The Dark Side of Operational Wi-Fi Calling Services

Tian Xie*, Guan-Hua Tu*, Chi-Yu Li†, Chunyi Peng‡, Jiawei Li*, Mi Zhang*
*Department of Computer Science and Engineering, Michigan State Univeristy
†College of Computer Science, National Chiao Tung University
‡Department of Computer Science, Purdue University
Email: *{xietian1, ghtu, lijiawe7, mizhang}@msu.edu, †chiyuli@cs.nctu.edu.tw, ‡chunyi@purdue.edu

***Abstract*—All of four major U.S. operators have rolled out nationwide Wi-Fi calling services. They are projected to surpass VoLTE (Voice over LTE) and other VoIP services in terms of mobile IP voice usage minutes in 2018. They enable mobile users to place cellular calls over Wi-Fi networks based on the 3GPP IMS (IP Multimedia Subsystem) technology. Compared with conventional cellular voice solutions, the major difference lies in that their traffic traverses untrustful Wi-Fi networks and the Internet. This exposure to insecure networks may cause the Wi-Fi calling users to suffer from security threats. Its security mechanisms are similar to the VoLTE, because both of them are supported by the IMS. They include SIM-based security, 3GPP AKA (Authentication and Key Agreement), IPSec (Internet Protocol Security), etc. However, are they sufficient to secure Wi-Fi calling services? Unfortunately, our study yields a negative answer.**

In this work, we explore security issues of the operational Wi-Fi calling services in three major U.S. operators' networks using commodity devices. We disclose that current Wi-Fi calling security is not bullet-proof. We uncover four vulnerabilities which stem from improper standard designs, device implementation issues and network operation slips. By exploiting them, we devise two proof-of-concept attacks: user privacy leakage and telephony harassment or denial of voice service (THDoS); they can bypass the security defenses deployed on both mobile devices and network infrastructure. We have confirmed their feasibility and simplicity using real-world experiments, as well as assessed their potential damages and proposed recommended solutions.

I. Introduction

Since 2016, all of four major operators in the U.S., T-Mobile, AT&T, Verizon and Sprint, have launched their nationwide Wi-Fi calling services [1][1]. The Wi-Fi calling technology utilizes the 3GPP IMS (IP Multimedia Subsystem) system [2] to provide a packet-switched voice service over Wi-Fi networks. It enables mobile users to dial outgoing calls, receive incoming calls, and send/receive text messages through their home/public Wi-Fi networks instead of cellular base stations. It is considered as a popular alternative voice solution for the mobile users with weak signals of base stations. A recent Cisco report [3] forecasts that the Wi-Fi calling is going to surpass VoLTE (Voice over LTE) and VoIP (Voice over IP, e.g., Microsoft Skype and Google Hangouts) services by 2018 in terms of voice usage minutes. By 2020, the Wi-Fi calling will take 53% of mobile IP voice service usage (about 9,000 billions of minutes per year), whereas the VoLTE and other VoIP services will have only 26% and 21%, respectively. As a result, any security loopholes of the Wi-Fi calling can lead to devastating consequences on a global scale due to its rapid global deployment [4]. We believe that there is a critical need for a security investigation on Wi-Fi calling services.

Technically, Wi-Fi calling services differ from the proprietary VoIP services such as Skype or other SIP-based (Session Initiation Protocol) voice services. Though its signaling protocol is also SIP-based, it is a 3GPP-specific version [5], [6]. For security reasons, both 3GPP and GSMA stipulate that Wi-Fi calling services shall use well-examined SIM-based security (i.e., storing each user's private secret key in a physical card) and authentication methods (i.e., 3GPP AKA (Authentication and Key Agreement) [7]), which are employed by the VoLTE. In addition, all the Wi-Fi calling signaling and voice/text packets shall be delivered through IPSec (Internet Protocol Security) channels between Wi-Fi calling devices and the cellular network infrastructure, since they may cross public, insecure networks. To defend against Wi-Fi DoS (Denial-of-Service) attacks (e.g., all the Wi-Fi calling packets are discarded by malicious Wi-Fi networks), the Wi-Fi calling employs a system-switch security mechanism, which switches Wi-Fi calling users back to cellular-based voice/text services when the users are unreachable through Wi-Fi networks.

Given these security mechanisms, which have been well studied in the VoLTE [11] and cellular networks [12] for years, it seems that the Wi-Fi calling should be as secure as the VoLTE. Unfortunately, it is not the case. We have identified several security threats in the Wi-Fi calling services deployed by T-Mobile, Verizon and AT&T in the U.S. The threats can be attributed to design defects of Wi-Fi calling standards, implementation issues of Wi-Fi calling devices, and operational slips of cellular networks. Specifically, we discover four vulnerabilities. First, Wi-Fi calling devices do not exclude insecure Wi-Fi networks which may impede their Wi-Fi calling services from their selection (V1). Second, they do not defend against ARP spoofing/poisoning attacks, which can be exploited to launch various MITM (Man-In-The-Middle) attacks (V2). Third, the Wi-Fi calling traffic, which is protected by the IPSec, is still vulnerable to side-channel attacks (e.g., privacy leakage) (V3). Fourth, even when the

[1]It is also named as VoWiFi (Voice over Wi-Fi).

Category	Vulnerability	Type	Root Cause
Device	V1: Wi-Fi calling devices do not exclude insecure Wi-Fi networks from their selection.	Design defect	Current 3GPP Wi-Fi network selection mechanism [1], [8] considers only the connectivity capabilities of Wi-Fi networks, but not their security risks (Section III-A).
	V2: Wi-Fi calling devices do not defend against ARP spoofing/poisoning attacks.	Implementation issue	Wi-Fi calling devices do not advance the Wi-Fi security for Wi-Fi calling services (Section III-A).
Infrastructure	V3: the Wi-Fi calling traffic, which is protected by the IPSec, is still vulnerable to side-channel attacks (e.g., privacy leakage)	Operation slip	For all three carriers, T-Mobile, AT&T, and Verizon, the IPSec session between mobile devices and the core network carries only Wi-Fi calling traffic, so traffic patterns can be learned to infer different events easily. (Section III-A).
	V4: even when the performance of a Wi-Fi calling call is bad (e.g., voice is muted), the mechanism of service continuity across the Wi-Fi calling and the cellular-based voice services is not effective in some scenarios.	Design defect	The triggers of 3GPP SRVCC/DRVCC [9], [2], [10] procedures, which keep service continuity across different radio access technologies, consider only radio quality but not service quality (Section III-B).

TABLE I
SUMMARY OF IDENTIFIED SECURITY VULNERABILITIES.

performance of a Wi-Fi calling call is bad (e.g., voice is muted), the mechanism of service continuity across the Wi-Fi calling and the cellular-based voice services is not effective in some scenarios (V4). They are summarized in Table I.

We exploit these four vulnerabilities to devise two proof-of-concept attacks: (1) user privacy leakage and (2) telephony harassment or denial of voice service attack (THDoS). They can bypass the existing security mechanisms on the Wi-Fi calling devices and network infrastructure. Note that in our threat model, the adversary has no control over the Wi-Fi network used by the victim, the victim's Wi-Fi calling device, or cellular network infrastructure. In the first attack, we develop a tool, a Wi-Fi calling analyzer *WiCA*, to analyze encrypted Wi-Fi calling packets and infer the user's call statistics, which have been widely used to infer user personality (e.g., conscientiousness [13]), mood (e.g., stressful [14]), and user behaviors (e.g., dialing spamming calls [15]). In the second attack, we devise four THDoS attacks: annoying-incoming-call, zombie-call, mute voice call, and telephony denial-of-voice-service. We further propose solutions to address the identified security issues. In summary, this paper makes three main contributions.

1) We conduct the first security study to explore the dark side of operational Wi-Fi calling services in three major U.S. operators' networks (i.e., T-Mobile, AT&T, and Verizon) using commodity devices. We identify four vulnerabilities which root in design defects of Wi-Fi calling standards, operational slips of operators, and implementation issues of Wi-Fi calling devices.
2) We devise two proof-of-concept attacks by exploiting the identified vulnerabilities and assess their impacts in those three U.S. carriers' networks.
3) We identify diversified root causes and propose recommended solutions. The lessons learned can facilitate and secure the global deployment of Wi-Fi calling services.

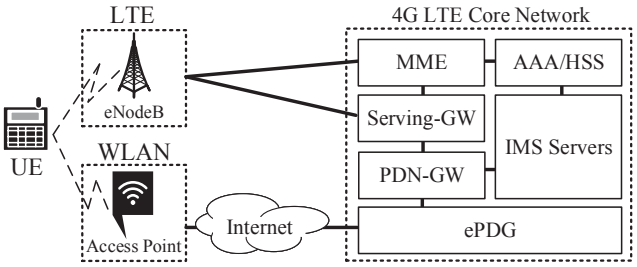


Fig. 1. Wi-Fi calling service network architecture.

II. WHAT MAY PROBABLY GO WRONG?

In this section, we introduce the Wi-Fi calling service support in the cellular network, as well as present possible security threats, the threat model and experimental methodology.

1) Wi-Fi calling Primer: The Wi-Fi calling technology enables mobile users to consume cellular-based voice services through Wi-Fi networks. It is similar to most VoIP (Voice over IP) services by using the SIP (Session Initiation Protocol) with 3GPP-specific modifications ([5], [6]) as its signaling protocol. The difference mainly lies in its service architecture, which relies on the cellular core network.

Figure 1 illustrates a simplified Wi-Fi calling service network architecture. It consists of two parts: Wi-Fi Access Point (AP), which is WiFi-RAN (WiFi-based Radio Access Network), and the LTE core network. The UE (User Equipment) consumes the voice service supported by the core through the Wi-Fi. The core network consists of four main components: the ePDG (Evolved Packet Data Gateway), the PDN-GW (Public Data Network Gateway), the AAA (Authentication, Authorization, and Accounting) server, and the IMS (IP Multimedia Subsystem) servers. To enable the Wi-Fi calling service, the ePDG first authenticates the user with the assistance of the AAA server, and then establishes an IPSec (Internet Protocol Security) tunnel to its UE. Afterwards, the ePDG routes packets between the UE, the PDN-GW and the IMS servers, which offer the Wi-Fi calling service.

2) Possible Security Threats: The Wi-Fi calling technology provides cellular-based voice services as traditional ones, and is considered as an alternative voice solution for the cases that cellular users are in the areas with bad signals of cellular base stations. Its radio access may rely on the Wi-Fi networks which are not controlled by operators, instead of cellular base stations owned by them. This naturally raises some security concerns. *Is the Wi-Fi calling technology as secure as the traditional voice service?* In particular, we start with the following two questions.

Q1. Do Wi-Fi calling devices still activate Wi-Fi calling services while associating with an insecure Wi-Fi network?
Q2. If yes, do any security mechanisms exist on the devices or/and the network infrastructure to defend against security threats? Moreover, can the threats be completely eliminated?

Unfortunately, we disclose that operational Wi-Fi calling services, as well as their technical support behind, are not bullet-proof. Wi-Fi calling devices do not avoid to activate Wi-Fi calling services in an insecure Wi-Fi network (Q1); Although Wi-Fi calling devices and infrastructure deploy some security mechanisms to defend against some malicious attacks (e.g., DoS attack). However, they do not completely eliminate the security threats (Q2). We then uncover four vulnerabilities from three aspects: design defects in the standard, operational slips from the operator's network, and the device's implementation issues. We elaborate on each vulnerability in Section III.

3) Threat model: In this work, the victim is a mobile user who associates with one Wi-Fi AP and is consuming the Wi-Fi calling service. The adversary is a user who has a networked device under the same subnet as the victim. S(he) does not need to associate with the Wi-Fi AP to which the victim connects. Take a campus' network as an example. The victim can be the one associating with any campus AP, whereas the adversary needs to connect to the campus' network but can be anywhere on the campus. They are thus under the same subnet of the campus' gateway. The adversary does not have any IPSec keys of the victim or any control of the victim's mobile device and the cellular network infrastructure.

4) Methodology: We validate our proposed vulnerabilities and attacks on three major U.S. carriers: T-Mobile, Verizon and AT&T. They together take more than 75% of market share [16] in the U.S. We conduct experiments using a software-based Wi-Fi AP on a MacBook Pro 2014 laptop, an ASUS RT-AC1900 Wi-Fi AP, and eight popular mobile devices with the Wi-Fi calling service, which include Samsung Galaxy S6/S7/S8/J7, Apple iPhone6/iPhone7/iPhone8, and Google Nexus 6P. The experiments are done in several campus Wi-Fi networks including Michigan State University, New York University, University of California Berkeley, and Northeastern University.

We understand that some feasibility tests and attack evaluations might be harmful to operators and/or users. So, we proceed with this study in a responsible manner by running experiments in fully controlled environments. In all the experiments, victims are our own mobile accounts and devices so that no other people get hurt.

III. HOW DOES IT GO WRONG?

We next answer the aforementioned two questions in details by examining vulnerabilities in the standard, operator networks and device implementations.

A. Q1: Do Wi-Fi calling devices still activate Wi-Fi calling services while associating with an insecure Wi-Fi network?

To answer this question, we seek to examine whether the selection policy of Wi-Fi networks on the Wi-Fi calling devices considers security impacts or not, in addition to the performance issues involving the quality of Wi-Fi links and Internet connectivity. However, it is not the case due to the discovery of three security vulnerabilities, so the answer to Q1 is yes. By exploiting these vulnerabilities, an adversary is able to intercept the packets of the Wi-Fi calling service for one victim, as well as then infer service events of his/her ongoing Wi-Fi calling call and manipulate the delivery of the service packets.

Vulnerability 1 (V1). We discover that Wi-Fi calling devices are unable to exclude an insecure Wi-Fi network from their Wi-Fi calling services due to its network selection mechanism stipulated by the Wi-Fi calling standards ([1], [8]). V1 can be thus considered as a design defect in the standard. Specifically, two Wi-Fi network selection modes are specified: manual and automatic modes. In the manual mode, the devices maintain a prioritized list of selected Wi-Fi networks, but how to do it is vendor-specific. In the automatic mode, the devices are guided to do the selection by the network infrastructure using the ANDSF (Access network discovery and selection function) function [17]. The selection is mainly based on the capabilities [18] and the radio quality (e.g., ThreshBeaconRSSIWLANLow [8]) of available Wi-Fi networks. Both modes do not consider security risks of selected Wi-Fi networks.

Vulnerability 2 (V2). We find that all of our test Wi-Fi calling devices suffer from the ARP (Address Resolution Protocol) spoofing, which is the prerequisite of various MITM attacks. An attacker sends spoofed ARP messages onto a local area network to associate his/her MAC address with the IP address of the default gateway and is thus able to intercept the victim's packets. It can be considered a common implementation issue of Wi-Fi calling devices. An adversary can leverage this spoofing technique to intercept all the packets belonging to Wi-Fi calling devices.

Vulnerability 3 (V3). It is observed that the Wi-Fi calling is the only service that is carried by the IPSec channel between the mobile device and the ePDG (shown in Figure 1), and it may be exploited to leak user privacy in terms of Wi-Fi calling service events (e.g., call status and text messaging status). It happens in all of our test operators, and can be attributed to their operational issues.

No.	Time	Source	Destination	Protocol	Length	Info
440	56.276919	208.54.16.4	192.168.2.5	ESP	176	ESP (SPI=0xbb21253b)
441	56.266969	208.54.16.4	192.168.2.5	ESP	176	ESP (SPI=0xbb21253b)
465	56.316883	192.168.2.5	208.54.16.4	ESP	176	ESP (SPI=0x0855c9c8)
468	56.337334	192.168.2.5	208.54.16.4	ESP	176	ESP (SPI=0x0855c9c8)
469	56.347763	208.54.16.4	192.168.2.5	ESP	176	ESP (SPI=0xbb21253b)
470	56.348012	208.54.16.4	192.168.2.5	ESP	176	ESP (SPI=0xbb21253b)

Fig. 2. A trace of intercepting Wi-Fi calling packets through the ARP spoofing.

1) Validation: We next validate the three vulnerabilities.

V1. We validate V1 by checking whether Wi-Fi calling devices keep connecting to the Wi-Fi APs which are under the ARP spoofing attack. In the attack, we let one computer masquerade as a victim's Wi-Fi calling device, and thus the downlink packets belonging to the victim are delivered to the computer. After intercepting those packets, we can still forward them to the victim after inspecting them, drop all of them, or take other actions. Our results show that all the test devices in three different operators' networks do not interrupt their connections with the APs under the attack while using their Wi-Fi calling services.

V2. To validate V2, we examine whether the test devices can be resistant to the ARP spoofing attack. We employ a tool, EtterCap, to send spoofed ARP messages, which claim one computer as the default gateway of a network, to all the Wi-Fi calling devices in the network. It is observed that all those devices accept the ARP commands carried by the spoofed messages and then send packets to the computer. The computer can thus intercept all the uplink Wi-Fi calling packets, as shown in Figure 2.

V3. We examine whether any information can be inferred based on the intercepted Wi-Fi calling packets, which are encrypted by the IPSec. After analyzing their patterns, we discover that there are six service events in all the three operators' Wi-Fi calling services: dialing/receiving a call, sending/receiving a text message, and activating/deactivating the service. Figure 3 shows the IPSec packets captured on our Wi-Fi AP from an experiment, where we trigger those six events on a test phone. We apply C4.5 algorithm [19] to studying and classifying encrypted Wi-Fi calling service packets into six events. To prepare a set of training data for the C4.5 algorithm, we repeat the aforementioned six Wi-Fi calling service operations on the test phone with 20 runs and collect all the IPSec packets of the phone on the Wi-Fi AP. Based on the training data, the C4.5 can generate a model to classify the Wi-Fi calling events. In 20 tests, we can get 100% accuracy. Note that we validate the result of the decision tree for each test by comparing it with the test phone's trace, as shown in Figure 4. The trace is collected from the phone, Nexus 6P, with T-Mobile Wi-Fi calling service.

We next check whether the classification method is viable for cross-phone/cross-carrier cases by testing it on different devices with the Wi-Fi calling services of three different carriers. It is observed that those six events in all the test cases can be classified accurately. Specifically, the classification model trained based on the Nexus 6P device with T-Mobile's Wi-Fi calling service can be largely applied to the other

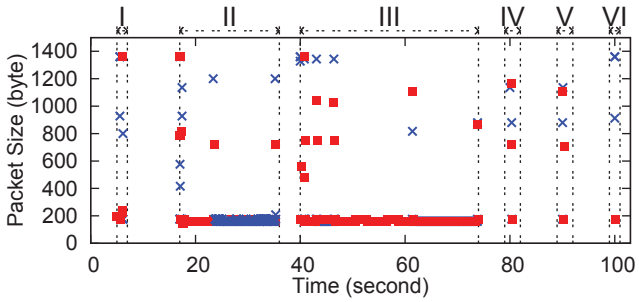


Fig. 3. The IPSec packets of six Wi-Fi calling events over time (×: uplink packets; ■: downlink packets; I/VI: Activating/Deactivating Wi-Fi calling; II/III: Receiving/Dialing a call; IV/V: Sending/Receiving a text).

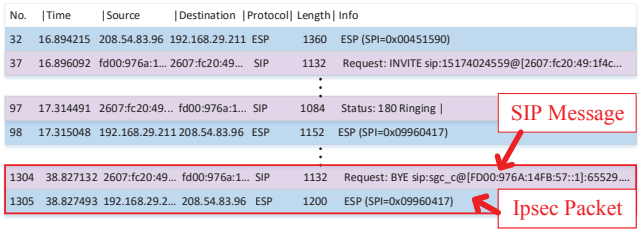


Fig. 4. A trace of Wi-Fi calling packets: SIP and IPSec packets collected on a test phone.

devices and carriers.

2) Rationale and security implications: It is not without reasons that the Wi-Fi calling standards consider only the quality of Wi-Fi links and Internet connectivity for the selection of Wi-Fi networks, since the Wi-Fi calling sessions have been protected by IPSec with end-to-end confidentiality and integrity protection. Though it is unlikely for an adversary to decrypt/alter the Wi-Fi calling packets, intercepting them and inferring user privacy based on their patterns are still possible according to the above three vulnerabilities. We believe that 3GPP and GSMA shall revisit current selection mechanisms of Wi-Fi networks for Wi-Fi calling services in terms of security.

B. Q2: Do any security mechanisms exist on the devices or/and the network architecture against security threats?

The answer is yes, but they do not completely eliminate the security threats. There exists a system-switch mechanism in which when a Wi-Fi calling device cannot be reached through its connected Wi-Fi network (e.g., the Wi-Fi calling signaling packets cannot be delivered to users successfully), it would switch back to the cellular network and use the cellular-based voice service (e.g., VoLTE (Voice over LTE)). This mechanism can protect the Wi-Fi calling device from the DoS attack where the Wi-Fi calling packets are discarded, because the attack impact will be considered as the case that the device is unreachable. However, it does not work for some attack cases, e.g., packets are dropped during an ongoing Wi-Fi calling service. The fundamental issue lies in that seamless service continuity across the Wi-Fi calling and the cellular-based voice services considers only the quality of Wi-Fi links (Vulnerability 4 (V4)). That is, once the link quality is good, the Wi-Fi calling device will not switch from the Wi-Fi calling service to the cellular-based voice even if all the Wi-Fi calling

```
guaranteed_birate_dlink_ext=unknown
 EsmQos delivery_order=without delivery order traffic_class=interactive
class QCI=5 delay_class=1 transfer_delay=unknown residual_BER=1e-05
[INFO] [LteNasAnalyzer]: Call flow status: VoLTE_PROCESSING
[INFO] [LteNasAnalyzer]: EPS_Id=7 EPS_ID=7 type=default:
 EsmQos peak_tput=4000 mean_tput=best effort max_bitrate_ulink=39
max_bitrate_dlink=39 guaranteed_birate_ulink=39
```

Fig. 5. A trace shows that a Wi-Fi calling phone initiates a VoLTE call under the DoS attack which is launched during the time that a call is being dialed.

packets keep being dropped. As a result, an adversary is able to get Wi-Fi calling users stuck in malicious Wi-Fi networks and cause them to suffer from poor voice services.

1) Validation: We conduct two experiments to validate the behaviors of the system-switch and service continuity mechanisms during DoS attacks. We discard all Wi-Fi calling packets on our test phone in two cases: (1) during the time that a call is being dialed, and (2) during the time that a call's conversation is ongoing. Note that the signal strength of the Wi-Fi link between the phone and the AP is strong. In the first case, the phone keeps sending the `SIP INVITE` message to the Wi-Fi calling server and waiting for the response (i.e., `SIP 100 Trying`). After six attempts, we observe that it switches back to the cellular-based voice service by initiating a VoLTE outgoing call. In 10 runs, all the VoLTE calls are successfully established. Figure 5 shows a low-level cellular network trace of switching back to the VoLTE call. Note that this trace is obtained on our test phone via the MobileInsight [20] tool. In the second case, the Wi-Fi calling voice call is interrupted within 8-10 seconds after the attack starts, but switching back to the VoLTE call does not happen.

2) Rationale and security implications.: Seemingly, it is an operational slip of operators, since cellular network standards have stipulated how to keep service continuity across different radio access technologies (e.g., Wi-Fi, 3G, and LTE). Specifically, SRVCC (Single Radio Voice Call Continuity) [9] and DRVCC (Dual Radio Voice Call Continuity) [2], [10] are proposed for this purpose. After second thought, it might not be the case. The SRVCC/DRVCC procedure is initiated by the network infrastructure and triggered based on the radio quality conditions of current serving base station (BS) and neighboring BSes. It implies that the SRVCC/DRVCC will not be triggered if the radio quality of the current serving BS is good. This design makes sense in normal scenarios, but does not work for some attack cases. Therefore, we believe that the design of service continuity shall consider not only radio quality but also service quality.

IV. Proof-of-concept Attacks

We devise two proof-of-concept attacks: (1) user privacy leakage; (2) telephony harassment or denial of voice service attack (THDoS).

A. User Privacy Leakage

We devise a tool, Wi-Fi calling Analyzer (WiCA), to infer a Wi-Fi calling user's call statistics including who initiates the call and who hangs up first, as well as ringing and conversation

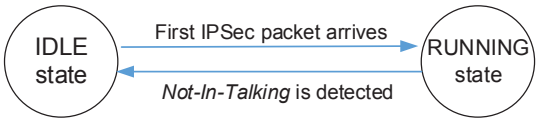


Fig. 6. State machine of WiCA.

Conditions		Identified Scenarios
$Num\_UL\_C_{Small}$	$Num\_DL\_C_{Small}$	
=0	>10	Ringing[a]
>10	>10	Talking
=0	=0	Not in Talking

TABLE II
USE $Num\_UL\_C_{Small}$ AND $Num\_DL\_C_{Small}$, WHICH ARE COLLECTED EVERY 2 SECONDS, TO DETERMINE *Ringing*, *Talking*, *Not in Talking* SCENARIOS FOR VERIZON, T-MOBILE AND AT&T.

[a]Only applicable for AT&T and T-Mobile since Verizon does not send small voice packets to the Wi-Fi calling user when the callee's phone is ringing.

time. The call statistics have been widely used to infer user privacy including personality (e.g., conscientiousness [13]), mood (e.g., stressful [14]), malicious behaviors (e.g., dialing spamming calls) [15], to name a few. In the following, we introduce how WiCA extracts the call statistics and evaluate its performance in the three U.S. operators.

1) Call Statistics Extraction: WiCA is an online pattern analyzer of Wi-Fi calling traffic. It is designed based on our observations on the traffic characteristics of Wi-Fi calling signaling and voice packets (the details are elaborated in Appendix A). Figure 6 illustrates its finite state machine, where the initial state is `IDLE`. It works as follows.

Step 1. At the `IDLE` state, when any IPSec packet is received, the system moves to the `RUNNING` state. By identifying whether it is sent from the UE, we can classify the subsequent call. If it is from the UE, the event, 'dialing a call', is identified. Otherwise, it is the event, 'receiving a call'.

Step 2. Right after entering the `RUNNING` state, the system considers the IPSec packets collected in the first two seconds and classifies them into the following three categories: (1) *C-Large*, where the packet size is larger than 800 bytes (for critical SIP call messages, e.g., `INVITE`, `RINGING`, etc.); (2) *C-Small*, where the packet size is smaller than 200 bytes (for voice packets); (3) *C-Middle*, where the packet size is between 200 and 800 bytes. We denote the 2-second IPSec packet collection as $Data_{2sec}(x)$, where x is the sequence of a series of 2-sec IPSec packet collection sets.

Step 3. Based on the rules specified in Table II, the system can discover three scenarios which are *Ringing*, *Talking* and *Not in Talking* by analyzing $Num\_UL\_C_{Small}$ and $Num\_DL\_C_{Small}$, which are the numbers of uplink and downlink *C-Small* packets respectively, within $Data_{2sec}(x)$. If no scenario is identified, the system buffers $Data_{2sec}(x)$ and goes back to `Step 2`.

If the *Ringing* is identified in $Data_{2sec}(x)$, the system revisits the collection, $Data_{2sec}(x-1)$, to discover the time that the last *C-Large* IPSec packet is captured. It is considered as the time that the ring starts, $T_{RingingStart}$.

If the *Talking* is identified and there is not any *Talking*

Time	T-Mobile		AT&T		Verizon	
	Mean	Std	Mean	Std	Mean	Std
Ringing	0.16s	0.11s	0.34s	0.11s	N/A	N/A [a]
Conversation	0.17s	0.07s	0.67s	0.13s	0.44s	0.2s

TABLE III
ERRORS OF RINGING AND CONVERSATION TIME ESTIMATION.

[a]Verizon does not send small voice packets to the Wi-Fi calling user when the callee's phone is ringing.

scenario before this, the system revisits $Data_{2sec}(x-1)$ to discover the time that the first *C-Large* IPSec packet (i.e., `SIP 200 OK`, which indicates the event of answering the call) is captured. It is considered as the time that the talk starts, $T_{TalkingStart}$.

If the *Not In Talking* is identified, the system revisits the collection, $Data_{2sec}(x-1)$, to discover the time that the first *C-Large* IPSec packet (i.e., `SIP BYE`) is captured. It is considered as the time of the call end, $T_{CallEnd}$. Moreover, the condition that a *C-Large* packet is sent by the UE means that s(he) hangs up first. When the call end is observed in the *Not in Talking* event, the pattern analyzer outputs who initiates the call, ringing time duration (i.e., $T_{TalkingStart} - T_{RingingStart}$ or $T_{CallEnd} - T_{RingingStart}$), conversation time duration (i.e., $T_{CallStop}$-$T_{TalkingStart}$), and who hangs up first. Afterwards, the system goes back to the `IDLE`. Note that the conversation duration estimation is not applicable to the calls which Wi-Fi calling users do not answer or hang up before the conversations start.

2) Evaluation: We next evaluate the WiCA's performance. On each test phone, we dial more than 50 outgoing calls and receive more than 50 incoming calls. In the mean time, we collect call-related traces on the phone during the experiments. We then compare the events identified by the WiCA and those extracted from the collected traces. Our results show that WiCA can correctly identify who initiates the call and who hangs up in all the experiments. For the ringing and conversation time, we evaluate them by the difference between the time estimated at WiCA and the time observed on the test phones. Table III summarizes the mean and standard deviation results. We do not observe significant estimation errors; all the values are less than 0.8s.

3) Negative Impacts: WiCA can be integrated into current security surveillance systems. With its call statistics, the systems can identify not only user identity by mature visual recognition techniques [21] but also the user device's IP address. It is because the actions that talking on the phones, putting the phone up to an ear, and removing the phone from the ear, are quite different from those of people surfing/reading/writing on their phones. Even more than one person is using Wi-Fi calling services, we can still narrow down the candidates since the time that they start talking or the time they hang up the call is more likely different. The WiCA-enabled surveillance systems can be used by universities to infer students' personality, recent mood or what network services and websites they surf, and take necessary actions against possible campus assaults. It can be also deployed at the airports. For some suspects or terrorists, the law enforcement agents can identify their identities and device addresses, and further install the malware on their phones for monitoring. Note that the distribution of malware can be achieved by exploiting public security vulnerabilities of the victims' devices (CVE database discloses many security issues discovered on various devices, https://cve.mitre.org/). We do not discuss the details here since it is not our focus in this paper.

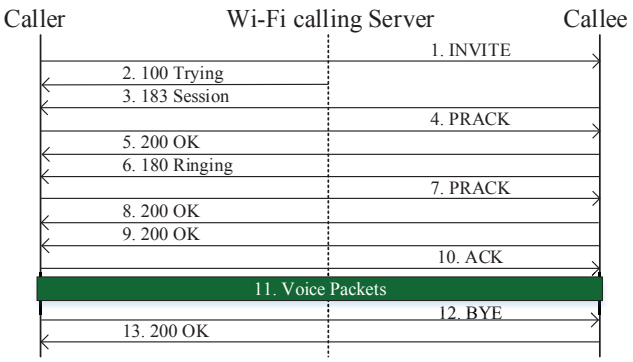


Fig. 7. Wi-Fi calling call flow diagram.

B. Telephony Harassment/Denial of voice Service Attack

We devise a telephony harassment or denial of voice service attack (THDoS) against Wi-Fi calling users. It can bypass the security defenses deployed on Wi-Fi calling devices and the infrastructure. The attack is based on the manipulation of the delivery of Wi-Fi calling signaling and voice packets for an ongoing call. It contains several variances, e.g., extra incoming calls, hiding callee's alerting tone, mute calls, etc. In the following, we first introduce the Wi-Fi calling call flow, and then present attack evaluation and real-world negative impacts.

1) Wi-Fi calling Call Flow: An outgoing call flow of the Wi-Fi calling is shown in Figure 7. To initiate a call, the caller sends an SIP `INVITE` message, which specifies the capabilities (e.g., voice codec) of the caller, to the callee. The Wi-Fi calling server at the IMS system replies an `100 Trying` message to indicate that the call setup is in progress. In the meantime, the callee also replies an `183 Session` message, which contains a list of chosen voice codecs, to the caller. Afterwards, the caller sends a `PRACK` (Provisional Acknowledge) message to inform the callee about the selected codec. After the callee's phone rings, it will send a `180 Ringing` message to the caller and then the caller's phone rings. Once the callee answers the call, two ends start to chat after exchanging `200 OK` and `ACK` messages. A `BYE` message is finally sent from the end who terminates the call and the other end acknowledges it with a `200 OK` message.

2) Attack Evaluation: We launch an ARP spoofing attack towards a Wi-Fi calling user and are thus able to intercept all of his/her Wi-Fi calling packets. Based on the identified traffic characteristics of Wi-Fi calling calls (see Appendix A), together with the Wi-Fi calling call flow, we are able to identify specific signaling packets and voice packets. We discover there are four attack variances by considering different patterns of

No.	Dropped Packets	Sender	Results
1	INVITE	Caller	Caller initiates a cellular-based call.
2	100 Trying	Server	No effect.
3	183 Session	Callee	Two outgoing calls arrive at callee.
4	PRACK	Caller	No effect.
5	200 OK	Callee	No effect.
6	180 Ringing	Callee	Caller will not enter conservation state. His/her phone gets stuck in the dialing screen.
7	PRACK	Caller	No effect.
8	200 OK	Callee	Caller keeps hearing the alerting tone.
9	200 OK	Callee	Caller keeps hearing the alerting tone.
10	ACK	Caller	No effect.
11	Voice Packets	Caller /Callee	Call drops or voice quality downgrades.
12	BYE	Caller	Callee gets stuck in the conversation state for 20s. Afterwards, the call is terminated.
13	200 OK	Callee	No effect.

TABLE IV
THE RESULTS OF DROPPING WI-FI CALLING SIGNALING AND VOICE PACKETS OF AN OUTGOING CALL.

Wi-Fi calling packets to drop. Table IV summarizes the attack results.

Annoying-Incoming-Call Attack. The callee is the victim and can receive multiple incoming calls from the caller. There are two approaches. First, the adversary can drop the `183 Session Progress` message sent by the callee, and then the caller's Wi-Fi calling device would initiate another VoLTE call towards the callee. Second, by discarding the `180 Ringing` message sent by the callee, the adversary can force the caller's Wi-Fi calling device to get stuck in the dialing screen. The caller cannot hear any alerting tone, but the callee's device would ring. The caller may thus keep redialling.

Zombie-Call Attack. The victim is the caller. The adversary can discard the `200 OK` message which is sent by the callee to indicate that the call has been answered, to force the caller's device to get stuck in the dialing screen and keep hearing the alerting tone. The conversation can never be started.

Mute Call Attack. Two parties of a Wi-Fi calling call are both victims. This attack does not aim to terminate the call but only mute the victims' speech. Our result shows that the adversary can mute the call up to 8 seconds by dropping voice packets. Note that if the voice suspension time is longer than 8 seconds, the voice call will be terminated by the network.

Telephony Denial-of-Voice-Service Attack. Both the caller and the callee can be victims. This attack can downgrade the voice quality a lot so that the conversation is hardly continued, but still keep the call on. Our experiment results show that the adversary can make it by controlling the drop rate of voice packets to be between 70% and 90%. Table V summarizes the negative impact on voice quality for different drop rates.

3) Negative Impacts: The real-world impact of our THDoS attack can be significant in practice. Most of U.S. universities have deployed campus Wi-Fi networks. However, our studies show that the campus Wi-Fi is the best attack surface for adversaries. Take Michigan State University as an example.

Drop Rate (%)	Voice Quality
20%	No clear impact.
40-60%	Some noises.
70-90%	Conversation is hardly continued.
100%	Call is terminated by the network.

TABLE V
VOICE QUALITY VARIES WITH THE DROP RATE OF VOICE PACKETS.

Its campus Wi-Fi (MSUNet) provides all students, the faculty, and the staff with free Wi-Fi access. In our 2-min experiment, we discover that more than 700 devices including smartphones, tablets, and computers, connect to MSUNet. All the devices are served by the same gateway which is vulnerable to ARP spoofing attack. Note that we validated the gateway's vulnerability through our own devices. As a result, adversaries are likely to intercept the packets of all Wi-Fi calling devices and launch the THDoS attack. Note that the campus Wi-Fi at Michigan State University is not the only Wi-Fi infrastructure with this issue. We find that it also exists in many universities' campus Wi-Fi, such as New York University, University of California Berkeley, Northeastern University, etc.

V. RECOMMENDED SOLUTIONS

In this section, we propose both short-term and long-term solutions. They can be respectively used to mitigate security threats quickly in practice and address the identified security vulnerabilities thoroughly.

Using Virtual Private Network (VPN) Service. This is a quick remedy for Wi-Fi calling users. We suggest that they can leverage the existing VPN services (e.g., AT&T IPSec VPN [22] and IPVanish VPN) to carry all the traffic, which can belong to Wi-Fi calling services and other applications, on their devices with the VPN tunnels. It can prevent adversaries from inferring Wi-Fi calling signaling and voice packets due to the mixed traffic. Note that Wi-Fi calling users may not need to pay extra cost since some popular VPN services (e.g., NordVPN, TunnelBear VPN and Golden Frog VyprVPN) provide users with free versions. We have confirmed that the attack proposed in Section III does not work when the VPN service is enabled on the victim's device. Though this quick remedy may not completely eliminate security threats, we believe that it can significantly reduce the real-world damages caused by Wi-Fi calling based attacks.

Upgrading Wi-Fi Calling Standards. We believe that the security should be the top-priority feature in the Wi-Fi calling standards, since the Wi-Fi calling traffic may cross public, insecure networks. Note that we do not aim to secure the public/private Wi-Fi networks which are out of operators' control but identify insecure Wi-Fi networks and possible Wi-Fi calling attacks, and take actions (e.g., switching back to cellular networks). We recommend three mechanisms for the upgrade of the standards.

First, Wi-Fi calling devices shall deploy necessary security defenses against common WiFi-based attacks (e.g., ARP spoofing attack), and examine whether the connected Wi-Fi networks are secure since not all of Wi-Fi calling attacks (e.g.,

discarding users' outgoing call request) can be identified by the infrastructure.

Second, both of Wi-Fi calling devices and infrastructure shall detect whether users are being under attacks by monitoring the quality of Wi-Fi calling services. Once any anomaly is detected, the devices can prompt the users to take actions (e.g., switching to another Wi-Fi network or conventional cellular services). Moreover, the infrastructure can update the ANDSF and RAN rules to exclude malicious Wi-Fi networks.

Third, the triggers of SRVCC/DRVCC shall be enhanced to consider not only radio quality but also service quality, since the service quality downgrade may be caused by adversaries. The first two mechanisms can address both V1 and V2, whereas the third one eliminates V4. Note that V3 does not exist when all the above three vulnerabilities are removed.

VI. Related Work

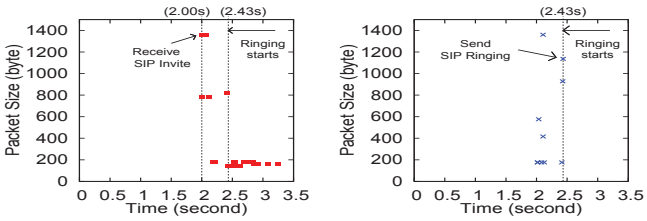
Side-Channel Attacks Against Mobile Systems. The side-channel information leakage against mobile systems has been a popular research area in recent years. Current studies [23], [24] target the side-channel information leaked by mobile users' traffic, which is generated by some particular Internet services, and then seek to infer users' activities. Two studies [23], [24] examine side-channel attacks on VoIP traffic. Different from them, we focus on the insecurity of the cellular Wi-Fi calling service, which is stipulated by 3GPP and is going to be deployed globally on billions of mobile devices in the near future.

Wi-Fi Calling Security. Wi-Fi calling security is a new research area and has not been fully studied by the academic yet, since carriers just deployed their Wi-Fi calling services in recent years. Current researchers mainly focus on the security vulnerabilities on Wi-Fi calling devices. Specifically, Beekman et. al pointed out that T-Mobile Wi-Fi calling devices (e.g., Samsung S2) are vulnerable to invalid server certificates [25]. Chalakkal studied SIM-related security issues on Wi-Fi calling devices [26]. However, our work examines the Wi-Fi calling security and discovers new vulnerabilities from all the three aspects: standards, operations and implementations.

VII. Conclusion

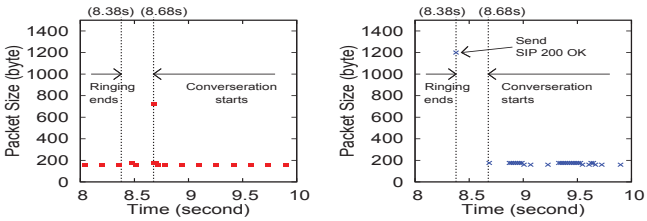
The Wi-Fi calling service is thriving and being deployed worldwide. In this work, we conduct the first study on the security implication of the operational Wi-Fi calling service over three major U.S. operators and state-of-the-art Wi-Fi calling devices (e.g., Apple iPhone8 and Samsung Galaxy S8). We discover four security vulnerabilities which stem from the design defects of Wi-Fi calling standards, operational slips of operators, and implementation issues of Wi-Fi calling devices. By exploiting them, adversaries can infer user privacy and launch the telephony harassment or denial of voice service attack. The fundamental issue is that the security defenses well examined in cellular network services are simply applied to the Wi-Fi calling service without considering its specific security threats. After identifying root causes, we propose two remedies to alleviate real-world damages. The ultimate solution calls for a concerted effort among all parties involved.

The Wi-Fi calling service is still at its early rollout, so the lessons learned from three major U.S. carriers can help secure mobile ecosystem and facilitate the global deployment, as well as provide new design insights for upcoming 5G networks. We hope that our initial study will stimulate more research efforts on the Wi-Fi calling service from both academia and industry.



(a) Downlink (sent by server) (b) Uplink (sent by callee)

Fig. 8. Packet arrivals for the event 'receiving a call with a ringtone'.



(a) Downlink (sent by server) (b) Uplink (sent by user)

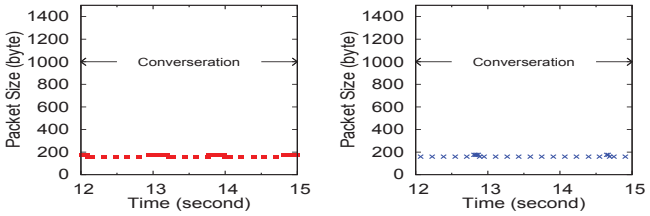
Fig. 9. Packet arrivals for 'answering the call'.

Appendix A
Identification of Traffic Characteristics for Wi-Fi Calling Signaling and Voice Packets

In this section, we use an example to illustrate the traffic characteristics of operational Wi-Fi calling signaling and voice packets observed on our test operators. First of all, the callee receives an incoming Wi-Fi calling call and answers it in 6 seconds after his phone rings. Afterwards, a 12-second voice conversation starts. Finally, the callee hangs the call up. We roughly split it into four events according to relative behaviors: (1) receiving a call with a ringtone; (2) answering the call; (3) talking; (4) ending the call. We next elaborate the characteristics of them.

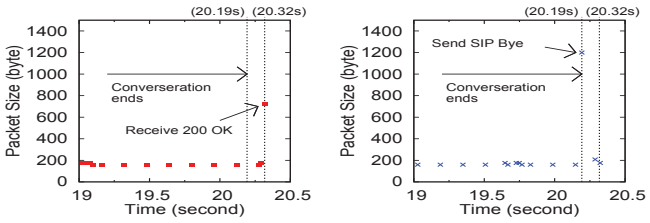
Event 1: Receiving a call with a ringtone (Figure 8). During this event, we collect both downlink (from the Wi-Fi calling server to the callee) and uplink packets over time on our WiFi AP. At the 2nd second, a 1360-byte IPSec packet is received by the callee. After decrypting it at the callee with root access, it is discovered to be an SIP `INVITE` message. At the 2.43th second, the callee sends an SIP `180 RINGING` message to the server. We observe that after the SIP `180 RINGING` is sent, the Wi-Fi calling server sends small 176-byte IPSec packets to the callee; however, the callee does not send any packets in response to them. These small IPSec packets are identified as RTP (Real-Time Protocol) packets carrying voice based on traces collected on the device.

Event 2: Answering the call (Figure 9). The callee answers the call at the 8.38th second and then sends an SIP `200 OK`



(a) Downlink (sent by server) (b) Uplink (sent by user)

Fig. 10. Packet arrivals for 'talking'.



(a) Downlink (sent by server) (b) Uplink (sent by user)

Fig. 11. Packet arrivals for 'ending the call'.

message to the server. At the 8.68th second, the server sends an acknowledgment in response to the message. Upon the receipt of the acknowledgment, the device starts to send voice packets to the server and the call conversation starts.

Event 3: Talking (Figure 10). During the call conversation, we observe that both the device and the server keep sending voice packets to each other and no SIP messages are observed. The callee will receive at least 10 voice packets from the Wi-Fi calling server every two seconds during the call conversation.

Event 4: Ending the call (Figure 11). The callee sends a `BYE` message to the server at the 20.19th second. Since the 20.32nd second, no IPSec packets are transmitted by the user and the server. This call is thus ended by the callee. Note that when the `BYE` is transmitted by the server, it means that the caller hangs up earlier than the callee.

Analysis Results. According to the above analysis, we have five observations. First, the sizes of IPSec packets carrying voice are smaller than 200 bytes. Second, the sizes of IPSec packets which carry critical Wi-Fi calling signaling (i.e., `INVITE`, `180 RINGING`, `200 OK`, `BYE`) are much larger than voice packets (e.g., 800-1360 bytes v.s. 176 bytes). Third, the callee receives voice packets from the Wi-Fi calling server after the `180 RINGING` message is sent. Fourth, no voice packets are sent out by the callee before the call conversation starts. Fifth, the callee keeps receiving at least 10 voice packets from the Wi-Fi calling server every two seconds after the call conversation starts. Note that they are observed in all the tested Wi-Fi calling service operators except that the third observation is only applicable to T-Mobile and AT&T.

REFERENCES

[1] GSMA, "IR.51 IMS OVER WI-FI V5.0," May 2017, //www.gsma.com/newsroom/all-documents/ir-51-ims-wi-fi-v5-0/.

[2] 3GPP, "TS23.237:IP Multimedia Subsystem (IMS) Service Continuity; Stage 2," 2017.

[3] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015V2020," 2016, https://www.cisco.com/c/dam/m/en_in/innovation/enterprise/assets/mobile-white-paper-c11-520862.pdf.

[4] Apple, "Wireless carrier support and features for iPhone over the world," 2017, https://support.apple.com/en-us/HT204039.

[5] M. Garcia-Martin, "Input 3rd-Generation Partnership Project (3GPP) Release 5 Requirements on the Session Initiation Protocol (SIP)," 2005, https://tools.ietf.org/html/rfc4083.

[6] R. Jesske, D. Telekom, K. Drage, and C. Holmberg, "Private Header (P-Header) Extensions to the Session Initiation Protocol (SIP) for the 3GPP," 2014, https://tools.ietf.org/html/rfc7315.

[7] 3GPP, "TS33.401:3GPP System Architecture Evolution (SAE);Security architecture," 2017.

[8] ——, "TS24.302:Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks," 2017.

[9] ——, "TS23.216:SRVCC V Single Radio Voice Call Continuity V Stage 2," 2017.

[10] ——, "TS24.237: IP Multimedia (IM) Core Network (CN) subsystem IP Multimedia Subsystem (IMS) service continuity; Stage 3," 2017.

[11] C.-Y. Li, G.-H. Tu, C. Peng, Z. Yuan, Y. Li, S. Lu, and X. Wang, "Insecurity of voice solution volte in lte mobile networks," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015, pp. 316–327.

[12] S. Bhattarai, S. Rook, L. Ge, S. Wei, W. Yu, and X. Fu, "On simulation studies of cyber attacks against lte networks," in *Computer Communication and Networks (ICCCN), 2014 23rd International Conference on*. IEEE, 2014, pp. 1–8.

[13] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting personality using novel mobile phone-based metrics," in *International conference on social computing, behavioral-cultural modeling, and prediction*. Springer, 2013, pp. 48–55.

[14] S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study," *BMC public health*, vol. 11, no. 1, p. 66, 2011.

[15] V. Balasubramaniyan, M. Ahamad, and H. Park, "Callrank: Combating SPIT using call duration, social networks and global reputation," in *CEAS'07*, 2007.

[16] statista, "Wireless subscriptions market share by carrier in the U.S. from 1st quarter 2011 to 3rd quarter 2017," 2017, https://www.statista.com/statistics/199359/market-share-of-wireless-carriers-in-the-us-by-subscriptions/.

[17] 3GPP, "TS24.312:Access Network Discovery and Selection Function (ANDSF) Management Object (MO)," 2017.

[18] "Ieee standard for wireless lan medium access control (mac) and physical layer (phy) specifications," *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pp. 1–2793, March 2012.

[19] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[20] UCLA, "Cellular Network Trace Collector: Spurring In-Phone Mobile Network Intelligence," 2017, http://www.mobileinsight.net/.

[21] G. Chéron, I. Laptev, and C. Schmid, "P-cnn: Pose-based cnn features for action recognition," in *IEEE ICCV*, 2015, pp. 3218–3226.

[22] AT&T, "Leverage IPSec in a hybrid VPN," 2017, https://www.business.att.com/solutions/Service/network-services/vpn/ipsec/.

[23] A. Compagno, M. Conti, D. Lain, and G. Tsudik, "Don't skype & type!: Acoustic eavesdropping in voice-over-ip," in *AsiaCCS*. ACM, 2017, pp. 703–715.

[24] J. Fang, Y. Zhu, and Y. Guan, "Voice pattern hiding for voip communications," in *Computer Communication and Networks (ICCCN), 2016 25th International Conference on*. IEEE, 2016, pp. 1–9.

[25] J. Beekman and C. Thompson, "Man-in-the-middle attack on t-mobile wi-fi calling," *Electrical Engineering and Computer Sciences University of California at Berkeley, https://www2.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-18.pdf*, 2013.

[26] S. Chalakkal, H. Schmidt, and S. Park, "Practical attacks on volte and vowifi," *ERNW Enno Rey Netzwerke, Tech. Rep*, 2017.