper both-plane coordination (§4.2)

Table 1: Summary of our main findings on VoLTE vulnerabilities and proof-of-concept attacks.

malware can *mute* an ongoing call when the session information is leaked from the side channels on the data plane or improper coordination between both data and control planes. Table 1 summarizes our findings, which are confirmed in two top-tier US carriers. We further propose remedies on both control and data planes, both device and network sides to secure VoLTE.

The identified root causes are also multi-faceted. The mobile technology standards, the device OS and hardware, and the network operations all contribute to the security weakness. First, mobile standards offer loose regulations to operations on both control and data planes. VoLTE carries both its control signaling and voice message over IP packets. However, both planes can be tricked to transmit or receive non-VoLTE packets. The mobile technology standards do not stipulate mechanisms to permit only authentic voice packets in VoLTE. Second, mobile devices cannot prohibit unintended access to VoLTE. Software and hardware rely on each other for protection. Non-VoLTE apps gain access via software to VoLTE, thus injecting packets into the hardware chipset. Meanwhile, the hardware always trusts the traffic coming from the software (OS and apps) and allows it to pass on without due check. Consequently, the device should also share the blame. Third, network operations should also be held accountable. The carrier trusts mobile devices (in fact, their chipsets), and proper defense on the network side is not entirely enforced.

In summary, we conduct the first empirical study on VoLTE insecurity by systematically considering all the dimensions: control plane, data plane, and their coordination. We further confirm our findings over two US carriers. The paper makes three contributions.

1. We identify seven vulnerabilities on its control plane, data plane and coordination between both. They span mobile devices (hardware and software) and carrier networks, ranging from access control, billing policy, to QoS schemes and VoLTE operations.

2. We devise proof-of-concept attacks (*e.g.*, free data access, data overcharging, data and voice DoS attacks) to exploit identified vulnerabilities. We assess their impact in operational networks.

3. We deduce root causes, recommend solutions, and share the learnt lessons. The mobile Internet industry is likely to benefit from such lessons, since it is still at its early stage for full-fledged, worldwide VoLTE deployment. Note that the security of control and data planes is a generic problem for different network services. The exposed issues and learned lessons from VoLTE can also be applied to them.

The rest of the paper is organized as follows. §2 gives the background of VoLTE, its potential vulnerabilities, and the attack model

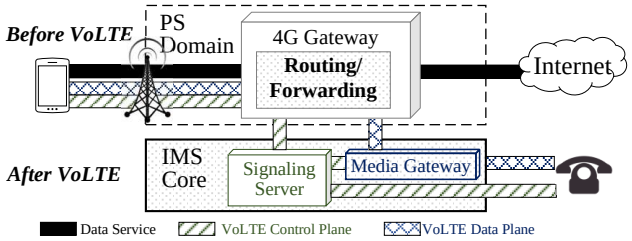


Figure 1: LTE network architecture for VoLTE.

of this work. In §3 and §4, we disclose vulnerabilities of VoLTE and sketch attacks in terms of VoLTE control and data planes, respectively. We then propose recommended fix in §5, discuss several remaining issues in §6, updates ongoing efforts to fix these problems in §7. § 8 and §9 present related work, and conclusion, respectively.

2. NEW SECURITY ISSUES WHEN VOLTE TURNS VOICE INTO "DATA"

In this section, we first review VoLTE and then identify its potential vulnerabilities. We describe the attack model and our assessment methodology.

2.1 VoLTE Primer

VoLTE is projected as the primary voice solution to the LTE users. It migrates the legacy, circuit-switched (CS) voice service to the packet-switched (PS) design.

Network architecture for VoLTE. Figure 1 depicts a simplified architecture to support VoLTE. Two subsystems are involved. The first is the PS delivery subsystem (top), which exists before VoLTE is enabled. Its role is to offer the PS connectivity to and from mobile devices, thus accommodating versatile data services. The core component is the 4G gateway, which forwards packets, akin to edge routers in the Internet. It also provides certain control utilities such as policy enforcement and data volume billing. The second is the IP Multimedia Subsystem (IMS, bottom), which supports IP telephony and multimedia services [9]. It consists of two key elements: the media gateway and the signaling server. The former is to deliver multimedia (*e.g.*, voice) traffic to VoLTE users or traditional telephony users. The latter is to perform call control functions among the device, the media gateway and the 4G gateway.

How VoLTE works? As illustrated in Figure 1, each VoLTE call maintains two communication sessions. The control-plane session is to exchange the call signaling messages through the popular Session Initiation Protocol (SIP) [30]; it is established and remains active as long as the VoLTE feature is on. The data-plane session

handles the voice packet delivery, *e.g.*, via the Internet Real-time Transport Protocol (RTP) [6]; it is established on demand by the control session. Note that no dedicated communication channel (circuit) is reserved between the caller and the callee. Instead, all the voice traffic and signaling messages are carried in packets and delivered over IP. As a result, the 4G gateway not only relays data packets to/from the Internet for ordinary mobile broadband services, but also routes packets on both control and data planes between the device and the IMS core.

To ensure carrier-grade call quality, VoLTE leverages the multiple service classes (*e.g.*, the guaranteed bit rate and different priorities [8]) offered by LTE. Its data plane is carried by the guaranteed-bit-rate class, which assures the lowest bit rate. Its control plane uses the non-guaranteed-bit-rate category, but with the highest priority. In both cases, VoLTE signaling and voice are delivered with higher priority than data services.

2.2 Potential Vulnerabilities

Ultimately, voice and data operate in the same, *connection-less* IP network. However, this paradigm shift is double-edged, exposing LTE networks and users to unanticipated vulnerabilities.

In this paper, we look into three security aspects.

1. How to trick VoLTE to gain PS data access, despite its designated role for voice? Technically, this relates to the access control fences to VoLTE at the device and the network.
2. How to learn private, critical information on voice calls from VoLTE? Note that VoLTE is IP based and more open and accessible than the legacy CS call service.
3. Will the voice-related policies and operations (*e.g.*, voice billing and QoS control) work well in the VoLTE context? If not, what are the imposed threats to LTE?

Our study covers three aspects of VoLTE operations: control-plane, data-plane, and the coordination between control and data planes. Such security issues span the device, the 4G Gateway and the IMS core. In the following sections, we disclose how the currently employed or newly developed mechanisms fail to harden VoLTE against attacks and how they are exploited to menace data services (§3) and voice calls (§4).

2.3 Attack Model and Methodology

The presumed attacker is a mobile user, whereas the victims can be the network operator or/and other mobile users. The adversary uses a commodity smartphone rooted to gain full programability. However, (s)he has no remote access, at least no privileged access to the victim phones. In some attacks (*i.e.*, data DoS, overbilling and voice DoS), an unprivileged malware is required to monitor basic activities and information (*e.g.*, when the data transfer starts and the IP information of network interfaces) on the victim phones. The voice DoS also requires the malware to generate spam traffic. In all cases, the attacker has no control over the carrier network. The network is not compromised.

To validate vulnerabilities and attacks, we conduct experiments in two top-tier US carriers denoted as OP-I and OP-II[1]. They together represent almost 50% of market share. We use two Android phone models that support VoLTE: Samsung Galaxy S5 and LG G3, running Android 4.4.4 and 4.4.2, respectively. Note that VoLTE functions on only a few recent models, because it requires phone hardware and software upgrades (its rollout in US started in 2014). Both rooted and unrooted ones are tested. We focus on the Android OS but we believe that the identified issues are applicable to any other OS. The results also apply to both carriers unless explicitly specified.

We bear in mind that some feasibility tests and attack evaluations might be detrimental to users or operators. Therefore, we conduct this study in a responsible manner through two measures. First, we use only our own phones as the victims. Second, in those tests beneficial to mobile users (*e.g.*, free data access), we carry out experiments without exploiting them for real benefit. Specifically, for each phone, the total volume of free data services we test is always less than that of data plan we have. We seek to disclose VoLTE vulnerabilities and effective attacks, but not to aggravate the damages. We do not claim that the attacks are not the most powerful ones to make damage.

[1]We hide their names to protect both carrier while working with them to fix the identified issues.

3. PERPETRATING MOBILE DATA SERVICE IN VOLTE CONTROL PLANE

The first uncovered problem is that VoLTE can be exploited to carry mobile data service, which is unintended by its designers. VoLTE is developed to support calls but it is *not restricted to voice* in operation. PS is employed to exchange VoLTE signaling messages on the control plane. However, it has not been hardened against access to non-VoLTE traffic (*i.e.*, normal PS data service). We discover that this can be indeed tricked to sustain two forms of Mobile-to-Internet and Mobile-to-Mobile data access in operational networks. More threateningly, the billing scheme and quality control for VoLTE can be abused to endanger LTE networks and users. In the former, such unintended data access can bypass its due charges, thereby yielding *free* data service, or result in overbilling at victims. In the latter, it unjustly gains the highest priority assigned to VoLTE signaling. It can be further manipulated for a DoS attack against normal data service.

3.1 Carrying Data In VoLTE Signaling

While VoLTE intends to use PS packets to carry signaling messages, it is *never forbidden* from turning PS data into VoLTE (signaling). Similar to the data service that retains a bearer (*i.e.*, IP connectivity), VoLTE also has a signaling bearer for its control-plane operation. As shown in Figures 2a and 2b, both need to first activate a bearer and obtain an IP connectivity within the LTE network. Afterwards, data packets can be delivered through this bearer once any service starts. The device sets the source address as that allocated by the 4G Gateway and the destination as the target host's. For VoLTE, upon any call request, SIP messages are exchanged between the device and the IMS core through the signaling bearer. It then on-demand invokes a voice bearer to carry conversation traffic if the call is accepted. When the call ends, the voice bearer is released.

With the packet-carrying capability, it is feasible to carry any data through the VoLTE signaling bearer under two vulnerabilities. First, on the device side, there is no access control to prevent non-VoLTE packets from being injected into the signaling bearer (*V1*). Second, on the network side, these injected packets are allowed to pass by (*e.g.*, routed to the destination by the 4G gateway, *V2*).

V1: Lack of Access Control at Phone Software & Hardware

Access control on the control plane is to ensure its exclusive use by *authentic* VoLTE signals. However, we discover that the device lacks bullet-proof access control to the VoLTE control plane. Figure 3 shows the current practice on the mobile device. We also plot the one for the CS voice for comparisons.

There are two options for access control at the phone: hardware (*i.e.*, 4G/3G chipset) and software (*i.e.*, OS and apps) based.

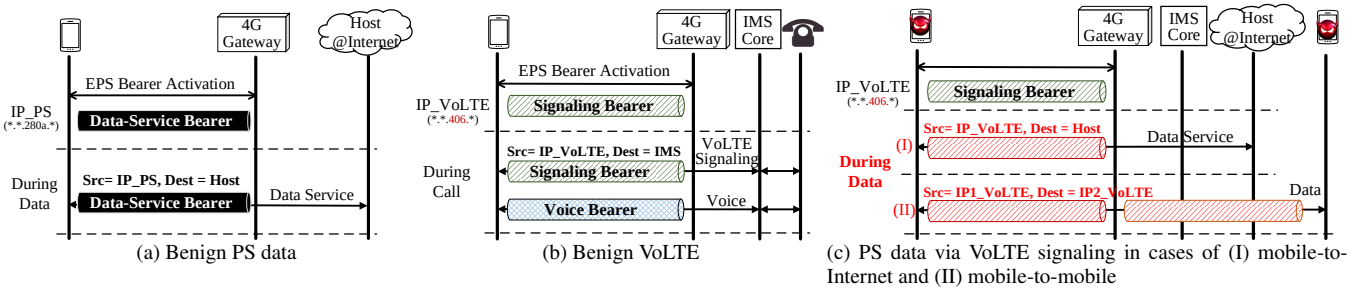


Figure 2: Typical procedures for (a) PS data service, (b) benign VoLTE and (c) PS data via VoLTE signaling.

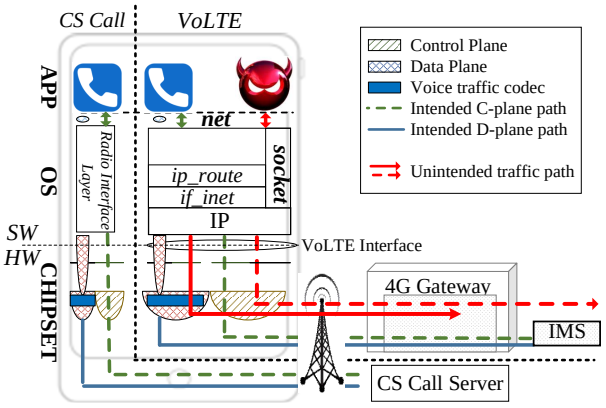


Figure 3: VoLTE Access control on the device side.

For CS calls, all signals are handled within the chipset, never exposed to OS and apps. There is no easy way to hijack them unless the hardware is compromised or special debugging mode is invoked by developers. In contrast, the VoLTE signaling access is exposed to the mobile OS, where a network interface, called VoLTE interface, is created for VoLTE only. There are a variety of reasons to take the software approach: VoLTE adopts the Internet protocols (IP and SIP) that are well supported in the OS (*e.g.*, `android.net.sip.*` library); the software scheme offers high flexibility (*e.g.*, easy upgrades) and rich information for the OS and apps to optimize performance. By design, only genuine signals can traverse the VoLTE interface and enter the underlying chipset.

However, the VoLTE interface has not been hardened against non-authorized access. Other unprivileged apps can easily obtain the VoLTE interface information, as they do to the mobile data interface (see Figure 4 (left) for a two-interface example where IPv6 is used). In fact, information can be directly retrieved from the `net` settings in the OS. For example, in our Android phones, IP address is obtained from `/proc/net/if_inet6`, and the signaling server's IP address from the routing table (`/proc/net/ipv6_route`). Moreover, it is feasible for the adversary to inject non-VoLTE data packets (the red dashed lines in Figure 3). The attacker without root privilege can specify its destination to any of the VoLTE-related servers. Given the default routing table with rules to them, unprivileged apps may inject packets into the VoLTE signaling bearer, and such packets are routed to those VoLTE servers. With root privilege, the adversary can add a routing rule to any destination for the VoLTE interface. He is thus able to inject packets to any target via the signaling bearer.

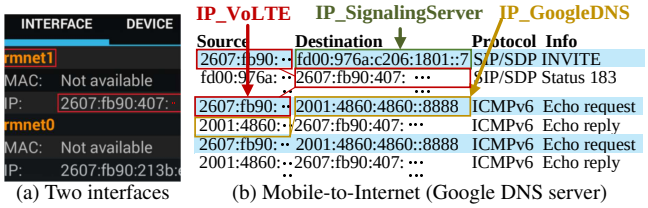
Note that we test all three popular types of traffic in the empirical studies of this section: UDP, TCP, and ICMP. The exposed vulnerabilities may only exist for certain protocols and ports. We use one type of traffic to show the feasibility of each vulnerability. All the other types of traffic are considered to have similar results if not explicitly specified.

Empirical validation. We confirm the above vulnerability through the following tests. First, an unprivileged app can obtain the interface for the VoLTE signaling bearer (`rmnet1`), as well as the one for PS data (`rmnet0`). We learn that `rmnet1` belongs to VoLTE because it appears/disappears when VoLTE is enabled/disabled. Figures 4 (left) and 5 (left) show the snapshots at two mobile phones in OP-I, captured by Network Info II, an Android app [3]. In both operators, IPv6 is used. IP addresses for both interfaces are different. Note that the roles of `rmnet0` and `rmnet1` may be swapped. We can infer them based on the routing table. The one assigned to the default routing rule is for the PS data service, and the other is for VoLTE. Second, we validate that the unprivileged applications are able to inject non-VoLTE traffic into the signaling bearer via the VoLTE interface. The test works as follows. We send a UDP packet with Hop Limit being set to 1, to the VoLTE signaling server and then receive an ICMP packet from the VoLTE gateway via the VoLTE interface. It is from the gateway because this IP address differs from that serving PS data services. This implies that this UDP packet is indeed sent through the signaling bearer, out of the phone.

Causes and lessons. We further examine why neither software nor hardware at the phone provides proper access control for the VoLTE signaling bearer. In general, the OS employs permission control or uses execution container to protect network access (*e.g.*, WiFi). However, it does not invoke system permission control for VoLTE. This is possibly because no corresponding mechanism in the OS distinguishes the network interface dedicated to VoLTE from that for the Internet data access. From the OS standpoint, accessing to both interfaces is identical, without any extra requirement from the underlying chipset. On the hardware side, the current practice fully relies on the software protection, and always trusts the traffic coming from the software's VoLTE interface. Concerted effort in both software and hardware to protect the VoLTE access is missing, when the chipset opens up the access to the OS.

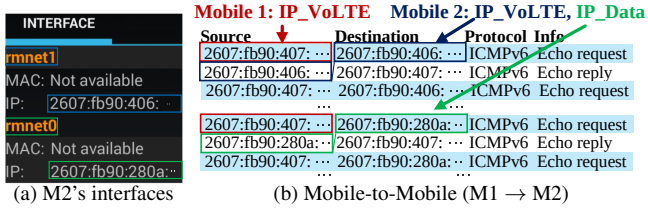
V2: Imprudent Routing and Forwarding in the Network

The next weakness lies on the network side in its routing and packet forwarding. It leads to two unexpected consequences. First, traffic carried through the VoLTE signaling bearer is not verified at runtime. Non-authentic control packets can be forwarded by the network. Without runtime filtering, the packets not destined to the VoLTE servers in the IMS core, can sneak in through the VoLTE bearer.



(a) Two interfaces (b) Mobile-to-Internet (Google DNS server)

Figure 4: An example of Mobile-to-Internet data service via the VoLTE signaling interface (`rmnet1`).



(a) M2's interfaces (b) Mobile-to-Mobile (M1 → M2)

Figure 5: An example of Mobile-to-Mobile data service from the VoLTE signaling interface (`rmnet1`) of Mobile 1 to Mobile 2. Both VoLTE-to-VoLTE (`rmnet1`) and VoLTE-to-DATA (`rmnet0`) are supported.

Second, routing rules within mobile networks are prone to abuse. VoLTE phones need to exchange signaling messages with each other via the VoLTE signaling server. When the routing rule toward each phone exists for the signaling bearer at the 4G gateway, phones can communicate with each other *without* reaching the signaling server. It thus enables direct, Mobile-to-Mobile communication. Meanwhile, it also opens the door to non-VoLTE Mobile-to-Mobile data communication. It is likely that routing rules to the rest of the Internet exist; this facilitates the Mobile-to-Internet data access through the VoLTE signaling bearer.

Empirical validation. The VoLTE signaling bearer can be misused to perform both (i) Mobile-to-Internet and (ii) Mobile-to-Mobile data service (Figure 2c). The former is only feasible in OP-I, whereas the latter works for both.

(i) Mobile-to-Internet: We observe message exchange between the phone and an external server through the VoLTE interface. We first uses `rmnet1` to `ping` a Google public DNS server (its address is `2001:4860:4860::8888`). Figure 4b shows partial packet traces collected via a `tcpdump`-like traffic sniffer, Shark [4]. The first two SIP messages indicate that the `rmnet1` interface is indeed used for VoLTE signaling. The following ping request and reply exchanged between the phone and the Google DNS server reveal that, it is viable to exploit the VoLTE signaling bearer to deliver normal data traffic. Both inbound (downlink) and outbound (uplink) data transfers are feasible. We also deploy an IPv6 server outside the mobile network and repeat the test, and similar results are observed.

(ii) Mobile-to-Mobile: We find that VoLTE can be exploited to directly communicate with another mobile device belonging to the same carrier. We send ICMP Echo Requests from the VoLTE interface of one phone (Mobile 1) to another phone (Mobile 2) via its VoLTE and data-service interfaces (see both interfaces in the left plot of Figure 5). The right plot shows Mobile 1 receives ICMP Echo Reply packets in both VoLTE-to-VoLTE and VoLTE-to-Data cases in OP-I. This confirms that both forms of Mobile-to-Mobile communication are feasible in OP-I. We find that OP-II supports only the VoLTE-to-VoLTE option for Mobile-to-Mobile communication, but permits only UDP traffic instead of ICMP. This implies that more defenses are implemented in the core network of OP-II, but they are still insufficient to guard against the VoLTE exploit.

We also examine the feasibility of protocol variants. Both UDP and ICMP work for OP-I, whereas only UDP is allowed for OP-II. The slight difference is that some, but not all UDP ports work in OP-I whereas almost all the ports can work in OP-II. In OP-I, the viable UDP port varies in the tests and requires pre-scanning. This divergence reflects the operator's freedom in making their own policy and implementation. However, TCP is not allowed for both carriers in any case. Later, we will demonstrate that the attack is always viable as long as at least one protocol is allowed to traverse the core network over VoLTE signaling. Any real traffic (TCP or UDP) can be encapsulated in ICMP/UDP tunneling.

Causes and lessons. The operator does not properly regulate routing and packet forwarding for the VoLTE signaling bearer. On the network side, the carrier does not enforce access protection for VoLTE, similar to CS voice calls and normal PS data. Once this bearer is allocated for the VoLTE control plane, the network relies on the phone to forward authentic signaling messages (unfortunately, it is not guaranteed). This imprudent practice ignores the distinction of VoLTE from CS calls and normal PS data (see V3 and V4 exploits later). Packets carried by the signaling bearer from the phone, should only reach the VoLTE signaling server, but not another phone or an Internet host, and vice versa.

3.2 Exploiting VoLTE for Free Data Access

Taking charging into account makes the unintended data access more threatening. The practice is that data/voice billing has never taken VoLTE signaling messages into account (*V3*). If the traffic is delivered through the signaling bearer, it is free of charge. This remains valid regardless of whether the traffic is destined to the signaling server or not.

V3: Abusing No Billing of VoLTE Signaling

VoLTE control signals are free of charge. Any packets via VoLTE signaling bearer are free, no matter whether their destination is the VoLTE signaling server or not. Therefore, such unintended data transfer is treated as VoLTE signaling and bypasses the billing mechanism for normal PS data access. Typically, mobile data access is charged based on volume (*i.e.*, the number of delivered bytes).

It is not surprising for operators to make VoLTE signaling delivery exempted from charging. Since VoLTE continues to offer call service, it is natural to adopt the time-based charging, following the common practice for traditional CS voice. Consequently, only the call duration on the data plane is collected for billing. VoLTE control messages are used to facilitate voice calls and should be free of charge. Moreover, some signals are even exchanged before the call is established, for example, `SIP-INVITE`, `SIP-INVITE-OK`, `SIP-INVITE-ACK` messages are used to set up a call. However, the practice to supply free VoLTE signaling does have loopholes. The operators does not enforce that all packets going through the VoLTE signaling bearer are indeed control messages. Even worse, there is no effective mechanism to limit the traffic volume going through it. As a result, this can be readily abused to make "free" data service.

Empirical validation. We first show that genuine signaling messages (through `rmnet1`) are free of charge. We generate excessive such messages by attempting to make many calls; every 15 seconds, we dial and immediately hang up before the call is answered.

QCI	Priority	Bearer Type	Delay	Loss	Traffic Example
1	2	Guaranteed	100 ms	10^{-2}	**VoLTE calls**
2	4	Bit Rate	150 ms	10^{-3}	video call
...	...	(GBR)	...	...	...
5	1	non-GBR	100ms	10^{-6}	**VoLTE signaling**
9	9		300ms	10^{-6}	Web, Email, *etc.*

Table 2: Bearer QoS configurations from [7].

This lasts for 10 hours. We deliver 42.4 MB control messages in total, but none is charged in the data bill. Moreover, no minutes are charged since no call is made through.

We further examine fake signals bypassing the server are still free. We test it with the VoLTE-to-VoLTE internal case in Figure 5. In an experimental run, we send 5000 ICMP Echo Requests (each carrying 1 KB) and receive 4914 ICMP Echo Replies; The traffic is about 10 MB (both uplink and downlink), but neither volume-based nor time-based billing is incurred. Similar results are observed in other test runs.

Causes and lessons. There is nothing wrong to practice the policy of free VoLTE signaling. However, people do have incentives to exploit any transfer that is free. This requires either bullet-proof access control or no free-of-charge policy, or both. Compared with another free data access via DNS tunneling disclosed in [23], VoLTE signaling faces more challenges. It aims to continue the traditional business model and VoLTE accounting only logs the time duration for the data-plane voice. The 4G gateway executes volume-based accounting for data access, but does not record the usage volume for the VoLTE signaling bearer.

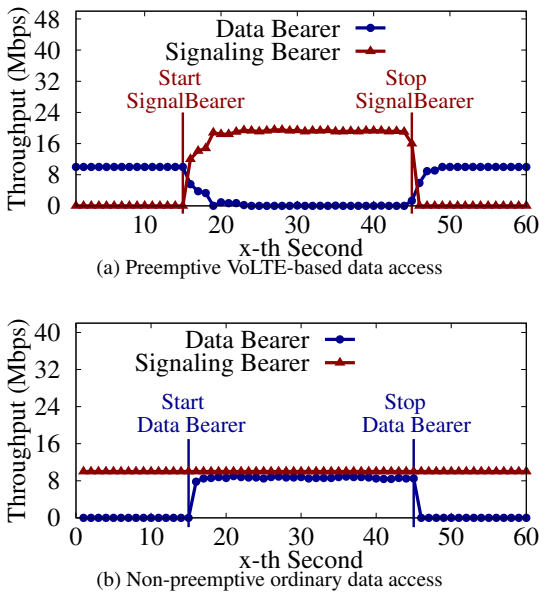
3.3 Manipulating Data Access Priority

Unsurprisingly, the VoLTE-exploited data access can obtain higher, yet underserved priority. The high priority is assigned to provide QoS for VoLTE. However, the expedited delivery at high priority hurts normal PS data services, particularly during network congestion.

V4: Abusing high QoS of VoLTE Signaling

A prominent feature of VoLTE is its capability to guarantee high quality for voice calls. Table 2 lists relevant bearer configurations specified in the 3GPP standard [7]. Each bearer is associated with a QoS class of identifier (QCI), which defines the IP packet characteristics in terms of priority level, bandwidth guarantee, packet delay and loss. The VoLTE signaling bearer has highest priority level (*i.e.*, 1) whereas the data bearer (*e.g.*, web, video streaming) has the lowest one (*i.e.*, 9). VoLTE-exploited data access can further suppress normal PS data with preemptive privileges. Note that both belong to the non-guaranteed bit rate category, whereas the voice bearer with $QCI = 1$ has guaranteed bit rate (GBR). GBR states that the bearer is granted dedicated network resources over both radio and wired links and assured voice quality by ensuring the average rate given a period of time. This will be exploited later in §4.1.

Empirical validation. We validate it through two comparative experiments: (1) During a long-lived downlink data session (10Mbps), we launch another VoLTE-exploited data access. The source rate is larger than the affordable downlink throughput (30 Mbps) and lasts between the 15^{th} second to 45^{th} (Figure 6a); (2) We swap the launch ordering for normal data session and the VoLTE-exploited one (Figure 6b). It is seen that the VoLTE-exploited data access has higher priority, thus suppressing the normal one with preemptive privileges. The data bearer throughput rapidly shrinks



(a) Preemptive VoLTE-based data access

(b) Non-preemptive ordinary data access

Figure 6: Preemptive VoLTE-exploited data access: (a) VoLTE-exploited data access (15s,45s) suppresses the ordinary PS data, (b) PS data can not affect VoLTE-exploited data access

to zero when the downlink resource is captured by the signaling bearer. On the contrary, the data session cannot affect the throughput of the signaling bearer. It can only grab the remaining resource and its peak throughput is throttled.

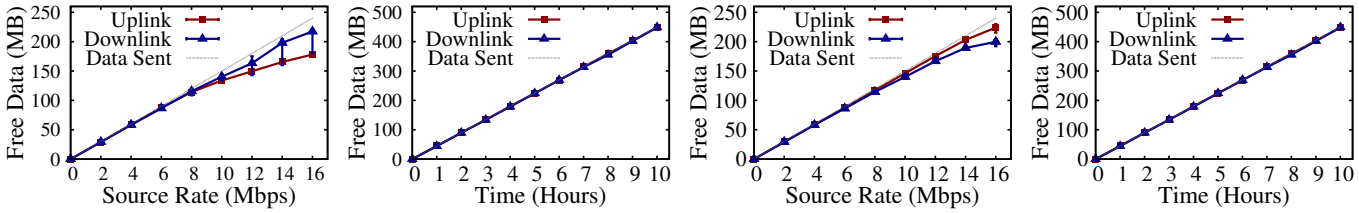
Causes and lessons. Similar to V3, nothing seems to be wrong for the operator to offer higher QoS to VoLTE. However, without prudent traffic filtering, it likely becomes an incentive for the adversary to exploit VoLTE.

3.4 Proof-of-concept Attacks

We devise three proof-of-concept attacks: (1) free data service; (2) data DoS; (3) data overcharging. They are illustrated in the top, middle, and bottom of Figure 7, respectively. The first one works for both operators, whereas the last two are feasible for only OP-I. All are easily launched and also damaging. Note that the last two attacks do not require root privilege at the victims.

Free-data attack. Clearly, the above loopholes can be exploited to gain free external (Mobile-to-Internet) and internal (Mobile-to-Mobile) data access. Note that the free external service works for only OP-I, but the free internal service is feasible for both. Take the OP-I as an example. The attacks work as follows. The adversary leverages ICMP tunneling to deliver data through the signaling bearer, since the ICMP packets are always allowed to be forwarded by the 4G gateway to the Internet or another mobile phone. Each data packet is encapsulated as an ICMP packet by using `Raw Socket`. Moreover, the routing table needs to be updated with the routing rules of designated destinations, so the ICMP packet can be sent via the signaling bearer to the destinations. These two operations can only be performed on a rooted phone. In the external case, we deploy a tunneling server out of mobile networks to run ICMP tunneling. In the internal case, the ICMP tunneling is between two VoLTE phones.

We test with various traffic source rates (up to 16 Mbps) and execution time (up to 10 hours) for both cases. Figure 8 shows that no sign of limit exhibits on the volume, throughput, and duration



(a) Mobile-to-Internet data w.r.t rate (b) Mobile-to-Internet data w.r.t time (c) Mobile-to-mobile data w.r.t rate (d) Mobile-to-mobile data w.r.t time

Figure 8: The volume of free data almost linearly increases with regards to (w.r.t) traffic source rate and run time in external (a,b) and internal (c,d) cases.

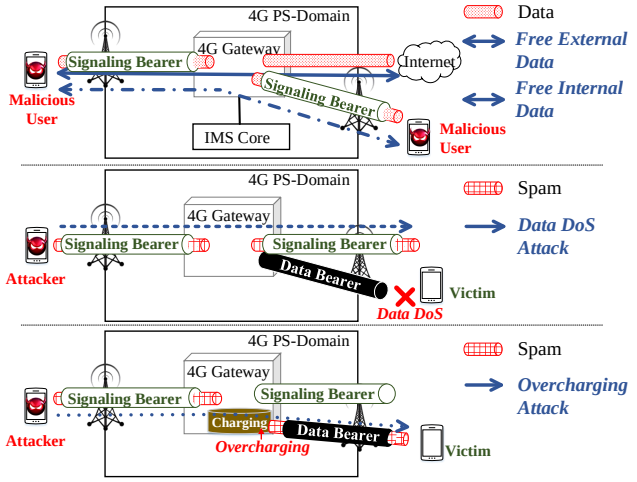


Figure 7: Illustration of three proof-of-concept attacks.

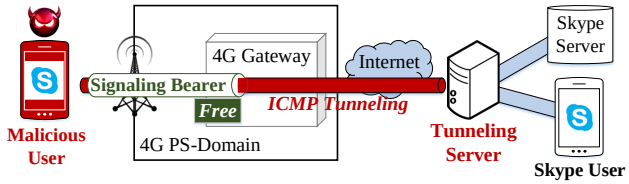


Figure 9: Illustration of the Skype service over ICMP tunneling for free VoLTE-exploited data access.

for free data service in the external and internal cases, respectively. In a test run, 450MB is observed free of charge.

Data DoS Attack. This attack aims to shut down any ongoing data service at the victim by leveraging higher-priority access yielded by VoLTE-exploited data transfer. The attacker injects high-rate spamming traffic through its signaling bearer, to the victim phone's signaling bearer. It can grab all the downlink bandwidth of the victim's data service, thereby causing data DoS. Note that the attacker and the victim are not charged on this spamming traffic, which is carried by the signaling bearers.

This requires an unprivileged malware on the victim device, which detects whether any data service starts, similar to the off-path TCP hijack attack [27, 28]. Once the victim starts any data service, this malware will send a message to an attacker server or an attack phone, leaking the IP address of the VoLTE interface. Afterwards, the attacker starts to inject high-rate spamming data to this IP. In the cases of rush-hour traffic (*e.g.*, 11am-1pm at a campus restaurant), it is observed that the data bearer throughput can be restrained to be zero, under a 10Mbps VoLTE-exploited flow. Note that the required volume of the VoLTE-exploited flow for data DoS varies with the traffic load of the network where the victim is.

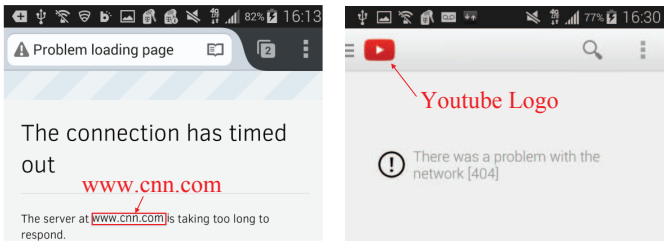
Overcharging Attack. The attacker can make the victim suffer excessive overcharge through injecting data from its signaling bearer into the victim phone's *data-service bearer*. There is only one difference from the above DoS attack. The DoS attack spams data toward the victim phone's *signaling bearer*. The chosen victim is an individual phone user, targeted or randomly picked. Given the victim's IP address, we uncover this data spamming can occur without consent from the victim. The IP address can be learned from a phishing Website or an unprivileged malware. Compared with other spamming attacks [18, 23, 24], this threat readily bypasses the firewall and security boxes. This is because they are always deployed at the border of mobile networks to prevent malicious traffic from the Internet. However, the spamming caused by VoLTE purely relies on the internal traffic without reaching the Internet. In one run in OP-I, the overcharged volume reached 449 MB, still showing no sign of limit.

3.5 Attacks on Real Apps

We further apply two attacks of free Mobile-to-Internet data service and data DoS to real apps. With the former, we use the Skype service with ICMP tunneling over mobile network for free-of-charge. The latter force both the Web browser and Youtube to abort at the victim's phone. Note that these two attacks are feasible only for OP-I.

Free Skype service over mobile network We build an ICMP tunnel between the phone and our external server, in order to exploit the free VoLTE-exploited data service. As shown in Figure 9, we deploy a tunneling server outside the mobile network. It sits between the phone, and application servers or other communicating hosts. At the phone, we create a virtual interface and modify the routing table to achieve two purposes. First, it is set to be the default interface for apps so that all the apps' packets are forwarded to it. Second, all the packets from the interface are redirected to our tunneling server. The server does encapsulation/decapsulation and the packets relay for the phone. Thus, the application servers or other communicating hosts do not require to be modified.

We then run the Skype app on top of it. We show that a malicious user with ICMP tunneling over the mobile network can have a 10-minute chat with another Skype user. Moreover, the data volume consumed during this time period is free-of-charge. Note that other apps can also be free-of-charge over mobile network by taking advantage of ICMP tunneling.



(a) Web DoS (b) Youtube DoS

Figure 10: Data DoS attacks against web browsing (cnn) (a) and Youtube (b).

Data DoS on Web browser and Youtube We launch Data DoS on a victim's phone while s(he) is loading the CNN webpage with the Web browser or watching Youtube. We send a 10Mbps VoLTE-exploited flow of spam to the phone while it is placed in the condition of rush-hour traffic. It is observed that both CNN browsing and Youtube watching are forced to abort, as shown in Figures 10a and 10b.

4. MUTING VOICE THROUGH SPAMS IN VOLTE DATA PLANE

We further investigate insecurity on the VoLTE data plane, as well as the coordination between it and the control plane. We discover that the data plane is not well protected, regardless of certain mechanisms to protect the confidential voice session information, compared with those on the control plane. Unprivileged apps, which do not require rooting or jailbreaking the phone, can inject non-voice junk into voice bearers. Moreover, the private information can be leaked from both the designated scheme for voice QoS and the coordination lapse. By exploiting them, we devise a novel voice DoS attack where a VoLTE call can be made through, but its voice is muted (*i.e.*, the caller and the callee cannot hear each other).

4.1 Injecting Packets into the Voice Bearer

In the data-plane operation, a voice bearer is built on-demand upon any call request, and then released after the call ends. Given the PS nature, it is also vulnerable to the injection of non-voice packets. However, compared with the signaling bearer, the voice bearer appears safer with two inherent protection mechanisms. First, voice packets are handled by the hardware (chipset) without software intervention. As shown in Figure 3, their payload must be encoded or decoded by the IMS (VoLTE) stack in the chipset (*e.g.*, Qualcomm Snapdragon Processors), without reaching mobile OS. Therefore, without hardware hack, it is unlikely for any app above the mobile OS to deliver *valid* data through the voice bearer. Second, each RTP session identifier (*i.e.*, the destination IP and a pair of ports for RTP and RTCP) is protected as a secret, not being exposed to the OS. This information is encrypted in the signaling messages and varies with each call. It is difficult for the adversary to get the session ID and forge the packet header (with correct session ID).

However, after thorough analysis, we find that the current defenses are still insufficient to protect the data plane. It is vulnerable to deliver *invalid* packets (junk data) through the voice bearer, though it fails to deliver valid ones. We unveil two vulnerabilities in the data plane. First, the VoLTE data-plane access control is problematic (*V5*) so that it is possible to inject data to the voice bearer, though the delivered bytes are beyond control by apps. Second, the confidential VoLTE data-plane session information can be inferred from its salient features (the guaranteed bit rate for its QoS configuration), thereby being exposed to malicious exploits (*V6*).

V5: Insufficient Data–Plane Access Defense at Phone

The data plane is also without sufficient access control. Compared with the control plane, it generates all voice packets within the hardware. Specifically, the voice codec that converts an analog voice signal to digitally encoded bytes, is implemented within the chipset, as illustrated in Figure 3. However, the seemingly secure hardware protection mechanism is still not sufficient. It never restricts the access to authentic VoLTE calls only (*i.e.*, the system-level dialer app). Instead, it accepts traffic injected by other apps, even those unprivileged ones, if they get the correct session information. One thing worth noting is that the hardware buffer for the voice bearer may overflow if the injected traffic exceeds the maximum bit rate (MBR) (*e.g.*, tens of kbps), which caps the voice bearer traffic. Genuine voice packets might be consequently discarded, thus degrading the voice quality.

Empirical validation. We confirm that an app without root privilege can inject high-rate traffic into the voice bearer. We run this application during an ongoing call at the callee, and generates packets with the voice RTP session identifier (*i.e.*, destination IP and RTP/RTCP ports) of the ongoing call at 10 Mbps rate, and sends them via the VoLTE interface. We will disclose how we exact the voice RTP session identifier in *V6* and V7. We run 20 tests and consistently observe that the callee's voice is muted at the caller (*i.e.*, no voice from the callee). This implies that the data packets created by the unprivileged app has been successfully injected into the voice bearer and the injected traffic indeed overflows the uplink buffer of the voice bearer at the callee, with most voice packets being discarded.

Causes and lessons. The current two defenses are still ill-equipped, without authenticating the origin of voice traffic. The first guard comes from the VoLTE common operation that the data plan follows the legacy design (*i.e.*, CS voice). Voice traffic is encoded/decoded by specific codec in the hardware, so it does not require software intervention. It inherently prevents the non-VoLTE apps from abusing the voice bearer. However, it merely reduces the hijacking likelihood but cannot avoid hijacking. The second defense relies on the secrecy of RTP session ID. Neither checks whether the traffic comes from the real VoLTE application. The hardware allows for the voice transfer from the OS, which still permits traffic from non-VoLTE apps.

V6: Side-channel Leakage of Session Privacy

The ID of each RTP session is regarded as a session secret [26]. It is carried by the signaling messages of the VoLTE application, and may be further encrypted. Without root privilege, other unprivileged apps should not be able to capture the signaling, thereby learning the session ID. Figure 11 shows an example of the session ID in a SIP message. Note that we get it by decrypting the encrypted SIP message from OP-I according to the method [2]. Note that the decryption requires root privilege.

We however propose an approach to obtain the ID through side-channel hints without root privilege or call operation permission. It contains two parts. First, the destination IP address (*i.e.*, the media gateway's IP) can be easily retrieved from the routing table. Second, the RTP and RTCP port numbers can be inferred from a unique pattern due to its guaranteed-bit-rate QoS scheme. Specifically, the standard regulates that the voice bearer should be guaranteed with

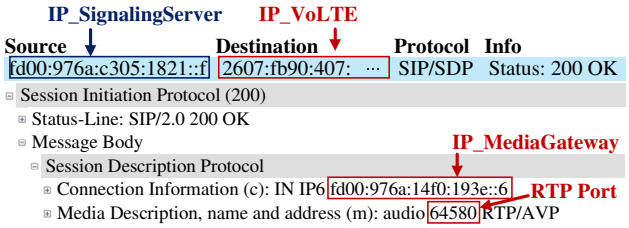


Figure 11: A decrypted SIP message which contains the description of the RTP session.

a minimum rate (*e.g.*, 8 KB/s) [10], regardless of whether packets over the highest-priority signaling bearer would be served.

Moreover, the VoLTE signaling and voice bearers use the same IP allocated to the VoLTE interface, and the packets for these two bearers are differentiated based on their corresponding ports. That is, only the packets with the RTP and RTCP ports of the session ID are delivered to the voice bearer, whereas the others are to the signaling bearer. Hence, it is possible to learn the session ports by scanning all ports (sending one packet via one port). In case of injecting heavy traffic into the signaling bearer, the ones over the voice bearer would have much smaller delay, since they have guaranteed resource. In fact, the two ports with the smallest delay should be the ones used by RTP and RTCP.

Empirical validation. We focus only on the destination ports of the uplink RTP session (*i.e.*, RTP and RTCP ports), since the destination IP can be obtained as disclosed in §3.1.

During an ongoing call, the application without root privilege does two things. First, it scans each port by sending one packet. Second, it keeps sending many packets to certain ports (*e.g.*, 80) which are definitely not for RTP and RTCP, in order to overwhelm the signaling bearer. We consider one-hop RTT for each port, where we sent a UDP packet with Hop Limit being set to 1, and receive ICMP response. The one-hop RTT is calculated based on the time difference between the sending time of the UDP packet and the receipt time of the ICMP response. Figure 12a plots the perceived delay in one test run. The packets with two destination ports, 64580 and 64581, have the smallest delay, 39 ms, whereas other ports experience larger delay ($>$ 90ms). These ports match those disclosed from the decrypted SIP messages. Figure 12b shows that the delay gap between the RTP/RTCP ports and other ports is consistently observed, with at least as large as 50 ms in the 20 tests. So it is viable to infer the RTP port numbers, though they vary in each run.

Causes and Lessons. Two factors may trespass session privacy. First, signaling and voice packets are dispatched from the VoLTE interface based on an exclusive rule in that, those with the voice session ID are delivered to the voice bearer, and others to the signaling bearer. Second, resource reservation for the voice bearer is not affected by the signaling bearer. This side information helps to differentiate the two bearers. Therefore, QoS is good for performance but can be bad for privacy.

4.2 Leakage in Coordination between Planes

We further disclose another weakness in the coordination between control and data planes. This makes it easier to leak the voice session ID.

V7: Side-channel Leakage by Improper Coordination

The session ID can also be leaked from improper coordination between planes. It can be obtained during call setup and call termination. At these two phases, the voice bearer on the data plane is established and released upon receiving certain signals on the control plane. When operations are not invoked in the correct timing sequence, voice packets can be erroneously delivered to the control plane.

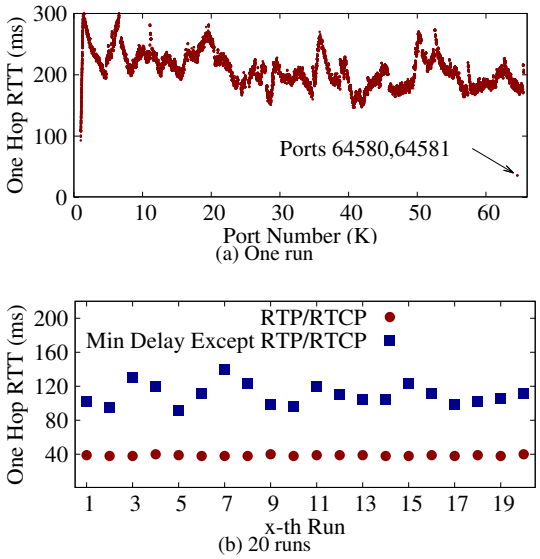


Figure 12: Latency of the packets using different ports.

We first see that some initial voice packets from the media gateway will be forwarded to the control plane when the voice bearer is not established in time. This is because the 4G gateway has only one forwarding rule for the VoLTE IP (*i.e.*, to/from the signaling bearer) before the voice bearer is created. For example, certain initial voice packets, such as alerting tone and early media (*e.g.*, CallerTune), should be delivered to the phone before the call is answered. The media gateway in the IMS core does have valid reasons to do so. This is a voice feature provided by most operators.

We further observe a few VoLTE voice packets from the control-plane interface after the call hangs up. It turns out that, the signaling server consists of two separate components, *i.e.*, the proxy server and the serving server. The former regulates the establishment/release of voice bearer at the 4G gateway, whereas the latter manages the start/stop of voice delivery at the media gateway. However, there is no explicit coordination procedure between them. Instead, implicit coordination is activated through the signaling messages passing them. When the phone terminates a call, it sends a `SIP BYE` message. This message first arrives at the proxy server, and is then forwarded to the serving server. Through this sequence, the former releases the voice bearer before the latter stops delivering voice packets. As a result, voice packets arriving at the 4G gateway must be forwarded to the VoLTE signaling bearer, since there is no voice bearer. When voice packets are forwarded to the signaling bearer, they reach the phone's VoLTE interface. Once captured by non-VoLTE apps, they can leak the session ID.

Empirical validation. Figure 13 shows IP packets collected from the VoLTE signaling interface on the mobile device via Shark [4]. Some VoLTE voice packets (here, UDP) are captured while the user dials out or hangs up. We further develop an unprivileged app to capture these early packets. The app binds to the VoLTE IP address and the UDP source port used by the voice RTP session. The UDP source port can be inferred; both operators have a simple selection rule. For OP-I, it starts from the number, 1234, after booting, and

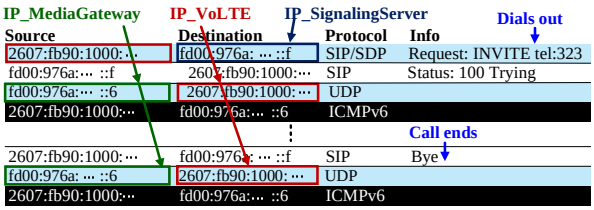


Figure 13: Packet trace collected from the VoLTE signaling interface while dialing and ending a call. UDP packets are those from Media Gateway in the data plane, but are delivered to the control plane.

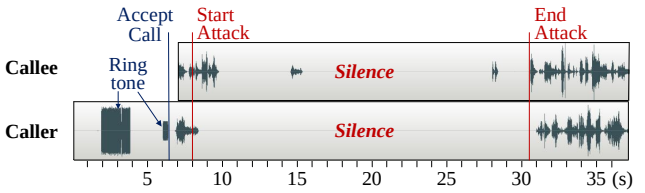


Figure 14: The voice wave of both caller and callee while the callee is under the enhanced voice-muted attack.

then monotonically increases by 10 for each call. For OP-II, the port is always 49158. We launch this app before making a VoLTE call. During the call setup and termination phases, it is observed to receive some UDP packets. We verify that they indeed belong to the ongoing call's RTP session.

Causes and lessons. VoLTE requires substantial upgrades by adopting IMS in the mobile network. It incurs complex operations, which can be exploited for unintended purposes.

4.3 Voice-Muted DoS Attack

We launch an attack to compel the call to be mute, rather than canceling the call service. Under this attack, the victim can always establish the call, but neither side can hear each other after the call is answered. This requires a malware without root privilege or voice call permission in the victim phone. Note that this malware can be embedded into network applications.

There are two major steps to launch this attack during an ongoing call. The first step is to learn the ports of the RTP session as soon as possible. We exploit *V6* and/or *V7* to do that. Exploiting V6 needs to scan many ports. However, without root privilege, the network socket returns only the error type of the ICMP messages instead of their content. So we cannot retrieve the scanned port corresponding to each received ICMP message within a socket. For this reason, we do a group-based search. We aggregate ports into many small groups and compare their smallest delay in each group. We then scan each port in the group with the smallest delay to locate RTP ports. It can be done within 20 seconds in our test phones. V7 can easily aggravate the damage of the voice-muted attack by offering a faster approach to probing session privacy. The session ID is directly learned from the initial voice packets, without spending time on searching for the session ID.

At the second step, the malware starts to hijack the voice bearer by injecting forged RTP packets with the correct session ID. They can mute both uplink and downlink voice, even though they are for only the uplink traffic. This is because the uplink traffic of RTP packets overloads one hardware component, Robust Header Compression (RoHC), which is required by VoLTE for voice packet (de)compression [5]. Figure 14 shows one run of the aggregated voice-muted attack when V7 is exploited. The malware starts to launch the DoS attack at the 8^{th} second, after the call is answered, and stops at 31^{th} second. During the attack period, the voice at both the caller and the callee become silent for most of the time. Once the ID is leaked, this attack remains effective until the call ends. The attack works no matter whether the malware is at the caller or the callee.

5. RECOMMENDED FIX

The proposed defenses cover both the network and the device. The network-side solution has three measures. Note that carriers exert little control over devices and have more incentives to fix these issues.

First, the 4G gateway enforces strict routing regulation for each bearer. This is to mandate that the traffic carried by the signaling/voice bearer is relayed by the 4G gateway only between the phone and the signaling server or the media gateway in the IMS core. It eliminates V2. Since the hijacked control-plane access on the device fails to reach any destination, it becomes less urgent to fix V1. However, without fixing V1, it is still vulnerable to data DoS attacks since precious radio resource might be wasted before these packets are dropped by the gateway. This requires the upgrade of the 4G gateway and can be done by adding filters for VoLTE bearers. It may need extra effort when adding all valid servers into the whitelist.

Second, the operator stops practicing free-signaling policy and charges signals similar to data traffic. This eliminates V3 and also largely reduces user incentives to exploit VoLTE for data access. However, this requires the upgrade of the billing system. The 4G gateway should also enable its accounting for signaling packets. The main challenge is not on the technical side but on the business practice. It requires every user to possess a data plan. Today, voice plans are independent of data plans. In LTE networks, it seems reasonable to let each consumer subscribe to a data plan for the Internet access. A supplementary solution is to enable a VoLTE data volume quota for each voice plan. When the volume is below this quota, it incurs no extra charge. This quota is based on the common usage case when the signaling volume for each call is typically small. This approach does not require users to be aware of the billing changes in benign scenarios, but abusing VoLTE for a large amount of volume delivery can be identified and charged for the extra.

Third, regarding DoS attacks, it calls for new mechanism to ensure resource allocation to authentic traffic only. A naive solution is to abandon high-priority QoS for VoLTE. However, it essentially ruins the VoLTE's appealing features. An alternative remedy is to enforce the deferral mechanism. Once the traffic is detected as fake or junk traffic, the volume will be accounted, and the source will be traced back if the volume exceeds certain threshold. At runtime, whenever the requested resource is larger than the quota, its priority is decreased. This scheme relieves, but does not eliminate preemptive resource waste. Ideally, the safeguard should be implemented at the device to prevent non-authentic traffic from being sent out. However, this may raise a new form of exploit of blocked access by leveraging the deferral mechanism if a malware is deployed at the victim.

On the device side, we suggest two remedies. First, the mobile OS should build a VoLTE permission that only allows the dialer app to access the VoLTE interface. Malware should not easily obtain private information. However, for a rooted phone, this protection might be bypassed. Second, the chipset enforces stringent access control. For the traffic from the OS, it needs to verify the traffic source and destination, as well as session port numbers, if

applicable. This resolves issues in most common settings, but cannot eliminate voice-muted attacks when a privileged malware on a rooted phone is in place. This latter case calls for defenses against mobile malware, which are complementary to above fences. Note that the device-based solutions really help but do not suffice. In some cases of legacy or compromised devices, the network-based solutions are still required.

Note that, the fundamental issues lie in the PS service and cannot be completely addressed by above fixes. The long-term solution calls for concerted effort to design security defenses for the given form of PS service with voice-specific policies (*e.g.*, resource consumption and billing). Our near-term solution aims to permit VoLTE for authentic voice only. This is to ensure backward compatibility with current practice. As more services are empowered by VoLTE (*e.g.*, HD video conferencing), it might not be the best choice any more. However, the rule is still applicable: When it works as PS data, it should be treated as PS data. Keeping a clean and consistent policy helps to eradicate unnecessary exploits.

6. DISCUSSION

We next clarify several remaining issues.

Bugs in early deployment? Hard-core optimists might claim that the revealed problems are implementation bugs at the early stage of deployment. They are likely to get fixed quickly as the deployment proceeds. Indeed, it turns out that the vulnerabilities we demonstrate are not hard to address. However, most of them are rooted in the VoLTE technology itself, but not carriers' specific misconfiguration. For example, both free-data and voice-muted attacks are feasible in both carriers. They are not simple bugs either, because where the vulnerability occurs is not equivalent to where the negative impact occurs: the device-side vulnerability (*e.g.* V1) results in network-side revenue loss, while the network-side vulnerability (*e.g.* V2 and V7) makes mobile device suffering from overcharging and voice mute attacks. Uncovering these vulnerabilities require a cooperative understanding of both the mobile device and the network. Because of this, it is not surprising that top carriers largely ignore them so far, and leave their deployed networks open to these attacks. Our ongoing interactions with such carriers indicate that, they are unaware of such loopholes even though some are implementation and operation glitches. The deeper reason is that, VoLTE has revolutionized the legacy voice service in mobile carriers, and complete understanding of its impact on the 4G system security is still missing. Since the US carriers act as leaders in the global market, the lessons from their early deployment are critical to the booming of secure VoLTE technology in the future.

Symptoms of incomplete transition? These problems can also be attributed to the incomplete transition from the old-fashioned mentality of mobile network to the Internet-style thinking. The former has been used to have control over all mobile devices since the early CS-based network. In contrast, the latter considers no control of devices, each of which can be potentially malicious. Since the mobile network adopts the Internet technologies (*e.g.*, IP), it has been giving more freedom to mobile devices and gradually having less control over them. However, its fundamental designs and operations are not developed by completely following the Internet's way of thinking. This way can cause the mobile network to be lack of protection against malicious mobile devices.

Blaming operators? While LTE operators and users suffer from these attacks, our position should be misinterpreted as a directed blame on network carriers (neither users). In fact, all parties contribute to the threats. Reshaping voice from CS to PS warrants substantial upgrades on all parties, including device OS and app developers, mobile chipset vendors, network equipment manufacturers and operators. The developers and vendors for the device OS, chipsets and network equipments should all share the responsibility without upgrading their access (permission) control in time. They together expose loopholes by the two IP interfaces for VoLTE and normal data. Without a holistic view how VoLTE works, each party takes a myopic approach to potential threats.

Incentives. Our focus so far is on vulnerabilities, but not on attack incentives. For certain attacks (*e.g.*, free data access), people are always motivated to leverage the loopholes. For other attacks (*e.g.*, DoS), it might be for fun or self interests for someone to block another's access to voice or data.

7. UPDATE

We are working with the industry to resolve the identified issues. We have already informed a major chipset vendor of the potential risks and are about to contact more vendors (including both OS and network equipment ones) regarding VoLTE access control vulnerabilities. We have also contacted both carriers to report and help to fix such vulnerabilities. So far, all data-relevant attacks, including free data attack, overbilling, preemptive data and data DoS attacks, have been fixed in OP-I carrier networks, since data packets through the VoLTE interface are not allowed to traverse the core network. The fix of voice DoS attacks, as well as the problems in another carrier is ongoing.

8. RELATED WORK

Several studies have explored security implications of the component solutions: IMS [22], SIP [31,36] and VoIP [17,19,38]. Park *et al.* model the threat and analyze possible issues of the IMS deployment [22]. Other studies [17, 19, 31, 38] examine SIP and VoIP in the Internet context, without addressing mobile network issues. They focus on caller ID spoofing or SIP message spoofing to launch DoS/DDoS attacks. Recent reports (*e.g.*, [1]) look into the VoLTE security but are limited to issues (*e.g.*, caller ID spoofing) addressed by previous studies. Our recent work revealed another voice DoS attack but it is launched through fine-grained manipulation of signaling messages in VoLTE [36]. In addition, most VoLTE research focuses on its performance analysis or deployment planning [20,21,37]. Our study differs from all prior arts. To the best of our knowledge, this is the first study on VoLTE security over operational networks. Our work covers both security analysis and real-world impact, whereas early findings are obtained by security analysis or limited experiments in the controlled environment. More importantly, the vulnerabilities and attacks discussed in this paper, have never been disclosed before.

Mobile network security has been an active research area in recent years. Peng *et al.* and Go *et al.* identify the loopholes in mobile data charging and devise free data access and overcharging attacks [18, 23–25, 35]. Enck and Traynor and *et al.* devise DoS attacks by overloading the control channel for SMS and other services [15,32,33]. Researchers also disclose vulnerabilities in other cellular-specific components, such as user authentication loopholes [11, 12], MMS spamming [29], information leakage at firewalls [27,28], and WiFi-calling loopholes in T-Mobile [13], to name a few. Our work is different since we look into VoLTE, an emerging voice service to 4G LTE networks. Mobile malware has been another well-covered topic (see [14, 39] for a few samples). Our DoS attacks require a malware without root permission, and we leave malware infection as an independent topic [16, 34].

9. CONCLUSION

VoLTE is still at its early phase for global rollout. It is natural to suffer from easy-to-fix mistakes during this period. However, we seek to sort out the fundamental issues beyond simple bugs and errors. Bearing the telecom-based design mindset, VoLTE calls for substantial upgrades on the infrastructure side (complex functions in the core), and device updates as well. In this work, we examine the security implications of VoLTE. We show that VoLTE can be exploited to launch attacks against both the network operator (thus benefiting mobile users) and an individual user. The user may gain free and high-priority data access by abusing the VoLTE signaling bearer to carry data packets. (S)he may also suffer from voice/data DoS attacks due to spamming over the voice/signaling bearer.

Two lessons can be learned from our work. First, VoLTE operates on both control and data planes. Its signaling and data are implemented in both software and hardware at the device, and carried by distinctive radio bearers within LTE. Consequently, to secure both planes, the solution calls for concerted effort between the network infrastructure and the end host, as well as the software and the hardware at the device. Second, VoLTE leverages the high priority services (compared with the low-priority, best-effort delivery) in mobile networks to ensure quality calls. The priority services supplemented by the LTE network may serve as an implicit side-channel to leak confidential information. As the voice solution becomes compatible with the Internet design, it is prudent to add more intelligence at the device and the network, to address the double-edged, security side-effects of PS and IP.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. This work is supported in part by the National Science Foundation under Grants No. CNS-1421933 and CNS-1422835, and an IBM PhD Fellowship (Guan-Hua Tu). The opinions, findings, and recommendations expressed in this material are those of the authors only and do not necessarily reflect the views of the National Science Foundation.

10. REFERENCES

[1] "2014: A VoLTE Security Nightmare?". http://tinyurl.com/p4rpm52.
[2] Decypt ISPEC packets. http://tinyurl.com/ptzurve.
[3] Network info ii. http://tinyurl.com/baa6jtu.
[4] Shark for Root. http://tinyurl.com/n7e9ubz.
[5] Voice over LTE. http://www.gsma.com/technicalprojects/volte.
[6] RFC 3550: RTP: A Transport Protocol for Real-Time Applications, Jul 2003.
[7] 3GPP. TS23.203: Policy and Charging Control Architecture, 2013.
[8] 3GPP. TS23.107:Quality of Service concept and architecture, 2014.
[9] 3GPP. TS23.228: IP Multimedia Subsystem (IMS);Stage 2, 2014.
[10] 3GPP. TS36.321:E-UTRA; Medium Access Control (MAC) protocol specification, Jan 2015.
[11] M. Arapinis, L. Mancini, E. Ritter, M. Ryan, N. Golde, K. Redon, and R. Borgaonkar. New Privacy Issues in Mobile Telephony: Fix and Verification. In *ACM CCS*, 2012.
[12] M. Arapinis, L. I. Mancini, E. Ritter, and M. Ryan. Privacy through pseudonymity in mobile telephony systems. In *NDSS*, 2014.
[13] J. Beekman and C. Thompson. Man-in-the-middle attack on t-mobile wi-fi calling. Technical Report UCB/EECS-2013-18, EECS Department, University of California, Berkeley, Mar 2013.
[14] S. Chakradeo, B. Reaves, P. Traynor, and W. Enck. MAST: Triage for Market-scale Mobile Malware Analysis. In *WiSec*, 2013.
[15] W. Enck, P. Traynor, P. McDaniel, and T. La Porta. Exploiting Open Functionality in SMS-Capable Cellular Networks. In *CCS*, 2005.
[16] C. Fleizach, M. Liljenstam, P. Johansson, G. M. Voelker, and A. Mehes. Can you infect me now?: malware propagation in mobile phone networks. In *ACM workshop on Recurring malcode*, 2007.
[17] C. Fuchs, N. Aschenbruck, F. Leder, and P. Martini. Detecting VoIP based DoS Attacks at the Public Safety Answering Point. In *ACM ASIACCS*, 2008.
[18] Y. Go, J. Won, D. F. Kune, E. Jeong, Y. Kim, and K. Park. Gaining Control of Cellular Traffic Accounting by Spurious TCP Retransmission. In *NDSS*, February 2014.
[19] A. D. Keromytis. A Look at VoIP Vulnerabilities. *Login*, 1:41–50, 2010.
[20] A. P. S. Louvros and A. Gkioni. Voice Over LTE (VoLTE): Service Implementation and Cell Planning Perspective. In *System-Level Design Methodologies for Telecommunication*, pages 43–62. Springer, 2014.
[21] N. S. Networks. From Voice over IP to Voice over LTE, 2013. http://tinyurl.com/q79vyu6.
[22] F. S. Park, D. Patnaik, C. Amrutkar, and M. T. Hunter. A Security Evaluation of IMS Deployments. In *IEEE IMSAA*, 2008.
[23] C. Peng, C. Li, G. Tu, S. Lu, and L. Zhang. Mobile Data Charging: New Attacks and Countermeasures. In *CCS*, Oct. 2012.
[24] C. Peng, C.-Y. Li, H. Wang, G.-H. Tu, , and S. Lu. Real Threats to Your Mobile Data Bills. In *ACM CCS*, Nov 2014.
[25] C. Peng, G. Tu, C. Li, and S. Lu. Can We Pay for What We Get in 3G Data Access? In *MobiCom*, Aug. 2012.
[26] J. Peterson. RFC 3323: A Privacy Mechanism for the Session Initiation Protocol (SIP), 2002.
[27] Z. Qian and Z. M. Mao. Off-Path TCP Sequence Number Inference Attack-How Firewall Middleboxes Reduce Security. In *S&P*, 2012.
[28] Z. Qian, Z. M. Mao, and Y. Xie. Collaborative TCP Sequence Number Inference Attack: How to Crack Sequence Number under a Second. In *ACM CCS*, 2012.
[29] R. Racic, D. Ma, and H. Chen. Exploiting MMS vulnerabilities to stealthily exhaust mobile phone's battery. In *SecureComm*, 2006.
[30] RFC3261: SIP: Session Initiation Protocol, June 2002.
[31] D. Sisalem, J. Floroiu, J. Kuthan, U. Abend, and H. Schulzrinne. *SIP security*. John Wiley & Sons, 2009.
[32] P. Traynor, W. Enck, P. McDaniel, and T. L. Porta. Exploiting open functionality in sms-capable cellular networks. *Journal of Computer Security*, 16:713–742, 2008.
[33] P. Traynor, P. McDaniel, and T. La Porta. On Attack Causality in Internet-Connected Cellular Networks. In *USENIX Security*, 2007.
[34] H. T. T. Truong, E. Lagerspetz, P. Nurmi, A. J. Oliner, S. Tarkoma, N. Asokan, and S. Bhattacharya. The company you keep: Mobile malware infection rates and inexpensive risk indicators. In *WWW'14*.
[35] G. Tu, C. Peng, C. Li, X. Ma, H. Wang, T. Wang, and S. Lu. Accounting for roaming users on mobile data access: Issues and root causes. In *MobiSys*, Jun. 2013.
[36] G.-H. Tu, C.-Y. Li, C. Peng, and S. Lu. How Voice Call Technology Poses Security Threats in 4G LTE Networks. In *IEEE Conference on Communications and Network Security (CNS)*, September 2015.
[37] L. L. Ying and S. W. Yuan. Forward handover for voice call continuity. In *NGMAST*, September 2012.
[38] R. Zhang, X. Wang, R. Farley, X. Yang, and X. Jiang. On the Feasibility of Launching the Man-In-The-Middle Attacks on VoIP from Remote Attackers. In *ACM ASIACCS*, 2009.
[39] Y. Zhou and X. Jiang. Dissecting Android Malware: Characterization and Evolution. In *IEEE S&P*, 2012.