Pretty Good Phone Privacy

Paul Schmitt
Princeton University

Barath Raghavan
University of Southern California

Abstract

A fundamental property of today's cellular architecture—in order to receive service, phones uniquely identify themselves to towers and thus to operators—is now a cause of major privacy violations. Over the past two years it has become clear that operators have sold and leaked identity and location data of hundreds of millions of mobile users.

In this paper, we examine how to improve privacy in modern mobile networks. We take an end-to-end perspective on today's cellular architecture and find key points of decoupling that enable a new type of operator to offer privacy-enhanced service with no changes to physical infrastructure and without direct cooperation from existing operators.

We describe Pretty Good Phone Privacy (PGPP) and demonstrate how our modified backend stack (EPC) works with real phones to provide ordinary yet privacy-preserving connectivity. We explore inherent privacy and efficiency tradeoffs in a simulation of a large metropolitan region. We show how PGPP maintains today's control overheads while significantly improving user identity and location privacy.

1 Introduction

Cellular phone and data networks are an essential part of global communications infrastructure. In the United States, there are 129 cellular subscriptions for every 100 people and the total number of cellular subscriptions worldwide now stands at over 7.9 billion [5]. Unfortunately, today's cellular architecture embeds privacy assumptions of a bygone era. In decades past, providers were highly regulated and centralized, few users had mobile devices, and data broker ecosystems were undeveloped. As a result, except for law enforcement access to phone records, user privacy was generally preserved. Protocols for cell communication embed an assumption of trusted hardware and infrastructure [2], and specifications for cellular backend infrastructure contain few formal prescriptions for preserving user data privacy. The result is that the locations of all users are constantly tracked as they simply carry a phone in their pocket, *without even using it*.

Privacy violations by carriers. In the last two years it has been extensively reported that mobile carriers have been selling and leaking mobile location data and call metadata of hundreds of millions of users [18, 19, 39, 65, 69]. This behavior appears to have been legal and has left mobile users without a means of recourse due to the confluence of a deregulated industry, high mobile use, and the proliferation of data brokers in the landscape. As a result, in many countries every mobile user can be physically located by anyone with a few dollars to spend. This privacy loss is ongoing and is *independent* of leakage by apps that users choose to install on their phones (which is a related but orthogonal issue).

While this major privacy issue has long been present in the architecture, the practical reality of the problem and lack of technical countermeasures against bulk surveillance is beyond what was known before. However there is a fundamental technical challenge at the root of this problem: even if steps were taken to limit the sale or disclosure of user data, such as by passing legislation, the cellular architecture generally and operators specifically would still seemingly need to know where users are located in order to provide connectivity. Thus users must trust that network operators will do the right thing with respect to privacy despite not having done so to date.

Architectural, deployable solution. We aim to remedy this state of affairs by identifying and leveraging points of decoupling in the architecture. Our solution is designed to be deployed by Mobile Virtual Network Operators (MVNOs), where the MVNO operates the evolved packet core (EPC) while the base stations (eNodeBs) are operated by a Mobile Network Operator (MNO). This presents us with architectural independence as the MVNO can alter its core functionality, so long as the EPC conforms to LTE/5G standards. Our approach is made feasible by the industry-wide shift toward software-based EPCs.

In our approach, users are protected even against tracking by their own carrier (the MVNO). We decouple network connectivity from authentication and billing, which allows the carrier to run EPC services that are unaware of the identity or location of their users but while still authenticating them for

network use. We shift authentication and billing functionality to outside of the cellular core and separate traditional cellular credentials from credentials used to gain global connectivity.

Since it will take time for infrastructure and legislation to change, our work is explicitly *not* clean slate. In addition, we assume that existing industry players are unlikely to adopt new technologies or have an interest in preserving user privacy unless legal remedies are instituted. As a result, we consider how privacy can be added on top of today's mobile infrastructure solely by new industry entrants (*i.e.*, MVNOs).

Contributions. We describe our prototype implementation, Pretty Good Phone Privacy (PGPP). In doing so, we examine several key challenges in achieving privacy in today's cell architecture. In particular, we consider: 1) which personal identifiers are stored and transmitted within the cellular infrastructure; 2) which core network entities have visibility into them (and how this can be mitigated); 3) which entities have the ability to provide privacy and with what guarantees; and 4) how we can provide privacy while maintaining compatibility with today's infrastructure and without requiring the cooperation of established providers.

We show PGPP's impact on control traffic and on user anonymity. We show that by altering the network coverage map we are able to gain control traffic headroom compared with today's networks; we then consume that headroom in exchange for improved anonymity. We analyze the privacy improvements against a variety of common cellular attacks, including those based on bulk surveillance as well as targeted attacks. We find that PGPP significantly increases anonymity where there is none today. We find that an example PGPP network is able to increase the geographic area that an attacker could believe a victim to be within by ~1,200% with little change in control load.

Our contributions are as follows:

- We conduct a measurement study to demonstrate privacy leakage that exists in today's mobile networks (§3).
- We design a new architecture that decouples connectivity from authentication and billing functionality, allowing us to alter the identifiers used to gain connectivity (§5.1) and enable PGPP-based operators to continue to authenticate and bill users (§5.1) without identifying them.
- We adapt existing mechanisms to grow control traffic broadcast domains, thus enhancing user location privacy while maintaining backwards compatibility (§5.2).
- We quantify the impacts of PGPP on both user privacy and network control traffic through simulation (§6) and demonstrate PGPP's feasibility in a lab testbed.

2 Background

Here we provide a brief overview of the cellular architecture and describe the inherent privacy challenges. For simplicity we focus on 4G LTE, though the fundamental challenges exist in 5G (discussed in §4.2) as well as legacy standards.

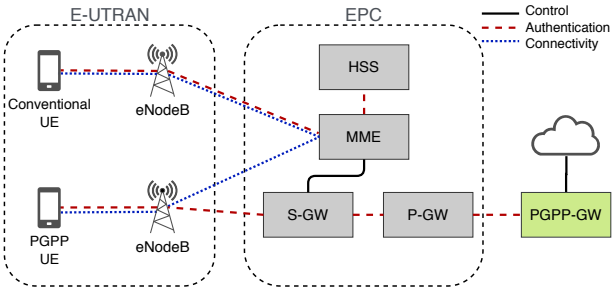


Figure 1: Simplified LTE architecture with and without PGPP. PGPP decouples authentication and connectivity credentials and shifts authentication to a new, external entity, the PGPP-GW. Details of the PGPP-GW are found in §5.1.

2.1 Cellular architecture overview

The 4G LTE architecture can be divided into two areas: the Evolved UMTS Terrestrial Radio Access Network (E-UTRAN), which is responsible for radio access; and the Evolved Packet Core (EPC), which includes the entities responsible for authentication and connectivity to the network core. Figure 1 shows a simplified architecture for both conventional cellular as well as with PGPP. PGPP moves authentication and billing to a new entity, the PGPP-GW, that is external to the EPC. We detail PGPP's specific changes in §5. We include a glossary of cellular terms in Appendix B.

E-UTRAN. The E-UTRAN is the network that facilitates connectivity between user devices (UEs)—commonly a cell phone with a SIM card installed—and the serving base station (eNodeB). The E-UTRAN is responsible for providing UEs a means of connecting to the EPC via eNodeBs.

EPC. The EPC is the core of the cellular network and includes entities that provide authentication, billing, voice, SMS, and data connectivity. The EPC entities relevant to our discussion are the Mobility Management Entity (MME), the Home Subscriber Server (HSS), and the Serving and Packet Data Network Gateways (S-GW and P-GW, respectively). The MME is the main point of contact for a UE and is responsible for orchestrating mobility and connectivity. UEs authenticate to the network by sending an identifier that is stored in the SIM to the MME. The HSS is then queried to verify that the UE is a valid subscriber. Once the UE is authenticated, the MME assigns the UE to an S-GW and P-GW, which offer an IP address and connectivity to the Internet. Note that LTE networks can include many copies of these entities and contain many more entities; however, for the purposes of our discussion this simplified model suffices.

MVNOs. We design our solution to be implemented by a Mobile Virtual Network Operator (MVNO). MVNOs are virtual in that they offer cellular service without owning the infrastructure itself. Rather, MVNOs pay to share capacity on the infrastructure that an underlying carrier operates. MVNOs can choose whether they wish to operate their own LTE core entities such as the MME, HSS, and P-GW, which is the type

of operation we propose. MVNOs that run their own core network are often called "full" MVNOs. Critically, our architecture is now feasible as the industry moves toward "whitebox" eNodeBs that connect to a central office that is a datacenter with virtualized EPC services, as in the Open Networking Foundation's M-CORD project [26] and in the upcoming 5G standard. Recent work has shown that dramatic performance gains are possible using such newer architectures [54, 55].

2.2 Identity in the cellular architecture

Maintaining user privacy has long been challenging in cellular networks as it is not a primary goal of the architecture. In order to authenticate users for access and billing purposes, networks use globally unique identifiers. Likewise, the infrastructure itself must always know the location of a user in order to minimize latency when providing connectivity. We briefly discuss cellular identifiers as well as location information available from the perspective of the network in this section. We use acronyms from the 4G LTE architecture; however, similar entities exist in all generations (2G, 3G, 5G).

User identifiers. There are multiple identifiers that can be used to associate network usage with a given subscriber. The International Mobile Subscriber Identity (IMSI) is the identifier used to gain access to the network when a phone (UE) performs initial attachment. The IMSI is globally unique, permanent, and is stored on the SIM card. Carriers maintain a HSS database containing the list of IMSIs that are provisioned for use on the network and subscription details for each.

Given the IMSI's importance and sensitivity, temporary identifiers are often used instead. The Globally Unique Temporary Identifier (GUTI) can be thought of as a temporary replacement for an IMSI. Once a phone attaches to the network, the Mobility Management Entity (MME) generates a GUTI value that is sent to the UE, which stores the value. The UE uses the GUTI rather than the IMSI when it attaches to the network in the future. The GUTI can be changed by the MME periodically. Prior work recently found that GUTIs are often predictable with consistent patterns, thus offering little privacy [31], but this can be remedied with a lightweight fix that we expect will be used going forward.

User location information. Cellular networks maintain knowledge of the physical location of each UE. Location information is necessary to support mobility and to quickly find the UE when there is an incoming call, SMS, or data for a user. The mechanism used to locate a UE is known as "paging" and it relies on logical groupings of similarly located eNodeB's known as "tracking areas" (TAs). Each eNodeB is assigned to a single TA. TAs can be thought of as broadcast domains for paging traffic. If there is incoming data for an idle UE, the paging procedure is used, where the network broadcasts a paging message to all eNodeBs in the user's last-known TA. Prior work has shown that the paging mechanism can be leveraged by attackers that know an identifier of the victim (*e.g.*, phone number, WhatsApp ID) to generate paging messages intended for the victim, which enables an unprivileged attacker to identify a specific user's location [42]. Cellular operators also often store location metadata for subscriber, giving them the ability to trace user movement and location history.

3 Measurement study

In this section we demonstrate the privacy leakage that exists in today's cellular architecture by conducting a measurement study while acting as a relatively weak attacker in a real-world environment. Recall from §2.2 that the IMSI is a globally unique, permanent identifier. Unfortunately for user privacy, the traditional cellular architecture uses IMSIs for authentication and billing, as well as providing connectivity, causing the IMSI to be transmitted for multiple reasons.

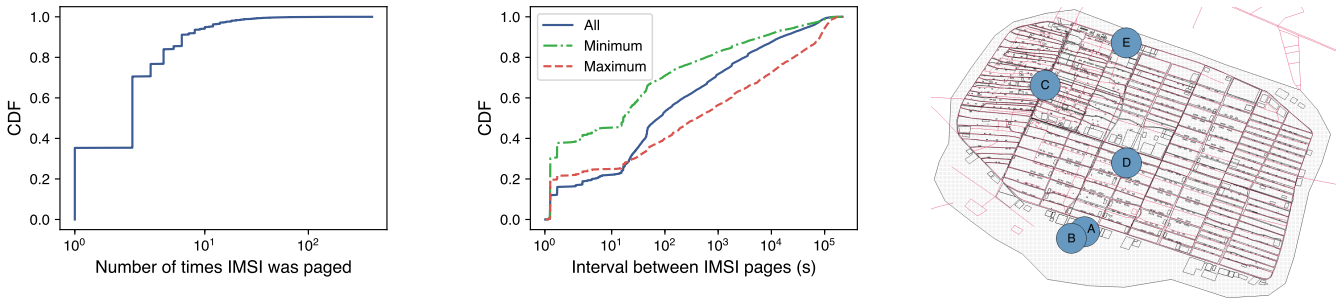
Because of its importance and permanence, the IMSI is seen as a high-value target for those who wish to surveil cellular users. For example, in recent years there has been a proliferation of cell-site simulators, also known as IMSI catchers. These devices offer what appears to be a legitimate base station (eNodeB) signal. Since UE baseband radios are naïve and automatically connect to the strongest signal, they attempt to attach to the IMSI catcher and offer their IMSI. IMSI catchers have been used extensively by law enforcement and state-level surveillance agencies, with and without warrants, to identify, track, and eavesdrop on cellular users [52].

Dataset. We analyze a dataset of cellular traces that our team gathered previously in a large refugee camp over a period of three days. The traces include messages that were sent on broadcast channels in plaintext for three cellular providers that offer service in the area. Traces were captured using software defined radios and mobile phones. The trace dataset provides a vantage point that is akin to an IMSI catcher.[1]

IMSIs are often broadcast in-the-clear. We discover that, while the architecture is designed to largely use temporary GUTIs once UEs are connected, IMSIs are often present in paging messages. Overall we see 588,921 total paging messages, with 38,917 containing IMSIs (6.6% of all pages). Of those messages we see 11,873 unique IMSIs. We track the number of times each individual IMSI was paged and plot a CDF in Figure 2a. As shown, more than 60% of IMSIs were paged more than once in the traces. Note that we count multiple pages seen within one second as a single page. Given this network behavior, even a passive eavesdropper could learn the permanent identifiers of nearby users.

IMSIs can be tracked over time. Given that IMSIs are regularly broadcast, an eavesdropper can track the presence or absence of users over time. We investigate the intervals between pages containing individual IMSIs. In Figure 2b we plot a CDF of intervals (greater than one second) between subsequent pages of individual IMSIs. Overall, we see that

[1] Trace collection methodology and analysis received IRB approval; extraneous details omitted for blind review.



(a) IMSI page counts. (b) Intervals between pages. (c) User locations over time.

Figure 2: Analysis of IMSI broadcasts based on cellular traces captured in measurement study.

IMSIs are repeatedly broadcast over time, even though the design of the architecture should dictate that IMSIs should be used sparingly in favor of temporary GUTIs.

Individuals can be tracked over time. Given that we can track IMSIs over time, a passive attacker can track individuals' movements. Figure 2c shows locations of base stations that broadcast the IMSI for a single user in the traces. As shown, we saw the user in multiple locations over the course of two days. Location A was recorded at 10am on a Monday; location B was thirty minutes later. The user connected to a base station at location C at noon that same day. Locations D and E were recorded the following day at noon and 1:30pm, respectively. From this we see that a passive observer unaffiliated with a cellular carrier can, over time, record the presence and location of nearby users. This attacker is weak, with a relatively small vantage point. In reality, carriers *can and do* maintain this information for *all* of their users.

4 Scope

We believe that many designs are possible to increase privacy in mobile networks, and no architecture, today or in the future, is likely to provide perfect privacy. Nevertheless, below we discuss various properties that PGPP strives to achieve.

Prior work examined the security vulnerabilities in modern cell networks [33, 42, 63] and revealed a number of flaws in the architecture itself. In addition, data brokers and major operators alike have taken advantage of the cellular architecture's vulnerabilities to profit off of revealing sensitive user data. We believe mobile networks should aim to, at a minimum, provide one or both of the following privacy properties.

Identity privacy. A network can aim to protect users' identity. Networks—as well as third party attackers—identify users through IMSIs, which are intended to be uniquely identifying.

Location privacy. A network can aim to protect information about the whereabouts of a phone.

Naturally, these privacy properties do not exist in isolation; they intersect in critical ways. For example, attackers often aim to learn not only who a user is but where a specific user is currently located, or where a user was when a specific call was made. Also, the definition of an attacker or adversary is a complex one, and depending on context may include individuals aiming to steal user data, mobile carriers and data brokers looking to profit off of user data, governments seeking to perform bulk surveillance, law enforcement seeking to monitor a user with or without due process, and many others. Due to context dependence, we do not expect all privacy-focused mobile networks to make the same choice of tradeoffs.

4.1 Cellular privacy threat model

Given the above discussion, we distinguish between bulk and targeted data collection. We define bulk collection to be the collection of information from existing cellular architecture traffic without the introduction of attack traffic; thus, bulk collection is passive. Bulk attacks commonly target user identities (*e.g.*, IMSIs). PGPP's core aim is to protect against bulk attacks. Targeted attacks are active and require injection of traffic to attack specific targets. Targeted attacks are often aimed at discovering a victim's location. We also delineate attacks by the adversary's capabilities, as they may have visibility into an entire network (global) versus, for an unprivileged attacker, some smaller subset of a network's infrastructure (local). Table 1 gives the taxonomy of attacks.

Carriers and governments are the most common **global-bulk** attackers. Such bulk surveillance is commonplace in cellular networks, and has been at the center of recent lawsuits and privacy concerns. Attacks that employ IMSI catchers or passively listen to broadcasts using software-defined radios are considered **local-bulk**. Here, an IMSI catcher is only able to monitor phones that connect directly to it, so its visibility is limited to its radio range. Similarly, SDR-based passive snooping (as in the example in §3) is only able to monitor nearby base stations and will miss portions of the network. We design PGPP with a primary focus on thwarting bulk attacks by nullifying the value of IMSIs (§5.1).

Local-targeted attacks can be carried out by ordinary users by generating traffic that causes a network to page a victim (*e.g.*, phone call to the victim). As local-targeted attackers

		Attack type	
		Bulk	**Targeted**
Visibility	**Global**	Carrier logs [18, 19, 39, 69] / Government Surveillance [9]	Carrier Paging
	Local	SDR [3, 50, 68] / IMSI Catcher [37, 52]	Paging attack [34, 42]

Table 1: Common cellular attacks.

do not have visibility into the entire network, they must rely upon knowledge of the geographic area that is encompassed by a tracking area. Due to the prevalence of such attacks, as an enhancement, an operator can provide functionality, in cooperation with the user, that reduces the efficacy of local-targeted attacks through the use of TALs (§5.2).

Global-targeted attacks represent a very powerful attacker who can actively probe a victim while having global visibility of the network. We envision defenses against such attacks would require fundamental changes to to communication models. PGPP does not mitigate global-targeted attacks as we focus on immediately deployable solutions; we leave this to future work.

4.2 Aims

Next we discuss the aims of PGPP by considering several common questions that arise.

What sort of privacy does PGPP provide? As its name suggests, PGPP aims to provide "pretty good" privacy; we don't believe there is a solution that provides perfect privacy, causes no service changes (*i.e.*, does not increase latency), and is incrementally deployable on today's cellular networks. The main focus is to offer privacy against global-bulk surveillance of mobility and location, a practice by carriers that is widespread and pernicious. We thwart this via eliminating the IMSI as an individual identifier and decoupling the authentication and connectivity mechanisms in the cellular architecture.

Isn't 5G more secure than legacy generations? We are currently on the brink of a new generation of cellular connectivity: 5G. While the ITU requirements for what can be called 5G have not been fully ratified (they are scheduled for this year), many preliminary components of the standard have achieved widespread agreement.

Encrypted IMSIs. 5G includes the addition of encrypted IMSIs, where public key cryptography, along with ephemeral keys generated on the SIM, is used to encrypt the IMSI when sending it to the network. This protects user IMSIs from eavesdroppers. However, encrypted IMSIs do not prevent the cellular provider *itself* from knowing the user's identity. An analogy for encrypted IMSIs can be found in TLS for web traffic: eavesdroppers cannot see unencrypted traffic, yet the endpoints (the web server for TLS, the cellular core in 5G) can. The goal of this work is to not only thwart local-bulk attacks, but also protect user privacy from mobile operators that would otherwise violate it (*i.e.*, global-bulk attacks).

Small cell location privacy. The 5G standard strives for reduced latencies as well as much higher data throughputs. This necessitates the use of cells that cover smaller areas in higher frequency spectrum in order to overcome interference compared with previous cellular generations that used macro-cells to provide coverage to large areas. A (likely unintended) byproduct of 5G's use of smaller cells is a dramatic *reduction* in location privacy for users. As the 5G network provider maintains state pertaining to the location in the network for a given user for the purposes of paging, smaller cells result in the operator, or attacker, knowing user locations at a much higher precision compared with previous generations.

What about active | traffic analysis | signaling attacks? While active, targeted attacks aren't our main focus, we improve privacy in the face of them by leveraging TALs to increase and randomize the broadcast domain for paging traffic, making it more difficult for attackers to know where a victim is located (analyzed in §6.2). Further, the goal of many active attacks is to learn users' IMSIs, and our nullification of IMSIs renders such attacks meaningless.

An attacker with a tap at the network edge could use traffic analysis attacks to reduce user privacy. We largely view this as out of scope as users can tunnel traffic and use other means to hide their data usage patterns.

Cellular networks rely on signaling protocols such as Signaling System 7 (SS7) and Diameter when managing mobility as well as voice and SMS setup and teardown. These protocols enable interoperability between carriers needed for roaming and connectivity across carriers. Unfortunately, these protocols were designed with inherent trust in the network players, and have thus been used to reduce user privacy and disrupt connectivity [24, 30, 49, 53, 62]. We design PGPP for 4G/5G data only, which renders legacy SS7 compatibility moot. Our PGPP design expects users to use outside messaging services rather than an in-EPC IMS system.

Can PGPP support roaming? Yes. While we envision that many PGPP users would explicitly not wish to roam, as roaming partners may not provide privacy guarantees, roaming is possible using a Diameter edge agent that only allows for home routed roaming, forcing traffic to route from the visited network's S-GW back to the PGPP operator's P-GW, rather than local breakout due to our authentication mechanism (§5.1). Roaming, and international roaming in particular, adds billing complexities for the PGPP operator. Typically, the visited network collects call data records for each roaming user on its network and calculates the wholesale charges payable by the home network. The visited network then sends a Transferred Account Procedure (TAP) file to the home network via a data clearing house. The home network then pays the visited network. In PGPP, the individual identity of the user that roamed is not known, yet the PGPP operator remains able to pay the appropriate fees to visited networks.

How does PGPP protect user privacy for voice or text service? Out of the box, PGPP doesn't provide protection for such service. Instead, PGPP aims provide privacy *from the*

cellular architecture itself, and in doing so users are free to use a third party VoIP provider (in which case the phone will operate identically to a normal phone for telephony service from a user's perspective) or use recent systems by Lazar et al. [44, 45] that provide strong metadata privacy guarantees for communications, or similar systems such as [16,17,46,67]. We view PGPP as complementary to such systems.

How does PGPP protect users against leaky apps? PGPP doesn't, as it is about providing protection in the cellular infrastructure. Even without leaky apps, users can always intentionally or inadvertently reveal their identity and location. Leaky apps make this worse as they collect and, sometimes, divulge sensitive user information. We see PGPP as complementary to work that has targeted privacy in mobile app ecosystems. Further, apps are not as fundamental as connectivity—users can choose whether to install and run a leaky app, and can constrain app permissions. However, phones are, by their nature, always connected to carrier networks, and those very networks have been selling user data to third parties.

If users can't be identified by carriers, how can carriers still make money? We introduce PGPP tokens in §5.1 as a mechanism for a PGPP operator to charge customers while protecting user anonymity.

Can't phone hardware be tracked as well? Phones have an International Mobile Equipment Identity (IMEI). The IMEI is assigned to the hardware by the manufacturer and identifies the manufacturer, model, and serial number of a given device. Some operators keep an IMEI database to check whether a device has been reported as stolen, known as an equipment identity register (EIR); IMEIs in the database are blacklisted.

For many devices, the IMEI can be changed through software, often without root access. We envision a PGPP MVNO would allow for subscribers to present their unchanged device IMEI, giving the PGPP operator the opportunity to check against a EIR to verify the phone has not been reported as stolen. At that point, the IMEI could be reprogrammed to a single value, similar to our changes to the IMSI. Note that different jurisdictions have different rules about whether, how, and by whom an IMEI can be changed, so only in some cases IMEI changes require cooperation with the MVNO.

Is PGPP legal? Legality varies by jurisdiction. For example, U.S. law (CALEA [1]), requires providers to offer lawful interception of voice and SMS traffic. A PGPP-based MVNO is data-only, with voice and messaging provided by third parties. CALEA requires the provider to offer content of communication data at the P-GW, *e.g.*, raw (likely-encrypted) network traffic. This is supported by PGPP.

5 Design

In this section we describe the mechanisms PGPP employs to increase user identity and location privacy. Ultimately, PGPP's design choices appear obvious in retrospect. We believe its simplicity is an asset, as PGPP is compatible with existing networks and immediately deployable.

In order to provide identity privacy against bulk attacks, we nullify the value of the IMSI, as it is the most common target identifier for attackers. In our design, we choose to set all PGPP user IMSIs to an identical value to break the link between IMSI and individual users. This change requires a fundamental shift in the architecture, as IMSIs are currently used for connectivity as well as authentication, billing, and voice/SMS routing. We design a new cellular entity for billing and authentication that preserves identity privacy. Fortunately, the industry push for software-based EPCs makes our architecture feasible. We describe the architecture in §5.1.

To provide location privacy from targeted attacks, PGPP leverages an existing mechanism (TALs) in the cellular specification in order to grow the broadcast domain for control traffic (§5.2). By changing the broadcast domain for every user, the potential location of a victim is broadened from the attacker's vantage point.

5.1 User identity privacy

As discussed in §2.2, IMSIs are globally unique, permanent identifiers. As such, they are routinely targeted by attackers, both legal and illegal. In this section we re-architect the network in order to thwart *bulk* attacks introduced in §4.1 that are based on identifying individuals via IMSI.

We decouple back-end connectivity from the authentication procedure that normally occurs at the HSS when a UE attaches to the network. Instead, the PGPP operator issues SIM cards with *identical* IMSIs to all of its subscribers. In this model, the IMSI is used only to prove that a user has a valid SIM card to use the infrastructure and, in turn, the PGPP network can provide an IP address and connectivity and offer the client a GUTI, providing the user with a unique identity necessary for basic connectivity.

LTE authentication is normally accomplished using IMSIs at the HSS; however, all PGPP users share a single IMSI. Thus, to authenticate a user, we designed a post-attach oblivious authentication scheme to ensure that the PGPP operator is able to account for the user without knowing who they are.

PGPP Gateway. In order to perform this authentication we create a new logical LTE entity called a PGPP Gateway (PGPP-GW), which sits between the P-GW and the public Internet. The P-GW is configured to have a fixed tunnel to a PGPP-GW, which can be located outside of the PGPP operator's network. Using this mechanism, the PGPP-GW only sees an IP address, which is typically NATed by the P-GW, and whether that IP address is a valid user. Notably, it does not have any information about the user's IMSI. The PGPP-GW design also allows for many different architectures. For instance, multiple PGPP-GWs could be placed in multiple datacenters or even use a privacy service such as Tor.[2]

Authentication properties. From the perspective of the PGPP-GW, there are multiple properties an authentication

[2]We leave exploration into such scenarios to future work.

Scheme	Customer?	Anonymous?	Unique?
Standard auth	•		
Group/ring sig	•	•	
Linkable ring sig	•		•
Cryptocurrency		•	•
PGPP tokens	•	•	•

Table 2: Three properties needed for user authentication in a privacy-preserving cell network and schemes to achieve them.

scheme must guarantee: (1) the gateway can authenticate that a user is indeed a valid customer[3]; (2) the gateway and/or any other entities cannot determine the user's identity, and thus cannot link the user's credentials/authentication data with a user identity; and (3) the gateway can determine whether a user is unique or if two users are sharing credentials.

As we show in Table 2, the challenge is that standard approaches for authentication only provide one of the three required properties and widely-studied cryptographic mechanisms only provide two of the three properties. For example, an ordinary authentication protocol (of which there are many [7,36]) can provide property 1) but not 2) and 3). A cryptographic mechanism such as group signatures [8, 12] or ring signatures [20, 59] can protect the user's identity upon authentication, providing properties 1) and 2), but not 3) as providing the last property would violate the security of the signature scheme. Similarly, traitor tracing schemes [14] (such as for broadcast encryption [25]) can provide all three properties but in practice cannot provide property 3) as the traitor tracing would require actual physical confiscation of the "traitor" phone by the MVNO, which is infeasible. A variation on ring signatures known as linkable ring signatures [48] provides the ability for a user's identity to be revealed if the user signs multiple messages with the same key. While this is useful in establishing that the user is unique and hasn't shared their credentials, it also partially violates the user's anonymity, as that key cannot be used again.

Effective authentication. There are two approaches that we view as viable, depending on the circumstances. An anonymity-preserving cryptocurrency can provide properties 2) and 3), but not 1) as a cryptocurrency would combine billing and authentication at the PGPP-GW. For MVNOs that are not required to know their customers, an anonymity-preserving cryptocurrency may be the ideal solution for both user authentication and payment, though even the best coins provide imperfect anonymity guarantees [38].

To provide all three properties, we develop a simple scheme called **PGPP tokens** that helps us sidestep the issues with alternative approaches. The choice of authentication scheme is deployment-context specific. With PGPP tokens, when paying a monthly bill a user retrieves authentication tokens that are blind-signed using Chaum's classic scheme [6, 11] by the billing system. Later, when authenticating to the service, the user presents tokens and the service (the PGPP-GW) verifies their signature before allowing the user to use the network. The token scheme ensures that the service can check the validity of tokens without identifying the user requesting access. The user then presents the next token in advance so as to ensure seamless service. Note that PGPP tokens disallow the post-pay model for cellular billing, as the network would be required to know the identity of users in order to accurately charge them for usage. Therefore, PGPP is pre-pay only, though this can be adjusted to emulate post-payment (*e.g.*, users pre-pay for tokens on an ongoing basis rather than only monthly, and tokens are valid for a longer time period, such as a year, rather than for only one billing period).

Each token represents a unit of access, as is appropriate for the service provider. Some providers may choose to offer flat-rate unlimited-data service, in which case each token represents a fixed period of time; this is the default approach that we use to describe the scheme below. Other providers may choose to offer metered service, in which case each token represents a fixed unit of data, such as 100 MB or 1 GB, rather than a period of time. Still others may choose to provide two-tiered service priority by marking each token with a priority bit, in addition to either unlimited data or metered data service; such prioritization does come with slight privacy loss, as the MVNO and MNO alike would be able to differentiate which priority level was in use. The privacy loss of two-tiered data priority can be partially mitigated by offering all users some amount of time or GB of high-priority service after which they must fall back to low-priority service; such a service plan structure is fairly standard in the industry today. In such a setting, each user would have both high-priority and low-priority tokens and thus would not be clearly stratified into two identifiable groups of users.

At the beginning of a billing period, the billing system defines s time slices (*e.g.*, corresponding to hours) or another unit of access (*e.g.*, a unit of data) and generates s RSA keypairs for performing blind signatures using Chaum's scheme. It then appends the public keys for this time period to a well-known public repository that is externally maintained (*e.g.*, on GitHub), and these are fetched by users. The user generates s tokens where each token takes the form $i \| r$ where i is the time slice index as a 256-bit unsigned value zero indexed from the beginning of the billing period, and r is a 256-bit random value chosen by the user. The user then blinds these tokens. The user pays the bill using a conventional means of payment (*e.g.*, credit card), and presents the blinded tokens to the billing system to be signed; the system signs each token with the corresponding time slice key and returns these values to the user. The user unblinds the response values and verifies the signatures for each.

Upon later authentication to the service, the user presents its signed token for the current time slice to the PGPP-GW,

[3] Due to "Know Your Customer" rules in some jurisdictions, the provider may need to have a customer list, necessitating that the user authentication scheme be compatible with periodic explicit customer billing.

which verifies the signature and if valid begins forwarding the user's traffic onto the Internet. Since the token signature was generated using Chaum's scheme, the service cannot determine which human user corresponds to which signed token. If the same token is used by two different users during the same time period then the service can conclude that a user has shared their credentials and is attempting to cheat.

The costs of this scheme to both the PGPP operator and the user are low. The operator stores the list of used tokens in a standard consistent and replicated cloud database, so the service can operate multiple PGPP-GWs, though it is likely that a small number of PGPP-GWs can serve a large number of users: we benchmarked the 2048-bit RSA signature verification used here at 31μs per call using Crypto++ [21] on a single core of a 2.6GHz Intel Xeon E5-2640 CPU, and thus with a single CPU core the PGPP-GW can handle token verification for tens of millions of users. The tokens themselves are small and the storage cost to the provider is about 1.5 MB / user per time period, which is a small amount for any user's phone to store and for a provider even hundreds of millions of tokens amounts to mere GBs of data in cloud storage.

User device agent. To automate the process of authenticating with the PGPP-GW, we create a simple agent that runs as background job on the user device. This agent leverages the Android JobScheduler API; in the event of cellular connectivity, the JobScheduler triggers PGPP-token-based authentication with the PGPP-GW. The agent establishes a TLS connection to the PGPP-GW and then sends the token for the current time slice. Once the user presents a valid token, the PGPP-GW begins forwarding traffic for that user, and thus this behavior is akin to a captive portal though the authentication is automatic and unseen by the user.

5.2 Location privacy

As described in §2.2, cellular operators track user location in the form of tracking areas for UEs in order to quickly find users when there is incoming content. PGPP leverages an existing mechanism in the cellular standard to reduce the effectiveness of *local-targeted* attacks described in §4.1.

Paging has been exploited in the past to discover user location by adversaries. However, the use of tracking areas is useful for the cellular provider in that it confines the signaling message load (*i.e.*, paging messages) to a relatively small subset of the infrastructure. Tracking areas reduce mobility signaling from UEs as they move through the coverage zone of a single tracking area. Note that emergency calling represents a special case in cellular networks. When a device dials 911, the phone and network attempt to estimate accurate location information. In this work we do not alter this functionality as we anticipate that users dialing 911 are willing to reveal their location.

TALs. In PGPP, we exploit the tracking area list (TAL) concept, introduced in 3GPP Release 8 [2]. Using TALs, a UE no longer belongs to a single tracking area, but rather is given a list of up to 16 tracking areas that it can freely move through without triggering a tracking area update, essentially creating larger tracking areas. Whereas prior work has focused on using TALs to pre-compute optimal tracking area combinations for users [56–58], in PGPP, we use TALs to provide provide improved location anonymity. Typically, TALs consist of groups of adjacent tracking areas that are pre-computed, essentially growing the tracking area for a UE to the union of all tracking areas in the TAL. We do not use TALs in this way. Instead, we generate TALs on-the-fly and generate them uniquely for each UE. When a UE attaches or issues a tracking area update message, the MME learns the eNodeB and tracking area the UE is currently attached to. The MME then generates a unique TAL by iteratively selecting at random some number (up to the TAL limit of 16) of additional, adjacent tracking areas. By generating unique TALs for each user, attackers are unable to know a priori which set of tracking areas (or eNodeBs) that victim is within. We explore tradeoffs in terms of TAL length, control traffic overhead, and location anonymity in the next section.

6 Analysis

To study the implications of a PGPP deployment, we create a simulation to model users, mobility, and cell infrastructure. We study the impact of PGPP's design on various cellular attacks that occur today. We then analyze the inherent tradeoffs from the PGPP operator's perspective, as improved privacy comes at the price of increased control traffic. Lastly, we examine PGPP in a lab testbed on real devices.

6.1 Simulation configuration

eNodeB dataset. We select Los Angeles County, California as the region for our simulation, which provides a mix of both highly urban areas as well as rural areas. For eNodeB location information, we use OpenCellID [43], an open database that includes tower locations and carrier information. To simplify the simulation, we select eNodeBs from the database that are listed as providing LTE from AT&T, the provider with the most eNodeBs (22,437) in the region.

Given their geographic coordinates, we estimate coverage areas for every eNodeB using a Voronoi diagram. During the simulation, a UE is assigned to the eNodeB that corresponds to the region the UE is located within. While such discretization is not likely in reality as UEs remain associated with an eNodeB based on received signal strength, this technique provides us with a tractable mobility simulation. A partial map of the simulation region is shown in Figure 3. ENodeB regions are shaded based on the tracking area value in the OpenCellID database.

Mobility traces. To simulate realistic mobility patterns (*i.e.*, users must follow available paths), we generate mobility traces using the Google Places [29] and Directions [28] APIs. First, we use the Places API to find locations in the simulation region that are available when searching for "post

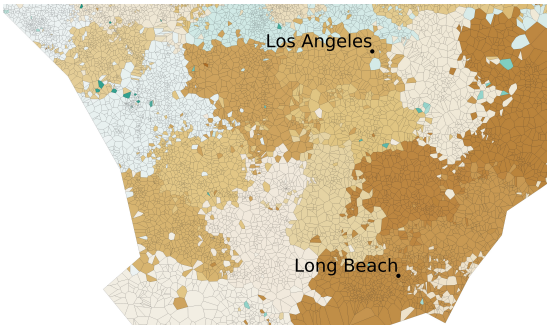


Figure 3: Partial simulation map. Cells are shaded by AT&T LTE tracking area.

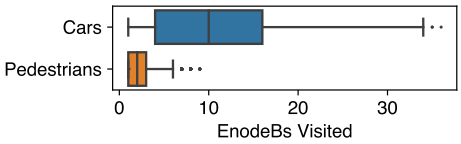
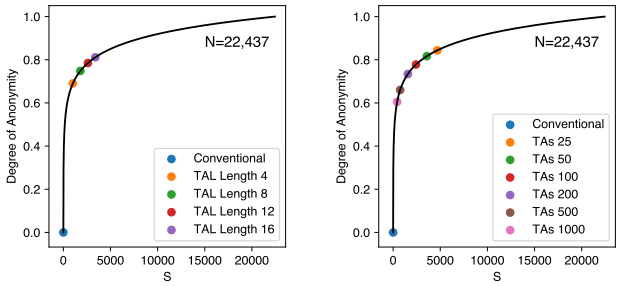


Figure 4: ENodeBs visited by simulated mobile users.

office." Each place is associated with latitudinal and longitudinal coordinates. We then generate mobility traces by randomly selecting start and end points, and use the Directions API to obtain a polyline with coordinates along with estimated times to reach points along the line. We generate 50,000 mobility traces: 25,000 cars and 25,000 pedestrians. We then use ns-3 to process the mobility traces and generate coordinates for each trace at 5-second intervals, in a method similar to [10]. We use this output, along with the eNodeB Voronoi diagram to assign each simulated UE to an eNodeB for every 5-second interval in the mobility trace. Figure 4 shows the distribution of the number of eNodeBs visited by UEs in the simulation. As expected, car trips result in a significantly higher number of eNodeBs for a UE compared with pedestrian trips.

Synthetic traffic. We simulate one hour. To create control traffic, at every 5-second interval we randomly select 5% of the user population to receive a "call." A call results in a paging message that is sent to all eNodeBs in the UE's tracking area. Each paged user enters a 3-minute "call" if it is not already in one, at which point further paging messages are suppressed for that user until the call is complete. We run the simulation with PGPP enabled as well as with the conventional infrastructure setup.

Custom TAs. As we detail further in §6.3, large TALs increase control traffic loads, which lowers the network's user capacity. Therefore, we generate new tracking areas in the underlying network in order to mitigate the control traffic burden. As tracking areas normally consist of groups of adjacent eNodeBs, we need a method by which we can cluster nearby eNodeBs into logical groupings. To do so, we use k-means clustering with the eNodeB geographic coordinates allowing



(a) TALs. (b) Custom TAs.

Figure 5: Degree of anonymity using TALs and custom TAs.

for Euclidean distance to be calculated between eNodeBs. We generate several underlying tracking area maps, with the number of TAs (*i.e.*, k-means centers) ranging from 25 to 1,000. For comparison, the AT&T LTE network in the simulation is composed of 113 TAs.

6.2 Cellular privacy attack analysis

Given the taxonomy we presented in §4.1, we analyze the identity and location privacy benefits of PGPP in the simulated environment.

Global-bulk attacks. By nullifying the value of IMSIs, separating authentication with connectivity, and increasing the broadcast domain for users, we increase user identity privacy even with an adversary that is capable of bulk surveillance over an entire network (*e.g.*, operators, governments).

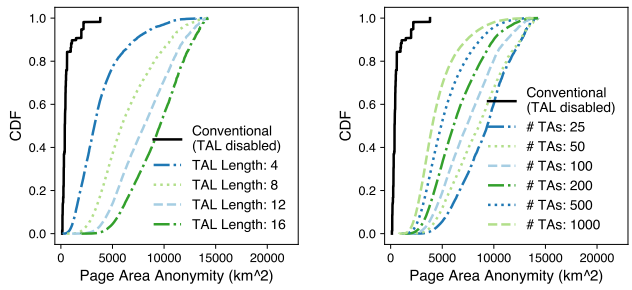
Anonymity analysis We measure the anonymity of a user when under bulk attacks using *degree of anonymity* [22]. The degree of anonymity value ranges from zero to one, with ideal anonymity being one, meaning the user could be any member of the population with equal probability. In this case, we consider the IMSI value to be the target identity. The size of the anonymity set for a population of N users will result in a maximum entropy of:

$$H_M = log_2(N) \tag{1}$$

The degree of anonymity is determined based on the size of the subset of user identities S that an attacker could possibly believe the victim to be:

$$d = \frac{H(X)}{H_M} = \frac{log_2(S)}{log_2(N)} \tag{2}$$

Given global visibility into the network, we can reason about the anonymity set using the number of eNodeBs that a victim could possibly be connected to. This is because a cellular carrier can know the exact base station that a user is connected to once the UE enters an active state. As a baseline, the anonymity set for traditional cellular is $\frac{log_2(1)}{log_2(22{,}437)} = 0$, as each IMSI is a unique value. With PGPP, IMSIs are identical, so from the perspective of the carrier, the victim could be



(a) TALs. (b) Custom TAs.

Figure 6: Area anonymity using TALs and custom TAs.

connected to any eNodeB that has at least one PGPP client connected to it. Using our simulated environment we collect, for each paging message, the number of eNodeBs that had users within their range and use the median value to calculate the degree of anonymity. Figures 5a and 5b show the degree of anonymity using different configurations of TALs and custom TAs, respectively. We see that high degrees of anonymity are attainable despite an attacker's global visibility. For instance, with TALs of length 8, the degree of anonymity is 0.748.

Local-bulk attacks. PGPP's use of identical IMSIs reduces the importance of IMSIs, and by extension the usefulness of local bulk attacks on user identity. An attacker that can view traffic at the eNodeB(s) can gain insight into nearby IMSIs.

In traditional cell networks, each user has a globally unique IMSI ($S = 1$), resulting in a degree of anonymity of zero as the victim could only be one user. In our measurement study (§3), we showed that IMSIs are routinely broadcast over cell networks, making an IMSI catcher or SDR attack powerful. The subset S in PGPP, on the other hand, is the size of the population of PGPP users in a given location, as all IMSI values are identical and a local bulk attacker cannot know the true identity of a single user. To get an idea of S, we can calculate the number of PGPP users connected to each eNodeB in the simulation. Over the course of the simulation, we find a mean value of 223.09 users connected to each eNodeB that has users, which results in a degree of anonymity $\frac{log_2(223.09)}{log_2(50{,}000)} = 0.50$. While this value is somewhat low compared to the ideal value of 1, it is a drastic improvement over conventional cellular architecture, and is dependent on the overall user population in the network. As more PGPP users exist, the degree of anonymity increases.

Local-targeted attacks. In PGPP, local-targeted attacks to discover a user's location are diminished in two ways: first, IMSIs are no longer a useful ID, so identifying an individual among all users is challenging; and second, we use TALs to increase the paging broadcast domain for a given UE. From an attacker's point of view, this broadens the scope of where the target UE may be located.

In Figure 6a, we plot the CDF of geographic areas in which pages are broadcast as we increase TAL lengths using the base map consisting of 113 tracking areas. We calculate the area by generating a bounding box around all eNodeBs that are included in the broadcast domain. As shown, large TALs result in drastically higher area anonymity compared with TALs disabled, particularly considering the number of UEs that could potentially be located in the larger geographic areas. For instance, the median area for the conventional simulation is 378.09 km^2 whereas TAL lengths of 8 and 16 result in median areas of 5,876.96 and 9,585.17 km^2, respectively.

We analyze anonymity with TALs of length 16 while the underlying map is varied using custom TAs. Figure 6b shows our results. We observe that as the number of tracking areas increase, resulting in smaller tracking areas, the area anonymity decreases. However, despite the decrease, the area anonymity remains considerably larger than anonymity with TALs disabled as TALs include additional tracking areas. For instance, the median area for the conventional case is 378.09 km^2 whereas the median area for a base map of 500 tracking areas with TAL 16 is 4891.08 km^2, a nearly 13-fold increase from the perspective of a local targeted attacker.

6.3 Impact of PGPP on network capacity

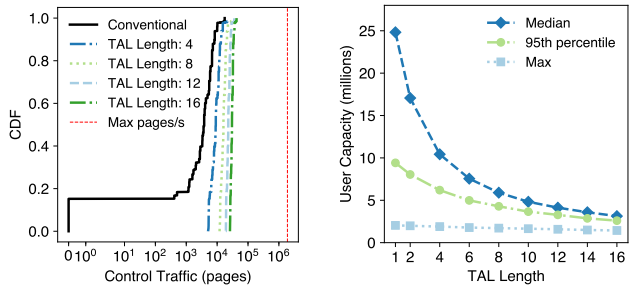
From an operational perspective, the privacy benefits delivered by PGPP must coincide with feasibility in terms of control overhead in order for it to be deployable. Control traffic determines network capacity in terms of the number of users that are serviceable in a given area. In this section, we explore control traffic load when using TALs.

6.3.1 Control overhead with PGPP TALs

We first seek to quantify control message overhead while we leverage tracking area lists to provide location anonymity against local-targeted attacks. Recall from §5.2 that we randomly select additional tracking areas from the simulated coverage area to create TALs, which increases the broadcast domain for a page. Increased control traffic impacts both eNodeBs and MMEs, however, from our experience with real cellular networks the control traffic capacity at eNodeBs is the bottleneck as MMEs have much higher capacity. Thus, we focus on eNodeB control load.

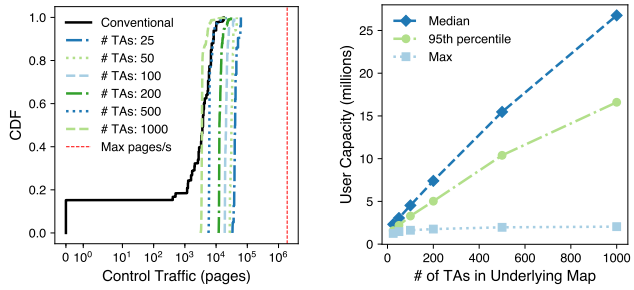
Figure 7a shows a cumulative distribution function (CDF) for the number of pages broadcast by the simulated eNodeBs. In the figure, "Conventional" corresponds to disabling TAL functionality. As expected, larger TAL lengths result in increased control traffic for eNodeBs as they are more likely to be included in the paging broadcast domain for a given UE.

To gain insight into the control limitations of real eNodeBs, we consider the capabilities of a Huawei BTS3202E eNodeB [32], which is limited to 750 pages per second. When capacity planning, it is commonplace to budget paging traffic headroom; accordingly, we estimate the maximum paging capacity for an eNodeB to be 525 pages per second (70% of the BTS3202E capacity). This value is depicted in the vertical red line in the figure (525 pages $\times$ 3600 seconds = 1,890,000 pages/hour). The simulation allows us to illustrate the user



(a) Control traffic with TALs. (b) Capacity with TALs.

Figure 7: Control traffic and system capacities leveraging PGPP TALs in the simulated environment.



(a) Custom TAs: Control traffic. (b) Custom TAs: Capacity.

Figure 8: Control traffic and system capacities with custom tracking areas in the simulated environment.

population that could be supported by the network, provided a population with similar mobility and traffic profiles as defined in §6.1. Recall that we simulate 50,000 users, both pedestrians and cars. We consider the paging load for the network and select the eNodeBs with the maximum paging load, the 95th percentile, and the median to estimate the number of users each could theoretically support by taking into account the max page limitation of the BS3202E. Figure 7b shows the user capacity as TAL lengths are increased. A TAL length of one shows the conventional network, as the TAL is composed of a single tracking area. As expected, larger TALs result in a reduction in the number of users the eNodeBs can handle compared with performance when TALs are disabled, due to increased paging load.

6.3.2 Control overhead with custom tracking areas

As we've demonstrated, large TALs result in eNodeBs with higher control traffic load, effectively reducing the user capacity the network. To explore whether we can re-gain control traffic we again consider new, custom tracking area maps that are generated using k-means where we vary the number of unique tracking areas in the simulated network.

We run the simulation with various custom tracking area maps, with all UEs using TAL lengths of 16. The results are shown in Figures 8a and 8b. We observe that a basemap

Figure 9: PGPP prototype test hardware.

consisting of 25 tracking areas leads to even higher control traffic compared with the conventional (*i.e.*, AT&T) tracking area map. A map consisting of more tracking areas results in TAs with fewer eNodeBs, thus reducing the paging load. We see that a map of 500 TAs, even with a TAL of length 16, results in similar paging load compared with the conventional map with TAL disabled. Correspondingly, the user capacity of the network with a higher number of tracking areas nears the conventional capacity from Figure 7b.

6.4 Testbed analysis

We study our PGPP design on a lab testbed in order to understand potential drawbacks. We implement a software-based EPC and connect commodity phones to the software-defined radio-based eNodeB.

Prototype. We create our prototype code on srsLTE [27], an open-source platform that implements LTE-compliant eNodeB and EPC functionality and can be run using software-defined radios. Our EPC / eNodeB testbed, shown in Figure 9, consists of an Intel Core i7 machine running Linux and a USRP B210 radio. We use off-the-shelf commodity phones (Moto X4, Samsung Galaxy S6, and two OnePlus 5s) with programmable SIM cards installed to allow the phones to connect to the PGPP LTE network.

The srsLTE MME maintains EPS mobility management (EMM) and EPS connection management (ECM) contexts for connected UEs. The contexts are stored as structs that include the UE IMSI in a simple key-value store, with the IMSI serving as the key. When the MME receives S1 application protocol (S1AP) messages (*e.g.*, due to mobility), it looks up the appropriate ECM or EMM contexts to handle the requests. We add an additional value, a PGPPIMSI, into the ECM and EMM structs. The PGPPIMSI is generated by combining the IMSI with a temporary value that is unique to the individual UE-eNodeB-MME connection. Accordingly, each UE has a unique PGPPIMSI, which then allows us to look up the correct context when managing states.

Identical IMSIs and Shared Keys. Given identical IMSI values for all users, the PGPP attach procedure can result in additional steps compared with the traditional attach. This is caused by sequence number synchronization checks dur-

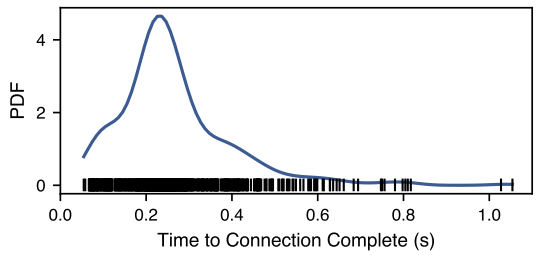


Figure 10: Connection delays due to sync_failure.

ing the authentication and key agreement (AKA) procedure, which is designed to allow the UE and the network to authenticate each other. The fundamental issue is that the HSS and the SIM maintain a sequence number (SQN) value that both entities increment with each successful attach. As multiple devices use the same IMSIs, the sequence numbers held at the HSS and on individual devices will no longer match, causing an authentication failure (known as a sync_failure). At that point the UE re-synchronizes with the HSS. We include an overview figure and details of the procedure in Appendix A.

We explore the delay introduced by sync_failures using our testbed. Figure 10 shows a PDF of the delays to connection completion for UEs that hold identical IMSIs and attempt to authenticate simultaneously. In order to trigger many simultaneous authentication requests, we use openairinterface5G [51] to create 100 simulated UEs. We observe in that the first successful UE usually takes roughly 200 ms to connect, while subsequent UEs that experienced sync_failures experience additional delays. In our relatively small experiment the UEs all successfully connect to the network within 1.1 seconds. In a large-scale production network the number of UEs that simultaneously attempt to connect would be larger. PGPP-based networks can mitigate the issue by using more HSSes, which would reduce the number of UEs that each HSS is responsible for. Fortunately, the push for 5G will lend itself to many HSSes as the core network entities are being redesigned to be virtualized and located nearer to UEs.

7 Related Work

Prior work on anonymous communications often traded off latency and anonymity [16, 17, 46, 67]. Likewise, Tor [23] and Mixnets [13] also result in increased latency while improving anonymity. However, such solutions are inappropriate for cellular systems as, apart from SMS, cellular use cases require low latency. Additionally, the architecture continues to utilize identifiers (*e.g.*, IMSI) that can expose the user to IMSI catcher attack or allow for location tracking by the operator.

There has been extensive prior work on finding security and privacy issues in cellular networks [33, 42, 47, 60, 63]. We decouple the IMSI from the subscriber by setting it to a single value for all users of the network. Altering the IMSI to specifically thwart IMSI catcher and similar passive attacks has been previously proposed [4, 40, 64, 66]. These techniques use pseudo-IMSIs (PMSIs), which are kept synchronized between the SIM and the HSS, or hypothetical virtual SIMs, allowing for user identification. We aim to go beyond thwarting IMSI catchers, and do so while considering active attacks without requiring fundamental changes on the UE; we protect users from the operator itself.

Hussain *et al.* introduce the TORPEDO attack [34], which allows attackers to identify the page frame index and using that, the presence or absence of a victim in a paging broadcast area (*i.e.*, a tracking area). However, our use of tracking area lists to provide additional paging anonymity (§5.2) increases the location in which a victim could potentially be, reducing the effectiveness of third-party paging-related localization attacks. The authors also define the PIERCER attack, which enables the attacker to reveal a victim's IMSI with only their phone number. PGPP nullifies this attack by making all IMSIs identical. Cellular signaling protocols have been demonstrated by multiple works to leave users' privacy vulnerable to attack [24, 30, 49, 53, 62]. Our initial design avoids signaling protocol vulnerabilities by providing data-only rather than voice/SMS, and roaming to other networks can be enabled by requiring home-routing rather than local breakout. Hussain *et al.* identifies a 5G vulnerability that allows an attacker to neutralize GUTI refreshment in [35]. However, this requires a MiTM attack (*e.g.*, IMSI catcher), which necessarily means the attacker knows the victim's location. Additionally, the GUTI is a temporary identifier, and is not associated with a specific user.

Choudhury and Køien alter IMSI values, however both require substantial changes to network entities [15, 41]. We argue that a privacy-preserving architecture must be fully compatible with existing infrastructure as the global telecom infrastructure is truly a network of networks, comprised of multiple operators that connect via well-known APIs.

8 Concluding Remarks

User privacy is a hotly contested topic today, especially as law enforcement organizations, particularly in authoritarian states, insist upon increasingly ubiquitous surveillance. In addition, law enforcement has long demanded backdoor access to private user devices and user data [61].

We do not believe that users of PGPP, in its current form, would be capable of withstanding targeted legal or extra-legal attacks by nation-state organizations (*e.g.*, the FBI or NSA), though PGPP would likely limit the ability of such organizations to continue to operate a regime of mass surveillance of user mobility. In addition, a more common and problematic form of privacy loss today is due to the surreptitious sale of user data by network providers; this is a matter PGPP addresses in a manner that aligns with user autonomy. Our aim is to improve privacy in line with prior societal norms and user expectations, and to present an approach in which privacy-enhanced service can be seamlessly deployed.

References

[1] 103rd Congress, 2nd Session, 1994. Communications Assistance for Law Enforcement Act (CALEA). 47 USC 1001-1010. Public Law 103-414.

[2] 3GPP. General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access. Technical Specification (TS) 23.401, 3rd Generation Partnership Project (3GPP), 01 2015.

[3] S. Aragon, F. Kuhlmann, and T. Villa. SDR-based network impersonation attack in GSM-compatible networks. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015.

[4] Myrto Arapinis, Loretta Mancini, Eike Ritter, Mark Ryan, Nico Golde, Kevin Redon, and Ravishankar Borgaonkar. New privacy issues in mobile telephony: Fix and verification. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, Raleigh, North Carolina, USA, 2012.

[5] World Bank. International telecommunication union, world telecommunication/ict development report and database. https://data.worldbank.org/indicator/IT.CEL.SETS, 2018.

[6] Mihir Bellare, Chanathip Namprempre, David Pointcheval, and Michael Semanko. The one-more-rsa-inversion problems and the security of chaum's blind signature scheme. *Journal of Cryptology*, 16(3), 2003.

[7] Mihir Bellare and Phillip Rogaway. Entity authentication and key distribution. In *CRYPTO*, 1993.

[8] Dan Boneh, Xavier Boyen, and Hovav Shacham. Short group signatures. In *CRYPTO*, 2004.

[9] Carpenter v United States. Number 16-402. Jun 2018.

[10] Tiago Cerqueira and Michele Albano. Routesmobility-model: Easy realistic mobility simulation using external information services. In *Proceedings of the 2015 Workshop on Ns-3*, WNS3 '15, 2015.

[11] David Chaum. Blind signatures for untraceable payments. In *CRYPTO*, 1983.

[12] David Chaum and Eugène Van Heyst. Group signatures. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 257–265. Springer, 1991.

[13] David L Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

[14] Benny Chor, Amos Fiat, and Moni Naor. Tracing traitors. In *CRYPTO*, 1994.

[15] Hiten Choudhury, Basav Roychoudhury, and Dilip Kr. Saikia. Enhancing user identity privacy in lte. In *Proceedings of the 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, TRUSTCOM '12, Washington, DC, USA, 2012.

[16] Henry Corrigan-Gibbs, Dan Boneh, and David Mazières. Riposte: An anonymous messaging system handling millions of users. In *Proceedings of the 2015 IEEE Symposium on Security and Privacy*, SP '15, 2015.

[17] Henry Corrigan-Gibbs and Bryan Ford. Dissent: accountable anonymous group messaging. In *Proceedings of ACM CCS*, 2010.

[18] Joseph Cox. I Gave a Bounty Hunter $300. Then He Located Our Phone. https://motherboard.vice.com/en_us/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile, January 2019.

[19] Joseph Cox. Stalkers and Debt Collectors Impersonate Cops to Trick Big Telecom Into Giving Them Cell Phone Location Data. https://www.vice.com/en_us/article/panvkz/stalkers-debt-collectors-bounty-hunters-impersonate-cops-phone-location-data, March 2019.

[20] Ronald Cramer, Ivan Damgård, and Berry Schoenmakers. Proofs of partial knowledge and simplified design of witness hiding protocols. In *CRYPTO*, 1994.

[21] Crypto++ 8.2, 2019. https://www.cryptopp.com/.

[22] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, PET'02, page 54–68, Berlin, Heidelberg, 2002. Springer-Verlag.

[23] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of USENIX Security*, 2004.

[24] Tobias Engel. Locating mobile phones using signalling system 7. In *25th Chaos communication congress*, 2008.

[25] Amos Fiat and Moni Naor. Broadcast encryption. In *Annual International Cryptology Conference*, pages 480–491. Springer, 1993.

[26] Open Networking Foundation. M-cord open source reference solution for 5g mobile wireless networks. https://www.opennetworking.org/m-cord/, 2019.

[27] Ismael Gomez-Miguelez, Andres Garcia-Saavedra, Paul D. Sutton, Pablo Serrano, Cristina Cano, and Doug J. Leith. srslte: An open-source platform for lte evolution and experimentation. In *WiNTECH '16*, New York City, New York, 2016.

[28] Google. Get started | directions api | google developers. https://developers.google.com/maps/documentation/directions/start, 2019.

[29] Google. Overview | places api | google developers. https://developers.google.com/places/web-service/intro, 2019.

[30] S Holtmans, B Kotte, and S Rao. Detach me not-dos attacks against 4g cellular users from your desk. In *Blackhat Europe 2016*, 2016.

[31] Byeongdo Hong, Sangwook Bae, and Yongdae Kim. Guti reallocation demystified: Cellular location tracking with changing temporary identifier. In *Network and Distributed System Security Symposium, NDSS*, San Diego, California, USA, Feb 2018.

[32] Huawei BTS3202E eNodeB, 2019. "http://support.huawei.com/hdx/hdx.do?docid=SE0000758199&lang=en".

[33] Syed Rafiul Hussain, Omar Chowdhury, Shagufta Mehnaz, and Elisa Bertino. LTEInspector: A systematic approach for adversarial testing of 4G LTE. In *Network and Distributed System Security Symposium, NDSS*, San Diego, California, USA, Feb 2018.

[34] Syed Rafiul Hussain, Mitziu Echeverria, Omar Chowdhury, Ninghui Li, and Elisa Bertino. Privacy Attacks to the 4G and 5G Cellular Paging Protocols Using Side Channel Information. In *Network and Distributed System Security Symposium, NDSS*, San Diego, California, USA, Feb 2019.

[35] Syed Rafiul Hussain, Mitziu Echeverria, Imtiaz Karim, Omar Chowdhury, and Elisa Bertino. 5GReasoner: A property-directed security and privacy analysis framework for 5G cellular network protocol. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, London, United Kingdom, 2019.

[36] Markus Jakobsson and David Pointcheval. Mutual authentication for low-power mobile devices. In *International Conference on Financial Cryptography*, pages 178–195. Springer, 2001.

[37] F Joachim and B Rainer. Method for identifying a mobile phone user or for eavesdropping on outgoing calls. *EPO Patent EP1051053*, 2003.

[38] George Kappos, Haaroon Yousaf, Mary Maller, and Sarah Meiklejohn. An empirical analysis of anonymity in zcash. *arXiv preprint arXiv:1805.03180*, 2018.

[39] Kate Kaye. The $24 Billion Data Business That Telcos Don't Want to Talk About. https://adage.com/article/datadriven-marketing/24-billion-data-business-telcos-discuss/301058/?mod=article_inline, October 2015.

[40] Mohammed Shafiul Alam Khan and Chris J Mitchell. Trashing imsi catchers in mobile networks. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '17, Boston, Massachusetts, 2017.

[41] G. M. Køien. Privacy enhanced mutual authentication in lte. In *2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Lyon, France, Oct 2013.

[42] Denis Foo Kune, John Koelndorfer, Nicholas Hopper, and Yongdae Kim. Location leaks on the gsm air interface. *ISOC NDSS (Feb 2012)*, 2012.

[43] Unwired Labs. Opencellid - open database of cell towers & geolocation. https://www.opencellid.org, 2019.

[44] David Lazar, Yossi Gilad, and Nickolai Zeldovich. Karaoke: Distributed private messaging immune to passive traffic analysis. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'18, Carlsbad, CA, USA, 2018.

[45] David Lazar, Yossi Gilad, and Nickolai Zeldovich. Yodel: Strong metadata security for voice calls. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles*, SOSP '19, Huntsville, Ontario, Canada, 2019.

[46] David Lazar and Nickolai Zeldovich. Alpenhorn: Bootstrapping secure communication without leaking metadata. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 571–586, Savannah, GA, 2016. USENIX Association.

[47] P. P. C. Lee, T. Bu, and T. Woo. On the detection of signaling dos attacks on 3g wireless networks. In *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007.

[48] Joseph K Liu, Victor K Wei, and Duncan S Wong. Linkable spontaneous anonymous group signature for ad hoc groups. In *Australasian Conference on Information Security and Privacy*, pages 325–335. Springer, 2004.

[49] G Lorenz, T Moore, G Manes, J Hale, and S Shenoi. Securing ss7 telecommunications networks. In *Workshop on Information Assurance and Security*, volume 2, page 1115, 2001.

[50] Stig F Mjølsnes and Ruxandra F Olimid. Easy 4G/LTE IMSI catchers for non-programmers. In *International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security*, pages 235–246. Springer, 2017.

[51] Navid Nikaein, Mahesh K Marina, Saravana Manickam, Alex Dawson, Raymond Knopp, and Christian Bonnet. Openairinterface: A flexible platform for 5G research. *ACM SIGCOMM Computer Communication Review*, 44(5):33–38, 2014.

[52] Kristin Paget. Practical cellphone spying. *Def Con*, 18, 2010.

[53] C. Peeters, H. Abdullah, N. Scaife, J. Bowers, P. Traynor, B. Reaves, and K. Butler. Sonar: Detecting SS7 Redirection Attacks with Audio-Based Distance Bounding. In *2018 IEEE Symposium on Security and Privacy (SP)*, May 2018.

[54] Zafar Ayyub Qazi, Phani Krishna Penumarthi, Vyas Sekar, Vijay Gopalakrishnan, Kaustubh Joshi, and Samir R. Das. Klein: A minimally disruptive design for an elastic cellular core. In *Proceedings of the Symposium on SDN Research*, SOSR '16, Santa Clara, CA, USA, 2016.

[55] Zafar Ayyub Qazi, Melvin Walls, Aurojit Panda, Vyas Sekar, Sylvia Ratnasamy, and Scott Shenker. A high performance packet core for next generation cellular networks. In *SIGCOMM '17*, Los Angeles, CA, USA, aug 2017.

[56] S. M. Razavi and D. Yuan. Reducing signaling overhead by overlapping tracking area list in lte. In *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)*, Vilamoura, Algarve, Portugal, May 2014.

[57] S. M. Razavi, D. Yuan, F. Gunnarsson, and J. Moe. Dynamic tracking area list configuration and performance evaluation in lte. In *2010 IEEE Globecom Workshops*, Miami,FL, Dec 2010.

[58] S. M. Razavi, D. Yuan, F. Gunnarsson, and J. Moe. Exploiting tracking area list for improving signaling overhead in lte. In *IEEE Vehicular Technology Conference*, VTC2010, Taipei, Taiwan, May 2010.

[59] Ronald L Rivest, Adi Shamir, and Yael Tauman. How to leak a secret. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2001.

[60] D. Rupprecht, K. Kohls, T. Holz, and C. Pöpper. Breaking LTE on layer two. In *2019 IEEE Symposium on Security and Privacy (SP)*, May 2019.

[61] Stefan Savage. Lawful device access without mass surveillance risk: A technical design discussion. In *ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, Toronto, Canada, Oct 2018.

[62] Hemant Sengar, Ram Dantu, Duminda Wijesekera, and Sushil Jajodia. Ss7 over ip: signaling interworking vulnerabilities. *IEEE Network*, 20(6):32–41, 2006.

[63] Altaf Shaik, Ravishankar Borgaonkar, N. Asokan, Valtteri Niemi, and Jean-Pierre Seifert. Practical attacks against privacy and availability in 4g/lte mobile communication systems. *CoRR*, abs/1510.07563, 2015.

[64] Keen Sung, Brian Neil Levine, and Marc Liberatore. Location privacy without carrier cooperation. In *IEEE Workshop on Mobile Security Technologies, MOST*, page 148, 2014.

[65] Jennifer Valentino-DeVries. Service Meant to Monitor Inmates' Calls Could Track You, Too. https://www.nytimes.com/2018/05/10/technology/cellphone-tracking-law-enforcement.html, May 2018.

[66] Fabian van den Broek, Roel Verdult, and Joeri de Ruiter. Defeating IMSI catchers. In *ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, Denver, Colorado, USA, Oct 2015.

[67] Jelle van den Hooff, David Lazar, Matei Zaharia, and Nickolai Zeldovich. Vuvuzela: Scalable private messaging resistant to traffic analysis. In *Proceedings of the 25th Symposium on Operating Systems Principles*, SOSP '15, Monterey, California, 2015.

[68] Kenneth van Rijsbergen. The effectiveness of a homemade IMSI catcher build with YateBTS and a BladeRF. *University of Amsterdam*, 2016.

[69] Zack Whittaker. US cell carriers are selling access to your real-time phone location data. https://www.zdnet.com/article/us-cell-carriers-selling-access-to-real-time-location-data/, May 2018.

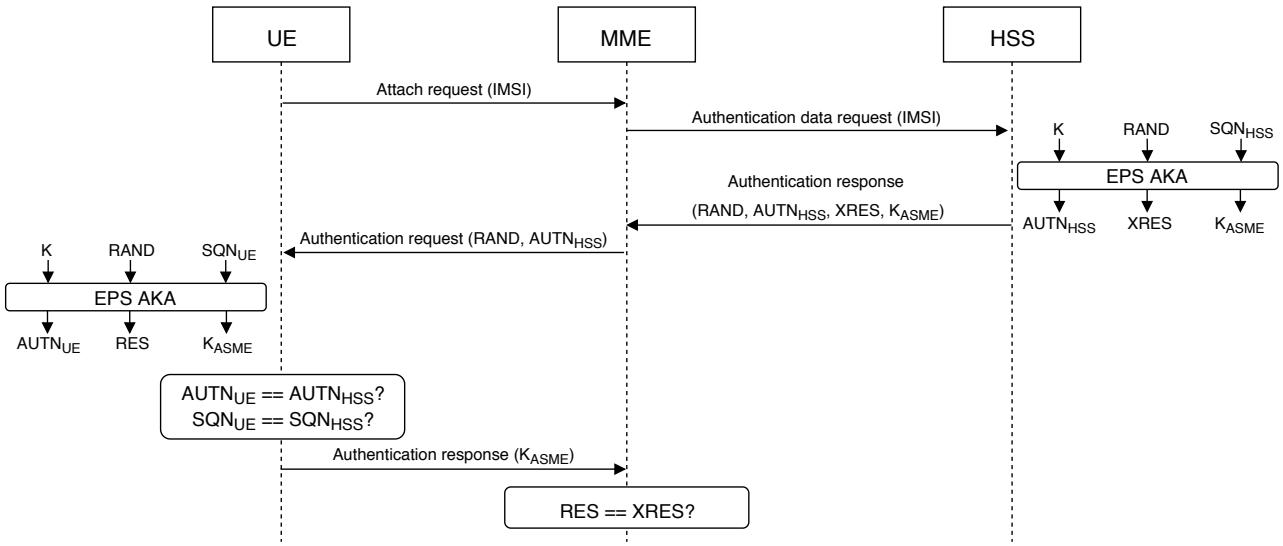


Figure 11: LTE authentication procedure. Given multiple users with identical IMSIs, the sequence number check will fail, causing multiple attach sequences.

A Sequence number check

When multiple UEs attach to the network using identical IMSIs, the sequence number check will fail in the network, triggering synch_failure messages and introducing delay before successful attachment. Figure 11 shows the message sequence of the LTE authentication procedure, used today on most celluar networks, and notes what occurs if duplicate IMSIs are present.

When a UE sends an attach request, the HSS feeds the following to its AKA function: a new nonce value RAND, the shared key K that is stored for the IMSI, and the locally-held sequence number SQN_{HSS} for that IMSI. The AKA function generates an authentication token ($AUTN_{HSS}$) that includes the SQN_{HSS}, an expected response (XRES), and a key (K_{ASME}) that is ultimately used for UE and MME authentication. The MME forwards the RAND and the $AUTN_{HSS}$ to the UE.

The UE then uses its own SQN_{UE}, its shared key K, and the RAND to generate its own $AUTN_{UE}$, a response (RES), and the key (K_{ASME}). The UE checks if the $AUTN_{UE}$ and $AUTN_{HSS}$ match and whether the its SQN_{UE} matches the SQN_{HSS} in the $AUTN_{HSS}$. In PGPP, this will fail for most attach procedures, as individual SIMs will not have the same SQN value as the HSS for the shared IMSI. When the SQN check fails the UE will send an authentication failure message with the cause: `sync_failure`, along with the UE's current SQN which allows the HSS to re-synchronize the SQN value. The UE then begins the attach sequence again, which will then succeed as the HSS and UE should begin the attach procedure with the same SQN values.

B Glossary

AKA

Authentication and Key Agreement. The process by which the UE and the HSS exchange information by which they can each verify a secret key held by the other, and calculate keys to be used for ciphering and integrity protection of data transmitted between the UE and the network. 11, 16

Diameter

The authentication, authorization, and accounting protocol used by 4G/5G cellular networks. Diameter is used to enable roaming between modern cellular networks. 5

E-UTRAN

Evolved UMTS Terrestrial Radio Access Network. Network that serves to connect UEs and eNodeBs. 2

ECM

EPS Mobility Management. The set of routines used by the core network to manage mobility procedures such as attach and tracking area updates. 11

EIR

Equipment Identity Register. A database that stores IMEIs of devices in the LTE systems. IMEIs can be white-listed, grey-listed or black-listed. The EIR allows a device's identity to be checked for blacklisting, (e.g., whether is has been reported stolen). 6

EMM

EPS Connection Management. The set of routines used by the core network to manage connectivity between the UE and the EPC. 11

eNodeB

Evolved NodeB. The base station in LTE. 1–3, 8–11

EPC

Evolved Packet Core. The core network in LTE. Main logical nodes of the EPC are the Packet Data Network Gateway (P-GW), Serving Gateway (S-GW), the Home Subscriber Server (HSS), and Mobility Management Entity (MME). 1, 2, 6, 11

GUTI

Globally Unique Temporary Identity. The GUTI is a temporary identifier that can be used in lieu of an IMSI to identify a subscriber to the core network. 3, 6, 12

HSS

Home Subscriber Server. The entity that holds subscription information to allow or deny access to the network. 2, 3, 6, 11, 12, 16

IMEI

International Mobile Equipment Identity. A globally unique, permanent device identifier which is allocated to each individual mobile device. It is set by the manufacturer. 6

IMS

IP Multimedia Subsystem. The entity that provides voice and messaging services for the network. 5

IMSI

International Mobile Subscriber Identity. A globally unique identifier associated with each mobile phone subscriber. It is stored in the SIM inside the phone and is sent by the phone to the network. 3–6, 9–12, 16

MME

Mobility Management Entity. The control entity that manages signaling between the UE and the core network. MME supports functions related to bearer and connection management and manages mobility between eNodeBs. 2, 3, 8, 10, 11, 16

MNO

Mobile Network Operator. A cellular service provider. 1, 7

MVNO

Mobile Virtual Network Operator. A cellular operator that does not necessarily own its own spectrum or all of the network equipment it operates upon. MVNOs run on top of MNO networks. 1, 2, 6, 7

P-GW

Packet Data Network Gateway. The gateway that provides global IP connectivity from the EPC. The P-GW typically offers NATed IP addresses. 2, 5, 6

PGPP-GW

PGPP Gateway. A proposed gateway for PGPP that sits between the P-GW and the global Internet. The PGPP-GW allows for billing without requiring the user's identity.. 6–8

RAND

A random nonce used during the authentication procedure. 16

S-GW

Serving Gateway. The serving gateway terminates the interface towards the cellular core and manages EPC signaling in response to data. Associated UEs are connected to a single S-GW at any point in time. 2, 5

S1AP

S1 Application Protocol. The signaling protocol used between the E-UTRAN and the EPC. 11

SIM

Subscriber Identity Module. An entity that holds the IMSI, which uniquely identifies a subscriber. SIMs are used to authenticate a user to the network. 2, 3, 5, 6, 11, 12, 16

SQN

Sequence Number. A value stored at the HSS and the SIM to maintain synchrony between the entities. 11, 16

SS7

Signaling System 7. The protocol standard used by entities on public switched telephone networks communicate with one another. SS7 is used to setup and tear down voice calls, deliver SMS, etc. SS7 has been largely replaced by Diameter in modern cellular standards. 5

TA

Tracking Area. A tracking includes one or many eNodeBs. Typically, the UE can move freely within eNodeBs in a tracking area without notifying the MME with a tracking area update. 3, 9–11

TAL

Tracking Area List. A list of tracking areas stored on the device that the device can enter without triggering a tracking area update. 4–6, 8–11

TAP

Transferred Account Procedure. A file detailing usage and wholesale charges due to roaming. 5

UE

User Equipment. The mobile device which allows a user to access network services, connecting to the UTRAN or E-UTRAN via the radio interface. Commonly understood to be a mobile phone. 2, 3, 6, 8–12, 16

XRES

Expected Response. A value generated by the core network used during the authentication procedure. 16