LOCA: A Location-Oblivious Cellular Architecture

Zhihong Luo *UC Berkeley*
Silvery Fu *UC Berkeley*
Natacha Crooks *UC Berkeley*
Shaddi Hasan *Virginia Tech*
Christian Maciocco *Intel*
Sylvia Ratnasamy *UC Berkeley*
Scott Shenker *UC Berkeley & ICSI*

Abstract

Cellular operators today know both the identity and location of their mobile subscribers and hence can easily profile users based on this information. Given this status quo, we aim to design a cellular architecture that protects the location privacy of users from their cellular providers. The fundamental challenge in this is reconciling privacy with an operator's need to provide services based on a user's identity (e.g., post-pay, QoS and service classes, lawful intercept, emergency services, forensics).

We present LOCA, a novel cellular design that, for the first time, provides location privacy to users without compromising on identity-based services. LOCA is applicable to emerging MVNO-based cellular architectures in which a virtual operator acts as a broker between users and infrastructure operators. Using a combination of formal analysis, simulation, prototype implementation, and wide-area experiments, we show that LOCA provides provable privacy guarantees and scales to realistic deployment figures.

1 Introduction

Providing users with *location* privacy is an important part of the larger challenge of online privacy. Unfortunately, today's cellular architecture offers little location privacy: network operators know the identity of a user and the geographic location of the access point to which that user connects and hence can trivially track a user's location in time. There is mounting concern over this situation as cellular providers are reported to routinely share their users' location profiles [28, 29, 62, 66, 105]. Moreover, 5G is likely to require smaller cell sizes [19] thus exposing much finer-grained location information and exacerbating the privacy problem.

Hiding a user's location from their network operator is challenging because connecting to an access point fundamentally reveals the user's location. One approach to improving privacy is to hide the user's *identity* from the network operator using so-called "blindly signed tokens" [23, 78, 86]. However, as discussed in §3, this approach comes at the cost of preventing network operators from providing *identity-based services*. These are services whose correct execution depends on the user's identity, such as post-pay [22], QoS prioritization [1] and lawful interception [3]. Such services are an essential part of today's networks and hence it is unlikely that operators can/will abandon them in exchange for improved user privacy. Thus, our question is whether we can enable location privacy

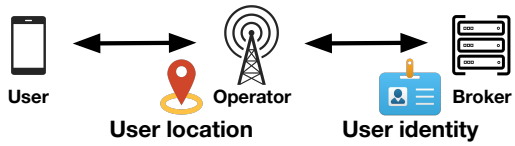

Figure 1: LOCA's overall architecture.

(*i.e.,* ensuring that network operators cannot easily track or infer a user's location) *without* compromising on identity-based services.

Privacy and identity-based services might seem to be fundamentally at odds. However, we see a way forward via mobile *virtual* network operators (MVNOs) such as Google Fi and Cricket [30, 49]. MVNOs are service providers that do not own radio infrastructure but instead provide user-facing services (sales, billing, *etc.*) while relying on business agreements with some number of traditional mobile network operators (MNOs) to provide the radio infrastructure. In this scenario, users pay MVNOs for service and MVNOs settle with MNOs on behalf of users. In other words, with MVNOs in the picture we can decouple infrastructure operation from user management and the MVNO acts as a broker between the user and the infrastructure operator.[1]

As shown in Fig. 1, our insight is that the existence of a broker between the user and operator enables us to reconcile privacy with identity-based services by strategically hiding different pieces of information from each party: the broker (*i.e.,* MVNO) knows the user's identity but not her location, while the operator (*i.e.,* MNO) knows the user's location but not her identity. With this arrangement, the broker can still tell the operator what identity-based services are to be applied to the user without revealing the user's identity, and the operator can implement the required services without knowing the identity of the user on whose behalf they are implemented.

However, hiding information in this manner is challenging for four reasons. First, in order to hide the user's identity from the operator, we must hide not just her identity but also her *trajectory* across multiple cell towers. This is because the operator could still infer the user's identity based on the sequence of towers she has visited, a form of privacy loss we refer to as *trajectory leakage* (§3.3).

Second, in order to hide the user's location from the broker, we must also hide the *identity of its operator* from the broker. This is because the locations of an operator's cell

[1] In this paper, we use the terms MVNO and broker interchangeably; we do the same with the terms MNO and operator.

tower deployments are public knowledge and hence can reveal a user's location [81]. The emergence of operators with small footprints, such as private and enterprise 5G networks, underscores the importance of this [13, 38, 52, 97].

The last two challenges arise because of this need to hide the identity of the operator from the broker. Brokers will always want to ensure that only authorized operators service their users. Since our approach hides the operator's identity from the broker, we now need a solution that allows the broker to verify the legitimacy of an operator *without revealing the operator's identity*. Lastly, when it comes time to settle payments, the operator should be able to claim payment from the broker *without* revealing what users it has served (since doing so would otherwise reveal user locations).

We design a privacy-preserving protocol that addresses the challenges above. Our contribution lies in developing new techniques (*e.g.,* aggregate claims) and synthesizing them with existing ones (*e.g.,* blind signatures, zero-knowledge proofs) into an end-to-end **Location-Oblivious Cellular Architecture (LOCA)**. To our knowledge, LOCA is the first system to enable location privacy for users while also supporting a provider's operational goals such as usage-based billing, QoS and service levels, lawful intercept, and so forth.

We evaluate the privacy and scalability of our protocol through formal analysis, simulation, prototype implementation, and wide-area experiments. We recognize that LOCA does introduce certain complexity and system overheads. However, our evaluation shows that these overheads are modest and within reach of what can be practically supported today. An important part of our contribution is thus in exposing the architectural complexity and performance tradeoffs that might be necessary to achieve our privacy goals.

Our work is based on certain assumptions about user and operator incentives. We assume that privacy concerns will influence some users in their selection of providers which will incentivize some operators to adopt the proposed techniques.[2] In addition, a growing number of jurisdictions have enacted policies that require providers to protect user privacy and, as discussed in §3, our architecture makes it easier for a provider to ensure compliance with these legal requirements. We do not assume that this motivation will apply to all users or operators: since our architecture can co-exist with the existing cellular infrastructure, we envision it will be applied to (by) the subset of users (providers) that are motivated by location privacy.

Finally, we recognize that there are many ways in which a user's location may be revealed through their online activities (*e.g.,* posting timestamped photos). We do not claim to prevent all forms of location leakage. Our focus is only on preventing the leakage of location information that today occurs every time a user connects to the cellular network.

In summary, the contributions of this paper are: (1) a new approach to preserve user location privacy while providing identity-based services; (2) the detailed design and implementation of a protocol (LOCA) based on this approach, and an evaluation of its performance and scalability; and (3) a formal analysis of the privacy provided by LOCA. Looking forward, we view LOCA as a first step towards privacy-preserving cellular infrastructure with room for improvement along multiple dimensions. We discuss these limitations extensively in the paper to motivate efforts on addressing these issues.

[2] Such market dynamics are already emerging in other contexts such as the smartphone market [10, 57, 85].

2 Background

The cellular ecosystem: MNOs and MVNOs Traditionally, the two main participants in a cellular network are the user with her device (called User Equipment, or UE) and the Mobile Network Operator (MNO). The MNO owns and operates cellular infrastructure and also provides user support services such as sales, billing and customer care. The user typically enters into a contractual agreement with one MNO which serves as her "home" provider. The user then consumes cellular services from her home provider or visited MNOs that her home provider has roaming agreements with.

In recent years, we've seen the rise of Mobile *Virtual* Network Operators (MVNOs). MVNOs are service providers that do not own radio infrastructure, but instead provide user-facing services (sales, billing, *etc.*), often focusing on serving specific underserved market segments [72, 91], while relying on business agreements with some number of MNOs to provide use of their radio infrastructure. In other words, the MVNO acts as a *broker* between the user and the infrastructure operator. In this scenario, the user contracts with an MVNO, and the MVNO in turn contracts with MNOs. Two well-known MVNOs in the US are Google Fi [49] and Cricket [30]. MVNOs can be involved in cellular operations to varying degrees, ranging from fully offloading to MNOs to operating their own core networks.

Identity-based services: These are services whose correct execution depends on the user's identity. An example of such services is lawful interception, a function that allows law enforcement agencies to selectively wiretap individual users [3, 4, 39]. In most countries, operators are legally required to support lawful interception. Additional examples of identity-based services include: (i) post-pay, which relies on identity-based accounting to charge a user based on her service consumption; (ii) QoS prioritization, where the network's treatment of a user's traffic depends on details of the user's subscription plan and past usage; (iii) deep packet inspection (DPI), where traffic is filtered based on the user's identity for purposes such as parental controls.

Location privacy in cellular networks: Location privacy, as defined in [17], is "the ability to prevent other parties from learning one's current or past locations". In the cellular context, this means that neither MVNOs nor MNOs should be able to learn a particular user's current or past locations. The exception is when location information must be revealed for legal purposes like emergency services and forensics.

3 Approach and Design Rationale

In this section, we briefly discuss the goals and assumptions that motivate LOCA's approach.

3.1 System Assumptions and Threat Model

System model: LOCA assumes a broker-centric architecture like today's MVNOs. This architecture involves three entities: (i) users, (ii) brokers, and (iii) operators. Operators own and operate cellular infrastructure. Brokers act as intermediaries between users and operators: a user subscribes to services from her broker, and the broker represents the user to operators, including handling settlements with each. LOCA requires brokers to authenticate their users.[3] The user need not be aware of the specific operator her device is attached to.

Threat model: We adopt a common threat model among privacy preserving systems that seek to prevent inadvertent information leakage between participants [25, 34, 54, 61, 79]. We assume brokers and operators are *semi-honest* (*i.e.,* honest-but-curious) and *non-colluding*: they follow the protocol but will attempt to extract user location information from the protocol execution, and that brokers and operators do not collude. We also assume that operators may attempt to *overbill* brokers by lying about session usages or what users they serve[4]. Attacks based on out-of-protocol information or collusion are out of scope but discussed in §5.

Incentives: One might ask why brokers and operators would implement the changes we propose. We believe that adopting our system is beneficial to them for both financial and legal reasons: as users are becoming more privacy-conscious [50, 74, 104], brokers that offer an opt-in location-oblivious service will be more attractive to customers. Second, doing so may soon become mandatory: regulations like GDPR recommend the privacy-by-design approach, which continues to place increasingly strong requirements on manipulating PII [16, 98, 107, 108]. By implementing a design such as ours, brokers and operators reduce their risk of inadvertently infringing privacy regulations. We explicitly assume that these benefits will outweigh the benefits of selling location data or implementing ad-hoc approaches to enforcing regulations, and thus we focus on the technical feasibility of a location-oblivious cellular architecture that also supports operational goals like usage-based billing and customized service levels.

3.2 Goals

Consider a user U, operator O, and broker B. We say that U's location privacy is violated when O and/or B know both U's identity and location. Today's cellular protocol trivially reveals both U's identity and her location. By protocol we mean the messages – their syntax and semantics – exchanged between U, B, and O as defined by the standard. Today, protocol messages carry U's identity, and the identity of the tower that U attaches to reveals U's location. Hence simply implementing the protocol allows an operator to track U's location with no special effort. In contrast, we are interested in modifying the existing cellular protocol standard to protect user privacy.

Arch	Operator (O)	Broker (B)	ID-based SVC
Today	UID, Location, Trajectory	UID, OID	Full
PGPP	Location, Trajectory	OID	Partial
LOCA	Location	UID	Full

Table 1: Comparison of today's MVNO architecture, PGPP and LOCA in terms of information revealed to participants and support for identity-based services (ID-based SVC); U/OID: U/O's identity.

3.3 Approach

In research, the state of the art is the recently proposed PGPP protocol [86] which tries to provide location privacy by *hiding* U's identity from O and B. In PGPP, users are identified by a "blindly signed token" [23, 78] which they obtain during a registration phase prior to consuming service.[5] I.e., a user prepays for a certain quota of service (*e.g.,* some number of minutes of connectivity at a specified data rate) and in return obtains a blindly-signed token. When connecting to the network, the user presents this token via which the broker can authenticate the user without learning her identity.

To our knowledge, PGPP is the first system that tries to provide location privacy for cellular users. However, as we detail in §8, PGPP faces two drawbacks. First, PGPP does not easily allow operators to support identity-based services, which are widely deployed in today's networks. Second, a user's *trajectory* across towers is still visible to operators and hence the protocol is vulnerable to "trajectory-based location leakage" in which the operator can learn the user's identity by correlating her trajectory with other out-of-band information.[6] In designing LOCA, we wished to avoid these limitations which, as we will see, leads to an altogether different approach.

In summary, our goal in LOCA is to design a cellular protocol that protects the location privacy of users by achieving the following properties: no party in the protocol (broker, or operator) should simultaneously know both the identity and the location of a user; the protocol should also not reveal the user's trajectory to either broker or operator. Finally, the protocol should support identity-based services including post-pay and lawful intercept. In this work, we propose LOCA, a new cellular protocol that achieves these stronger privacy guarantees while supporting identity-based services.

We briefly comment on the scope and limitations of LOCA

[3] For MVNOs who by default offload all cellular operations, they can still support LOCA users by deploying their own authentication servers.

[4] One might ask whether we need to protect against over-billing if the operator is semi-honest. The reason we do so is because, as we'll see, *once we have privacy*, it becomes much easier for an operator to overbill since the broker cannot tell which users were serviced by the operator and hence cannot check the operator's billing claims. Hence, an operator can follow the protocol and yet overbill with impunity. To avoid this, we assume operators may overbill and design our protocol to prevent this.

[5] Such a token is *blindly* signed by the broker who can later verify the signature without being able to link it back to the original signing request.

[6] For example, consider a user that regularly travels between their home and office location: the operator could narrow down the identity of the user by correlating this trajectory with residential information in billing records.

as presented in this paper. Our goal is to safeguard users' location privacy at the *protocol* layer. This raises the bar relative to today's protocols but isn't sufficient to safeguard against violations that might occur outside the protocol, at other layers. For example: at the application layer, a user's identity might be revealed by inspecting their packets [14, 88], or physical-layer characteristics (*e.g.,* signal patterns) might be exploited to track a specific device [33, 47]. Such attacks are possible but (to our knowledge) not exploited today. However, if cellular protocols evolve to protect privacy, such app/physical layer leakages could become a more important issue. Fortunately, the research literature provides solutions to such attacks [43, 56, 60, 103, 110, 114] that we believe can coexist with protocol-layer solutions like LOCA. We elaborate on this in §5.2 but leave an in-depth exploration to future work.

There is an obvious tension between guaranteeing location privacy and offering identity-based services: connecting to a cellular tower fundamentally reveals a user's location, while customizing service to a user requires knowing the user's identity. Our insight is that we can extend broker-centric architectures to create a situation in which the broker knows the user's identity but not their location, while the operator knows the user's location but not their identity; neither broker nor operator knows the user's trajectory.

How do we achieve this? First, to hide U's location from B, we hide the identity and location of the *operator* O from B. Recall that U attaches to the network (and hence to B) via O's infrastructure and hence, if B cannot tell where O is located, then it cannot tell where U is located either. Hiding O's location is not sufficient: we must also hide O's identity from B, as knowing O's identity might be sufficient to narrow down O's location (and hence U's location). An operator's tower locations are public knowledge and, moreover, we're seeing an increasing deployment of small-scale cellular networks due to the emergence of private and enterprise 5G networks, as well as various forms of community networks [13, 38, 52, 97].

As we will describe in §4, we hide O's identity from B by having O obtain an unlinkable token from B during an offline registration process.[7] O later uses this token (denoted $\hat{O}$) as its identifier when interacting with B. By the properties of blind signatures, B can verify that $\hat{O}$ is a pre-authorized operator but cannot link $\hat{O}$ to O. In addition, O hides its IP address from B by using anonymous communication solutions.

The above suffices to hide U's location from B. The other half of our arrangement is to hide U's identity from O. This is easily achieved since O does not need to know U's identity to service U; since B knows U's identity, B can tell O what services are required (rate limits, filtering rules, *etc.*) thus enabling identity-based services without revealing U's identity. Thus, U simply uses a temporary pseudonym (denoted $\hat{U}$) in her interactions with O. Finally, by periodically changing U's temporary pseudonym and randomizing attachment timing, we limit O's ability to track a particular user's trajectory.

[7] The use of such a token is similar to PGPP but used by O instead of U which we will see leads to a very different set of considerations.

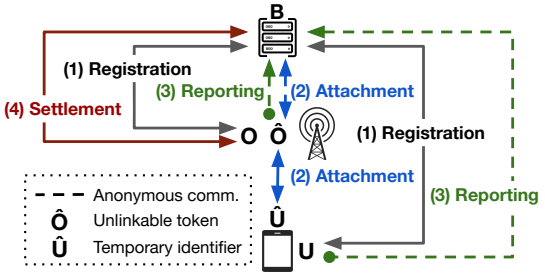

Figure 2: An overview of LOCA's protocols.

As summarized in Table 1, the above approach offers U location privacy while still supporting identity-based services. However it gives rise to a new challenge: how does O receive payment for its services to U? In today's architecture, B directly settles with O based on the service that U received. We wish to preserve this direct billing system between O and B. Yet, our protocol intentionally hides O's identity from B. To address this issue, we devise a solution that allows O to reveal its true identity *only* when claiming payment from B. Our solution leverages zk-proof techniques to design a novel *aggregate claiming* procedure via which (i) O claims payment for an aggregate of the user sessions it has serviced, and (ii) B can verify the correctness of O's claim without revealing the identity of the users that O serviced.

4 Design

At a high level, the process of obtaining cellular services can be broken down into four phases or steps: (i) *registration*, during which the various parties (U, B, O) enter into pairwise contractual relationships: U signs up with B for service, and B with O as an operator for its users; (ii) *attachment* involves the protocol by which U discovers and connects to a tower in O's infrastructure, (iii) *mobility* involves the handover protocols via which U is migrated from one tower to another as needed, and (iv) *settlement* refers to the norms and processes via which B pays O for the service that O has provided B's users.

Of the above, *attachment* and *mobility* are defined by today's 3GPP standard while *registration* and *settlement* are out-of-band processes. Our goal is to implement LOCA with minimal disruptions to today's protocols, and without involving any new entities in the registration or settlements process.

Next, we describe LOCA's operation in these phases, an overview of which is given in Fig. 2.We briefly summarize how each phase is typically implemented in today's networks and then present the changes that LOCA introduces. Finally, we elaborate on how identity-based services work in LOCA.

4.1 Registration

Today: In today's networks, when U signs up with a broker B, they exchange shared secret keys (*SSK*s) that will be used for mutual authentication during the attachment process. In 5G, B also shares its public key (PK_B) with U so that U can encrypt her identity in later attachment requests.

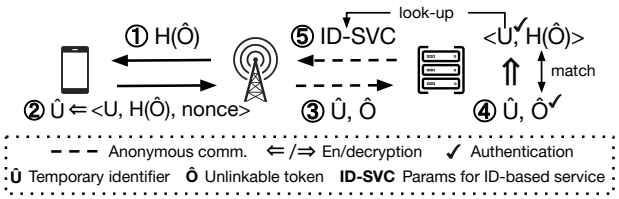

Figure 3: LOCA's attachment procedure.

LOCA: With LOCA, B and U continue to exchange PK_B and *SSK*. Like today, these keys will be used for mutual authentication between U and B (§4.2) and to hide U's identity from O. The main change LOCA introduces is in the registration process between B and O. When B and O sign up with each other, LOCA requires that they participate in a blind signature protocol [23, 78] as a result of which O obtains unlinkable tokens (denoted as $\hat{O}$) that are blind-signed by B. When $\hat{O}$ is later presented to B, the blinding process ensures that B can verify the signature but cannot link $\hat{O}$ to O. Thus blind tokens allow B to authenticate O without learning O's identity. LOCA uses a standard blind-signing protocol [23] (summarized in Appendix A). In addition to blind tokens, B and O also exchange a shared hash function H that will be used in our attachment and settlement processes as described later.

4.2 Attachment

Today: Attachment today involves three main steps. First, O broadcasts its identity on the radio control channel that U listens on to discover O. Next, after discovering O, U sends an *attachment request* to O who forwards the request to B for authentication. In 5G, U uses an encrypted identifier (termed SUCI [2]) in this attachment request. Finally, once U has been authenticated, B responds to O authorizing service. B's response includes U's permanent identifier (termed SUPI [5]).

Thus today's attachment process reveals O's identity to U in the first step. In the second step, B learns O's identity (and hence U's location) from both the contents of the attachment request and the act of receiving it from O (which reveals O's IP address). Finally, O learns U's identity via the authorization response it receives from B.[8] Thus today's attachment reveals U's identity and location to both B and O.

LOCA: We describe LOCA's attachment process with an emphasis on how we prevent (i) B from learning O's identity and (ii) O from learning U's identity. As mentioned in §3.3, we achieve the former by having O interact with B as $\hat{O}$ (O's unlinkable tokens) via anonymous communication channels. LOCA achieves (ii) by encrypting U's identity (with PK_B) and never exposing it outside of B. As shown in Fig. 3, LOCA's attachment process consists of the following five steps.

(i) Operator discovery. Instead of its actual identity, O broadcasts the hash of its token (*i.e.,* H($\hat{O}$)) on the control channel.

(ii) User preparation. U sends an attachment request to O (formatted as a NAS message [7]). This request includes B's identity, U's identity (IMSI) plus a nonce, and H($\hat{O}$). The last two – (IMSI+nonce) and H($\hat{O}$) – are encrypted by B's public key and serve as a temporary identifier for U which we denote as $\hat{U}$. We assume that B has a large user group so that its identity leaks little information on U's identity. The nonce ensures that $\hat{U}$ is different every time U attaches to the same O which helps prevent O from tracking U's trajectory (§4.3).

(iii) Operator preparation. On receiving U's attachment request, O forwards the request to B over an anonymous communication channel and uses its unlinkable token $\hat{O}$ to identify itself. Typical solutions for anonymous communication are Tor [99] and VPN [80] with different performance/security trade-offs, which we will discuss in §6.3. This anonymous channel can be set up offline, prior to attachment, whenever O changes token $\hat{O}$. Thus B does not see O's true identity nor the IP address from which $\hat{O}$ sends the request. The latter is necessary as several studies have shown that IP addresses can often be geo-located with high accuracy [26].

(iv) Broker authorization. On receiving the attachment request, B first verifies the $\hat{O}$ token thus ensuring that the request comes from an operator that B has previously authorized during the registration phase. Next B decrypts the request, and authenticates U via today's challenge-response protocol based on the shared secret key *SSK* [6]. In addition, B verifies that $\hat{O}$ is indeed the operator to which U wants to attach; B can verify this by validating H($\hat{O}$) (using the shared hash function established when O registered with B) and thus prevents replay or hijacking attacks. Once B has authenticated and verified the request, it looks up the parameters associated with U's service plan (as today): *e.g.,* rate limits, QoS parameters, whether to intercept U's traffic, and so forth. B then crafts a response authorizing the attachment (including the proper service parameters, security parameters that allow U to authenticate the network, etc), signs it, and returns it to $\hat{O}$.

(v) Access attachment. B's response authorizes $\hat{O}$ to service U as per the parameters from B. Beyond this point, $\hat{O}$ (*i.e.,* O) serves U as in today's networks. We elaborate on how O provides identity-based services to U in §4.4. Note that O can still perform functions like establishing radio bearers that require binding U's identifier to temporary identifiers like GUTI and RNTI; O simply uses $\hat{U}$ instead of U.

4.3 Mobility

Today: In current networks, mobility is implemented via a handover process, where O initiates U's migrations by directing U to switch from a tower T1 to another T2. This approach ensures a seamless mobility experience for U because U's IP address remains unchanged after the migration. However, as O initiates U's migrations, O trivially observes U's trajectory across handovers, jeopardizing U's location privacy.

LOCA: Trajectory leakages are inevitable if O fully controls U's mobility like today: although LOCA already hides U's identity from O during attachments, O can still track $\hat{U}$'s trajectory and use that to infer U's identity, making LOCA vulnerable to trajectory analysis. To mitigate this fundamental

[8] Prior to 5G, U's permanent identifier (IMSI) was included in the initial attachment request, allowing O to directly discover U's identity. Since 5G, U's attachment request uses an encrypted temporary identifier over the air to defend against IMSI catchers [93]. Nonetheless, O still learns U's permanent (SUPI) identifier from B's authorization response (step 3).

issue, we leverage a user-driven mobility approach proposed in [77]. In this approach, U initiates migrations across towers by simply detaching from T1 and then attaching to T2. U then relies on modern transport protocols like MPTCP [82] and QUIC [71] to maintain connections despite changing IP addresses. Prior work has shown that this user-driven approach does not degrade service even when reattaching on a *per-tower* basis [77]. LOCA adopts and extends this approach to minimize trajectory leakages with two techniques: (i) periodic reattachment and (ii) randomized attachment timings.

First, U will detach and reattach periodically (not at every tower) with a new temporary identifier. Thus, O cannot trivially track U across new sessions based on U's identifiers. The reattachment frequency is a configurable parameter that bounds the length of U's trajectory that is visible to O where length might be measured in time (*e.g.,* valuable for a mostly stationary user), in towers, or some combination thereof.

Even with periodic reattachment, O may still attempt to infer U's trajectory by doing a timing analysis over her detach and attach events. In particular, such analysis would be effective in a naive implementation that uses a fixed interval between when U detaches from T1 and subsequently attaches to T2. To address this issue, we have U wait for a randomized but bounded duration of time before issuing her attachment. When possible, we can also leverage make-before-break attachments[9] in which U may attach to T2 *before* detaching from T1 thus increasing the time window over which U can randomize their attach/detach events which makes inference harder. Together with periodic reattachment, this randomization of U's attachment times limits O's ability to correctly infer U's trajectory, because U's (re)attachments are obfuscated by the periodic (re)attachments from other nearby users.

We recognize that user-driven mobility introduces some complexity as well as dependencies on newer transport stacks, however this tradeoff is fundamentally necessary if we are to prevent trajectory leakages, and supporting these techniques incurs a minimal impact on the user's performance (§6.3). As we will detail in §5.1.3, the obfuscation effect of our approach depends on the specific configurations, *i.e.,* reattachment frequency and attachment time window; as well as the deployment scenarios, *i.e.,* the number of nearby users and the length of U's trajectory. Overall, under realistic deployment scenarios and configurations, the probability that O can correctly infer U's trajectory is negligible.

4.4 Identity-based Services

LOCA ensures that operators and brokers can continue to provide critical identity-based services, including allowing law enforcement agencies to locate specific users when required.

The key reason LOCA can support identity-based services is that brokers continue to know the identity of their users. This enables B and O to collaborate on identity-based services. For instance, during attachment, B can select the service level associated with U's plan and indicate that to O in its authorization response – *e.g.,* via the QoS Class Identifier (QCI) parameter [109]. O then simply enforces the QCI for the duration of its session with $\hat{U}$ without knowing U's true identity.

To realize services such as lawful interception, law enforcement agencies work with B and O. As today, O runs a lawful interception (LI) system — *e.g.,* installing an interception gateway [96]. A law enforcement agency notifies B of the user whose communication it wants to intercept. B passes on this notification to O during the attachment process, and then O's LI systems report the required information to the agency.

Emergency services (*e.g.,* 911 calls) work in a similar manner. A law enforcement agency knows U's identity and needs to learn U's location. The agency reaches out to B; B looks up U's current temporary identifier $\hat{U}$, and asks $\hat{O}$ (via their anonymous communication channel) to reveal $\hat{U}$'s location to law enforcement. Thus, the agency can collect U's current location without violating LOCA's privacy guarantees (*i.e.,* O does not know U's identity while B does not know O's identity or location). The same approach can be used to recover U's past locations based on the records logged at B and O.

4.5 Settlements

Today: In today's MVNO networks, B pays O based on U's service parameters and the resources consumed, as reported by O to B.While differing in the details, existing settlement processes all require that B knows which users/sessions were serviced by O, thus potentially violating user location privacy.

LOCA: To settle O's payments while preserving U's location privacy, LOCA's settlement process contains two phases: a *reporting* phase, where U and O report session usage to B; and a *claiming* phase, where O claims settlement from B.

Reporting phase: In LOCA, we define a *session* as the user-operator association that starts when U completes the attachment process with $\hat{O}$ and ends when U detaches from the same. At some point after a session ends, U and $\hat{O}$ independently send traffic reports to B. Note that O continues to hide its identity and location when sending its report to B. U reveals its identity to B but also sends its reports over an anonymous channel because its IP address can reveal its whereabouts. The traffic report from U lists the sessions in which U participated; $\hat{O}$ does the same for its sessions. Each entry in the list contains a session identifier (SID), usage metrics (*e.g.,* bytes, duration), and QoS metrics (*e.g.,* packet loss rate). In addition, O appends a nonce to each session in its report. These nonces are generated from the shared hash function H known to both O and B, and taking secret inputs that are only known to O. We call these inputs "embedded secrets", and as we will see, O later uses these secrets to claim its settlement from B.

B then compares the reports from U and $\hat{O}$, generates bills for U and publishes a *session table* to start the claiming phase. The table includes the usage calculated based on the reports from $\hat{O}$ and U, for all sessions during the last billing cycle.

[9] The support for make-before-break, so-called dual active protocol stack (DAPS) handovers has been introduced in 5G 3GPP specifications [8, 45, 95].

LOCA: Aggregate Claiming Protocol

O: operator, **B**: broker; F: claiming function

Procedures:

Setup:

1. **B** performs (PK, VK) = VerifierSetup(F)
2. **B** keeps verifier key VK and sends the prover key PK to **O**

Per billing cycle:

3. **B** publishes a session table ST
4. **O** generates (z, π) = VCProve(F, ST, Secrets, PK)
5. **B** verify the claim (z, π) with VCVerify(ST, z, π)

ZK-Proof-based Verifiable Computation:

VerifierSetup(F)
- **B** compiles F into an arithmetic circuit C
- **B** performs zk-SNARK preprocessing on C, generates PK and VK, and returns (PK, VK)

VCProve(F, x, w, PK)
- **O** computes z = F (x, w)
- **O** uses zk-SNARK to generate a ZK-proof π based on (x, z) as the primary input, w as the witness, using PK, and returns (z, π)

VCVerify(x, z, proof, VK)
- **B** uses zk-SNARK to verify the proof with (x, z) as the primary input, using VK

Figure 4: A summary of the aggregate claiming protocol.

When generating statistics in the session table, B can consider factors other than reported usages such as QoS metrics.

Claiming phase: Every billing cycle, O reveals its identity and claims settlement from B but does so without revealing which sessions O has serviced. To achieve this, we must solve three problems: (i) ***No over-claims.*** How does B verify that O is claiming only the sessions O actually serviced? (ii) ***No mis-claims.*** How do we ensure that O can claim the sessions but no one else? (iii) ***Session oblivious.*** How does O claim settlement without revealing to B which sessions it is claiming? We combine zero-knowledge proofs with the above mentioned "embedded secrets" to address (i) and (ii); and "aggregate" claims to address (iii).

Embedded secrets serve as the basis for O proving its session ownership to B. However, naively having O reveal its secrets fails the session oblivious requirement because B now knows what users O has serviced. This leads to our aggregate claiming protocol that fulfills all three requirements:

Aggregate claiming with ZK-proof: First, we observe that in order to generate O's payments, B does not need to know *individual* session ownership; instead, it only needs to know the session ownership in *aggregation*, *i.e.,* the aggregate usages for payments for a specific O. Based on this insight, our *aggregate claiming* mechanism works as follows: the claiming begins with B publishing a session table readable to all Os. O then reveals its identity and claims its payment from B:

Intuitively, O's claim takes the form: "I have sessions that add up to X bytes." Because the number of different sessions that could add up to X is large, it is difficult for B to infer whether an individual session is part of O's claim or not, thus obfuscating the session ownership. In §5.1.2, we show that the expected number of session combinations that add up to the same X grows *exponentially* w.r.t. the total number of sessions in the table via both theoretical and empirical analysis.

Note that this naive aggregate claiming suffices if we assume O will not overbill B. However, it is important to realize that without additional mechanisms (like the zk-proof that follows), O can more easily overbill B without being detected in LOCA than in today's (non-privacy preserving) architecture simply because B does not know what users O serves.

Hence, since naive aggregate claiming allows O to overbill, we extend our solution such that O can *prove* its claim by showing that O knows the embedded secrets corresponding to its claim. For this, we leverage *proof-based verifiable computation* [102], a cryptographic tool that uses zero-knowledge proof to enable one party to prove to another that it has run a computation $z = f(x, w)$, where f is the function, x is the public input, w is the prover's private input and z is the output, without revealing any information about w. Proof-based verifiable computation systems have two components: (i) a zk-SNARK backend [84] that proves and verifies satisfiability of *arithmetic circuits*, and (ii) a compiler frontend that translates program executions to arithmetic circuits. Such an arithmetic circuit is also referred to as "a set of *constraints*".

Fig. 4 describes LOCA's aggregate claiming protocol. First, B performs *VerifierSetup*, where B compiles a claiming function F into an arithmetic circuit, and uses zk-SNARK to preprocess the circuit and generate prover key PK and verifier key VK. This verifier setup step needs to be performed only once, after which B keeps VK and sends PK to each participating O. The claiming function F takes two inputs: a session table with at most N sessions as the public input, and a set of (at most K) secrets as the private inputs. F computes the hashes of the provided secrets, iterates all the sessions in the session table, adds a session's usage to the aggregate usage if one of the precomputed hashes matches the nonce of that session, and finally returns the aggregate usage.

Next, once per billing cycle, each O performs *VCProve*, which involves two steps: (i) O executes the claiming function F with the session table and its embedded secrets, which returns the aggregate usage z for O's sessions. (ii) O passes to zk-SNARK the session table, its secrets, the computed aggregate z and the prover key PK, to generate a zero-knowledge proof π, which allows O to prove to B that it has secrets for sessions that add up to z, without leaking any information about individual session ownership. O then sends a *claim* including the aggregate usage z and proof π to B.

For each O's claim, B performs *VCVerify*, where B uses zk-SNARK to verify the proof π with the session table, the claimed aggregate z and the verifier key VK. If the verification passes, given the soundness property of the zk-SNARK proof system [84], B can confidently approve O's claim and generate O's payment according to the claimed aggregate usage and other factors such as O's reputation. The duration of a billing cycle is configurable: longer cycles lead to larger session tables, which in turn indicates stronger privacy protections (§5.1.2) at the cost of more expensive operations (§6.2).

Session group: The design presented above assumes a single session per token, which may not scale to large deployments:

O generates a proof every billing cycle, and proving with zk-SNARK is expensive [111]. In our setup, as we will show in §6.2, the time complexity to prove a circuit for the claiming function F is $O(K*N)$, where K is the maximum number of sessions O can claim and N the total number of sessions in the session table. Such proving time would be prohibitively long when there are a large number of sessions to claim.

To address this scalability challenge, we introduce the notion of a *session group*, which includes all the sessions that are associated with the same unlinkable token. By grouping multiple sessions into a single session group, we can reduce the number of entries in the session table. To support session groups, we made the following extensions to our protocol:

- **Attachment:** We allow O to use a single token and the corresponding anonymous communication channel for multiple sessions as the same session group.
- **Reporting:** We allow O to send a traffic report containing all the sessions of the session group.
- **Claiming:** We allow B to publish a session group table with one session group for each row. O claims session groups the same way as it claims sessions before.

The size of the session group is tunable in LOCA and determines how many sessions each token is used for. Tuning the group size allows LOCA to explicitly trade off between privacy and scalability: (i) a smaller session group is better for privacy, because it minimizes indirect location leakages (detailed in §5.3), which occur when a user of a session within a session group has her locations leaked, in which case users of other sessions within the group also suffer a privacy loss; (ii) larger session groups are desirable in terms of scalability of zk-SNARK, as it takes longer to actually generate a session group (with users' attachment), while proving cost remains the same, as N is the same, so zk-SNARK proving becomes relatively faster. Fortunately, modern zk-SNARK is fast enough that a balance between privacy and scalability can be achieved: as we will show in our evaluation (§6.2), LOCA can scale to large deployments with sufficiently small session groups and thus introduces only minimal privacy loss.

5 Privacy Analysis

Safeguarding location privacy requires fulfilling three properties: (i) O does not knows U's identity, (ii) B does not know U's location, and (iii) neither B nor O knows U's trajectories. To our knowledge, LOCA is the first protocol to meet these requirements. In this section, we analyze the conditions and assumptions under which LOCA meets these requirements. We show that LOCA achieves all three properties under the assumptions of our threat model which are that participants are semi-honest and do not collude (§5.1). We then briefly consider attacks beyond our threat model and show that LOCA offers substantial protection even when participants use out-of-protocol information (§5.2) or collude (§5.3).

5.1 Semi-honest and Non-colluding

We first analyze LOCA's privacy properties under our threat model of semi-honest and non-colluding participants (§3.1).

5.1.1 Hiding U's identity from O

LOCA hides U's identity from O. Specifically, U's identity is encrypted using B's public key. B is thus the only party that can decrypt and observe U's identity in plaintext. B also never exposes U's identity to O, even after U successfully attaches.

5.1.2 Hiding U's location from B

LOCA hides U's location by (i) hiding O's identity and location when O interacts with B on behalf of U and (ii) hiding which users were serviced by O when O reveals its identity to claim its settlement. Next, we show how LOCA achieves (i) via the security properties of existing cryptographic constructs (*i.e.,* anonymous communication and blind signature) and achieves (ii) via aggregate claiming; we establish the latter property via formal analysis and empirical simulations.

For (i), LOCA leverages anonymous communication such that two parties can communicate without revealing their identities to one another. Similarly, LOCA builds on a blind signature scheme that allows a participant to authenticate another without learning its identity. Taken together, these existing cryptographic constructs allow operators to register and report sessions to brokers without revealing their identities.

Discussing the security of aggregate claiming requires more care. We break this process down into two halves (i) the security of the claiming mechanism itself, (ii) the information leaked by revealing the aggregate value to B. The former follows directly from the security of our zero-knowledge proof construction. We do not discuss this further. Instead, we focus on the impact of B learning the aggregate value of the claimed sessions. Specifically, we show that B has an exponentially small likelihood to correctly infer what sessions/users O has serviced based on the aggregate value.[10] Our core intuition is simple. Let us assume that for N session groups with a uniform distribution of session group usage from 1 to m, operators will claim the aggregate usage of K session groups, which sum to aggregate value S. The total number of possible session group combinations grows exponentially as a function of N. In contrast, the number of possible claimed values only grows linearly ($m*N$). In expectation, there will consequently be exponentially many possible session group combinations that could have summed to S. We formally prove that this result holds as long as the ratio between K (the number of session groups belonging to an operator) and the total number of session groups N falls within a specific range. We identify this range formally below, and show through simulation that these bounds can be further improved and are wide enough to support realistic deployment scenarios.

Theoretical proof: We formulate the aforementioned problem as follows. Consider arrays X and Y, one of size $N-K$,

[10] The general reasoning extends to when B analyzes multiple claims from different operators but we don't get into the details in this paper.

and one of size K, where each cell contains a value from 1 to m drawn from the discrete uniform distribution. Let S be the sum of all elements in Y. We derive a bound on the expected number of possible subsets of elements in X that sum to S.

Theorem 5.1. *Considering two independent arrays X and Y, consisting of $N-K$ and K iid random variables from $U\{1,m\}$, there exists $L(m),U(m)$ such that the expected number of subsets in X, whose sums are equal to the sum of Y, is exponential w.r.t N, if $L(m) \leq \frac{K}{N} \leq U(m)$. Note that $L(m),U(m)$ depend on m, and $0 < L(m) \leq U(m) < 1$, $\forall m \in Z_{>0}$*

The proof, at a high level, works by (i) deriving the closed-form distributions of the sum and the subset sum of an array of discrete uniform variables similar to prior theoretical work [20], (ii) expressing the expected number of matched subsets with these two closed-form distributions, and finally (iii) reducing to an exponential lower bound for the expression. More details of the proof can be found in the appendix.

The reductions in step (iii) are highly conservative. Hence the proven feasible range of ratio $[L(m),U(m)]$ is narrow, and the exponential bound is small. We confirm through simulation that this bound holds analytically for a significantly wider range and encompasses many real-world scenarios:

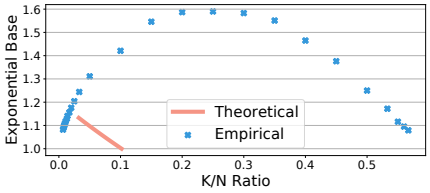

Figure 5: Exponential bounds for different K/N ratios with $m = 5$.

Empirical simulations: In these experiments, our goal is to understand within what range of ratio, the number of matched subsets grows exponentially w.r.t N. For each ratio K/N, we scale N while increasing K proportionally according to the ratio and estimate the expected number of matched subsets for the (K,N). More details about our simulation setup are in Appendix B. Now that we have estimates for multiple (K,N)'s of the ratio K/N, we fit the results with an exponential curve of N by performing linear fittings on the logs of the estimates:

$$R = a * b^N \rightarrow \log(R) = \log(b) * N + \log(a)$$

The slope of the fitted linear curve is thus the log of the exponential base. Our fitted linear curves closely match the logs of estimates with an adjusted R-squared value of over 0.99, which suggests a significant exponential relation between our estimates and N. Fig.5 shows the exponential bounds of different K/N ratios for uniform distribution with m=5. Compared with the theoretical results, the empirical results suggest much larger exponential bounds over a wide range of ratios: exponential base over 1.1 for ratios from $1/150$ to over $1/2$. We observe similar behavior with other values of m and with other non-uniform session group usage distributions.

5.1.3 Hiding U's trajectory

LOCA hides U's trajectory from O via (i) periodic reattachment and (ii) randomized attachment timing. The former prevents O from directly observing U's trajectory, and the later makes accurate timing-based trajectory inference infeasible:

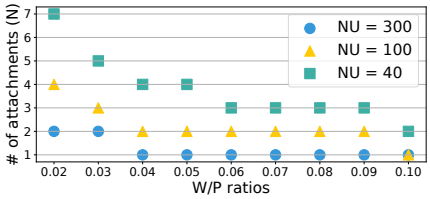

Figure 6: The longest trajectories beyond which the likelihood of correct inference is less than 1% for different NUs and W/Ps.

With periodic reattachments, O is unaware of which attachments belong to U and hence O can only *infer* U's trajectory by correlating between detachment and subsequent attachment. By randomizing attachment timing, any detachment that arises within a time window before and after (with make-before-break handovers) an attachment is equally likely to correlate with that attachment. We call this set of detachments "candidate detachments", and since all users periodically reattach, there is a *lower bound* on the number of candidate detachments. Lastly, to recover U's trajectory, O has to select the correct detachments for all of U's attachments along the trajectory, which becomes *exponentially* harder for longer trajectories. Modelling all these factors, we can analyze the difficulty of trajectory inference in LOCA: denoting time window as W, the reattachment period as P, the number of nearby users as NU, the number of candidate detachments as ND, the number of attachments in U's trajectory as N, we can derive the likelihood of O correctly inferring the trajectory $Prob$:

$$ND \geq 1 + NU * \frac{W}{P}, \quad Prob \approx (\frac{1}{ND})^N$$

This formulation tells us why accurate trajectory inference is infeasible: (i) since $Prob$ decays exponentially w.r.t N, even with a ND of 2 (only one alternative candidate detachment), O has a less than 1% likelihood of inferring a trajectory with more than 6 attachments; (ii) The ratio between the time window and re-attachment period (*i.e.,* $\frac{W}{P}$) is configurable, and a larger ratio increases ND and thus the inference difficulty. Fig.6 shows the longest trajectories that O can infer with a likelihood larger than 1% for different NUs and $\frac{W}{P}$s. For $\frac{W}{P}$ larger than 0.03, O is unable to infer long trajectories ($N > 4$), even if the number of nearby users is small ($NU = 40$).

We've shown that LOCA safeguards user location privacy at the protocol layer and under the assumptions of our threat model. We believe this raises the bar relative to the status quo however, as discussed earlier, LOCA would still be vulnerable to attacks that exploit either: (i) out-of-protocol information or (ii) information from other participants via collusion. We will next discuss such attacks, their impact, and potential mitigation strategies, but leave an in-depth study to future work.

5.2 With out-of-protocol information

Next, we show that (i) as a protocol-layer solution, LOCA does not prevent attacks based on out-of-protocol information, (ii) the impact of these attacks on LOCA is minimal and, (iii) mitigation strategies for these attacks can coexist with LOCA.

Attacks: B and O can compromise U's location privacy by exploiting out-of-protocol information. Here we enumerate some attacks that violate each of the three privacy properties: (i) If O has access to a resident directory near its cell towers, it can nail U down to a smaller user group. O might also learn U's identity by inspecting U's data traffic. (ii) If B is capable of network monitoring, it might learn O's identity by conducting a traffic analysis, where it observes traffic at each operator and correlates that with incoming traffic it receives. (iii) O might track U's trajectory by profiling U's physical-layer characteristics, such as its signal patterns and strengths.
Impact: LOCA's design limits the impacts of out-of-protocol attacks on user's location privacy: (i) Attacks that allow O to uncover $\hat{U}$'s identity only incur *per-hop* leakages: U's identity remains unknown to O when she reattaches with a different $\hat{U}$. (ii) Attacks that allow B to uncover $\hat{O}$'s identity only incur *per-token* leakages: O's identity remains unknown to B when O switches tokens. This means that locations of users who are served by O with a different token from the revealed one remain unknown to B. (iii) Lastly, inter-operator attachments can minimize impacts of attacks that allow O to track U's trajectory. Firstly, instead of having her entire trajectories leaked, U suffers only *per-operator* leakages. Secondly, as U moves in and out of O, it is challenging for O to link all of U's trajectories within its footprint, because O is unaware of U's locations when U connects to other operators.
Mitigation: LOCA can coexist with countermeasures designed for different out-of-protocol information. For instance, for attacks based on traffic characteristics, end-to-end encryptions of U's traffic can help counter packet inspections by O; and communication systems that are robust to traffic analysis like Vuvuzela [101] could be adopted for communications between O and B. For attacks based on physical-layer signals, one could use defense mechanisms such as randomizing transmission coefficients [89] and injecting artificial noises [56].

5.3 With collusion

In the following, we show that (i) there are forms of collusion that lead to violations of user location privacy, and (ii) except for direct collusion between brokers and operators that serve the user, other forms of collusion only incur minimal leakages.
Attacks: Collusion between B and O reveals both U's identity and location. Note that this is the case for any MVNO-based architecture where B knows U's identity (for offering identity-based services) and O knows U's location (as it provides connectivity). Therefore, we focus on showing what other forms of collusion also impair user location privacy. For O, colluding with participants other than B does not provide it with extra information on U's identity or trajectories. For B, however, it can gain additional knowledge regarding U's location by colluding with (i) other users or (ii) other operators. The former is due to the use of session groups. Specifically, B knows that sessions in a session group belong to the same O, hence that users of these sessions have visited the same location at a similar time. Therefore, if some users who share session groups with U reveal their locations to B, then B knows U's location via such collusion. We call these "indirect location leakages". The latter is due to operators sharing the session table in the claiming phase. Specifically, B now effectively has a "smaller" session table consisting of only sessions from non-colluding operators, which is detrimental to the privacy guarantee provided by aggregate claiming.
Impact & Mitigation: While brokers gain extra user location information via collusion with other users or operators, the actual impacts are minimal and can be further reduced with different mitigation strategies. First, the impact of indirect location leakages is bounded by the size of session groups, which in turn depends on how fast the zk-SNARK backend is. Fortunately, even with a single-core backend, aggregate claiming can scale to large deployments with a session group that lasts as little as 20 *s* (§6.2). One could adopt faster backends like distributed zk-SNARK [111] to further reduce the size of session groups and hence leakages. Secondly, since the obfuscation effect of aggregate claiming is *exponential* w.r.t. the size of the session table (§5.1.2), a smaller table still grants sufficient protections. One could use a longer billing period to ensure a large enough session table even with collusion.

6 Implementation and Evaluation

In this section, we present the implementation of our LOCA prototype (§6.1) and investigate the two key questions regarding the feasibility of LOCA: (i) can LOCA scale to realistic deployment sizes? and (ii) how much overhead does LOCA introduce compared to existing cellular protocols? We answer the first question by performing a scalability analysis of the privacy building blocks (§6.2); and the second by conducting a performance analysis with wide-area experiments (§6.3).

6.1 Implementation

We prototyped LOCA as an extension to the CellBricks system [21] which is itself built from open source cellular platforms (Magma [41] and srsLTE [92]). We extended the operator and broker modules with the following: (i) the token generation and verification procedures implemented with rsablind [32]; (ii) the anonymous communication channel between the operator and broker implemented with Tor-socks [51,99] and NordVPN [80] and (iii) the claiming procedure implemented with Pequin [83,102] that has a single-core libsnark [69] as the zk-SNARK backend. In total, our extension includes 478 LoC in C (for claiming), 144 LoC in Go (for unlinkable token), and 16 LoC shell scripts (for anonymous communication and various setup). We prototyped LOCA with these languages as they were used in the original implementations that we extended. We built a testbed with two x86 machines: one as the user's device and the other as the operator's cell and core. We connect each machine to an SDR device (USRP B205-mini [40]) for radio connectivity. Lastly, the broker's service is deployed on AWS instances [15].

As an opt-in service, LOCA can be incrementally deployed and adopted starting with a small number of LOCA-

compatible users, brokers, and operators: users can have partial privacy by signing up with brokers that support LOCA and by using LOCA-based operators when available and falling back to legacy ones otherwise. We leave an evaluation of the privacy benefits under incremental adoption to future work.

6.2 Scaling analysis

LOCA must be able to scale to a large number of operators serving many users. Therefore, we evaluate whether the three privacy building blocks that we adopt can scale to large deployments, on the order of today's large MVNOs.

6.2.1 Blind signature

Blind signatures are used for generating and verifying unlinkable tokens. We measure a blind signature generation throughput of 522/sec and a verification throughput of 17202/sec on a 2.6GHz Intel I7-8850H CPU. These single-core throughputs are significant: generating 50 tokens for 10 operators per second. Moreover, brokers can easily achieve higher throughput with more cores or machines, hence we conclude that scaling blind signature operations will not be a problem.

6.2.2 Anonymous communication

For anonymous communication schemes in LOCA, an operator must have sufficient network capacity to send attachment requests to brokers. We measure the average network throughput of a Tor circuit to be 4.2 Mbps uplink and 6.1 Mbps downlink (consistent with Tor's reports [73]). Such throughput can support ≈ 400 attachment requests per second. Operators can easily scale up the throughput by establishing multiple Tor circuits with the same token. Alternatively, operators can use other anonymous communication schemes that have higher network throughput, such as VPNs (§6.3).

6.2.3 Aggregate Claiming with zk-SNARK

zk-SNARK has a long setup and proving time [111]. Given our aggregate claiming protocol is based on zk-SNARK, we evaluate whether the protocol can scale to large deployments. Since the generated keys are reused across billing cycles, zk-SNARK setup is performed offline only once, which excludes the setup time from the performance critical path. Hence we focus on the zk-SNARK proving time, which is invoked by each operator at every billing cycle to claim its session groups.

As noted in §4.5, LOCA allows claiming sessions in groups with a configurable size: smaller session groups offer stronger privacy guarantees as they minimize indirect location leakages. However, due to the slow zk-SNARK proving, operators may need to use large session groups so that they can *claim session groups faster than the rate of session group creation* and not develop a backlog of unclaimed sessions, at the cost of some privacy loss. To evaluate the amount of such privacy loss, we answer the following question: how small can session groups be while allowing operators to claim them fast enough? Specifically, we would like to obtain a *lower bound* for the average duration of a session group T[11]. As we will show next, even a single-core zk-SNARK implementation is fast enough to support session groups of small T, hence aggregate claiming will not be a scalability bottleneck.

As noted, we let K represent the maximum number of session groups an operator can claim and N represent the maximum number of session groups in the broker's session group table. If we denote $P(K,N)$ as the time it takes for zk-SNARK to prove the circuit of the claiming function parameterized by K and N, we have the following lower bound for T:

$$T \geq \frac{P(K,N)}{K}$$

To obtain the lower bound, we evaluate the proving time of our implementation for the claiming procedure $P(K,N)$. As mentioned in §4.5, proof-based verifiable computation has a compiler frontend and a zk-SNARK backend. Therefore, to evaluate $P(K,N)$, we need to answer two questions: (i) for a given K and N, how many constraints will the claiming function be compiled into? and (ii) how long will zk-SNARK take to prove these circuits of different sizes?

To answer the first question, we compile claiming functions with different Ks and Ns, and find the following formula that closely matches the numbers of constraints:

$$\#\ of\ constraints = K * (128 * N + 35394)$$

Terms in this formula are tied to the logic of the claiming function. As mentioned in §4.5, the claiming function contains two steps: (i) calculating hashes of the K provided secrets, and (ii) iterating through the N rows in the session table, checking whether the hash matches with one of the K precomputed hashes and adding it to the aggregate if so. Therefore, step (i) generates $35394 * K$ constraints, where 35394 is the number of constraints for computing a single SHA256 hash, consistent with prior work [65]; step (ii) contains an outer loop of N and an inner loop of K, which gets unrolled by the compiler into $128 * K * N$ constraints. Therefore, for a large enough N, the number of constraints scale almost linearly with $K * N$.

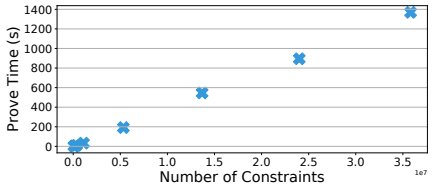

Figure 7: Proving time under varied number of constraints.

To answer the second question, we evaluate the proving time of compiled circuits with different numbers of constraints with a single-core libsnark backend on a 2.5GHz Intel 8259CL CPU. As shown in Fig 7, consistent with prior work [84, 111], the proving time increases linearly with the number of constraints: about 38 seconds per 1 million constraints.

Since we have shown that (i) the number of constraints of the claiming circuit increases linearly w.r.t $K * N$, and (ii) the proving time is linear w.r.t the number of constraints, we know that the *zk-SNARK proving time increases linearly w.r.t* $K * N$, *i.e.,* $P(K,N) = O(K * N)$. The constant factor c depends on the specific compiler frontends and zk-SNARK backends. For

[11] One can calculate the average number of sessions in a group as $T * r$, where r is the deployment-dependent rate of attachments for an operator.

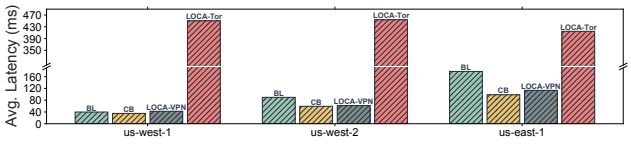

Figure 8: Average attachment latency of Magma baseline (BL), CellBricks (CB), LOCA-VPN and LOCA-Tor.

our implementations, $c \approx 128 * 38\ us = 4.894\ ms$. Therefore,

$$T \geq \frac{P(K,N)}{K} \approx \frac{c*K*N}{K} = c*N$$

This means that the lower bound on the duration of the session group grows *linearly w.r.t.* N. As stated earlier, we are mostly interested in cases of large Ns (*i.e.,* larger numbers of smaller session groups) as these lead to stronger privacy guarantees (§5.1.2). Fortunately, even with only the single-core libsnark backend, the lower bound of T for large N is reasonably small. As an example, the largest circuit that we evaluated (K=64, N=4096) has proving time $P(64,4096)$=1369 s; this translates to a lower bound of T=$P(64,4096)/64$=21.4 s. The asymptotic expression of $T = c*N = 4.864\ ms * 4096 \approx 20\ s$ matches with the measurement. The gap is due to ignoring the $35394 * K$ term, which will reduce as N goes even bigger.

Therefore, with $N = 4096$, the smallest session group that a single-core zk-SNARK can support has a duration of 20 s. This means that users who attach more than 20 s apart cannot reveal any information about each other's location, even if one user's location were leaked to the broker. We do not evaluate circuits with more than $35M$ constraints due to the scaling limit of the libsnark implementation. Recent work [111] on distributed zk-SNARK allows faster proving of much larger circuits, the evaluation of which is left to future work.

6.3 Performance analysis

Lastly, we would like to understand the performance that users receive with LOCA. Procedures like token generation and aggregate claiming happen off the critical path of users receiving services, thus do not affect user experience. Instead, we focus on the attachment procedure, since LOCA's attachment is both more complex and more frequent than today's protocols. We thus measure the additional latency overhead that LOCA adds to the attachment procedure.

We replicate the wide-area test setup from CellBricks [77]: the user equipment and the operator's cell and cellular core are always located in our local testbed, and we run experiments with the subscriber database (in the case of Magma) and the broker hosted on AWS EC2 [9]. This matches deployment practice where certain core network components are run in the carrier's datacenter. For each setup, we repeat the same attachment request using different cellular implementations 100 times and report the average performance.

Fig.8 shows the attachment latency after removing the time spent in lower radio layers (*i.e.,* RRC layer and below) for different placements of the subscriber database and broker. We compare four schemes: (i) unmodified Magma (baseline, denoted BL, that captures today's cellular architecture), (ii) CellBricks (denoted CB), LOCA's attachment protocol with (iii) VPN (denoted LOCA-VPN) and (iv) Tor (denoted LOCA-Tor) as the anonymous communication channel.

We make two observations from these results. First, the choice of anonymous communication scheme introduces a tradeoff between trust assumptions and attachment latency: LOCA-VPN requires trusting the VPN provider but achieves faster attachments than LOCA-Tor. In fact, LOCA-VPN is only 5 to 15 *ms* slower than CellBricks and still faster than today's attachment (*i.e.,* Magma). The reason we outperform Magma's attachment latency is because today's attachment procedure requires two round trips to the cloud, while CellBricks optimized this process to a single round-trip; since we build on CellBricks, we inherit this performance gain.

Our second observation is that even the slower LOCA-Tor is sufficiently fast for periodic reattachments: prior work [77] shows that attachment latencies of up to 500 *ms* have a minimal impact on application performance, even when users reattach on a per-tower basis. Hence LOCA-Tor, with a constant 400 *ms* latency due to the overhead of Tor [73], can support frequent reattachments with minimal disruptions.

7 Discussion

Viewing LOCA as a first step towards privacy-preserving cellular infrastructure, we next discuss two notable areas for improvement and potential directions to achieving them: (i) supporting beyond semi-honest and non-colluding participants, and (ii) improving non-privacy-related aspects of LOCA.

7.1 Beyond semi-honest and non-colluding

As stated in §3.1, there are both financial and legal reasons for brokers and operators to be semi-honest and not collude. However, relaxing these assumptions can certainly facilitate adoption. We next discuss directions towards such relaxation.

Semi-honest: LOCA suffers from privacy leakages in the face of various active attacks, *e.g.,* those based on out-of-protocol information (§5.2), which restricts it to semi-honest participants. We see two orthogonal directions towards supporting more aggressive participants. First, one could adopt specific defense mechanisms for different attacks (*e.g.,* traffic analysis, device fingerprinting) that have been proposed in prior work [43, 56, 60, 103, 110, 114]. LOCA, as a protocol-layer solution, can coexist with these mechanisms. Second, instead of averting attacks, one can detect these attacks and punish the misbehaving participants. The detection mechanism can involve multiple parties. For instance, operator over-reporting usage can be detected by brokers cross comparing the operator's reports with the ones from users. For the punishment mechanism, a promising approach is to build up a reputation system [77], where misbehaviors are factored into participant's reputation scores. Participants with poor reputation then receive degraded treatments: *e.g.,* a broker can decline to authorize an operator in the registration phase (§4.1). Such an approach is appealing in the cellular context, where brokers and operators need to remain operational for long enough to see a profit, allowing their track records to be built up.

Non-colluding: As elaborated in §5.3, except for direct collusion between brokers and operators that serve the user, other forms of collusion only incur minimal leakages in LOCA. An interesting question is then whether we could relax this requirement of no broker-operator collusion. Intuitively, preserving location privacy with *arbitrary* collusion seems unattainable: if a broker colludes with all the operators, it easily knows both the user's identity and all of her locations. Instead, we believe it is both feasible and interesting to investigate whether one could provide *partial* privacy guarantee if only a *subset* of operators collude with brokers. Under such a scenario, the coverage of non-colluding operators forms a region where little location information is revealed. Such a region is referred to as a *mix zone* and widely studied for location privacy in non-cellular contexts [17, 18, 53], and future work could leverage the insights of these work for cellular privacy.

7.2 Beyond privacy

Another area for improvement is the design and evaluation on non-privacy-related aspects of LOCA, such as performance and operational support. For performance, in §6.3, we measure LOCA's attachment latency to be less than 500 ms even with slower anonymous communication channel (*i.e.,* Tor), which was evaluated in [77] to have minimal performance impacts to applications like voice calls, video streaming and web browsing. It would be interesting to evaluate on more challenging applications such as video conferencing. Moreover, besides reducing trajectory leakages (§5.1.3), make-before-break handovers are expected to have better performance as well, the evaluation of which in LOCA is left to future work.

For operational support, LOCA supports tasks like identity-based services by having brokers offload these tasks to authorized but identity unknown operators (§4.4). However, there might be tasks that require knowledge of the operator's identity, such as recording misbehaving operators (for the aforementioned reputation system) and performing on-site inspections. To support these tasks, one potential approach is to involve a trusted third party when generating unlinkable tokens (§4.1). The goal is that upon legitimate requests, this third party can later assist in revealing the operator's identity for a token. One promising direction towards achieving this goal is to extend the registration phase with cryptographic constructs like secure multi-party computation (MPC) [35, 48, 113].

8 Related Work

Cellular: There has been extensive prior work on mitigating privacy violations by third parties other than network operators [11, 46, 58, 63, 68, 87, 93, 94, 100]. Our work instead focuses on protecting a user's location privacy from the network operator itself. To our knowledge, PGPP [86] is the only prior work that systematically studies this issue. As discussed earlier, PGPP adopts a different approach based on hiding users' identities from the network operator, which however compromises the network's ability to provide identity-based services and does not address the issue of trajectory-related leakages. One advantage of PGPP is higher tolerance for collusions, as it hides user's identity from both operators and brokers. However, it also assumes semi-honest participants who will not actively thwart its privacy mechanisms.

CellBricks [77] is a new cellular architecture that aims to democratize cellular access by enabling users to easily leverage small-scale operators. LOCA borrows the idea of user-driven mobility, although we use it for privacy reasons while CellBricks requires it to give users the ability to dynamically select an operator of their choice. CellBricks does not address the issue of location privacy and hence is similar to 3GPP protocols in this regard. In fact, we note that the importance of hiding O's identity from B is greater under the CellBricks vision of larger numbers of smaller-scale operators.

General location privacy: There is extensive prior work on location privacy in non-cellular contexts [17, 36, 67, 76, 90, 106, 112]. These reveal four general methods for protecting location privacy: (i) regulatory strategies – government rules to regulate the use of personal information; (ii) privacy policies – trust-based agreements between individuals and whoever is receiving their location data; (iii) anonymity – use a pseudonym and create ambiguity by grouping with other people. (iv) obfuscation – temporal or spatial degradation of the location data. Regulatory strategies and privacy policies are orthogonal to computational countermeasures like techniques adopted in LOCA. In the cellular context, neither obfuscation nor anonymity is desirable: obfuscation is not feasible, because a user's location data is generated by the infrastructure, the temporal or spatial resolution of which is not determined by the user; anonymity is the approach adopted by PGPP [86] which, as discussed earlier, compromises on identity-based services. LOCA exploits the unique role of brokers and adopts a novel approach to preserving location privacy while supporting identity-based services. LOCA's approach of strategically hiding different pieces of information from each party has been investigated for preserving privacy in other contexts as well, such as Apple's private relay [31].

Applications of LOCA's privacy building blocks: Blind signatures have been applied for e-voting [59, 64, 75]. Anonymous communication has been used in social networking and web browsing [44, 55, 99]. Proof-based verifiable computation has been used in outsourced computing [24, 27, 70]. LOCA synthesizes these building blocks to support cellular procedures like attachment and aggregate claiming.

9 Conclusion

We presented LOCA, a novel cellular architecture that provides location privacy while supporting identity-based services such as usage-based billing, QoS, and lawful intercept.

We view our work as a first step towards enabling privacy-preserving communication infrastructure and hope that future work will extend our design to address additional threat models and reduced overheads, as well as explore the applicability of LOCA's design to other access technologies.

References

[1] 3GPP. Lte;telecommunication management; performance management (pm); performance measurements evolved universal terrestrial radio access network (e-utran). Technical Specification (TS) 32.425, 3rd Generation Partnership Project (3GPP), 08 2016. Version 13.5.0.

[2] 3GPP. 5g; security architecture and procedures for 5g system. Technical Specification (TS) 33.501, 3rd Generation Partnership Project (3GPP), 10 2018. Version 15.2.0.

[3] 3GPP. Lawful Interception (LI);Handover interface for the lawful interception of telecommunications traffic. https://www.etsi.org/deliver/etsi_es/201600_201699/201671/03.02.01_50/es_201671v030201m.pdf, 2018.

[4] 3GPP. Lawful interception architecture and functions. Technical Specification (TS) 33.107, 3rd Generation Partnership Project (3GPP), 07 2019. Version 15.6.0.

[5] 3GPP. 5g; security architecture and procedures for 5g system. Technical Specification (TS) 23.501, 3rd Generation Partnership Project (3GPP), 10 2020. Version 16.6.0.

[6] 3GPP. Lte; 3gpp system architecture evolution (sae); security architecture. Technical Specification (TS) 33.401, 3rd Generation Partnership Project (3GPP), 03 2020. Version 15.11.0.

[7] 3GPP. Non-Access Stratum. https://www.3gpp.org/technologies/keywords-acronyms/96-nas, 2020.

[8] 3GPP. Nr and ng-ran overall description. Technical Specification (TS) 38.300, 3rd Generation Partnership Project (3GPP), 07 2020. Version 16.2.0.

[9] Amazon Web Service. Aws ec2 regions. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.RegionsAndAvailabilityZones.html, 2021.

[10] Apple. Privacy - apple. https://www.apple.com/privacy/, 2021.

[11] 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*, pages 205–216, 2012.

[12] Richard Arratia and Louis Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of mathematical biology*, 51(1):125–131, 1989.

[13] AT&T. AT&T private cellular networks. https://www.business.att.com/products/att-private-cellular-networks.html, 2020.

[14] AT&T. Deep packet inspection explained. https://cybersecurity.att.com/blogs/security-essentials/what-is-deep-packet-inspection, 2021.

[15] AWS. Amazon ec2 instance types. https://aws.amazon.com/ec2/instance-types/.

[16] Leda Bargiotti, Inge Gielis, Bram Verdegem, Pieter Breyne, Francesco Pignatelli, Paul Smits, Ray Boguslawski, et al. Guidelines for public administrations on location privacy: European union location framework. Technical report, Joint Research Centre (Seville site), 2016.

[17] Alastair R Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive computing*, 2(1):46–55, 2003.

[18] Alastair R Beresford and Frank Stajano. Mix zones: User privacy in location-aware services. In *IEEE Annual conference on pervasive computing and communications workshops, 2004. Proceedings of the Second*, pages 127–131. IEEE, 2004.

[19] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Teja Sukhavasi, Chirag Patel, and Stefan Geirhofer. Network densification: the dominant theme for wireless evolution into 5g. *IEEE Communications Magazine*, 52(2):82–89, 2014.

[20] Camila CS Caiado and Pushpa N Rathie. Polynomial coefficients and distribution of the sum of discrete uniform variables. In *Eighth Annual Conference of the Society of Special Functions and their Applications, Pala, India, Society for Special Functions and their Applications*, 2007.

[21] CellBricks. Cellbricks. https://cellbricks.github.io/, 2021.

[22] Mobile Internet Resource Center. Top pickfeatured overview: postpaid consumer plans by verizon (cellular data plans). https://www.rvmobileinternet.com/gear/the-verizon-plan/, 2021.

[23] David Chaum. Blind signatures for untraceable payments. In *Advances in cryptology*, pages 199–203. Springer, 1983.

[24] Seung Geol Choi, Jonathan Katz, Ranjit Kumaresan, and Carlos Cid. Multi-client non-interactive verifiable computation. In *Theory of Cryptography Conference*, pages 499–518. Springer, 2013.

[25] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 41–50. IEEE, 1995.

[26] Gene Connolly, Anatoly Sachenko, and George Markowsky. Distributed traceroute approach to geographically locating ip devices. In *Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2003. Proceedings*, pages 128–131. IEEE, 2003.

[27] Craig Costello, Cédric Fournet, Jon Howell, Markulf Kohlweiss, Benjamin Kreuter, Michael Naehrig, Bryan Parno, and Samee Zahur. Geppetto: Versatile verifiable computation. In *2015 IEEE Symposium on Security and Privacy*, pages 253–270. IEEE, 2015.

[28] Joseph Cox. I gave a bounty hunter $300. then he located our phone. https://www.vice.com/en/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile, 2019.

[29] Joseph Cox. Stalkers and debt collectors impersonate cops to trick big telecom into giving them cell phone location data. https://www.vice.com/en/article/panvkz/stalkers-debt-collectors-bounty-hunters-impersonate-cops-phone-location-data, 2019.

[30] Cricket. Cricket wireless. https://www.cricketwireless.com/, 2021.

[31] Jason Cross. icloud+ private relay faq: Everything you need to know. https://www.macworld.com/article/348965/icloud-plus-private-relay-safari-vpn-encryption-privacy.html, 2021.

[32] CryptoBallot. Rsa blind signing using a full domain hash. https://github.com/cryptoballot/rsablind, 2021.

[33] Boris Danev, Davide Zanetti, and Srdjan Capkun. On physical-layer identification of wireless devices. *ACM Computing Surveys (CSUR)*, 45(1):1–29, 2012.

[34] Emma Dauterman, Eric Feng, Ellen Luo, Raluca Ada Popa, and Ion Stoica. {DORY}: An encrypted search system with distributed trust. In *14th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 20)*, pages 1101–1119, 2020.

[35] Wenliang Du and Mikhail J Atallah. Secure multi-party computation problems and their applications: a review and open problems. In *Proceedings of the 2001 workshop on New security paradigms*, pages 13–22, 2001.

[36] Matt Duckham and Lars Kulik. Location privacy and location-aware computing. *Dynamic & mobile GIS: investigating change in space and time*, 3:35–51, 2006.

[37] Steffen Eger. Stirling's approximation for central extended binomial coefficients. *The American Mathematical Monthly*, 121(4):344–349, 2014.

[38] Ericsson. Evolving cellular IoT for industry digitalization. https://www.ericsson.com/en/internet-of-tRFWirelessWorldhings/iot-connectivity/cellular-iot, 2020.

[39] ETSI. Lawful intercept ETSI. https://www.etsi.org/technologies/lawful-interception, 2020.

[40] Ettus. Usrp b205mini. https://www.ettus.com/all-products/usrp-b205mini-i/, 2020.

[41] Facebook. Magma. https://www.magmacore.org/, 2021.

[42] Nour-Eddine Fahssi. Some identities involving polynomial coefficients. *arXiv preprint arXiv:1507.07968*, 2015.

[43] Kassem Fawaz, Kyu-Han Kim, and Kang G Shin. Protecting privacy of {BLE} device users. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 1205–1221, 2016.

[44] Eran Gabber, Phillip B Gibbons, Yossi Matias, and Alain Mayer. How to make personalized web browsing simple, secure, and anonymous. In *International Conference on Financial Cryptography*, pages 17–31. Springer, 1997.

[45] Ruchi Garg. Dual active protocol stack handover (daps ho). https://www.linkedin.com/pulse/dual-active-protocol-stack-handover-daps-ho-ruchi-garg/, 2021.

[46] M Køien Geir et al. Privacy enhanced mutual authentication in lte. In *2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 614–621. IEEE, 2013.

[47] Hadi Givehchian, Nishant Bhaskar, Eliana Rodriguez Herrera, Héctor Rodrigo López Soto, Christian Dameff, Dinesh Bharadia, and Aaron Schulman. Evaluating physical-layer ble location tracking attacks on mobile devices. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1690–1704. IEEE, 2022.

[48] Oded Goldreich. Secure multi-party computation. *Manuscript. Preliminary version*, 78(110), 1998.

[49] Google. Google-Fi. https://fi.google.com/about/, 2021.

[50] Swish Goswami. The rising concern around consumer data and privacy. https://www.forbes.com/sites/forbestechcouncil/2020/12/14/the-rising-concern-around-consumer-data-and-privacy/?sh=6e6200a6487e, 2020.

[51] David Goulet. Torsocks. https://github.com/dgoulet/torsocks, 2021.

[52] GSMA. Enabling neutral host: CCS case study. https://www.gsma.com/futurenetworks/wp-content/uploads/2018/09/180920-CCS_GSMA_Case_Study-FINAL_NE-Modelling-removed.pdf, 2020.

[53] Nan Guo, Linya Ma, and Tianhan Gao. Independent mix zone for location privacy in vehicular networks. *IEEE Access*, 6:16842–16850, 2018.

[54] Trinabh Gupta, Natacha Crooks, Whitney Mulhern, Srinath Setty, Lorenzo Alvisi, and Michael Walfish. Scalable and private media consumption with popcorn. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, pages 91–107, 2016.

[55] Nguyen Phong Hoang and Davar Pishva. Anonymous communication and its importance in social networking. In *16th International Conference on Advanced Communication Technology*, pages 34–39. IEEE, 2014.

[56] Jinsong Hu, Shihao Yan, Feng Shu, Jiangzhou Wang, Jun Li, and Yijin Zhang. Artificial-noise-aided secure transmission with directional modulation based on random frequency diverse arrays. *IEEE Access*, 5:1658–1667, 2017.

[57] Huawei. Huawei privacy. https://consumer.huawei.com/en/privacy/, 2021.

[58] Syed Hussain, Omar Chowdhury, Shagufta Mehnaz, and Elisa Bertino. Lteinspector: A systematic approach for adversarial testing of 4g lte. In *Network and Distributed Systems Security (NDSS) Symposium 2018*, 2018.

[59] Subariah Ibrahim, Maznah Kamat, Mazleena Salleh, and Shah Rizan Abdul Aziz. Secure e-voting with blind signature. In *4th National Conference of Telecommunication Technology, 2003. NCTT 2003 Proceedings.*, pages 193–197. IEEE, 2003.

[60] Marc Juarez, Mohsen Imani, Mike Perry, Claudia Diaz, and Matthew Wright. Toward an efficient website fingerprinting defense. In *European Symposium on Research in Computer Security*, pages 27–46. Springer, 2016.

[61] Seny Kamara, Payman Mohassel, and Mariana Raykova. Outsourcing multi-party computation. *IACR Cryptol. Eprint Arch.*, 2011:272, 2011.

[62] 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, 2019.

[63] 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*, pages 207–218, 2017.

[64] Malik Sikandar Hayat Khiyal, Aihab Khan, Saba Bashir, Farhan Hassan Khan, and Shaista Aman. Dynamic blind group digital signature scheme in e-banking. *International Journal of Computer and Electrical Engineering*, 3(4):514–519, 2011.

[65] Ahmed Kosba, Charalampos Papamanthou, and Elaine Shi. xjsnark: A framework for efficient verifiable computation. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 944–961. IEEE, 2018.

[66] KrebsOnSecurity. Tracking firm locationsmart leaked location data for customers of all major u.s. mobile carriers without consent in real time via its web site. https://krebsonsecurity.com/2018/05/tracking-firm-locationsmart-leaked-location-data-for-customers-of-all-major-u-s-mobile-carriers-in-real-time-via-its-web-site/, 2018.

[67] John Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

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

[69] SCIPR Lab. Libsnark. https://github.com/scipr-lab/libsnark, 2021.

[70] Junzuo Lai, Robert H Deng, HweeHwa Pang, and Jian Weng. Verifiable computation on outsourced encrypted data. In *European Symposium on Research in Computer Security*, pages 273–291. Springer, 2014.

[71] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, et al. The quic transport protocol: Design and internet-scale deployment. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pages 183–196, 2017.

[72] Sangwon Lee, Sylvia M Chan-Olmsted, and Hsiao-Hui Ho. The emergence of mobile virtual network operators (mvnos): An examination of the business strategy in the global mvno market. *The International Journal on Media Management*, 10(1):10–21, 2008.

[73] Karsten Loesing, Steven J. Murdoch, and Roger Dingledine. A case study on measuring statistical data in the Tor anonymity network. In *Proceedings of the Workshop on Ethics in Computer Security Research (WECSR 2010)*, LNCS. Springer, January 2010.

[74] Natasha Lomas. Uh oh! european carriers are trying to get into 'personalized' ad targeting. https://techcrunch.com/2022/06/24/trustpid/, 2022.

[75] Lourdes López-García, Luis J Dominguez Perez, and Francisco Rodríguez-Henríquez. A pairing-based blind signature e-voting scheme. *The Computer Journal*, 57(10):1460–1471, 2014.

[76] Zhaojun Lu, Gang Qu, and Zhenglin Liu. A survey on recent advances in vehicular network security, trust, and privacy. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):760–776, 2018.

[77] Zhihong Luo, Silvery Fu, Mark Theis, Shaddi Hasan, Sylvia Ratnasamy, and Scott Shenker. Democratizing cellular access with cellbricks. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, pages 626–640, 2021.

[78] Anna Lysyanskaya, Ronald L Rivest, Amit Sahai, and Stefan Wolf. Pseudonym systems. In *International Workshop on Selected Areas in Cryptography*, pages 184–199. Springer, 1999.

[79] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*, pages 19–38. IEEE, 2017.

[80] NordVPN. NordVPN. https://nordvpn.com, 2021.

[81] OpenCellID. The world's largest open database of cell towers. https://www.opencellid.org/, 2021.

[82] Christoph Paasch and Sebastien Barre. Multipath TCP. https://www.multipath-tcp.org, 2021. Accessed: 2020-04-29.

[83] The Pepper Project. Pequin: An end-to-end toolchain for verifiable computation, snarks, and probabilistic proofs. https://github.com/pepper-project/pequin, 2021.

[84] Charles Rackoff and Daniel R Simon. Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In *Annual International Cryptology Conference*, pages 433–444. Springer, 1991.

[85] Samsung. Samsung's approach to privacy. https://www.samsung.com/us/account/our-approach-to-privacy/, 2021.

[86] Paul Schmitt and Barath Raghavan. Pretty good phone privacy. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 1737–1754, 2021.

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

[88] Justine Sherry, Chang Lan, Raluca Ada Popa, and Sylvia Ratnasamy. Blindbox: Deep packet inspection over encrypted traffic. In *Proceedings of the 2015 ACM conference on special interest group on data communication*, pages 213–226, 2015.

[89] Yi-Sheng Shiu, Shih Yu Chang, Hsiao-Chun Wu, Scott C-H Huang, and Hsiao-Hwa Chen. Physical layer security in wireless networks: A tutorial. *IEEE wireless Communications*, 18(2):66–74, 2011.

[90] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *2011 IEEE symposium on security and privacy*, pages 247–262. IEEE, 2011.

[91] Pham Hai Son, Sudan Jha, Raghvendra Kumar, Jyotir Moy Chatterjee, et al. Governing mobile virtual network operators in developing countries. *Utilities Policy*, 56:169–180, 2019.

[92] srsRAN. srsLTE: Your own mobile network. https://www.srslte.com/, 2020.

[93] Daehyun Strobel. Imsi catcher. *Chair for Communication Security, Ruhr-Universität Bochum*, 14, 2007.

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

[95] Techplayon. 5g nr dual active protocol stack (daps) handover – 3gpp release 16. https://www.techplayon.com/5g-nr-dual-active-protocol-stack-daps-handover-3gpp-release-16/, 2020.

[96] TelcoBridges. Lawful intercept solutions. https://www.telcobridges.com/solutions/operators/lawful-intercept, 2021.

[97] Telecoms. Neutral host networks and how to support them. https://telecoms.com/opinion/neutral-host-networks-and-how-to-support-them/, 2020.

[98] Tessian. 22 biggest gdpr fines of 2019, 2020, and 2021 (so far). https://www.tessian.com/blog/biggest-gdpr-fines-2020/, 2021.

[99] Tor. Tor. https://www.torproject.org/, 2021.

[100] Fabian Van Den Broek, Roel Verdult, and Joeri de Ruiter. Defeating imsi catchers. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, pages 340–351, 2015.

[101] 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*, pages 137–152, 2015.

[102] Riad S Wahby, Srinath TV Setty, Zuocheng Ren, Andrew J Blumberg, and Michael Walfish. Efficient ram and control flow in verifiable outsourced computation. In *NDSS*, 2015.

[103] Tao Wang, Xiang Cai, Rishab Nithyanand, Rob Johnson, and Ian Goldberg. Effective attacks and provable defenses for website fingerprinting. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 143–157, 2014.

[104] Lance Whitney. Data privacy is a growing concern for more consumers. https://www.techrepublic.com/article/data-privacy-is-a-growing-concern-for-more-consumers/, 2021.

[105] 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/, 2018.

[106] Björn Wiedersheim, Zhendong Ma, Frank Kargl, and Panos Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. In *2010 Seventh international conference on wireless on-demand network systems and services (WONS)*, pages 176–183. IEEE, 2010.

[107] Josephine Wolff and Nicole Atallah. Early gdpr penalties: Analysis of implementation and fines through may 2020. *Journal of Information Policy*, 11:63–103, 2021.

[108] Ben Wolford. What are the gdpr fines? https://gdpr.eu/fines/, 2021.

[109] RF Wireless World. LTE QoS quality of service, class identifier(QCI), QoS in LTE. https://www.rfwireless-world.com/Tutorials/LTE-QoS.html, 2021.

[110] Charles V Wright, Scott E Coull, and Fabian Monrose. Traffic morphing: An efficient defense against statistical traffic analysis. In *NDSS*, volume 9. Citeseer, 2009.

[111] Howard Wu, Wenting Zheng, Alessandro Chiesa, Raluca Ada Popa, and Ion Stoica. {DIZK}: A distributed zero knowledge proof system. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 675–692, 2018.

[112] Hui Zang and Jean Bolot. Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 145–156, 2011.

[113] Chuan Zhao, Shengnan Zhao, Minghao Zhao, Zhenxiang Chen, Chong-Zhi Gao, Hongwei Li, and Yu-an Tan. Secure multi-party computation: theory, practice and applications. *Information Sciences*, 476:357–372, 2019.

[114] Yulong Zou, Jia Zhu, Xianbin Wang, and Victor CM Leung. Improving physical-layer security in wireless communications using diversity techniques. *IEEE Network*, 29(1):42–48, 2015.

Appendix

A Unlinkable token

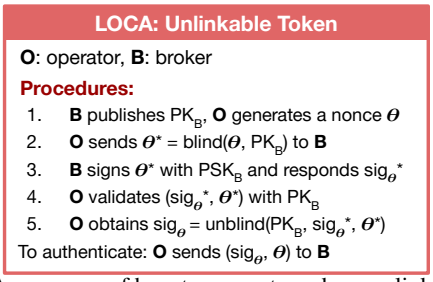

Figure 9: A summary of how to generate and use unlinkable tokens.

Fig.9 summarizes how O obtains tokens. The protocol starts with B publishing its public key PK_B and O generating a nonce θ (*i.e.,* the token). To get B *blindly-sign* the θ, O first *blinds* θ using PK_B and sends the blinded token θ^* to B, requesting B to sign the token. O reveals its identity to B who can decide whether to accept this request. Next, B signs θ^* using its private key PSK_B and returns the blind signature sig^*_θ to O which O can validate using PK_B.
Next, O obtains the *unblinded signature* of the token using PK_B, sig^*_θ, and θ^*. To authenticate itself to B, O sends θ and the unblinded signature sig_θ to B. B can then verify the token's authenticity with sig_θ as a normal digital signature. Note that B *cannot* link θ to O since θ was blindly-signed and never seen by B (only the blinded θ^* was).

B Aggregate Claiming

Simulation setup: For each ratio K/N, we scale N while increasing K proportionally according to the ratio and estimate the expected number of matched subsets for the (K,N). For example, for the ratio K/N=1/10, we run experiments for (K,N)=$(1,10),(2,20),...,(5,50)$. For each (K,N), we again make the simplification to not consider subsets that contain any of the K session groups. By doing so, we can independently sample arrays Xs of length $(N-K)$ and Ys of length K, count the number of subsets in X that have the same sum as Y for each pair of (X,Y), and report the average of all the pairs as the estimate for (K,N), denoted as R. To ensure that the estimate is accurate, we use a large number of samples for each (K,N), up to 2^{25} so that the simulation can finish within a reasonable time frame.
Theoretical proof: For this proof, we make use of *polynomial coefficients*, also named as extended binomial coefficients, which are natural extensions of the well-known binomial coefficient. For $n,m \in Z_{>0}$ Polynomial coefficients $\binom{n}{k}_m$ is the coefficient of x^k in the following expansion:

$$(1+x+...+x^m)^n = \sum_{k=0}^{k=mn} \binom{n}{k}_m x^k$$

Note that $\binom{n}{k}_m = 0$ for $k \notin \{0,..,mn\}$. Binomial coefficient is the special case where $m=1$. An equivalent definition of $\binom{n}{k}_m$ is:

$$\binom{n}{k}_m = \sum_{\substack{k_0\geq 0,...,k_m\geq 0 \\ k_0+...+k_m=n \\ 0\cdot k_0+...+m\cdot k_m=k}} \binom{n}{k_0,...,k_m} \tag{1}$$

It is known that polynomial coefficients are symmetric: $\binom{n}{k}_m = \binom{n}{mn-k}_m$, and $\binom{n}{k}_m$ is a non-decreasing function of k for $0 \leq k \leq \lfloor \frac{mn}{2} \rfloor$ and a non-increasing function for $\lceil \frac{mn}{2} \rceil \leq k \leq mn$ [42].
Prior work has shown that the sum of N iid random variables from the discrete uniform distribution of $\{0,...,m\}$ ($U\{0,m\}$), denoted as S_N, has the following closed-form distribution expressed with polynomial coefficients [20]:

$$\begin{aligned}
P(S_N=y) &= \sum_{\substack{a_0\geq 0,...,a_m\geq 0 \\ a_0+...+a_m=N \\ 0\cdot a_0+...+m\cdot a_m=y}} P(a_0,...,a_m) && (a_i \text{ stands for the number of elements equal to } i) \\
&= \sum_{\substack{a_0\geq 0,...,a_m\geq 0 \\ a_0+...+a_m=N \\ 0\cdot a_0+...+m\cdot a_m=y}} (\frac{1}{m+1})^N \binom{N}{a_0,...,a_m} && (\text{each element can be putted into bucket } i \text{ with a likelihood of } \frac{1}{m+1}) \\
&= (\frac{1}{m+1})^N \binom{N}{y}_m && (\text{according to definition (1)})
\end{aligned}$$

Lemma B.1. *The distribution of sums of N iid random variables from $U\{1,m\}$ has the following closed-form expression:* $P_N(y) = (\frac{1}{m})^N \binom{N}{y-N}_{m-1}$

Proof. The proof is similar to the proof above for sums of N iid random variables from $U\{0,m\}$, with some minor adjustment:

$$\begin{aligned}
P_N(y) &= \sum_{\substack{a_1\geq 0,\ldots,a_m\geq 0\\ a_1+\ldots+a_m=N\\ 1\cdot a_1+\ldots+m\cdot a_m=y}} (\frac{1}{m})^N \binom{N}{a_1,\ldots,a_m} \\
&= \sum_{\substack{a_1\geq 0,\ldots,a_m\geq 0\\ a_1+\ldots+a_m=N\\ 0\cdot a_1+\ldots+(m-1)\cdot a_m=y-N}} (\frac{1}{m})^N \binom{N}{a_1,\ldots,a_m} && \text{(align with the format of (1))} \\
&= (\frac{1}{m})^N \binom{N}{y-N}_{m-1} && \text{(according to definition (1))}
\end{aligned}$$

□

Lemma B.2. *The distribution of subset sums of N iid random variables from $U\{1,m\}$ has the following closed-form expression:* $Q_N(y) = (\frac{1}{2m})^N \sum_{k=0}^{N} \binom{N}{k} m^k \binom{N-k}{y-(N-k)}_{m-1}$

Proof. Subset sums of $U\{1,m\}$ can be equivalently treated as sums of elements X_i that has the following distribution:

$$P(X_i) = \begin{cases} \frac{1}{2}, & \text{if } X_i = 0 \\ \frac{1}{2m}, & \text{if } X_i \in \{1,\ldots,m\} \\ 0, & \text{otherwise} \end{cases}$$

Therefore, we can calculate the probability of a subset sum by multiplying the probability of having different numbers of zeros with the probability of adding up to the sum with the remaining non-zero elements:

$$\begin{aligned}
Q_N(y) &= \sum_{k=0}^{N} \binom{N}{k} (\frac{1}{2})^k \sum_{\substack{a_1\geq 0,\ldots,a_m\geq 0\\ a_1+\ldots+a_m=N-k\\ 1\cdot a_1+\ldots+m\cdot a_m=y}} (\frac{1}{2m})^{N-k} \binom{N-k}{a_1,\ldots,a_m} \\
&= \sum_{k=0}^{N} \binom{N}{k} (\frac{1}{2})^k \sum_{\substack{a_1\geq 0,\ldots,a_m\geq 0\\ a_1+\ldots+a_m=N-k\\ 0\cdot a_1+\ldots+(m-1)\cdot a_m=y-(N-k)}} (\frac{1}{2m})^{N-k} \binom{N-k}{a_1,\ldots,a_m} && \text{(similar to Lemma B.1)} \\
&= \sum_{k=0}^{N} \binom{N}{k} (\frac{1}{2})^k (\frac{1}{2m})^{N-k} \binom{N-k}{y-(N-k)}_{m-1} && \text{(according to definition (1))} \\
&= (\frac{1}{2m})^N \sum_{k=0}^{N} \binom{N}{k} m^k \binom{N-k}{y-(N-k)}_{m-1}
\end{aligned}$$

□

Lemma B.3. *Here we define a "head" function $H(hN;N,p)(0\leq h\leq 1)$ and a "tail" function $T(tN;N,p)(0\leq t\leq 1)$, where:*

$$H(hN;N,p) = \sum_{k=0}^{hN} \binom{N}{k} p^k; \qquad T(tN;N,p) = \sum_{k=tN}^{N} \binom{N}{k} p^k$$

Then we can prove the following: if $h < \frac{p}{1+p}$, $H(hN;N,p) \leq (b_H)^N$, with $b_H < (1+p)$. Similarly, if $t > \frac{p}{1+p}$, $T(tN;N,p) \leq (b_T)^N$, with $b_T < (1+p)$.

Proof. Here we show the proof for the head function, the proof for the tail function is very similar. The idea is to rewrite the head function to follow the format of a binomial distribution, so that we could use tail bounds for binomial distributions to

provide a lower bound. Specifically:

$$\begin{aligned}
H(hN;N,p) &= \sum_{k=0}^{hN} \binom{N}{k} p^k \\
&= (1+p)^N \sum_{k=0}^{hN} \binom{N}{k} (\frac{p}{1+p})^k (\frac{1}{1+p})^{N-k} \\
&= (1+p)^N \sum_{k=0}^{hN} \binom{N}{k} (p')^k (1-p')^{N-k} && \text{(with } p' = \frac{p}{1+p}\text{)} \\
&= (1+p)^N F(hN;N,p')
\end{aligned}$$

where $F(hN;N,p')$ refers to the probability of having at most hN successes in a Binomial trial $B(N,p')$. For $F(hN;N,p')$, it's known that if $\frac{hN}{N} = h \leq p'$, we have the following tail bounds [12]:

$$F(hN;N,p') \leq \exp[-Nf(h,p')] \qquad \text{with } f(h,p') = \begin{cases} 2(h-p')^2, & \text{with Hoeffding's inequality} \\ D(h||p'), & \text{with Chernoff bound} \end{cases}$$

where $D(a||p)$ is the relative entropy between a *Bernoulli*(a) (a-coin) and a *Bernoulli*(p) (p-coin):
$D(a||p) = (a)\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}$. For either of this, $f(h,p') > 0$ for $h < p'$. Therefore, for $H(hN;N,p)$, if $h < p' = \frac{p}{1+p}$, we have:

$$\begin{aligned}
H(hN;N,p) &= (1+p)^N F(hN;N,p') \\
&\leq (1+p)^N \exp[-Nf(h,p')] \\
&= \{(1+p)\exp[-f(h,p')]\}^N \\
&= (b_H)^N && \text{(with } b_H = (1+p)\exp[-f(h,p')] < (1+p)\text{)}
\end{aligned}$$

□

Lemma B.4.

$$\frac{1}{m^2+m+1} < 1 - \frac{ln(m)}{ln(m+1)}, \quad \forall m \in Z_{>0}$$

Proof. This is equivalent as showing

$$h(m) = (m^2+m+1)ln(m) - (m^2+m)ln(m+1) < 0, \quad \forall m \in Z_{>0}$$

Taking derivative, we have

$$\begin{aligned}
h'(m) &= (2m+1)\ln(\frac{m}{m+1}) + 1 + \frac{1}{m} \\
&\leq (2m+1)(\frac{m}{m+1} - 1) + 1 + \frac{1}{m} && \text{(with } \ln(x) \leq x-1\text{)} \\
&= \frac{-m^2+m+1}{m(m+1)} < 0 && \text{(for } m \geq 2 > \frac{1+\sqrt{5}}{2}\text{)}
\end{aligned}$$

Therefore, because (i) $h(m)$ monotonically decreases for $m \geq 2$, and (ii) $h(1), h(2) < 0$, we have $h(m) < 0, \forall m \in Z_{>0}$ □

$f[n]$ is said to be exponential w.r.t n iff:

$$\exists M \in R_{>0}, c \in R_{>1}, \quad \lim_{n\to\infty} \frac{f[n]}{c^n} = M$$

Therefore, a sufficient condition for $f[n]$ to be exponential is that

$$\exists M \in R_{>0}, c \in R_{>1}, N_0 \in Z, \quad f[n] \geq M \cdot c^n, \quad \forall n \geq N_0$$

With this we could prove the following lemma:

Lemma B.5. *For any integer $K \geq 1$, $h_K[n] = a^n - \sum_{i=1}^{K} b_i^N$ is exponential w.r.t n, if $a > 1$ and $a > \max_{i:i\in\{1,..,K\}} b_i$*

Proof. $h_K[n]$ can be rewritten as $h_K[n] = \sum_{i=1}^{K} \frac{1}{K}a^n - b_i^n$. We can prove that $\frac{1}{K}a^n - b_i^n$ are exponential for all b_i's. Specifically, we can show that.

$$\forall c \in \{x \in R \mid \max(b_i, 1) < x < a\}, M \in R_{>0}, \exists N(a,c,M) \in Z, \quad \frac{1}{K}a^n - b_i^n > Mc^n, \quad \forall n \geq N(a,c,M)$$

This because $\frac{1}{K}a^n - b_i^n > Mc^n \Leftrightarrow \frac{1}{k}(\frac{a}{c})^n > M + (\frac{b_i}{c})^n$. By having $n \geq N(a,c,M) = \lceil log_{\frac{a}{c}}[k(M+1)] \rceil$, we have $\frac{1}{k}(\frac{a}{c})^n > M+1 > M + (\frac{b_i}{c})^n$. Therefore, $h_K[n] = \sum_{i=0}^{K} \frac{1}{K}a^n - b_i^n$ is obviously exponential:

$$\forall c \in \{x \in R \mid \max(\max_{i:i\in\{1,..,K\}} b_i, 1) < x < a\}, M \in R_{>0}, \exists N(a,c,M) \in Z, \quad h_K[n] > KMc^n, \quad \forall n \geq N(a,c,M) \qquad \square$$

We are now ready to prove the main theorem, which, in the context of LOCA, states that with the usage of each session as a uniform distribution of $\{1,...,m\}$, a bTelco's number of sessions as N_B and the total number of sessions a broker receives from all bTelcos as N_A, if $\frac{N_B}{N_A}$ meets certain requirements (depending on m), the expected number of session subsets that have the same aggregate usage as the bTelco's N_B sessions grows exponentially w.r.t the total number of sessions, N_A.
Next, we prove that a lower bound of this expected number, considering subsets consisting of only the remaining $N_A - N_B$ sessions, is already exponential w.r.t N_A:

Theorem B.6. *Considering two independent arrays X and Y, consisting of $N_A - N_B$ and N_B iid random variables from $U\{1,m\}$, there exists $L(m), U(m)$ such that the expected number of subsets X, whose sums are equal to the sum of Y, is exponential w.r.t N_A, if $L(m) \leq \frac{N_B}{N_A} \leq U(m)$. Note that $L(m), U(m)$ depend on m, and $0 < L(m) \leq U(m) < 1$, $\forall m \in Z_{>0}$*

Proof. We prove this theorem by showing a valid $L(m), U(m)$ pair. We denote this expected number as $E(m, N_A, N_B)$, and derive its closed-form expression by using the distributions of sums and subset sums, which were computed in Lemma B.1 and Lemma B.2. Specifically, for a random array B, the probability that its sum equals to y is $P_{N_B}(y)$, and the expected number of subsets in A that add to y is $2^{N_A-N_B}Q_{N_A-N_B}(y)$:

$$\begin{aligned}
E(m,N_A,N_B) &= 2^{N_A-N_B}\sum_{y=N_B}^{mN_B} P_{N_B}(y)Q_{N_A-N_B}(y) \\
&= 2^{N_A-N_B}\sum_{y=N_B}^{mN_B} (\frac{1}{m})^{N_B}\binom{N_B}{y-N_B}_{m-1} (\frac{1}{2m})^{N_A-N_B}\sum_{k=0}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\binom{N_A-N_B-k}{y-(N_A-N_B-k)}_{m-1} \qquad (\text{B.1, B.2}) \\
&= (\frac{1}{m})^{N_A}\sum_{y=N_B}^{mN_B}\binom{N_B}{y-N_B}_{m-1}\sum_{k=0}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\binom{N_A-N_B-k}{y-(N_A-N_B-k)}_{m-1} \\
&= (\frac{1}{m})^{N_A}\sum_{k=0}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\sum_{y=N_B}^{mN_B}\binom{N_B}{y-N_B}_{m-1}\binom{N_A-N_B-k}{y-(N_A-N_B-k)}_{m-1}
\end{aligned}$$

By using identities of polynomial coefficients [42], we can transform the last term as:

$$\begin{aligned}
\sum_{y=N_B}^{mN_B}\binom{N_B}{y-N_B}_{m-1}\binom{N_A-N_B-k}{y-(N_A-N_B-k)}_{m-1} &= \sum_{y=N_B}^{mN_B}\binom{N_B}{(m-1)N_B-(y-N_B)}_{m-1}\binom{N_A-N_B-k}{y-(N_A-N_B-k)}_{m-1} && (\text{symmetry}) \\
&= \binom{[N_B]+[N_A-N_B-k]}{[(m-1)N_B-(y-N_B)]+[y-(N_A-N_B-k)]}_{m-1} && (\text{Vandermonde}) \\
&= \binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1}
\end{aligned}$$

Therefore, we have a closed-form expression for $E(m,N_A,N_B)$:

$$E(m,N_A,N_B) = (\frac{1}{m})^{N_A}\sum_{k=0}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1}$$

According to the definition of polynomial coefficients:

$$\binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1} \geq 1 \text{ if } 0 \leq (m+1)N_B-N_A+k \leq (m-1)(N_A-k) \Leftrightarrow N_A-(m+1)N_B \leq k \leq N_A-(1+\frac{1}{m})N_B \tag{2}$$

Therefore, we could obtain a lower bound for E:

$$\begin{aligned} E(m,N_A,N_B) &\geq (\frac{1}{m})^{N_A}\left\{\sum_{k=0}^{N_A-N_B}\binom{N_A-N_B}{k}m^k - \sum_{k=0}^{N_A-(m+1)N_B-1}\binom{N_A-N_B}{k}m^k - \sum_{k=N_A-(1+\frac{1}{m})N_B+1}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\right\} \\ &= (\frac{1}{m})^{N_A}\left\{(1+m)^{N_A-N_B} - H(N_A-(m+1)N_B-1;N_A-N_B,m) - T(N_A-(1+\frac{1}{m})N_B+1;N_A-N_B,m)\right\} \end{aligned}$$

According to Lemma B.3, we know that with

$$\frac{N_A-(m+1)N_B}{N_A-N_B} \leq \frac{m}{1+m} \quad \Longrightarrow \quad \frac{N_B}{N_A} \geq \frac{1}{m^2+m+1} \tag{3}$$

$$\frac{N_A-(1+\frac{1}{m})N_B}{N_A-N_B} \geq \frac{m}{1+m} \quad \Longrightarrow \quad \frac{N_B}{N_A} \leq \frac{m}{2m+1} \tag{4}$$

We have:

$$\begin{aligned} E(m,N_A,N_B) &\geq (\frac{1}{m})^{N_A}\left\{(1+m)^{N_A-N_B} - H(N_A-(m+1)N_B-1;N_A-N_B,m) - T(N_A-(1+\frac{1}{m})N_B+1;N_A-N_B,m)\right\} \\ &\geq (\frac{1}{m})^{N_A}\left\{(1+m)^{N_A-N_B} - (b_H)^{N_A-N_B} - (b_T)^{N_A-N_B}\right\} \qquad \text{with } b_H < (1+m), b_T < (1+m) \end{aligned}$$

Therefore, according to Lemma B.5, we can ignore the head and tail, and E has an exponential lower bound w.r.t N_A as long as the base of $(\frac{1}{m})^{N_A}(1+m)^{N_A-N_B}$ is larger than 1. Since:

$$(\frac{1}{m})^{N_A}(1+m)^{N_A-N_B} = \left[\frac{(1+m)^{1-\frac{N_B}{N_A}}}{m}\right]^{N_A}$$

to ensure base larger than 1, we need

$$\frac{(1+m)^{1-\frac{N_B}{N_A}}}{m} > 1 \quad \Longrightarrow \quad \frac{N_B}{N_A} < 1 - \frac{\ln(m)}{\ln(m+1)} \tag{5}$$

Combining constraints (3), (4) and (5), we show an exponential lower bound for $E(m,N_A,N_B)$ with the following $\frac{N_B}{N_A}$:

$$\frac{1}{m^2+m+1} \leq \frac{N_B}{N_A} < 1 - \frac{\ln(m)}{\ln(m+1)}$$

As proved in Lemma B.4, such a range is always valid for any $m \in Z_{>0}$. Therefore, we show that for $L(m) = \frac{1}{m^2+m+1}, U(m) = 1-\frac{\ln(m)}{\ln(m+1)}$, $E(m,N_A,N_B)$ has an exponential lower bound of $\left[\frac{(1+m)^{1-\frac{N_B}{N_A}}}{m}\right]^{N_A}$.

We believe one could achieve a much large exponential bounds and/or a wider range for feasible $\frac{N_B}{N_A}$ by carefully reexamining the two reductions that we made:

- Only the remaining N_A-N_B elements are considered for subsets that have the same sum. This is to simplify the problem so that we can treat it as a problem involving two independent arrays of length N_A-N_B and N_B. This, however, significantly underestimates the number of matched subsets, especially when $\frac{N_B}{N_A}$ is high.
- The reduction we made in (2): $\binom{N_A-k}{(m-1)N_B-N_A+k}_{m-1} \geq 1$ is obviously coarse grained. □

Lastly, we can prove the other side of the story: if $\frac{N_B}{N_A}$ is too large or too small, the expected number of subsets out of $N_A - N_B$ elements that have the same sum as the N_B elements *does not* grow exponentially w.r.t to N_A:

Theorem B.7. *Considering two independent arrays X and Y, consisting of $N_A - N_B$ and N_B iid random variables from $U\{1,m\}$, there exists $LL(m), UU(m)$ such that the expected number of subsets in X, whose sums are equal to the sum of Y, can not be exponential w.r.t N_A, if $\frac{N_B}{N_A} \leq LL(m)$ or $\frac{N_B}{N_A} \geq UU(m)$. Note that $0 < LL(m) \leq UU(m) < 1$, $\forall m \in Z_{>0}$*

Proof. Firstly, we show that such a $UU(m)$ exists. We start with the closed-form expression of $E(m,N_A,N_B)$ that is derived in the proof above.

$$\begin{aligned}
E(m,N_A,N_B) &= (\frac{1}{m})^{N_A} \sum_{k=N_A-(m+1)N_B}^{N_A-(1+\frac{1}{m})N_B} \binom{N_A-N_B}{k} m^k \binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1} \\
&\leq (\frac{1}{m})^{N_A} \sum_{k=0}^{N_A-(1+\frac{1}{m})N_B} \binom{N_A-N_B}{k} m^k \binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1} \\
&\leq (\frac{1}{m})^{N_A} \sum_{k=0}^{N_A-(1+\frac{1}{m})N_B} \binom{N_A-N_B}{k} m^k \binom{N_A}{(m+1)N_B-N_A+k}_{m-1} \qquad (\binom{n}{k}_m \geq \binom{n-\Delta}{k}_m \ \forall \Delta \in Z_{\geq 0})
\end{aligned}$$

As noted above $\binom{n}{k}_m$ is a non-increasing function of k for $\lceil \frac{mn}{2} \rceil \leq k \leq mn$ [42], therefore, if we have

$$(m+1)N_B - N_A \geq \lceil \frac{(m-1)N_A}{2} \rceil \quad \Longrightarrow \quad \frac{N_B}{N_A} > \frac{1}{2}$$

$\binom{N_A}{(m+1)N_B-N_A+k}_{m-1}$ decreases as k increases from 0 to $N_A - (1+\frac{1}{m})N_B$, thus:

$$\begin{aligned}
\binom{N_A}{(m+1)N_B-N_A+k}_{m-1} &\leq \binom{N_A}{(m+1)N_B-N_A}_{m-1} \qquad (\text{for } k \in \{0,..,N_A-(1+\frac{1}{m})N_B\}) \\
&= d^{N_A}(m,N_A,\frac{N_B}{N_A}) \qquad (\text{with } d(m,N_A,\frac{N_B}{N_A}) = \left[\binom{N_A}{(m+1)N_B-N_A}_{m-1}\right]^{\frac{1}{N_A}}) \qquad (6)
\end{aligned}$$

where:

$$d(m,N_A,\frac{N_B}{N_A}) = \left[\binom{N_A}{(m+1)N_B-N_A}_{m-1}\right]^{\frac{1}{N_A}} = \left[\binom{N_A}{[(m+1)\frac{N_B}{N_A}-1]N_A}_{m-1}\right]^{\frac{1}{N_A}}$$

We denote

$$f(m,\frac{N_B}{N_A}) = \lim_{N_A\to\infty} d(m,N_A,\frac{N_B}{N_A})$$

There are two properties of $g(m,\frac{N_B}{N_A})$ that we leverage here: (i) $1 \leq f(m,\frac{N_B}{N_A}) \leq m$, this is because $\binom{N_A}{[(m+1)\frac{N_B}{N_A}-1]N_A}_{m-1} < m^{N_A}$ and $\binom{N_A}{[(m+1)\frac{N_B}{N_A}-1]N_A}_{m-1}$ is a non-decreasing function w.r.t N_A; (ii) $f(m,\frac{N_B}{N_A})$ is a non-increasing function w.r.t $\frac{N_B}{N_A}$ for $\frac{1}{2} < \frac{N_B}{N_A} < \frac{m}{m+1}$, this is because $d(m,N_A,\frac{N_B}{N_A})$ is a decreasing function w.r.t $\frac{N_B}{N_A}$. Moreover, $f(m,\frac{m}{m+1}) = 1$ and $f(m,\frac{1}{2}) = m$. The later is because, as shown in [37], $\binom{n}{\frac{m}{2}n}_m \sim \frac{(m+1)^n}{\sqrt{2\pi n\frac{m(m+2)}{12}}}$ as $n \to \infty$, which indicates that $\lim_{n\to\infty}[\binom{n}{\frac{m}{2}n}_m]^{\frac{1}{n}} = m+1$. Therefore, $f(m,\frac{1}{2}) = \lim_{N_A\to\infty}[\binom{N_A}{\frac{m-1}{2}N_A}_{m-1}]^{\frac{1}{N_A}} = m$. With (6), we now have the following upper bound for $E(m,N_A,N_B)$ (to simply notation, we use d for $d(m,N_A,\frac{N_B}{N_A})$):

$$E(m,N_A,N_B) \leq (\frac{d}{m})^{N_A} \sum_{k=0}^{N_A-(1+\frac{1}{m})N_B} \binom{N_A-N_B}{k} m^k$$

Similar to Lemma B.3, we can then bound the term $\sum_{k=0}^{N_A-(1+\frac{1}{m})N_B}\binom{N_A-N_B}{k}m^k$ by using tail bounds of binomial distribution. Specifically, as mentioned in the proof of Lemma B.3:

$$\sum_{k=0}^{N_A-(1+\frac{1}{m})N_B}\binom{N_A-N_B}{k}m^k \leq \left\{(1+m)\exp[-D(\frac{N_A-(1+\frac{1}{m})N_B}{N_A-N_B}||\frac{m}{1+m})]\right\}^{N_A-N_B}$$
$$\text{if} \quad \frac{N_A-(1+\frac{1}{m})N_B}{N_A-N_B} \leq \frac{m}{1+m} \quad \Rightarrow \quad \frac{N_B}{N_A} \geq \frac{m}{2m+1} \tag{7}$$

where $D(a||p)$ is the relative entropy between a $Bernoulli(a)$ (a-coin) and a $Bernoulli(p)$ (p-coin): $D(a||p) = (a)\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}$. Here we denote

$$\begin{aligned} g(m,\frac{N_B}{N_A}) &= (1+m)\exp[-D(\frac{N_A-(1+\frac{1}{m})N_B}{N_A-N_B}||\frac{m}{1+m})] \\ &= (1+m)\exp[-D(\frac{1-(1+\frac{1}{m})\frac{N_B}{N_A}}{1-\frac{N_B}{N_A}}||\frac{m}{1+m}))] \end{aligned}$$

Note that $g(m,\frac{N_B}{N_A})$ is a decreasing function of $\frac{N_B}{N_A}$ for $\frac{N_B}{N_A} \geq \frac{m}{2m+1}$. Moreover, $\lim_{\frac{N_B}{N_A}\to\frac{m}{m+1}} g(m,\frac{N_B}{N_A}) \to 0$, and $g(m,\frac{1}{2}) > 1$ for $m \geq 2$. The later is because $g(m,\frac{1}{2}) = (1+m)\exp[-D(\frac{m-1}{m}||\frac{m}{1+m}))]$ is an increasing function with m, thus $g(m,\frac{1}{2}) \geq g(2,\frac{1}{2}) = 3\exp[-D(\frac{1}{2}||\frac{2}{3}))] > 1$.
With (7), we now have:

$$\begin{aligned} E(m,N_A,N_B) &\leq (\frac{d}{m})^{N_A}[g(m,\frac{N_B}{N_A})]^{N_A-N_B} \\ &= \left\{(\frac{d}{m})[g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}\right\}^{N_A} \end{aligned}$$

Considering the limit of the base:

$$\begin{aligned} &\lim_{N_A\to\infty}[\frac{d(m,N_A,\frac{N_B}{N_A})}{m}][g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}} \\ &= [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}\lim_{N_A\to\infty}[\frac{d(m,N_A,\frac{N_B}{N_A})}{m}] \\ &= [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}[\frac{f(m,\frac{N_B}{N_A})}{m}] \end{aligned}$$

With the properties of $f(m,\frac{N_B}{N_A})$ and $g(m,\frac{N_B}{N_A})$ that we mentioned, one could prove that there exists a ratio $R(m)$ such that:

$$h(m,\frac{N_B}{N_A}) = [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}[\frac{f(m,\frac{N_B}{N_A})}{m}] \leq 1 \quad \forall \quad \frac{N_B}{N_A} \geq R(m)$$
$$\text{where } h(m,R(m)) = 1$$

This means that $E(m,N_A,N_B)$ is not exponential w.r.t N_A for $\frac{N_A}{N_B} \geq R(m)$
We could prove that there is a unique $R(m)$ with $h(m,R(m)) = 1$, and $\frac{m}{2m+1} < R(m) < \frac{m}{m+1}$, with the following properties of $f(m,\frac{N_B}{N_A})$ and $g(m,\frac{N_B}{N_A})$: (i) both of them are decreasing functions w.r.t $\frac{N_B}{N_A}$; (ii) $1 \leq f(m,\frac{N_B}{N_A}) \leq m$, $f(m,\frac{m}{m+1}) = 1$ and $f(m,\frac{1}{2}) = m$; (iii) $g(m,\frac{1}{2}) > 1$, and $\lim_{\frac{N_B}{N_A}\to\frac{m}{m+1}} g(m,\frac{N_B}{N_A}) \to 0$.
Firstly, because $\frac{f(m,\frac{N_B}{N_A})}{m} \leq 1$, for $\frac{N_B}{N_A}$ such that $g(m,\frac{N_B}{N_A}) < 1 \Rightarrow [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}} < 1$, we have
$h(m,\frac{N_B}{N_A}) = [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}[\frac{f(m,\frac{N_B}{N_A})}{m}] < 1$. Moreover, since $\lim_{\frac{N_B}{N_A}\to\frac{m}{m+1}} g(m,\frac{N_B}{N_A}) \to 0$, we know that $R(m) < \frac{m}{m+1}$

Secondly, for $\frac{N_B}{N_A}$ such that $g(m,\frac{N_B}{N_A})>1$, we have $h(m,\frac{N_B}{N_A})$ as a decreasing function w.r.t $\frac{N_B}{N_A}$. Moreover, since $g(m,\frac{1}{2})>1$, $f(m,\frac{1}{2})=m$, we have $h(m,\frac{1}{2})=[g(m,\frac{1}{2})]^{1-\frac{1}{2}}[\frac{f(m,\frac{1}{2})}{m}]>1$. Therefore, we know that $R(m)>\frac{1}{2}$.

Therefore, we have proved a valid $UU(m)$: $\frac{1}{2}<UU(m)=R(m)<\frac{m}{m+1}$, where $R(m)$ is defined as:

$$
\begin{aligned}
R(m) &= \arg_{\frac{1}{2}<\frac{N_B}{N_A}<\frac{m}{m+1}} h(m,\frac{N_B}{N_A})=1 \\
\text{where } h(m,\frac{N_B}{N_A}) &= [g(m,\frac{N_B}{N_A})]^{1-\frac{N_B}{N_A}}[\frac{f(m,\frac{N_B}{N_A})}{m}] \\
f(m,\frac{N_B}{N_A}) &= \lim_{N_A\to\infty}\left[\binom{N_A}{[(m+1)\frac{N_B}{N_A}-1]N_A}_{m-1}\right]^{\frac{1}{N_A}} \\
g(m,\frac{N_B}{N_A}) &= (1+m)\exp[-D(\frac{1-(1+\frac{1}{m})\frac{N_B}{N_A}}{1-\frac{N_B}{N_A}}||\frac{m}{1+m})]
\end{aligned}
$$

and $E(m,N_A,N_B)$ is not exponential if $\frac{N_B}{N_A}\geq UU(m)$.
Next, we show a valid $LL(m)$ by using a similar approach. We use E to represent $E(m,N_A,N_B)$ in the follows.

$$
\begin{aligned}
E &\leq (\frac{1}{m})^{N_A}\sum_{k=N_A-(1+m)N_B}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\binom{N_A-k}{(m+1)N_B-N_A+k}_{m-1} \\
&\leq (\frac{1}{m})^{N_A}\sum_{k=N_A-(1+m)N_B}^{N_A-N_B}\binom{N_A-N_B}{k}m^k\binom{(m+1)N_B}{(m+1)N_B-N_A+k}_{m-1} && (\binom{n}{k}_m\geq\binom{n-\Delta}{k}_m\ \forall\Delta\in Z_{\geq 0}) \\
&< (\frac{1}{m})^{N_A}m^{(m+1)N_B}\sum_{k=N_A-(1+m)N_B}^{N_A-N_B}\binom{N_A-N_B}{k}m^k && (\binom{(m+1)N_B}{(m+1)N_B-N_A+k}_{m-1}<m^{(m+1)N_B}) \\
&\leq (\frac{1}{m})^{N_A}m^{(m+1)N_B}\left\{(1+m)\exp[-D(\frac{N_A-(1+m)N_B}{N_A-N_B}||\frac{m}{1+m})]\right\}^{N_A-N_B} && (\frac{N_A-(m+1)N_B}{N_A-N_B}\geq\frac{m}{1+m}\Rightarrow\frac{N_B}{N_A}\leq\frac{1}{m^2+m+1}) \\
&= \left\{m^{(m+1)\frac{N_B}{N_A}-1}\left\{(1+m)\exp[-D(\frac{1-(1+m)\frac{N_B}{N_A}}{1-\frac{N_B}{N_A}}||\frac{m}{1+m})]\right\}^{1-\frac{N_B}{N_A}}\right\}^{N_A}
\end{aligned}
$$

We could prove that for $l(m,\frac{N_B}{N_A})=m^{(m+1)\frac{N_B}{N_A}-1}\left\{(1+m)\exp[-D(\frac{1-(1+m)\frac{N_B}{N_A}}{1-\frac{N_B}{N_A}}||\frac{m}{1+m})]\right\}^{1-\frac{N_B}{N_A}}$, there exists a $0<S(m)<\frac{1}{m^2+m+1}$, which is the $LL(m)$ that we try to prove, that $l(m,\frac{N_B}{N_A})\leq 1$ for any $\frac{N_B}{N_A}\leq S(m)$. This is because (i) $l(m,\frac{N_B}{N_A})$ is an increasing function of $\frac{N_B}{N_A}$, and (ii) $l(m,\frac{1}{m^2+m+1})=m^{-\frac{m^2}{m^2+m+1}}(1+m)^{\frac{m^2+m}{m^2+m+1}}>(1+m)^{\frac{m}{m^2+m+1}}>1$ and $\lim_{\frac{N_B}{N_A}\to 0}l(m,\frac{N_B}{N_A})=0<1$.
Therefore, we have proved a valid $LL(m)$: $0<LL(m)=S(m)<\frac{1}{m^2+m+1}$, where $S(m)$ is defined as:

$$
\begin{aligned}
S(m) &= \arg_{0<\frac{N_B}{N_A}<\frac{1}{m^2+m+1}} l(m,\frac{N_B}{N_A})=1 \\
\text{where } l(m,\frac{N_B}{N_A}) &= m^{(m+1)\frac{N_B}{N_A}-1}\left\{(1+m)\exp[-D(\frac{1-(1+m)\frac{N_B}{N_A}}{1-\frac{N_B}{N_A}}||\frac{m}{1+m})]\right\}^{1-\frac{N_B}{N_A}}
\end{aligned}
$$

and $E(m,N_A,N_B)$ is not exponential if $\frac{N_B}{N_A}\leq LL(m)$. □