End-to-End Secure Messaging with Traceability *Only* for Illegal Content

James Bartusek[1], Sanjam Garg[1,3], Abhishek Jain[2], and Guru-Vamsi Policharla[1]

[1] University of California, Berkeley
[2] Johns Hopkins University
[3] NTT Research

Abstract. As end-to-end encrypted messaging services become widely adopted, law enforcement agencies have increasingly expressed concern that such services interfere with their ability to maintain public safety. Indeed, there is a direct tension between preserving user privacy and enabling content moderation on these platforms. Recent research has begun to address this tension, proposing systems that purport to strike a balance between the privacy of "honest" users and traceability of "malicious" users. Unfortunately, these systems suffer from a lack of protection against malicious or coerced service providers. In this work, we address the privacy vs. content moderation question through the lens of pre-constrained cryptography [Ananth et al., ITCS 2022]. We introduce the notion of *set pre-constrained* (SPC) *group signatures* that guarantees security against *malicious key generators*. SPC group signatures offer the ability to trace users in messaging systems who originate pre-defined illegal content (such as child sexual abuse material), while providing security against malicious service providers. We construct concretely efficient protocols for SPC group signatures, and demonstrate the real-world feasibility of our approach via an implementation. The starting point for our solution is the recently introduced Apple PSI system, which we significantly modify to improve security and expand functionality.

1 Introduction

End-to-end encrypted services offer users the ability to communicate information, with the guarantee that even the service provider itself cannot access the raw information that it is storing or transmitting. Billions of people worldwide are now using end-to-end encrypted systems such as WhatsApp and Signal.

However, the strong data privacy guarantees offered by end-to-end encryption (E2EE) technology have not been universally celebrated. Law enforcement and national security agencies have argued that such services interfere with their ability to prosecute criminals and maintain public safety [19,30]. In particular, E2EE appears to directly conflict with the goals of *content moderation*, which refers to the ability to screen, monitor, or trace the origin of user-generated content.

One prominent example of the use of content moderation is in fighting the proliferation of child sexual abuse material, or CSAM. In the United States, the proposed EARN IT act [28] would enable legal action to be taken against internet service providers that fail to remove CSAM material from their service. It has been argued that the proposed legislation would inhibit the use of E2EE, which prevents service providers from detecting in the first place if they are hosting or transmitting CSAM [37]. In fact, a 2019 open letter to Facebook signed by then U.S. Attorney General William Barr along with international partners explicitly requested that Facebook not proceed with its planned implementation of E2EE, due to its tension with CSAM detection [29].

One can imagine that this "encryption debate" polarizes to two conceivable outcomes: a world with E2EE but without any content moderation, or a world without E2EE but with content moderation. Since neither of these outcomes seems to be truly satisfactory, it becomes vital to explore the space in between, or more fundamentally, to identify if any such space even exists. Indeed, the past few years have seen researchers paying increased attention to this very question, as covered for example by a recent report [31] released by the Center for Democracy and Technology, a technical report on the risks of client side scanning [2] and a recent talk about the question of CSAM detection vs. E2EE given at Real World Crypto 2022 [35].

In this work, we explore the viability of using cryptographic techniques to balance the need for both user privacy and illegal content moderation in messaging systems. Along the way, we also study content moderation in the context of encryption systems used by cloud service providers. This might be of independent interest.

Prior solutions. In the setting of encrypted messaging systems, the principle goal of illegal content moderation is to identify the existence of illegal content in the system and uncover the identity of the originator of such content. The desirable privacy goals are to (i) hide the messages exchanged in the system, even from the server, and (ii) preserve the anonymity of the originator of any *harmless* content that is forwarded through the system. Note that this latter property is crucial in many real-world scenarios, e.g., whistleblowers may desire to use the protection provided by E2EE without the threat of being de-anonymized. A recent proposal [32] in this direction fails to adequately balance these goals, allowing a malicious server to de-anonymize *any* user, thereby completely violating the fundamental guarantee of E2EE.

We also note that some recent works have attempted to address the fundamentally different but related question of content moderation for *misinformation*, and we refer to Section 1.3 and Section 1.4 for discussion on this.

The problem. The main problem with existing proposals (including the "traceback" systems for addressing misinformation that we discuss later) is that they suffer from a glaring lack of protection against a server who wishes to use the system beyond its prescribed functionality. This is a serious problem, not only because the server itself might have malicious intent, but also because of the threat of coercion from powerful actors that may want to use the technology for surveillance or censorship.

This lack of built-in protection fundamentally damages the transparency of E2EE, reducing the incentive for users to adopt the systems for their communication. While these works have indeed tried to strike a balance between privacy and content moderation, we believe that, for the deterrence of *pre-defined*,[4] *illegal* content (such as CSAM), they have over-compromised on privacy. In this work, we seek to build systems that offer similar content tracing functionality, while offering greater transparency and rigorous cryptographic guarantees about the possible scope of server behavior.

1.1 Summary of Our Contributions

We present novel definitions and efficient protocols for illegal content moderation in the setting of encrypted messaging.

Set Pre-Constrained (SPC) Group Signatures. We propose a new notion of *set pre-constrained group signatures* which can be implemented in an end-to-end secure messaging application. This allows tracing users who send illegal content while ensuring privacy for everyone else.

– *Definition*: In SPC group signatures, a database D (of illegal content) can be encoded within the group's public key. The key requirement is that the signer of any message $m \in D$ can be de-anonymized by the group manager but signers of messages $m \notin D$ remain anonymous *even to the group manager*. Our definitions model *malicious* group managers and ensure that the group's public key encodes a database D that is authorized by a third-party such as the US National Center for Missing and Exploited Children (or more generally, multiple third parties). Furthermore, the public key is *publicly-verifiable*, so all clients in the system can verify for themselves (without knowing D) whether the group manager's public key encodes an acceptable D.[5]
– *Construction*: We provide a *concretely efficient* construction of SPC group signatures based on standard bilinear map assumptions, in the Random Oracle model. In this construction, we allow the group manager's public key to grow with the size of D. Crucially, however, the running time of the signing algorithm (with oracle access to the public key) as well as verification and tracing is *independent* of the size of D.

[4] By pre-defined, we mean any content that has been classified as "illegal", for example by a governmental body, *before* the parameters of the cloud storage or messaging system are sampled. Updating parameters to include new content classified as illegal is an interesting question in this context, which we discuss further in Section 1.3.

[5] In the body, we generalize our definition to consider general *functionalities* F as opposed to just the set-membership function specified by D. However, all of our constructions in this work target the special case of sets D, and we restrict our attention to such functionalities in the overview.

SPC Encryption. Along the way to constructing SPC group signatures, we define and construct efficient set pre-constrained (SPC) *encryption* schemes. Our construction builds and improves upon the recent Apple PSI protocol [10]: (1) We identify a gap in their proof of security against a malicious server and show how to efficiently build on top of their protocol in order to close this gap. (2) Further, we augment their construction to achieve a stronger notion of security that provides guarantees on the integrity of the database embedded in the public key (analogous to SPC group signatures).

Our SPC encryption scheme has public keys of size linear in the database D and constant encryption and decryption times. We demonstrate that this asymptotic efficiency trade-off is likely the "best-possible" in that further improvements would imply the elusive notion of doubly-efficient private information retrieval [12,11], which is not known to exist under standard cryptographic assumptions.

Evaluation. We implement our SPC group signature scheme and provide benchmarks in the full version [8]. We find that signing and verification take tens of milliseconds, and signature size is in the order of a few kilobytes[6]. When instantiated over the BN254 curve, the communication overhead for typical image sizes of 400 KB is under 1% and the additional computation incurs a $\sim 15\%$ overhead on top of message delivery time. We view these results as strong initial evidence that illegal content moderation in E2EE messaging systems – with security against malicious servers – can indeed be performed in the real world. While our current focus is on illegal content moderation, we believe that the efficiency properties of our SPC group signature and encryption schemes make them attractive tools for other applications that involve membership testing against a private "blocklist". Examples include privacy-preserving DNS blocklisting [25] where the blocklist could be proprietary, and anonymous credential systems where it is desirable to hide revocation attributes.

1.2 Our Approach

In this work, we aim to build a messaging system that satisfies, at the very least, the following set of requirements.

1. The system is end-to-end encrypted. In particular, the server cannot learn anything at all about the content transmitted in the system unless it receives some side information from a user participating in the system.
2. The originator of any piece of content remains anonymous to any user that receives the forwarded content.
3. If a user receives some illegal content, they can report it to the server, who can then determine the identity of the user who originated the content. This holds even if the content has been forwarded an arbitrary number of times before being reported.

[6] More precisely, for the BN254 curve, this translates to 3.5 Kilobytes per SPC group signature.

4. The originator of any *harmless* content remains anonymous, *even from the perspective of the server* who may receive a report about the content.

Naïve Approaches. To demonstrate the challenges in realizing all four properties, we first consider some existing approaches.

As a first attempt, we could try simply using end-to-end encryption. While this may satisfy the first two properties, it clearly does the support the third constraint, which we refer to as *traceability*.

A natural next attempt would be to use a *group signature* scheme [15,9] underneath E2EE in order to recover this property of traceability. In a group signature scheme, there is a group manager that generates a master public key mpk and a master secret key msk. A new client enters the system by interacting with the group manager in order to receive a client-specific secret key sk. Any client can use their sk to produce a signature σ on a message m, which can be verified by anyone that knows mpk. On the one hand, the identity of the signer remains anonymous from anyone that knows σ but not msk. On the other hand, knowing msk allows the group manager to determine which client produced σ. Thus, we can satisfy the first three goals above by having the messaging service provider additionally take on the role of the group manager. Each user in the system would then obtain a signing key sk from the server, and then attach a signature to any piece of content that they send (where the signature is also transmitted under the encryption). Unfortunately, this solution does not prevent the server from colluding with a user to identify the originator of *any* piece of content received by that user. That is, this solution appears to be fundamentally at odds with the crucial fourth requirement, or *anonymity*, stated above.

Despite some prior attempts at recovering a notion of anonymity in group signature (see Section 1.3 from some more discussion), we conclude that existing frameworks are insufficient for capturing the security that we demand. In order to address this issue, we must somehow *constrain* the ability of the group manager to de-anonymize anyone in the system.

SPC group signatures: Definitions. This motivates our first contribution, which is the definition of a *set pre-constrained group signature*, or SPC group signature. In this primitive, the group manager's master public key will be computed with respect to some set D of *illegal content* (which should remain hidden from clients even given the master public key). The novel security property we desire is that the anonymity of a client who produces a signature on some message $m \notin D$ remains intact, *even from the perspective of the group manager.*

More concretely, we ask for the following (informally stated) set of security properties.

- *Traceability*: the identity of a client who signs a message $m \in D$ should be recoverable given the signature and the master secret key.
- *Client-server anonymity*: the identity of a client who signs a message $m \notin D$ should be hidden, even given the master secret key.

– *Set-hiding:* the master public key should not reveal the set D.[7]
– *Unframeability:* no party, not even the master secret key holder, should be able to produce a signature that can be attributed to an honest client.
– *Client-client anonymity*: the identity of a client who signs any message m should be hidden from the perspective of any party who does not have the master secret key.

At this point, we must stop to consider the meaningfulness of the above security definitions as stated. In particular: who decides D? Clearly, if D is set to be the whole universe of messages, then this is no more secure than a standard group signature. And if an adversarial group manager is trying to break the client-server anonymity of the above scheme, what is preventing them from generating their master public key with respect to this "trivial" set D?

In order to constrain D in a meaningful way, we introduce a predicate $\mathcal{P}$ into the definition of client-server anonymity. The description of $\mathcal{P}$ will be fixed at setup time along with some public parameters pp known to everybody in the system and secret parameters sp known only to the group manager (we will discuss below the reason we include secret parameters). We will model client-server anonymity using an ideal functionality $\mathcal{F}_{\mathsf{anon}}$ that takes a set of items D as input from the group manager and a sequence of pairs of identities and messages $(\mathsf{pk}_1, m_1), \ldots, (\mathsf{pk}_k, m_k)$ from the client (who represents all clients in the system). If $\mathcal{P}(\mathsf{pp}, \mathsf{sp}, D) = 0$, the functionality aborts, and otherwise it delivers $\{m_i\}_{i \in [k]}, \{\mathsf{pk}_i\}_{i : m_i \in D}$ to the group manager.

This gives us a generic framework for specifying how to constrain the possible D used by the group manager. In particular, we are able to delegate the responsibility of constraining D to a *third-party* (e.g. the National Center for Missing and Exploited Children, or NCMEC), who is tasked with setting up the parameters $(\mathsf{pp}, \mathsf{sp})$ for the predicate $\mathcal{P}$. That is, we can gracefully split the responsibility of implementing / maintaining the encrypted messaging system (by e.g. WhatsApp) and the responsibility of specifying what constitutes illegal content (by e.g. NCMEC or a collection of such agencies).

Perhaps the most natural example of $\mathcal{P}$ is the "subset" predicate, which is parameterized by a set D^* of "allowed" messages (e.g. the entire database of illegal content as defined by NCMEC), and accepts only if $D \subseteq D^*$. In this case, since D^* itself represents illegal content, we do not want to make it public. Thus, we set $\mathsf{sp} = D^*$, and $\mathsf{pp} = |D^*|$. We refer to security with respect to this subset predicate as *authenticated-set security*.

In our full definition, we explicitly consider the third-party Auth as a participant in the system, who begins by setting up a pp and sp of their choice. Then, we require security against an adversary that corrupts *either* the client (and thus cannot learn anything about D), the group manager (and thus can only learn $\{\mathsf{pk}_i\}_{i:m_i \in D}$ for some "valid" D), or the third-party Auth (and thus cannot learn anything about *any* of the identities pk_i). Note that security is only vacuous if the adversary manages to corrupt both the group manager and Auth

[7] Note that if we want to prevent even the group manager from seeing / storing the illegal content, we can set D to be hashes of the content itself.

at the very beginning of the protocol, and thus is able to set sp and D as it wishes. While this seems like a potential limitation, our framework is general enough to support a *de-centralized* Auth. That is, we could consider many third-parties $\mathsf{Auth}_1, \ldots, \mathsf{Auth}_\ell$ who each specify a database D_i^*, and set $\mathcal{P}$ to accept D only if (for example) $D \subseteq D_1^* \cap \cdots \cap D_\ell^*$. Thus, in order to compromise the system, an adversary would have to corrupt the group manager and *all* third-party authorities *simultaneously*, while the key generation procedure is occurring.

SPC group signatures: Construction. We next investigate the feasibility and efficiency of constructing SPC group signatures. To do so, we abstract out the basic "pre-constraining" property we need from the group signature scheme, and re-state it in the context of an *encryption scheme*.

That is, we first define a scheme for what we call *set pre-constrained encryption*, or SPC encryption, with the following properties.

- The public key pk is generated with respect to some database D of items.
- The public key pk should not reveal D, since D may consist of sensitive or harmful content.
- Any user, given pk, can encrypt a message m with respect to an item x such that the key generator (using sk) can recover m if $x \in D$, but learns *nothing* about m if $x \notin D$.

We note that our terminology is inspired by the recent work of Ananth et al. [4] who proposed the notion of pre-constrained encryption. However, our definitions and constructions are quite different; see Section 1.4 for further discussion.

Our security definition for set pre-constrained encryption mirrors the anonymity definition explained above, where the key generator for the encryption scheme now plays of role of the group generator. Specifically, we can still parameterize security by a predicate $\mathcal{P}$ and parameters $(\mathsf{pp}, \mathsf{sp})$ set up by a third-party Auth.

Now, we describe a generic construction of an SPC group signature scheme from an SPC encryption scheme plus standard crytographic tools: a one-way function F, a digital signature scheme, and a zero-knowledge non-interactive argument of knowledge.

The group manager will take as input some set D and sample a public key for the SPC encryption scheme computed with respect to D. It will also include a verification key for the signature scheme in its master public key. A client can join the system by sampling a secret s, setting $\mathsf{id} = F(s)$ to be their public identity, and obtaining a signature on id from the group manager. Now, to sign a message m, the client first encrypts their identity id with respect to item m using the SPC encryption scheme, producing a ciphertext ct. Then, they produce a zero-knowledge proof π that

"I know some id, a signature on id, and s such that $\mathsf{id} = F(s)$, such that ct is an SPC encryption of id with respect to m"

Observe that given any valid signature (ct, π) on a message $m \in D$, the group manager should be able to recover the id that produced (ct, π) by decrypting ct.

We refer to this property as *traceability*. One subtle issue that emerges here is that π can only attest that ct is in the space of valid ciphertexts encrypting id under item m, and cannot show that ct was *sampled* correctly. Thus, we will need to require that the SPC encryption is *perfectly correct*, that is, ct is perfectly binding to id when $m \in D$.

Next, we see that any signature (ct, π) on a message m hides id from any other client, which gives us the client-client anonymity property. More specific to our case, we can also show that any signature (ct, π) on a message $m \notin D$ hides id, *even from the server*, which we capture using our simulation-based security definition.

Finally, we highlight the notion of *unframeability*, which requires that a malicious server cannot produce a signature (ct, π) that can be opened to the id of any honest client. Intuitively, this follows because the server will not know the pre-image s of id, and so cannot produce a valid proof π.

SPC Encryption: Construction. With this generic compiler in hand, we provide a *concretely efficient* construction of SPC encryption, and then a *concretely efficient* instantiation of the generic compiler described above. This results in a *practical* proposal for SPC group signatures, which is our main constructive result.

Our construction of SPC encryption builds on top of the Apple PSI protocol [10]. This protocol already satisfies the basic syntax that we require, namely, the ability to embed a set D in the public key pk of an encryption scheme. However, their security notion is much weaker than the authenticated-set security we desire, and described above. Nevertheless, we can capture the security they do claim to achieve using our generic framework, and we refer to it as *bounded-set security*. In more detail, in their scheme, the key generator is completely free to choose the set D, as long as the size of D is below some public bound n. That is, $\mathsf{pp} = n$, sp is empty, and $\mathcal{P}(n, , D) = 1$ if $|D| \leq n$.

Building on their basic scheme, we provide three new contributions.

- We observe that the proof of security (for bounded-set security) given in the Apple PSI paper [10] only holds when the bound n is large enough with respect to other system parameters. This results in a large gap between correctness (the number of items that an honest server programs into its public key) and security (the number of items that a malicious server can potentially program into its public key). We show how to remedy this in a concretely efficient manner, completely closing this gap and achieving essentially no difference between the correctness and security bounds.
- We build on top of the protocol in a different manner in order to establish an efficient protocol that satisfies our novel (and much stronger) definition of *authenticated-set security*.
- We show how to tweak these schemes in order to obtain the *perfect correctness* guarantee needed to make our compiler from SPC encryption to SPC group signatures work. Interestingly, we lose an "element-hiding" property of the scheme in this process. Luckily, we don't require this property for our compiler, since elements correspond to messages in the SPC group signature

scheme, which we are not worried about leaking to the server in the event of a user report.

An in-depth overview of the Apple PSI protocol and the technical ideas involved in our improved constructions are given in Section 3.1.

Finally, we derive a *concretely efficient* instantiation of the SPC encryption to SPC group signature compiler, which makes use of structure-preserving signatures [1] and the Groth-Sahai proof system [24]. We provide an overview of the technical ideas involved in our constructions in Section 4.3. We also implement the resulting SPC group signature scheme and provide further discussion and benchmarking in the full version [8].

SPC Encryption: Limitations. As a separate contribution, we investigate generic *asymptotic* efficiency properties of SPC encryption. We identify three desirable "succinctness" properties with respect to the database size n: succinct public-key size, succinct encryption time, and succinct decryption time, where in each case, succinctness refers to poly-logarithmic complexity in n. The Apple-PSI-based protocols have non-succinct public-key size, but succinct encryption and succinct decryption. A natural question is whether it is also possible to achieve *succinct* public key. We observe the following, and provide more details in the full version.

- There are techniques in the literature [3] that can achieve succinct public key and succinct encryption with either (i) non-succinct decryption with element-hiding, or (ii) succinct decryption without element-hiding, from standard cryptographic assumptions. However, these constructions are impractical and not suitable for real-world deployment.
- An "optimal" SPC encryption scheme with succinct public key, succinct encryption, succinct decryption, and element-hiding implies the elusive notion of *doubly-efficient private-information retrieval* [12,11], which is not known to exist under any standard cryptographic assumption.

Thus, while the Apple PSI paper is not explicit about why they settled for a protocol with a non-succinct public key, our analysis validates this choice.

1.3 Discussion

CSAM deterrence vs. misinformation. As mentioned above, CSAM deterrence and combating misinformation are two of the most prominent applications of online content moderation. While both applications indeed fall under the umbrella of content moderation, they each introduce unique challenges from a cryptosystem perspective. The pre-constraining techniques that we make use of in this paper are designed specifically for the deterrence of illegal content, such as CSAM. On the other hand, the "traceback" systems introduced in prior works such as [43,34,38] are arguably geared more towards the application of combating misinformation.

Perhaps the biggest distinction between these applications from a cryptographic perspective is their amenability to *pre-definition*. As already discussed, illegal content must be pre-defined in some sense, for example by a governmental body. It is crucial to take advantage of this pre-definition in designing cryptosystems for illegal content deterrence. Indeed, since the description of the illegal content itself can be baked into the parameters of the system, we can hope to obtain rigorous guarantees about *which* content is being tracked and monitored by the system administrator.

On the other hand, it is not even clear in the first place how to define misinformation, or even who has the authority to define it. Plus, new content that could potentially be classified as misinformation is constantly being created and distributed. Thus, it is less clear how to obtain rigorous security guarantees against potentially malicious servers in the setting of misinformation deterrence. A potential approach could be to allow new content (such as new misinformation or abuse) to be added to the "constrained" set, so that the originators of prior messages containing this content could be traced. This feature is reminiscent of "retrospective" access to encrypted data as considered in [22] in a somewhat different context. They show that such access requires the use of powerful (and currently very inefficient) cryptographic tools, and it would be interesting to see if the same implications hold in the setting of tracing in end-to-end encrypted messaging systems.

Deniability vs. unframeability. Another difference between illegal content and misinformation from a cryptographic perspective is reflected in the technical tension between the notions of *deniability* and *unframeability*. Deniability essentially asks that messages between users can be simulated without any user-specific secrets, where indistinguishability from real messages holds from the perspective of an entity with *full information*, including user and even server secrets. This can certainly be a desirable property of encrypted messaging systems, especially when there is a threat of coercion from powerful outside sources. However, this property conflicts with *unframeability* against malicious servers, since it enables servers to produce these simulated messages [42]. While deniability has been a sought-after feature of encrypted systems with traceback functionality [38], it actually appears to be *counter-productive* in systems that are meant to detect originators of CSAM or other illegal content. Indeed, it is important that not only can the server identify the originator, but also that the server can convince law enforcement of the identity of the content originator. On the other hand, we view unframeability against malicious servers as a crucial property of CSAM deterrence systems, since users can face dramatic consequences if framed for the generation or dissemination of illegal content. Thus, our techniques are tailored to obtain the strongest notion of unframeability and no deniability,[8] while prior work [38] that focused on combating misinformation took the opposite approach.

[8] Though we note that one could potentially alter our group signature scheme to obtain deniability at the cost of unframeability, by including in the zero-knowledge argument a clause along the lines of "OR I know the master secret key".

On security against malicious servers. In this work, we took steps towards ensuring privacy and anonymity against malicious (or even honest-but-curious) servers in encrypted systems with support for content moderation. As mentioned earlier in the introduction, it is absolutely vital to explore the space of solutions to the "encrytion debate" that don't give up fully on either end-to-end encryption or content moderation. There is much more work to be done in this space, and we view our techniques as one tool in an ever-expanding toolbox of techniques meant to address the broad question of privacy vs. content moderation.

In particular, while we remove the need to trust *service providers* (think, WhatsApp), the notion of authenticated-set security essentially moves this trust to a third party (think NCMEC). We consider this progress, since it splits the responsibility of providing a messaging service and defining illegal content. Moreover, as discussed earlier, our scheme would immediately extend to support *multiple* third parties that can each attest to the validity of the server's public parameters, further splitting the trust. However, we acknowledge that there is opportunity to further improve the transparency and trust in such content moderation systems.

Additional challenges and future directions. We conclude our discussion with a few directions for future work. First, a desirable property of encrypted illegal content moderation systems is the ability to *update* public parameters to include new illegal content. As discussed in the Apple PSI paper [10], a simple way to handle updates is to redo setup and release the updated public key as part of system update. Achieving more efficient updates, however, is an interesting direction for future work. For example, if an update only corresponds to locations that are changed, it may start leaking the positions that correspond to database elements. This suggests the need for creative solutions, for example the use of differential privacy techniques to hide this leakage.

Next, we did not consider *thresholding* in this work, which would protect the privacy of content or anonymity of users until *multiple matches* were found in the database. While this is straightforward to incorporate into SPC encryption, it is not as immediate for SPC group signatures, at least if the goal is to maintain concrete efficiency. We leave an exploration of this to future work.

Next, we chose to use Groth-Sahai proof systems in order to demonstrate that SPC group signatures could be constructed with reasonable efficiency. However, there are other tools available, such as efficient SNARGs (succinct non-interactive arguments) that may result in better verification time at the cost of increased signer work. We leave further investigation of this to future work.

Finally, we mention broader considerations that would come with using our system in the real world. In the system, the actual database D would likely not consist of the actual CSAM images themselves, but rather *hashes* of CSAM images computed using a perceptual hash function, such as Apple's NeuralHash [5]. This introduces the possibility of adversarial use of the hash function, for example targeted collision-finding. We view this as an important attack vector to consider, especially when using these hash functions in conjunction with cryptographic protocols meant to provide privacy against malicious servers. Exploration of this topic is outside the scope of the current work, and we refer the reader to [40] and references therein for current research on the topic.

1.4 Related Work

Pre-Constrained Cryptography. Our work borrows the terminology of pre-constrained cryptography from Ananth et al. [4] because of sharing a similar vision – that of putting pre-specified restrictions on the key generation authority. Our definitions and constructions, however, are different from [4]. First, we note that the notion of (set) pre-constrained group signatures is *new* to our work, while Ananth et al. [4] only focus on (pre-constrained) encryption systems. In the setting of pre-constrained encryption, the notion of malicious security in [4] is weaker than ours and allows the authority to choose *any* "constraint" from a class of constraints. This weaker notion is not meaningful in our setting, as it allows the service provider (think, WhatsApp) to use an *arbitrary set* of their choice. Ananth et al. propose constructions for different flavors of pre-constrained encryption; the one that comes closest to our setting relies on indistinguishability obfuscation [7], and is presently only of theoretical interest. In contrast, we provide concretely efficient constructions for our setting.

Traceback Systems. While our work focuses on moderation for *pre-defined illegal content*, there has also been much recent work on the adjacent question of moderation for *misinformation* or *abusive* content. Solutions for this problem typically build "traceback" mechanisms into end-to-end encrypted systems [43,34,38,27], extending the reach of so-called "message franking" systems [26,17,42]. These solutions rely on user reporting to identify the existence of harmful content. Once a report is received by the server, the server and reporting user can work together to identify the originator of the harmful message. Unfortunately, these systems suffer from various drawbacks [21]: (1) They allow a colluding server and users to de-anonymize the originator of *any* message, even if the content is harmless. (2) Initial solutions in this space additionally require the help of users on the traceback path to identify the originator, and do not maintain their anonymity. While the latter drawback was addressed in the recent work of [38], no known solution provides security guarantees against malicious servers. Our system addresses both of these shortcomings, for our specific setting of illegal content moderation.

Group Signatures. Finally, we mention a related line of work on group signatures with *message-dependent opening* (GS-MDO) [18,33]. Here, trust is split between the group manager and an additional entity called the "admitter". The identity of a group member that produces a signature on a message m can be revealed only if the group manager and admitter combine their private information. Unlike SPC group signatures, GS-MDO does *not* require any "commitment" to, or "pre-constraining" of, the set of messages that can be de-anonymized. This means that even after the system parameters are set up, the group manager and admitter can in principle work together to de-anonymize *every* signature while still acting "semi-honestly" w.r.t. the protocol specification. In particular, clients of the system will not have the peace of mind guaranteed by public parameters that are publicly "authenticated" to only allow de-anonymization of a particular set of illegal content specificied by some trusted (collection of) third party(ies).

2 Preliminaries

The security parameter is denoted by $\lambda \in \mathbb{N}$. A function $f : \mathbb{N} \to \mathbb{N}$ is said to be polynomial if there exists a constant c such that $f(n) \leq n^c$ for all $n \in \mathbb{N}$, and we write $\mathsf{poly}(\cdot)$ to denote such a function. A function $f : \mathbb{N} \to [0,1]$ is said to be negligible if for every $c \in \mathbb{N}$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $f(n) < n^{-c}$, and we write $\mathsf{negl}(\cdot)$ to denote such a function. A probability is *noticeable* if it is not negligible, and *overwhelming* if it is equal to $1 - \mathsf{negl}(\lambda)$ for some negligible function $\mathsf{negl}(\lambda)$. For a set $\mathcal{S}$, we write $s \leftarrow \mathcal{S}$ to indicate that s is sampled uniformly at random from $\mathcal{S}$. For a random variable $\mathcal{D}$, we write $d \leftarrow \mathcal{D}$ to indicate that d is sampled according to $\mathcal{D}$. An algorithm $\mathcal{A}$ is PPT (probabilistic polynomial-time) if its running time is bounded by some polynomial in the size of its input. For two ensembles of random variables $\{\mathcal{D}_{0,\lambda}\}_{\lambda \in \mathbb{N}}$, $\{\mathcal{D}_{1,\lambda}\}_{\lambda \in \mathbb{N}}$, we write $\mathcal{D}_0 \approx_c \mathcal{D}_1$ to indicate that for all PPT $\mathcal{A}$, it holds that

$$\big| \Pr_{d \leftarrow \mathcal{D}_{0,\lambda}}[\mathcal{A}(d) = 1] - \Pr_{d \leftarrow \mathcal{D}_{1,\lambda}}[\mathcal{A}(d) = 1] \big| \leq \frac{1}{2} + \mathsf{negl}(\lambda).$$

2.1 Basic cryptographic primitives and assumptions

We will use a standard symmetric-key encryption scheme $(\mathsf{Enc}, \mathsf{Dec})$ with key space $\mathcal{K}$ that satisfies *random key robustness*, which states that for any message m, $\Pr_{k,k' \leftarrow \mathcal{K}}[\mathsf{Dec}(k', \mathsf{Enc}(k, m)) = \bot] = 1 - \mathsf{negl}(\lambda)$. We will also make use of a standard digital signature scheme $(\mathsf{Gen}, \mathsf{Sign}, \mathsf{Verify})$ that is *existentially unforgeable under chosen message attacks* (EUF-CMA).

2.2 Non-interactive arguments of knowledge

Let $\mathcal{L}$ be an NP language and let $\mathcal{R}$ be the associated binary relation, where a statement $x \in \mathcal{L}$ if and only if there exists a witness w such that $(x, w) \in \mathcal{R}$. A non-interactive argument system for $\mathcal{R}$ consists of algorithms $\mathsf{Setup}, \mathsf{Prove}, \mathsf{Verify}$, where $\mathsf{Setup}(1^\lambda)$ outputs a string crs, $\mathsf{Prove}(\mathsf{crs}, x, w)$ outputs a proof π, and $\mathsf{Verify}(\mathsf{crs}, x, \pi)$ outputs either 1 to indicate accept or 0 to indicate reject. We say that a non-interactive argument system for a relation $\mathcal{R}$ that satisfies the standard notions of *completeness*, *knowledge extraction*, and *zero-knowledge*, is a *zero-knowledge non-interactive argument of knowledge* (ZK-NIAoK) for $\mathcal{R}$. We will use the fact that the following relations all have highly efficient ZK-NIAoKs in the ROM. Let $\mathbb{G}$ be a group of order q with generator g.

– The relation $\mathcal{R}_{\mathsf{DLog}} = \{((g, h), \alpha) : h = g^\alpha\}$. A ZK-NIAoK for $\mathcal{R}_{\mathsf{DLog}}$ follows from applying the Fiat-Shamir heuristic [20] to Schnorr's sigma protocol [41].

– The relation $\mathcal{R}_{\mathsf{DH}} = \{((g, h_1, h_2, h_3), \alpha) : (h_1 = g^{\alpha}) \wedge (h_3 = h_2^{\alpha})\}$. A ZK-NIAoK for $\mathcal{R}_{\mathsf{DH}}$ follows from applying the Fiat-Shamir heuristic to Chaum and Pederson's sigma protocol [14].
– For any n and $k \leq n$, the relation $\mathcal{R}_{\mathsf{DLog}_n^k} = \{((g, h_1, \ldots, h_n), (S, \{\alpha_i\}_{i \in S})) : (|S| = k) \wedge (\forall i \in S, h_i = g^{\alpha_i})\}$. A ZK-NIAoK for $\mathcal{R}_{\mathsf{DLog}_n^k}$ follows from applying the Fiat-Shamir heuristic to the protocol of [16]. Moreover, an efficient *succinct* argument system for this language whose size is logarithmic in n, was shown recently by [6].

2.3 Groth-Sahai Proofs

Let $\mathcal{G}$ be a bilinear group generator that on input 1^{λ} returns $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{T}, e, g_1, g_2)$, where $\mathbb{G}_1, \mathbb{G}_2, \mathbb{T}$ are groups of order p, where p is a λ-bit prime. g_1 is a generator of $\mathbb{G}_1$, g_2 is a generator of $\mathbb{G}_2$, and e is a non-degenerate bilinear map. That is, $e(g, g)$ is a generator of $\mathbb{T}$, and for all $a, b \in \mathbb{Z}_p$, it holds that $e(g_1^a, g_2^b) = e(g_1, g_2)^{ab}$. The DDH assumption is assumed to hold in each of $\mathbb{G}_1$ and $\mathbb{G}_2$. In other words, the SXDH (symmetric external Diffie-Hellman) assumption is assumed to hold.

Groth and Sahai [24] constructed efficient non-interactive zero-knowledge proof systems for statements that involve equations over bilinear maps. "GS proofs" can prove certain statements that consist of the equations over variables $X_1, \ldots, X_m \in \mathbb{G}_1, Y_1, \ldots, Y_n \in \mathbb{G}_2, x_1, \ldots,$ and $x_{m'}, y_1, \ldots, y_{n'} \in \mathbb{Z}_p$. Although, the GS proof system can handle many types of equation, we restrict our attention to two categories. The first type is *pairing product equations* – $\prod_{i=1}^{n} e(A_i, Y_i) \prod_{i=1}^{m} e(X_i, B_i) \prod_{i=1}^{m} \prod_{j=1}^{n} e(A_i, B_j)^{c_{ij}} = 1_{\mathbb{T}}$, for constants $A_i \in \mathbb{G}_1, B_i \in \mathbb{G}_2, c_{ij} \in \mathbb{Z}_p$, where $1_{\mathbb{T}}$ is the identity in $\mathbb{T}$ and b) *multi-scalar exponentiations* – $\prod_{i=1}^{n'} A_i^{y_i} \prod_{i=1}^{m} X_i^{b_i} \prod_{i=1}^{m} \prod_{j=1}^{n'} X_i^{c_{ij} y_j} = T_1$, for constants $A_i, T_1 \in \mathbb{G}_1, b_i, c_{ij} \in \mathbb{Z}_p$ and analogous statements for multi-scalar exponentiation in $\mathbb{G}_2$.

GS proofs are in the *common random string* model, and satisfy the completeness and zero-knowledge properties described in Section 2.2. However, they only satisfy a weaker notion of knowledge extraction which has been referred to as *partial* knowledge extraction [23]. This property states that if the witness consists of both group elements and exponents, only the group elements are extractable.

2.4 Cuckoo hashing

A cuckoo hashing scheme consists of the algorithms $(\mathsf{Setup}, \mathsf{Hash})$, and is parameterized by a universe $\mathcal{U}$ of elements.

– $\mathsf{Setup}(\lambda, n, \epsilon) \to (n', h_0, h_1)$: the setup algorithm takes as input an integer parameter λ, an integer bound n, and $\epsilon \geq 0$, and outputs an integer n' and two hash functions $h_0, h_1 : \mathcal{U} \to [n']$, where n' is a deterministic function of λ, n, and ϵ.

– $\mathsf{Hash}(h_0, h_1, D) \to T$: the (deterministic) hashing algorithm takes hash functions $h_0, h_1 : \mathcal{U} \to [n']$ and a set $D \subseteq \mathcal{U}$, and outputs a table $T = [T_1, \ldots, T_{n'}]$, where each T_i is either an element in D or $\perp$.

For correctness, we demand that for every $x \in \mathcal{U}$, $h_0(x) \neq h_1(x)$. We will assume that this is the case for every pair of even adversarially chosen hash functions.[9] Each non-$\perp$ element of T is distinct. Finally, for any n, ϵ and set $D \subseteq \mathcal{U}$ of size n, it holds that with probability $1 - \mathsf{negl}(\lambda)$ over $(m, h_0, h_1) \leftarrow \mathsf{Setup}(\lambda, n, \epsilon)$, there exists a set $D' \subseteq D$ such that $|D'| \geq (1 - \epsilon)|D|$ and such that for any $x \in D'$, either $T_{h_0(x)} = x$ or $T_{h_1(x)} = x$, where $T \coloneqq \mathsf{Hash}(h_0, h_1, D)$.

3 Set Pre-Constrained Encryption

In this section, we define and construct set pre-constrained (SPC) encryption. We start by providing an overview in Section 3.1. We then present formal definitions of SPC encryption in Section 3.2, and constructions in Section 3.3. In the full version [8] we demonstrate that an *optimal* version of SPC encryption implies doubly-efficient private information retrieval and also prove security of our protocols.

3.1 Overview

The basic Apple PSI protocol. We start by recalling the basic Apple PSI protocol, viewed as an *encryption scheme*. "Basic" here refers to the protocol without the extra threshold or synthetic match functionalities, which we will not consider explicitly in this work.

A key technique used in Apple's protocol is the Naor-Reingold Diffie-Hellman random self reduction [36]. Let $\mathbb{G}$ be a cyclic group of order q with generator g, and let h_1, h_2, h_3 be three other group elements. Suppose that $\beta, \gamma \leftarrow \mathbb{Z}_q$ are sampled as uniformly random exponents, and $h_2' \coloneqq g^\beta \cdot h_2^\gamma, h_3' \coloneqq h_1^\beta \cdot h_3^\gamma$. Then it holds that (i) if (g, h_1, h_2, h_3) is a Diffie-Hellman tuple (that is, there exists α such that $g^\alpha = h_1$ and $h_2^\alpha = h_3$), then (g, h_1, h_2', h_3') is a Diffie-Hellman tuple, and (ii) if (g, h_1, h_2, h_3) is *not* a Diffie-Hellman tuple, then (h_2', h_3') are fresh uniformly random group elements.

Now, this self-reduction can be used to construct a set pre-constrained encryption scheme for a single-item set $\{x\}$ as follows. Let H be a hash function that hashes items to group elements (H will be treated as a random oracle in the security proof). The key generator, on input an item x, will sample $\alpha \leftarrow \mathbb{Z}_q$ and publish $(A = g^\alpha, B = H(x)^\alpha)$ as the public key. Note that $(g, A, H(x), B)$ is a Diffie-Hellman tuple, while for any $x' \neq x$, $(g, A, H(x'), B)$ is *not* a Diffie-Hellman tuple. This suggests a natural encryption scheme. Given the public key, an item y, and a message m, the encryption algorithm will run the Naor-Reingold self-reduction on $(g, A, H(y), B)$ to produce group elements (Q, S), and

[9] For example, h_1 can be defined to first hash x and then check if the hash is equal to $h_0(x)$ and if so add 1.

then treat S as a secret key for encrypting the message m. That is, the ciphertext will consist of $(Q, \mathsf{SEnc}_S(m))$, where SEnc is a symmetric-key encryption scheme. If $y \neq x$, then S will be uniformly random, even from the key generator's perspective, so m remains hidden. On the other hand, if $y = x$, then (g, A, Q, S) is a Diffie-Helman tuple, and the element $S = Q^\alpha$ can be computed by the key generator and used to recover m.

This scheme can easily be extended to support larger set sizes, by having the key generator publish $(A, H(x_1)^\alpha, \ldots, H(x_n)^\alpha)$ as the public key, where $x_1, \ldots, x_n$ is its input set. However, the naive extensions of the encryption and decryption algorithms described above will have running time that grows with the size n of the set. The authors of the Apple PSI system make use of a technique called *cuckoo hashing* to significantly reduce this running time. Concretely, the key generator will hash the set $(x_1, \ldots, x_n)$ into a table T of size $n' = (1 + \epsilon)n$ for some constant ϵ, using randomly sampled hash keys h_0, h_1. The guarantee is that with high probability, for most x_i, either $T_{h_0(x_i)} = x_i$ or $T_{h_1(x_i)} = x_i$. Note that T will have $n' - n$ empty entries, which we denote with $\perp$. The key generator will then publish $(A, B_1, \ldots, B_{n'})$ as the public key, where for each $i \in [n']$, if $T_i = x$ then $B_i = H(x)^\alpha$, while if $T_i = \perp$ then $B_i = g^r$ for a random exponent r. Now, to encrypt a message m with respect to an item y, one only has to produce two pairs $(Q_0, \mathsf{SEnc}_{S_0}(m)), (Q_1, \mathsf{SEnc}_{S_1}(m))$, where (Q_b, S_b) is the result of applying the Naor-Reingold self-reduction to $(g, A, H(y), B_{h_b(y)})$.

This results in a set pre-constrained encryption scheme that can handle pre-constraining sets of size n with a public key of about $n' = (1 + \epsilon)n$ group elements, and encryption and decryption algorithms whose running times do not grow with the size of n. One can show (in the random oracle model) that this scheme already satisfies set-hiding under the DDH assumption, and can be made to satisfy element-hiding from DDH, as long as the two pairs $(Q_0, \mathsf{SEnc}_{S_0}(m))$, $(Q_1, \mathsf{SEnc}_{S_1}(m))$ that constitute the ciphertext are randomly permuted, and $B_1, \ldots, B_{n'}$ are all distinct group elements.

Achieving bounded-set and authenticated-set security. Next, we show that augmenting the above template with *simple* and *efficient* zero-knowledge arguments suffices to achieve first bounded-set and next authenticated-set security. While the potential utility of adding zero-knowledge arguments to Apple's PSI system has previously been discussed informally [13,39], we view our formalization and efficient realization of rigorous security definitions as a necessary and important contribution in this space.

The Apple PSI paper [10] actually already claims to achieve *bounded-set* security, which guarantees that a malicious key generator can only decrypt messages that are encrypted with respect to some set of items of size at most B. However, it is left unclear what B is, and how it depends on other parameters in the system. In fact, their proof completely breaks down if $B < n'$. In particular, their proof relies on extracting the input set X of the key generator by observing random oracle queries, potentially adding one item x to X for each group element B_i in the public key. If the resulting X is such that $|X| > B$, then the ideal functionality aborts, and the malicious key generator would not receive encryptions from the client. However, this behavior does not reflect what would happen in the real world, where the client would not be able to tell how large the key generator's "effective input" actually is.

This issue in the proof occurs with good reason, since a malicious key generator can indeed publish $(A, H(x_1)^\alpha, \ldots, H(x_{n'})^\alpha)$ for n' items $x_1, \ldots, x_{n'}$ without being detected. However, correctness for honest key generators is only guaranteed to hold for up to n items (due to the cuckoo hashing). Thus, in the best case, we hope for a scheme that achieves bounded-set security with bound n.

We show how to achieve this by instructing the key generator to append to their key $(A, B_1, \ldots, B_{n'})$ a zero-knowledge non-interactive proof of knowledge that they know the discrete logarithm α of A and at least $n' - n$ discrete logarithms $\{r_i\}$ of the elements $B_1, \ldots, B_{n'}$. Highly efficient proofs supporting these languages are known [16,6]. Intuitively, the $n' - n$ group elements B_i for which the generator knows r_i such that $B_i = g^{r_i}$ are "useless" for decrypting encrypted messages. To see why, recall that, due to the Naor-Reingold self-reduction, B_i can only be used to decrypt with respect to an item x such that $(g, A, H(x), B_i)$ forms a Diffie-Hellman tuple. However, if the generator knows an x, α, and r_i such that this holds, they can break the discrete logarithm problem, since $H(x) = g^{r/\alpha}$ and $H(x)$ can be programmed by a reduction. Thus, only at most n of the elements $(B_1, \ldots, B_{n'})$ will actually be useful for decrypting messages, which we leverage to show bounded-set security with a bound of n.

Next, we consider our notion of *authenticated-set* security, which introduces a third party that chooses the set D. In our scheme, the third party first sends D to the key generator. Then, the key generator prepares a public key $(A, B_1, \ldots, B_{n'})$. In the honest case, for each i it either holds that there exists $x \in D$ such that $(g, A, H(x), B_i)$ form a Diffie-Hellman tuple, or the generator knows r_i such that $B_i = g_i^r$. Now, these are claims that the generator can prove efficiently in zero-knowledge to the third party. The third party will then checks these proofs, and if all verify, will sign the set of group elements $(A, B_1, \ldots, B_{n'})$ under its public verification key. We show in Section 3.3 that this is sufficient for achieving authenticated-set security.

3.2 Definitions

A set pre-constrained encryption (SPCE) scheme $\Pi_{\mathsf{SPCE}}[\mathcal{U}, \mathcal{M}, n, \epsilon]$ consists of algorithms (Gen, Enc, Dec), and is parameterized by a universe $\mathcal{U}$ of elements, a message space $\mathcal{M}$, a set size n, and a correctness parameter ϵ. $\mathcal{U}, \mathcal{M}, n, \epsilon$ may actually be infinite families parameterized by the security parameter λ, though we suppress mention of this for ease of notation.

- $\mathsf{Gen}(1^\lambda, D) \rightarrow (\mathsf{pk}, \mathsf{sk})$: the parameter generation algorithm takes as input a security parameter 1^λ and a set $D \subseteq \mathcal{U}$ of size at most n, and outputs a public key pk and a secret key sk.
- $\mathsf{Enc}(\mathsf{pk}, x, m) \rightarrow \mathsf{ct}$: the encryption algorithm takes as input a public key pk, an item $x \in \mathcal{U}$, and a message $m \in \mathcal{M}$, and outputs a ciphertext ct.
- $\mathsf{Dec}(\mathsf{sk}, \mathsf{ct}) \rightarrow \{m, \perp\}$: the decryption algorithm takes as input a secret key sk and a ciphertext ct and outputs either a message $m \in \mathcal{M}$ or a symbol $\perp$.

We note that any SPC encryption scheme can be utilized for encrypted cloud storage as follows. The server initially publishes pk, and whenever the client wants to upload some content x, they would sample an (element-hiding) SPC encryption of (x, m), where m is arbitrary "associated data" (e.g. the name of the client). Then, if $x \in D$, the server would be able to use sk to recover the associated data m. Otherwise (x, m) will remain hidden from the server.

Efficiency. By default, all algorithms in an SPC encryption scheme should be polynomial-time in the size of their inputs, and n,$|x|$,$|m|$ should be polynomial-size in λ. However, we will want to consider a more fine-grained notion of efficiency with respect to the size n of the set D, which may be a large polynomial. We say that the scheme has *succinct public-key* if $|\mathsf{pk}| = \mathsf{poly}(\lambda, \log n)$, *succinct encryption* if the running time of Enc is $\mathsf{poly}(\lambda, \log n)$, and *succinct decryption* if the running time of Dec is $\mathsf{poly}(\lambda, \log n)$.

Correctness. We define notions of correctness for an SPC encryption scheme. We first consider the following notion of ϵ-correctness, where the parameter ϵ essentially determines an upper bound on the fraction of the set D that is "dropped" by the Gen algorithm.[10]

Definition 1. *An* SPC *encryption scheme* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is* ϵ*-correct for some* $\epsilon \geq 0$ *if for any* $\lambda \in \mathbb{N}$ *and* $D \subseteq \mathcal{U}$*, it holds that with probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{Gen}(1^\lambda, D)$*, there exists a* $D' \subseteq D$ *such that* $|D'| \geq (1-\epsilon)|D|$ *and for any* $x \in D'$ *and* $m \in \mathcal{M}$*,* $\Pr[\mathsf{Dec}(\mathsf{sk}, \mathsf{Enc}(\mathsf{pk}, x, m)) = m] = 1 - \mathsf{negl}(\lambda)$.

Next, we define the stronger notion of *perfect* ϵ-correctness that will be useful for our application of SPC encryption to building SPC group signatures in Section 4.

Definition 2. *An* SPC *encryption scheme* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is perfectly* ϵ*-correct for some* $\epsilon \geq 0$ *if the following two properties hold for any* $\lambda \in \mathbb{N}$ *and* $D \subseteq \mathcal{U}$*.*

- *With probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{Gen}(1^\lambda, D)$*, there exists a* $D' \subseteq D$ *such that* $|D'| \geq (1-\epsilon)|D|$ *and for any* m *and* $x \in D'$*,* $\Pr[\mathsf{Dec}(\mathsf{sk}, \mathsf{Enc}(\mathsf{pk}, x, m)) = m] = 1$.
- *For all* $(\mathsf{pk}, \mathsf{sk}) \in \mathsf{Gen}(1^\lambda, D)$*,* x, m*,* $\Pr[\mathsf{Dec}(\mathsf{sk}, \mathsf{Enc}(\mathsf{pk}, x, m)) \in \{m, \perp\}] = 1$.

Security. We define security using the simulation framework, via an ideal functionality described in Fig. 1. The ideal functionality $\mathcal{F}^{\mathcal{P}}_{\mathsf{SPCE}}$ takes place between a *server*, who runs Gen and Dec, a *client*, who runs Enc, and a third party Auth, whose role will be described below. In full generality, the server's input is a function F, but in our applications, we will always parse F as a description of a database D of items. The client's input is a sequence of items and messages $(x_1, m_1), \ldots, (x_k, m_k)$. The client should learn nothing about D, Auth should learn nothing about the messages $m_1, \ldots, m_k$, and the server should learn only $\{m_i\}_{i : x_i \in D}$ (and potentially the elements $\{x_i\}_{i \in [k]}$).

[10] Traditionally, one might expect ϵ to be negligible, and thus suppressed in the definition. However, our protocols will make use of cuckoo hashing which may introduce an inverse-polynomial ϵ.

To make security against the server meaningful, we must place some restriction on D. We do this (in a modular way) by parameterizing the functionality with a predicate $\mathcal{P}$. This predicate may depend on some public parameters pp (known to both client and server) and some secret parameters sp (known only to the server). It is the job of Auth to set up these parameters. We allow a malicious adversary to corrupt either the server, the client, or Auth. We note that one could also consider collusions between any pair of parties, but in each case security becomes vacuous, so we do not consider this in our proofs of security.

Below, we describe the instantiations of $\mathcal{P}$ that we will consider in this work: one will define what we call *bounded-set security* and the other will define what we call *authenticated-set security*.

$\mathcal{F}^{\mathcal{P}}_{\mathsf{SPCE}}$

Parties: server S, client C, and authority Auth.
Parameters: universe $\mathcal{U}$, message space $\mathcal{M}$.

- Obtain input $(\mathsf{pp}, \mathsf{sp})$ from Auth. Deliver pp to both C and S, and sp to S.
- Obtain input $F = (F_S, F_{\mathsf{Auth}})$ from S and deliver F to Auth. Abort and deliver $\perp$ to all parties if $\mathcal{P}(\mathsf{pp}, \mathsf{sp}, F) = 0$.
- Obtain input $(x_1, m_1), \ldots, (x_k, m_k)$ from client, where each $x_i \in \mathcal{U}$ and each $m_i \in \mathcal{M}$.
- Deliver $F_S(\{x_i, m_i\}_{i \in [k]})$ to server and $F_{\mathsf{Auth}}(\{x_i, m_i\}_{i \in [k]})$ to Auth.

Fig. 1. Ideal functionality for SPC encryption. $\mathcal{P}$ is a predicate that takes as input some public parameters pp, secret parameters sp, and a pair of functions $F = (F_S, F_{\mathsf{Auth}})$, and outputs a bit.

Bounded-set security. Here, we define two predicates $\mathcal{P}[\mathsf{BS}]$ and $\mathcal{P}[\mathsf{BS\text{-}EH}]$, where BS stands for bounded-set, and EH stands for element-hiding. For each, the public parameters pp are parsed as an integer n, there are no secret parameters sp, and F is parsed as the description of a database $D \subseteq \mathcal{U}$. The predicate then outputs 1 if and only if $|D| \leq n$. For $\mathcal{P}[\mathsf{BS}]$,

$$F_S(\{x_i, m_i\}_{i \in [k]}) = \{x_i\}_{i \in [k]}, \{m_i\}_{i : x_i \in D}, \quad F_{\mathsf{Auth}}(\{x_i, m_i\}_{i \in [k]}) = \{x_i\}_{i \in [k]},$$

and for $\mathcal{P}[\mathsf{BS\text{-}EH}]$,

$$F_S(\{x_i, m_i\}_{i \in [k]}) = k, \{m_i\}_{i : x_i \in D}, \quad F_{\mathsf{Auth}}(\{x_i, m_i\}_{i \in [k]}) = k.$$

Authenticated-set security. Here, we define two predicates $\mathcal{P}[\mathsf{AS}]$ and $\mathcal{P}[\mathsf{AS\text{-}EH}]$, where AS stands for authenticated-set. For each, the public parameters pp are parsed as an integer n, the secret parameters are parsed as a database $D^* \subseteq \mathcal{U}$

of size n, and F is parsed as a database $D \subseteq \mathcal{U}$. The predicate then outputs 1 if and only if $D \subseteq D^*$. For $\mathcal{P}[\mathsf{AS}]$,

$$F_S(\{x_i, m_i\}_{i\in[k]}) = \{x_i\}_{i\in[k]}, \{m_i\}_{i:x_i\in D}, \quad F_{\mathsf{Auth}}(\{x_i, m_i\}_{i\in[k]}) = \{x_i\}_{i\in[k]},$$

and for $\mathcal{P}[\mathsf{AS\text{-}EH}]$,

$$F_S(\{x_i, m_i\}_{i\in[k]}) = k, \{m_i\}_{i:x_i\in D}, \quad F_{\mathsf{Auth}}(\{x_i, m_i\}_{i\in[k]}) = k.$$

One can also define a game-based notion of security against *outsiders* by extending standard notions of semantic security to capture the indistinguishability of ciphertexts corresponding to encryptions of two different messages. We defer this to the full version [8].

3.3 Construction

We begin by giving templates for SPC encryption based on Apple's PSI protocol [10]. These schemes will have succinct encryption and succinct decryption, but *non-succinct* public-key. We will first describe a scheme $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}$ (Protocol 2) that satisfies ϵ-correctness and security against outsiders with element-hiding. Then, we describe a related scheme $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$ (Protocol 3) that satisfies *perfect* ϵ-correctness but only security against outsiders *without* element-hiding, and is tailored to support encrypting messages that are group elements. This latter scheme will be useful for our construction of set pre-constrained group signatures in Section 4. Following these basic templates, we will then show how to (efficiently) upgrade each to obtain bounded-set and authenticated-set security, resulting in schemes $\Pi_{\mathsf{SPCE}}^{\mathsf{BS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{BS-PC}}, \Pi_{\mathsf{SPCE}}^{\mathsf{AS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{AS-PC}}$.

Ingredients:

- A cyclic group $\mathbb{G}$ of prime order q in which the DDH problem is assumed to be hard.
- A symmetric-key encryption scheme ($\mathsf{RobEnc}, \mathsf{RobDec}$) with keyspace $\mathcal{K}$ that satisfies *random key robustness* (Section 2.1).
- Hash functions $H : \mathcal{U} \to \mathbb{G} \setminus \{0\}$ and $G : \mathbb{G} \to \mathcal{K}$ modeled as random oracles, where G maps the uniform distribution over $\mathbb{G}$ to (negligibly close to) the uniform distribution over $\mathcal{K}$.
- A cuckoo hashing scheme ($\mathsf{CH.Setup}, \mathsf{CH.Hash}$) (Section 2.4).

Achieving bounded-set security. It can in fact be shown that $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}[\mathcal{U}, \mathcal{M}, n, \epsilon]$ (resp. $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}[\mathcal{U}, \mathbb{G}, n, \epsilon]$) already securely emulates $\mathcal{F}_{\mathsf{SPCE}}^{\mathcal{P}[\mathsf{BS\text{-}EH}]}$ (resp. $\mathcal{F}_{\mathsf{SPCE}}^{\mathcal{P}[\mathsf{BS}]}$) with set size $\mathsf{pp} = n'$ where n' is such that $(n', \cdot, \cdot) \leftarrow \mathsf{CH.Setup}(\lambda, n, \epsilon)$. However, n' may be much larger than n, which means a large gap between correctness (an honest server would be able to decrypt with respect to $(1-\epsilon)n$ items) and security (a dishonest server would potentially be able to decrypt with respect to up to n' items).

Below, we show that an efficient tweak to the basic schemes results in schemes $\Pi_{\mathsf{SPCE}}^{\mathsf{BS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{BS-PC}}$ (Protocol 4) that completely close this gap. That is, for any n, the schemes $\Pi_{\mathsf{SPCE}}^{\mathsf{BS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{BS-PC}}$ securely emulate $\mathcal{F}_{\mathsf{SPCE}}^{\mathcal{P}[\mathsf{BS\text{-}EH}]}, \mathcal{F}_{\mathsf{SPCE}}^{\mathcal{P}[\mathsf{BS}]}$ with $\mathsf{pp} = n$.

$\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}[\mathcal{U}, \mathcal{M}, n, \epsilon]$

Parameters: universe $\mathcal{U}$, message space $\mathcal{M}$, set size n, correctness parameter ϵ, and security parameter λ.
Setup: description of the group $\mathbb{G}$ with generator g, and random oracles H, G.
$\mathsf{Gen}(1^\lambda, D)$:

- Run $(n', h_0, h_1) \leftarrow \mathsf{CH.Setup}(\lambda, n, \epsilon)$ and then $T := \mathsf{CH.Hash}(h_0, h_1, D)$.
- Sample $\alpha \leftarrow \mathbb{Z}_q$ and set $A := g^\alpha$.
- Define $\widetilde{T}$ as follows. For each $i \in [n']$, if $T_i = \bot$ then sample $r_i \leftarrow \mathbb{Z}_q$ and set $\widetilde{T}_i := g^{r_i}$, and otherwise set $\widetilde{T}_i := H(T_i)^\alpha$.
- Output $\mathsf{pk} := (h_0, h_1, A, \widetilde{T})$ and $\mathsf{sk} := \alpha$.

$\mathsf{Enc}(\mathsf{pk}, x, m)$:

- Parse pk as $(h_0, h_1, A, \widetilde{T})$ and abort if there are any duplicate entries in $\widetilde{T}$.
- For $b \in \{0,1\}$, sample $\beta_b, \gamma_b \leftarrow \mathbb{Z}_q$, and compute $Q_b := g^{\beta_b} \cdot H(x)^{\gamma_b}$, $S_b := A^{\beta_b} \cdot \widetilde{T}_{h_b(x)}^{\gamma_b}$, $\mathsf{ct}_b := \mathsf{RobEnc}(G(S_b), m)$. Sample $b \leftarrow \{0,1\}$ and output $\mathsf{ct} := (Q_b, \mathsf{ct}_b, Q_{1-b}, \mathsf{ct}_{1-b})$.

$\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$:

- Parse sk as α and ct as $(Q_0, \mathsf{ct}_0, Q_1, \mathsf{ct}_1)$.
- For $b \in \{0,1\}$ and compute $m_b := \mathsf{RobDec}(G(Q_b^\alpha), \mathsf{ct}_b)$. If exactly one of m_0 or m_1 is not $\bot$, then output this message, and otherwise output $\bot$.

Fig. 2. Basic SPC encryption with element-hiding

Observe that the correctness properties of $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$ are preserved by this transformation, due to the completeness of the ZK-NIAoKs, and the security against outsiders properties are also preserved, due to the zero-knowledge of the ZK-NIAoKs.

Achieving authenticated-set security. Next, we describe schemes $\Pi_{\mathsf{SPCE}}^{\mathsf{AS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{AS-PC}}$ (Protocol 5) that satisfy authenticated set security. In order to achieve this notion, we will relax Gen to be an *interactive protocol* between the server and Auth, with the following syntax. $\mathsf{Gen}\langle\mathsf{Server}, \mathsf{Auth}(D)\rangle(1^\lambda) \rightarrow (\mathsf{pk}, \mathsf{sk})$ where the parameter generation protocol takes place between a server and Auth with input a set $D \subseteq \mathcal{U}$, and outputs to the server a public key pk and a secret key sk.

Observe that the correctness properties of $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$ are preserved by the transformation, due to the completeness of the ZK-NIAoKs and correctness of Sig, and the security against outsiders properties are also preserved, due to the zero-knowledge of the ZK-NIAoKs.

$\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}[\mathcal{U}, \mathbb{G}, n, \epsilon]$

Parameters: same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}$, except that the message space is the set of group elements in $\mathbb{G}$.
Setup: Same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}$.
$\mathsf{Gen}(1^\lambda, D)$: Same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}$, except that h_0, h_1, T are included in sk, and we abort if there does not exist a $D' \subseteq D$ such that $|D'| \geq (1-\epsilon)|D|$ and such that for any $x \in D'$, either $T_{h_0(x)} = x$ or $T_{h_1(x)} = x$.
$\mathsf{Enc}(\mathsf{pk}, x, m)$:

- Parse pk as $(h_0, h_1, A, \widetilde{T})$ and abort if there are any duplicate entries in $\widetilde{T}$.
- For $b \in \{0,1\}$, sample $\beta_b, \gamma_b \leftarrow \mathbb{Z}_q$, and compute $Q_b := g^{\beta_b} \cdot H(x)^{\gamma_b}, S_b := A^{\beta_b} \cdot \widetilde{T}_{h_b(x)}^{\gamma_b}$. Output $\mathsf{ct} := (x, Q_0, S_0 \cdot m, Q_1, S_1 \cdot m)$.

$\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$:

- Parse sk as (h_0, h_1, T, α) and ct as $(x, Q_0, S_0', Q_1, S_1')$.
- If there exists exactly one $b \in \{0,1\}$ such that $T_{h_i(x)} = x$, then output $m = S_b'/Q_b^\alpha$. Otherwise, output $\perp$.

Fig. 3. Basic SPC encryption with perfect correctness

$\Pi_{\mathsf{SPCE}}^{\mathsf{BS-EH}}[\mathcal{U}, \mathcal{M}, n, \epsilon], \Pi_{\mathsf{SPCE}}^{\mathsf{BS-PC}}[\mathcal{U}, \mathbb{G}, n, \epsilon]$

Parameters: Same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$. Note that the parameters λ, n, and ϵ determine a maximum hash table size n', where $(n', \cdot, \cdot) \leftarrow \mathsf{Setup}(\lambda, n, \epsilon)$.

Setup: Let $(\mathsf{Prove}_{\mathsf{DLog}}, \mathsf{Verify}_{\mathsf{DLog}})$ be a ZK-NIAoK for $\mathcal{R}_{\mathsf{DLog}}$ and let $(\mathsf{Prove}_{\widetilde{\mathsf{DLog}}}, \mathsf{Verify}_{\widetilde{\mathsf{DLog}}})$ be a ZK-NIAoK for $\mathcal{R}_{\mathsf{DLog}_{n'}^{n'-n}}$ (Section 2.2). Both of these proof systems are in the random oracle model and have no additional setup, so there is no additional setup required for $\Pi_{\mathsf{SPCE}}^{\mathsf{BS-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{BS-PC}}$.

$\mathsf{Gen}(1^\lambda, D)$: Same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$, except that proofs $\pi_A \leftarrow \mathsf{Prove}_{\mathsf{DLog}}((g, A), \alpha)$ and $\pi_{\widetilde{T}} \leftarrow \mathsf{Prove}_{\widetilde{\mathsf{DLog}}}((g, \widetilde{T}), \{r_i\}_{i:T_i=\perp})$ are computed and appended to the public key pk.

$\mathsf{Enc}(\mathsf{pk}, x, m)$: Same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$, except that the algorithm aborts if either of π_A or $\pi_{\widetilde{T}}$ fails to verify, or the number of group elements in $\widetilde{T}$ is greater than n'.

$\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$: same as $\Pi_{\mathsf{SPCE}}^{\mathsf{Basic-EH}}, \Pi_{\mathsf{SPCE}}^{\mathsf{Basic-PC}}$.

Fig. 4. SPC encryption with bounded-set security

$\mathit{\Pi}^{\mathsf{AS-EH}}_{\mathsf{SPCE}}[\mathcal{U}, \mathcal{M}, n, \epsilon], \mathit{\Pi}^{\mathsf{AS-PC}}_{\mathsf{SPCE}}[\mathcal{U}, \mathbb{G}, n, \epsilon]$

Parameters: same as $\mathit{\Pi}^{\mathsf{Basic-EH}}_{\mathsf{SPCE}}, \mathit{\Pi}^{\mathsf{Basic-PC}}_{\mathsf{SPCE}}$.

Setup: let $(\mathsf{Sig.Gen}, \mathsf{Sig.Sign}, \mathsf{Sig.Verify})$ be a EUF-CMA secure signature scheme (Section 2.1). Before the protocol begins, Auth will sample $(\mathsf{vk}_{\mathsf{Auth}}, \mathsf{sk}_{\mathsf{Auth}}) \leftarrow \mathsf{Sig.Gen}(1^\lambda)$ and broadcast $\mathsf{vk}_{\mathsf{Auth}}$ to all parties. Also, let $(\mathsf{Prove}_{\mathsf{DLog}}, \mathsf{Verify}_{\mathsf{DLog}})$ be a ZK-NIAoK for $\mathcal{R}_{\mathsf{DLog}}$ and $(\mathsf{Prove}_{\mathsf{DH}}, \mathsf{Verify}_{\mathsf{DH}})$ be a ZK-NIAoK for $\mathcal{R}_{\mathsf{DH}}$ (Section 2.2). Both of these proof systems are in the random oracle model, so require no additional setup beyond $\mathit{\Pi}^{\mathsf{Basic-EH}}_{\mathsf{SPCE}}, \mathit{\Pi}^{\mathsf{Basic-PC}}_{\mathsf{SPCE}}$.

$\mathsf{Gen}\langle\mathsf{Server}, \mathsf{Auth}(D)\rangle(1^\lambda)$:

- Auth sends D to Server.
- Server first runs the Gen algorithm of $\mathit{\Pi}^{\mathsf{Basic-EH}}_{\mathsf{SPCE}}, \mathit{\Pi}^{\mathsf{Basic-PC}}_{\mathsf{SPCE}}$ on input $(1^\lambda, D)$ to obtain output $(h_0, h_1, A, \widetilde{T}), \alpha$, along with table T and randomness $\{r_i\}_{i:T_i=\perp}$. Next, compute $\pi_A \leftarrow \mathsf{Prove}_{\mathsf{DLog}}((g, A), \alpha)$. Finally, for each $i \in [n']$, where n' is the size of $T, \widetilde{T}$:
  - If $T_i = \perp$, compute $\pi_i \leftarrow \mathsf{Prove}_{\mathsf{DLog}}((g, \widetilde{T}_i), r_i)$.
  - If $T_i \neq \perp$, compute $\pi_i \leftarrow \mathsf{Prove}_{\mathsf{DH}}((g, A, H(T_i), \widetilde{T}_i), \alpha)$.

  Send $(A, T, \widetilde{T}, \pi_A, \{\pi_i\}_{i\in[n']})$ to Auth.
- Auth runs $\mathsf{Verify}_{\mathsf{DLog}}((g, A), \pi_A)$ and for each $i \in [n']$: if $T_i = \perp$, runs $\mathsf{Verify}_{\mathsf{DLog}}((g, \widetilde{T}_i), \pi_i)$ and if $T_i \neq \perp$, check that $T_i \in D$ and runs $\mathsf{Verify}_{\mathsf{DH}}((g, A, H(T_i), \widetilde{T}_i), \pi_i)$. If all checks pass, compute $\sigma \leftarrow \mathsf{Sig.Sign}(\mathsf{sk}, (A, \widetilde{T}))$, and return σ.
- Server outputs $\mathsf{pk} := (h_0, h_1, A, \widetilde{T}, \sigma)$ and $\mathsf{sk} := \alpha$.

$\mathsf{Enc}(\mathsf{pk}, x, m)$: same as $\mathit{\Pi}^{\mathsf{Basic-EH}}_{\mathsf{SPCE}}, \mathit{\Pi}^{\mathsf{Basic-PC}}_{\mathsf{SPCE}}$, except that it first runs $\mathsf{Sig.Verify}(\mathsf{vk}_{\mathsf{Auth}}, (A, \widetilde{T}), \sigma)$[a] and aborts if the signature fails to verify.
$\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$: same as $\mathit{\Pi}^{\mathsf{Basic-EH}}_{\mathsf{SPCE}}, \mathit{\Pi}^{\mathsf{Basic-PC}}_{\mathsf{SPCE}}$.

[a] Note that this verification only needs to be done once per user and not every time Enc is run, since the public key does not change.

Fig. 5. SPC encryption with authenticated-set security

4 SPC Group Signatures

In this section, we define and construct SPC group signatures. We present formal definition of SPC group signatures in Section 4.1, and constructions in Section 4.2 and Section 4.3. We defer proofs of security to the full version [8].

4.1 Definitions

A set pre-constrained group signature (SPCGS) scheme $\Pi_{\mathsf{SPCGS}}[\mathcal{M}, \mathcal{P}, n, \epsilon]$ consists of algorithms $\mathsf{Gen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open}$, along with an interactive protocol KeyGen. We refer to the party that runs Gen as the *group manager* GM, and the KeyGen protocol is run by GM and a client C. It is parameterized by a message space $\mathcal{M}$, an identity (or public key) space $\mathcal{P}$, a set size n, and a correctness parameter ϵ.

- $\mathsf{Gen}(1^\lambda, D) \rightarrow (\mathsf{mpk}, \mathsf{msk})$. The parameter generation algorithm takes as input a security parameter 1^λ and a set $D \subseteq \mathcal{M}$ of size at most n, and outputs a master public key mpk and a master secret key msk.
- $\mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C} \rangle \rightarrow (\mathsf{pk}, \mathsf{sk})$. The KeyGen protocol is run by the group manager GM with input msk and a client C. It delivers an identity $\mathsf{pk} \in \mathcal{P}$ to both GM and C, and an identity signing key sk to C.
- $\mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}, m) \rightarrow \sigma$. The signing algorithm takes as input the master public key mpk, an identity signing key sk, and a message $m \in \mathcal{M}$, and outputs a signature σ.
- $\mathsf{Verify}(\mathsf{mpk}, m, \sigma) \rightarrow \{\top, \bot\}$. The verification algorithm takes as input the master public key mpk, a message $m \in \mathcal{M}$, and a signature σ, and outputs either $\top$ or $\bot$, indicating accept or reject.
- $\mathsf{Open}(\mathsf{msk}, \sigma) \rightarrow \{\mathsf{pk}, \bot\}$. The opening algorithm takes as input the master secret key msk and signature σ, and outputs either an identity $\mathsf{pk} \in \mathcal{P}$ or $\bot$.

We imagine using an SPC group signature scheme for encrypted messaging as follows. We assume that there is already a standard end-to-end encrypted messaging system in place, and the server additionally publishes mpk for the SPCGS scheme. Each client runs a KeyGen protocol with the server in order to obtain their identity pk and their secret key sk. Then, whenever they want to send a message m, they additionally compute a signature σ on m, and send the message (m, σ) *under the end-to-end encryption.* Any message received that does not have a properly verifying signature is immediately discarded by the client algorithm. Finally, if an honest client receives a pair (m, σ) for some illegal content m, they can report this to the server, who can run the Open algorithm in order to determine which identity produced the signature σ.

We now port the definitions of bounded-set and authenticated-set security against malicious servers (as previously defined for SPC encryption) to the group signature setting. Further, we follow standard definitions of *traceability*, *anonymity*, and *unframeability* for group signatures.

Definition 3 (Correctness). *An* SPC *group signature scheme* $(\mathsf{Gen}, \mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open})$ *is correct if for any* $\lambda \in \mathbb{N}, D \subseteq \mathcal{M}$*, and message* $m \in \mathcal{M}$*, it holds with probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Gen}(1^\lambda, D), (\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C} \rangle$*, and* $\sigma \leftarrow \mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}, m)$ *that* $\mathsf{Verify}(\mathsf{mpk}, m, \sigma) = 1$*.*

Security. We formulate several notions of security for an SPC group signature scheme. First, we define a notion of *traceability*, which ensures that any signature on a message $m \in D$ that is accepted by the verification algorithm will leak the identity of the signer to the master secret key holder.

Definition 4 (Traceability). *An* SPC *group signature scheme* $(\mathsf{Gen}, \mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open})$ *is* ϵ*-traceable for some* $\epsilon \geq 0$ *if for any PPT adversary* $\mathcal{A}$, $\lambda \in \mathbb{N}$, *and* $D \subseteq \mathcal{M}$, *it holds that with probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Gen}(1^\lambda, D)$, *there exists a* $D' \subseteq D$ *with* $|D'| \geq (1-\epsilon)|D|$, *such that*

$$\Pr\left[\begin{array}{l}\mathsf{IsValid}[\mathsf{mpk}](m, \sigma, \mathsf{pk}) = 1 \\ \wedge\ (m \in D') \\ \wedge\ (\mathsf{pk} \notin \mathcal{I}_{\mathsf{Adv}})\end{array} : \begin{array}{r}(m, \sigma) \leftarrow \mathcal{A}^{\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{Open}}, \mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}} \\ \mathsf{pk} \leftarrow \mathsf{Open}(\mathsf{msk}, \sigma)\end{array}\right] = \mathsf{negl}(\lambda),$$

where the oracles $\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{Open}}, \mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}$, *set* $\mathcal{I}_{\mathsf{Adv}}$ *and predicate* $\mathsf{IsValid}[\mathsf{mpk}]$ *are defined as follows.*

- $\mathcal{O}_{\mathsf{AKG}}$ *(KeyGen initiated by the Adversary) has* msk *hard-coded and, when initialized, acts as the group manager in the* KeyGen *protocol. Define* $\mathcal{I}_{\mathsf{Adv}}$ *to be the set of identities obtained by* $\mathsf{GM}(\mathsf{msk})$ *as a result of the interactions between* $\mathcal{A}$ *and* $\mathcal{O}_{\mathsf{AKG}}$.
- $\mathcal{O}_{\mathsf{Open}}$ *has* msk *hard-coded, and on input a signature* σ, *outputs* $\mathsf{Open}(\mathsf{msk}, \sigma)$.
- $\mathcal{O}_{\mathsf{HKG}}$ *(KeyGen initiated by an Honest party) has* msk *hard-coded and, when queried, runs* $\mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle \rightarrow (\mathsf{pk}, \mathsf{sk})$, *and returns* pk *(and not* sk*). Define* $\mathcal{I}_{\mathsf{Hon}}$ *to be the set of* pk*'s output by* $\mathcal{O}_{\mathsf{HKG}}$.
- $\mathcal{O}_{\mathsf{Sign}}$ *takes a message* m *and an identity* pk *as input. If* $\mathsf{pk} \notin \mathcal{I}_{\mathsf{Hon}}$, *return nothing, and otherwise let* sk *be the secret key associated with* pk *and return* $\mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}, m)$. *Define* $\mathcal{J}$ *to be the set of* (m, pk) *queried to* $\mathcal{O}_{\mathsf{Sign}}$.
- $\mathsf{IsValid}[\mathsf{mpk}](m, \sigma, \mathsf{pk})$ *outputs* $(\mathsf{Verify}(\mathsf{mpk}, m, \sigma) = 1) \wedge ((m, \mathsf{pk}) \notin \mathcal{J})$. *That is, it accepts if the adversary produced a valid message signature pair that was not a query to its signing oracle.*

Next, we define the notion of *unframeability*, which ensures that an adversary cannot produce a verifying signature with respect to some identity pk for which they do not hold the corresponding sk, even if they know the master secret key.

Definition 5 (Unframeability). *An* SPC *group signature scheme* $(\mathsf{Gen}, \mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open})$ *satisfies unframeability if for any PPT adversary* $\mathcal{A}$ *and* $D \subseteq \mathcal{M}$,

$$\Pr\left[\begin{array}{l}\mathsf{Verify}(\mathsf{mpk}, m, \sigma) = 1 \\ \wedge\ (\mathsf{mpk}, \mathsf{msk}) \in \mathsf{Gen}(1^\lambda, D) \\ \wedge\ ((m, \mathsf{pk}) \notin \mathcal{J}) \wedge\ (\mathsf{pk} \in \mathcal{I}_{\mathsf{Hon}})\end{array} : \begin{array}{r}(\mathsf{mpk}, \mathsf{msk}, m, \sigma) \leftarrow \mathcal{A}^{\mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}}(1^\lambda, D) \\ \mathsf{pk} \leftarrow \mathsf{Open}(\mathsf{msk}, \sigma)\end{array}\right] = \mathsf{negl}(\lambda),$$

where the oracles $\mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}$ *and sets* $\mathcal{I}_{\mathsf{Hon}}, \mathcal{J}$ *are defined as in Definition 4.*

Now, we consider the notion of *anonymity*, which protects the identity of any signer who produces a signature on a message $m \notin D$, even against the group manager. Here, we will follow our simulation-based notion of security for SPC encryption. The ideal functionality $\mathcal{F}^{\mathcal{P}}_{\mathsf{anon}}$ takes place between a *group manager* GM who runs Gen, interacts in KeyGen, and runs Open, a *client*, who interacts in KeyGen and runs Sign, and an authority Auth, whose role will be described below. In full generality, the group manager's input is a function F, but in our applications, we will always parse F as a description of a database D of messages. The client's input is a sequence of identities and messages $(\mathsf{pk}_1, m_1), \ldots, (\mathsf{pk}_k, m_k)$. The client should learn nothing about D, Auth should learn nothing about the identities $\{\mathsf{pk}_i\}_{i\in[k]}$, and the group manager should learn nothing about the identities $\{\mathsf{pk}_i\}_{i:m_i\notin D}$, except perhaps how many "repeats" there are (if we don't require the property of *unlinkability*).

To make security against the server meaningful, we must place some restriction on D. We do this (in a modular way) by parameterizing the functionality with a predicate $\mathcal{P}$. This predicate may depend on some public parameters pp (known to both client and group manager) and some secret parameters sp (known only to the group manager). It is the job of Auth to set up these parameters.

Below, we describe the instantiations of $\mathcal{P}$ that we will consider in this work: one will define what we call *bounded-set security* and the other will define what we call *authenticated-set security.*

$\mathcal{F}_{\mathsf{anon}}^{\mathcal{P}}$

Parties: Group manager and client.
Parameters: message space $\mathcal{M}$, identity space $\mathcal{I}$.

- Obtain input $(\mathsf{pp}, \mathsf{sp})$ from Auth. Deliver pp to both group manager and client, and sp to group manager.
- Obtain input $F = (F_{\mathsf{GM}}, F_{\mathsf{Auth}})$ from group manager and deliver F to Auth. Abort and deliver $\perp$ to all parties if $\mathcal{P}(\mathsf{pp}, \mathsf{sp}, F) = 0$.
- Obtain input $(\mathsf{pk}_1, m_1), \ldots, (\mathsf{pk}_k, m_k)$ from client, where each $\mathsf{pk}_i \in \mathcal{I}$ and each $m_i \in \mathcal{M}$.
- Deliver $F_{\mathsf{GM}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]})$ to group manager and $F_{\mathsf{Auth}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]})$ to Auth.

Fig. 6. Ideal functionality for SPC group signatures with anonymity. $\mathcal{P}$ is a predicate that takes as input some public parameters pp, and a pair of functions $F = (F_S, F_{\mathsf{Auth}})$, and outputs a bit.

Bounded-set security. Here, we define two predicates $\mathcal{P}[\mathsf{BS}]$ and $\mathcal{P}[\mathsf{BS\text{-}link}]$, where link stands for linkability. For each, the parameters pp are parsed as an integer n, there are no secret parameters sp, and F is parsed as a description of a database $D \subseteq \mathcal{M}$. The predicate then outputs 1 if and only if $|D| \leq n$. For $\mathcal{P}[\mathsf{BS}]$,

$$F_{\mathsf{GM}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{\mathsf{pk}_i\}_{i:m_i\in D}, \{m_i\}_{i\in[k]}, \quad F_{\mathsf{Auth}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{m_i\}_{i\in[k]},$$

and for $\mathcal{P}[\mathsf{BS\text{-}link}]$,

$$F_{\mathsf{GM}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{\mathsf{pk}_i\}_{i:m_i\in D}, \mathsf{Aux}(\{\mathsf{pk}_i\}_{i:m_i\notin D}), \{m_i\}_{i\in[k]},$$

where for any multiset S, $\mathsf{Aux}(S)$ consists of the number of distinct elements in S along with how many times each appears, and

$$F_{\mathsf{Auth}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{m_i\}_{i\in[k]}.$$

Authenticated-set security. Here, we define two predicates $\mathcal{P}[\mathsf{AS}]$ and $\mathcal{P}[\mathsf{AS\text{-}EH}]$. For each, the public parameters pp are parsed as an integer n, the secret parameters are parsed as a database $D^* \subseteq \mathcal{M}$ of size n, and F is parsed as a database $D \subseteq \mathcal{M}$. The predicate then outputs 1 if and only if $D \subseteq D^*$. For $\mathcal{P}[\mathsf{AS}]$,

$$F_{\mathsf{GM}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{\mathsf{pk}_i\}_{i:m_i\in D}, \{m_i\}_{i\in[k]}, \quad F_{\mathsf{Auth}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{m_i\}_{i\in[k]},$$

and for $\mathcal{P}[\mathsf{AS\text{-}link}]$,

$$F_{\mathsf{GM}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{\mathsf{pk}_i\}_{i:m_i\in D}, \mathsf{Aux}(\{\mathsf{pk}_i\}_{i:m_i\notin D}), \{m_i\}_{i\in[k]},$$

where for any multiset S, $\mathsf{Aux}(S)$ consists of the number of distinct elements in S along with how many times each appears, and

$$F_{\mathsf{Auth}}(\{\mathsf{pk}_i, m_i\}_{i\in[k]}) = \{m_i\}_{i\in[k]}.$$

Finally, we consider "client-client" anonymity and unlinkability, which considers the security of signatures against other clients. Here, we can hope for stronger security properties, since clients do not hold the master secret key and thus might not be able to de-anonymize signatures even on messages $m \in D$. Thus, we give separate (game-based) definitions of anonymity and unlinkability against adversarial clients.

Definition 6 (Anonymity). *An* SPC *group signature scheme* $(\mathsf{Gen}, \mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open})$ *satisfies client-client anonymity if for any PPT adversary* $\mathcal{A}$*,* $\lambda \in \mathbb{N}$*,* $D \subseteq \mathcal{M}$*, and* $m \in \mathcal{M}$*, it holds that with probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Gen}(1^\lambda, D), (\mathsf{pk}_0, \mathsf{sk}_0) \leftarrow \mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle, (\mathsf{pk}_1, \mathsf{sk}_1) \leftarrow \mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle$*,*

$$\Pr\left[\mathcal{A}^{\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}}\begin{pmatrix}\mathsf{mpk}, \mathsf{pk}_0, \\ \mathsf{pk}_1, \sigma\end{pmatrix} = b \;:\; \begin{array}{r} b \leftarrow \{0,1\} \\ \sigma \leftarrow \mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}_b, m)\end{array}\right] \leq \frac{1}{2} + \mathsf{negl}(\lambda),$$

where the oracles $\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{HKG}}$*, and* $\mathcal{O}_{\mathsf{Sign}}$ *are defined as in Definition 4.*

Definition 7 (Unlinkability). *An* SPC *group signature scheme* $(\mathsf{Gen}, \mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify}, \mathsf{Open})$ *satisfies client-client unlinkability if for any PPT adversary* $\mathcal{A}$*,* $\lambda \in \mathbb{N}$*,* $D \subseteq \mathcal{M}$*, and messages* $m_0, m_1 \in \mathcal{M}$*, it holds that with probability* $1 - \mathsf{negl}(\lambda)$ *over* $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Gen}(1^\lambda, D), (\mathsf{pk}_0, \mathsf{sk}_0) \leftarrow \mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle$*,* $(\mathsf{pk}_1, \mathsf{sk}_1) \leftarrow \mathsf{KeyGen}\langle \mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle$*,*

$$\Pr\left[\mathcal{A}^{\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{HKG}}, \mathcal{O}_{\mathsf{Sign}}}\begin{pmatrix}\mathsf{mpk}, \mathsf{pk}_0, \\ \mathsf{pk}_1, \sigma_0, \sigma_1\end{pmatrix} = b \;:\; \begin{array}{r} \sigma_0 \leftarrow \mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}_0, m_0) \\ b \leftarrow \{0,1\} \\ \sigma_1 \leftarrow \mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}_b, m_1)\end{array}\right] \leq \frac{1}{2} + \mathsf{negl}(\lambda),$$

where the oracles $\mathcal{O}_{\mathsf{AKG}}, \mathcal{O}_{\mathsf{HKG}}$*, and* $\mathcal{O}_{\mathsf{Sign}}$ *are defined as in Definition 4.*

4.2 Generic construction

We show how to construct an SPC group signature scheme generically from an SPC encryption scheme that satisfies certain properties, plus a few standard cryptographic tools. Our construction is given in the random oracle model, though we note that if we were willing to assume an additional *simulation-soundness* property of the ZK-NIAoK, then we would not require a random oracle. It is presented in Protocol 7.

Ingredients:

- An SPC encryption scheme $\Pi_{\mathsf{SPCE}} = (\mathsf{SPCE.Gen}, \mathsf{SPCE.Enc}, \mathsf{SPCE.Dec})$ that satisfies *perfect ϵ-correctness*, *security against outsiders*, and either *bounded-set* security or *authenticated-set* security (Section 3).
- A one-way relation $(\mathcal{R}, \mathcal{R}.\mathsf{Gen}, \mathcal{R}.\mathsf{Sample})$ (Section 2.1). Let $\mathcal{P}$ denote the set of instances.
- An EUF-CMA secure signature scheme $\mathsf{Sig} = (\mathsf{Sig.Gen}, \mathsf{Sig.Sign}, \mathsf{Sig.Verify})$ with message space $\mathcal{P}$ (Section 2.1).
- A ZK-NIAoK scheme $\mathsf{ZK} = (\mathsf{ZK.Setup}, \mathsf{ZK.Prove}, \mathsf{ZK.Verify})$ in the common random string model for general NP relations (Section 2.2).
- A random oracle H.

4.3 An efficient instantiation

We now summarize our approach for a concretely efficient instantiation of the above generic template, based on constructions of SPC encryption schemes from Section 3. Note that we will need a concretely efficient instantiation of a zero-knowledge argument system that can be used to prove statements that involve verifying signatures and the correctness of SPC encryption. Our goal here is to avoid *non-black-box* use of the cryptography needed for signatures and SPC encryption. Thus, we use *bilinear maps*, and make use of the Groth-Sahai proof system [24], which can efficiently prove statements that involve certain operations in pairing groups.[11] We combine the GS proof system with the use of *structure-preserving signatures* [1], which support messages, verification keys, and signatures that consist solely of group elements. We provide details of the scheme, implementation and benchmarking in the full version [8].

Our construction will make use of a bilinear map $\mathcal{G} = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{T}, e, g_1, g_2)$ where the SXDH assumption is assumed to hold, as described in Section 2.3. The group signature scheme $\Pi_{\mathsf{SPCGS}}[\mathcal{M}, \mathbb{G}_1, n, \epsilon]$ will have an arbitrary message space $\mathcal{M}$ and identities consisting of group elements in $\mathbb{G}_1$. The four ingredients are instantiated as follows.

[11] We remark that, although GS proofs only satisfy *partial* knowledge extraction (see Section 2.3), this is sufficient for our construction. Indeed, the signatures extracted in order to show ϵ-traceability and unframeability only consist of group elements, and the one-way relation witness extracted during the proof of unframeability is also a group element.

$\Pi_{\mathsf{SPCGS}}[\mathcal{M}, \mathcal{P}, n, \epsilon]$

Parameters: message space $\mathcal{M}$, identity space $\mathcal{P}$, set size n, correctness parameter ϵ, and security parameter λ.

Setup: $\mathsf{pp} \leftarrow \mathcal{R}.\mathsf{Gen}(1^\lambda)$ and a random oracle H.

$\mathsf{Gen}(1^\lambda, D)$: run $(\mathsf{pk}_{\mathsf{SPCE}}, \mathsf{sk}_{\mathsf{SPCE}}) \leftarrow \mathsf{SPCE.Gen}(1^\lambda, D)$[a] and $(\mathsf{vk}_{\mathsf{Sig}}, \mathsf{sk}_{\mathsf{Sig}}) \leftarrow \mathsf{Sig.Gen}(1^\lambda)$. Set $\mathsf{mpk} := (\mathsf{pk}_{\mathsf{SPCE}}, \mathsf{vk}_{\mathsf{Sig}})$ and $\mathsf{msk} := (\mathsf{sk}_{\mathsf{SPCE}}, \mathsf{sk}_{\mathsf{Sig}})$.

$\mathsf{KeyGen}\langle\mathsf{GM}(\mathsf{msk}), \mathsf{C}\rangle$: the client C samples random coins s, computes $(\mathsf{pk}, w) := \mathcal{R}.\mathsf{Sample}(\mathsf{pp}; s)$, and sends pk to GM. GM parses msk as $(\mathsf{sk}_{\mathsf{SPCE}}, \mathsf{sk}_{\mathsf{Sig}})$ and then computes and sends $\sigma_{\mathsf{id}} \leftarrow \mathsf{Sig.Sign}(\mathsf{sk}_{\mathsf{Sig}}, \mathsf{pk})$. C sets $\mathsf{sk} := (s, \sigma_{\mathsf{id}})$.

$\mathsf{Sign}(\mathsf{mpk}, \mathsf{sk}, m)$: parse mpk as $(\mathsf{pk}_{\mathsf{SPCE}}, \mathsf{vk}_{\mathsf{Sig}})$ and sk as $(s, \sigma_{\mathsf{id}})$, compute $(\mathsf{pk}, w) := \mathcal{R}.\mathsf{Sample}(\mathsf{pp}; s)$, sample random coins r, and compute $\mathsf{ct} := \mathsf{SPCE.Enc}(\mathsf{pk}_{\mathsf{SPCE}}, m, \mathsf{pk}; r)$. Let $\mathsf{crs} := H(m, \mathsf{ct})$, and compute $\pi \leftarrow \mathsf{ZK.Prove}(\mathsf{crs}, (\mathsf{pp}, \mathsf{pk}_{\mathsf{SPCE}}, \mathsf{vk}_{\mathsf{Sig}}, m, \mathsf{ct}), (\mathsf{pk}, s, w, \sigma_{\mathsf{id}}, r))$ for the relation that checks that

- $\mathsf{ct} = \mathsf{SPCE.Enc}(\mathsf{pk}_{\mathsf{SPCE}}, m, \mathsf{pk}; r)$,
- $(\mathsf{pk}, w) = \mathcal{R}.\mathsf{Sample}(\mathsf{pp}; s)$,
- and $\mathsf{Sig.Verify}(\mathsf{vk}_{\mathsf{Sig}}, \mathsf{pk}, \sigma_{\mathsf{id}})$.

Output $\sigma := (\mathsf{ct}, \mathsf{crs}, \pi)$.

$\mathsf{Verify}(\mathsf{mpk}, m, \sigma)$: parse mpk as $(\mathsf{pk}_{\mathsf{SPCE}}, \mathsf{vk}_{\mathsf{Sig}})$ and σ as $(\mathsf{ct}, \mathsf{crs}, \pi)$, check that $H(m, \mathsf{ct}) = \mathsf{crs}$ and if so output $\mathsf{ZK.Verify}(\mathsf{crs}, (\mathsf{pp}, \mathsf{pk}_{\mathsf{SPCE}}, \mathsf{vk}_{\mathsf{Sig}}, m, \mathsf{ct}), \pi)$.

$\mathsf{Open}(\mathsf{msk}, \sigma)$: parse msk as $(\mathsf{sk}_{\mathsf{SPCE}}, \mathsf{sk}_{\mathsf{Sig}})$ and σ as $(\mathsf{ct}, \mathsf{crs}, \pi)$, and output $\mathsf{SPCE.Dec}(\mathsf{sk}_{\mathsf{SPCE}}, \mathsf{ct})$.

[a] If the SPC encryption scheme satisfies authenticated-set security, this will be an interactive procedure between GM and Auth.

Fig. 7. Generic construction of SPC group signatures.

- SPC encryption: Either the scheme $\Pi^{\mathsf{BS-PC}}_{\mathsf{SPCE}}[\mathcal{M}, \mathbb{G}_1, n, \epsilon]$ or $\Pi^{\mathsf{AS-PC}}_{\mathsf{SPCE}}[\mathcal{M}, \mathbb{G}_1, n, \epsilon]$ from Section 3.
- One-way relation: The Diffie-Hellman relation in $\mathbb{G}_1$, where $\mathcal{R}$ is the set of tuples $(g, g^\alpha, g^\beta, g^{\alpha\cdot\beta}) \in \mathbb{G}_1^4$. $\mathcal{R}.\mathsf{Gen}$ outputs $(g, g^\alpha) = (g, h)$, and $\mathcal{R}.\mathsf{Sample}$ chooses randomness β and outputs (g^β, h^β). This relation is one-way from the hardness of the computational Diffie-Hellman problem in $\mathbb{G}_1$.
- Signature scheme: The structure-preserving signature scheme from [1].
- ZK-NIAoK: The Groth-Sahai proof system (Section 2.3).

Details of our concretely efficient scheme can be found in the full version [8].

Acknowledgments

First, second and fourth authors were supported in part by DARPA under Agreement No. HR00112020026, AFOSR Award FA9550-19-1-0200, NSF CNS Award 1936826, and research grants by the Sloan Foundation, and Visa Inc. The first and the fourth author were also supported by a grant from the CLTC. The third author was supported in part by NSF CNS-1814919, NSF CAREER 1942789, Johns Hopkins University Catalyst award, AFOSR Award FA9550-19-1-0200, Office of Naval Research Grant N00014-19-1-2294, JP Morgan Faculty Award, and research gifts from Ethereum, Stellar and Cisco. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government or DARPA.

References

1. Abe, M., Fuchsbauer, G., Groth, J., Haralambiev, K., Ohkubo, M.: Structure-preserving signatures and commitments to group elements. In: Rabin, T. (ed.) CRYPTO 2010. LNCS, vol. 6223, pp. 209–236. Springer, Heidelberg (Aug 2010). https://doi.org/10.1007/978-3-642-14623-7_12
2. Abelson, H., Anderson, R., Bellovin, S.M., Benaloh, J., Blaze, M., Callas, J., Diffie, W., Landau, S., Neumann, P.G., Rivest, R.L., Schiller, J.I., Schneier, B., Teague, V., Troncoso, C.: Bugs in our pockets: The risks of client-side scanning (2021). https://doi.org/10.48550/ARXIV.2110.07450
3. Alamati, N., Branco, P., Döttling, N., Garg, S., Hajiabadi, M., Pu, S.: Laconic private set intersection and applications. In: Theory of Cryptography Conference. pp. 94–125. Springer (2021). https://doi.org/10.1007/978-3-030-90456-2_4
4. Ananth, P., Jain, A., Jin, Z., Malavolta, G.: Pre-constrained encryption. In: ITCS. LIPIcs, vol. 215, pp. 4:1–4:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2022). https://doi.org/10.4230/LIPIcs.ITCS.2022.4
5. Apple: CSAM Detection Technical Summary (2021)
6. Attema, T., Cramer, R., Fehr, S.: Compressing proofs of k-out-of-n partial knowledge. In: Malkin, T., Peikert, C. (eds.) CRYPTO 2021, Part IV. LNCS, vol. 12828, pp. 65–91. Springer, Heidelberg, Virtual Event (Aug 2021). https://doi.org/10.1007/978-3-030-84259-8_3
7. Barak, B., Goldreich, O., Impagliazzo, R., Rudich, S., Sahai, A., Vadhan, S.P., Yang, K.: On the (im)possibility of obfuscating programs. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 1–18. Springer, Heidelberg (Aug 2001). https://doi.org/10.1007/3-540-44647-8_1
8. Bartusek, J., Garg, S., Jain, A., Policharla, G.V.: End-to-end secure messaging with traceability only for illegal content. Cryptology ePrint Archive, Paper 2022/1643 (2022), https://eprint.iacr.org/2022/1643
9. Bellare, M., Micciancio, D., Warinschi, B.: Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 614–629. Springer, Heidelberg (May 2003). https://doi.org/10.1007/3-540-39200-9_38
10. Bhowmick, A., Boneh, D., Myers, S., Tarbe, K., Talwar, K.: The apple psi system (2021)

11. Boyle, E., Ishai, Y., Pass, R., Wootters, M.: Can we access a database both locally and privately? In: TCC (2). Lecture Notes in Computer Science, vol. 10678, pp. 662–693. Springer (2017). https://doi.org/10.1007/978-3-319-70503-3_22
12. Canetti, R., Holmgren, J., Richelson, S.: Towards doubly efficient private information retrieval. In: TCC (2). Lecture Notes in Computer Science, vol. 10678, pp. 694–726. Springer (2017). https://doi.org/10.1007/978-3-319-70503-3_23
13. Canetti, R., Kaptchuk, G.: The broken promise of apple's announced forbidden-photo reporting system – and how to fix it (2021), https://www.bu.edu/riscs/2021/08/10/apple-csam/
14. Chaum, D., Pedersen, T.P.: Wallet databases with observers. In: Brickell, E.F. (ed.) CRYPTO'92. LNCS, vol. 740, pp. 89–105. Springer, Heidelberg (Aug 1993). https://doi.org/10.1007/3-540-48071-4_7
15. Chaum, D., van Heyst, E.: Group signatures. In: Davies, D.W. (ed.) EUROCRYPT'91. LNCS, vol. 547, pp. 257–265. Springer, Heidelberg (Apr 1991). https://doi.org/10.1007/3-540-46416-6_22
16. Cramer, R., Damgård, I., Schoenmakers, B.: Proofs of partial knowledge and simplified design of witness hiding protocols. In: Desmedt, Y. (ed.) CRYPTO'94. LNCS, vol. 839, pp. 174–187. Springer, Heidelberg (Aug 1994). https://doi.org/10.1007/3-540-48658-5_19
17. Dodis, Y., Grubbs, P., Ristenpart, T., Woodage, J.: Fast message franking: From invisible salamanders to encryptment. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018, Part I. LNCS, vol. 10991, pp. 155–186. Springer, Heidelberg (Aug 2018). https://doi.org/10.1007/978-3-319-96884-1_6
18. Emura, K., Hanaoka, G., Kawai, Y., Matsuda, T., Ohara, K., Omote, K., Sakai, Y., Megias, D.: Group signatures with message-dependent opening: Formal definitions and constructions. Sec. and Commun. Netw. **2019** (jan 2019). https://doi.org/10.1155/2019/4872403, https://doi.org/10.1155/2019/4872403
19. Federal Bureau of Investigation: Going Dark: Are Technology, Privacy, and Public Safety on a Collision Course? [speech] (2014)
20. Fiat, A., Shamir, A.: How to prove yourself: Practical solutions to identification and signature problems. In: Odlyzko, A.M. (ed.) CRYPTO'86. LNCS, vol. 263, pp. 186–194. Springer, Heidelberg (Aug 1987). https://doi.org/10.1007/3-540-47721-7_12
21. Green, M.: Thinking about "traceability" (2021), https://blog.cryptographyengineering.com/2021/08/01/thinking-about-traceability/
22. Green, M., Kaptchuk, G., Laer, G.V.: Abuse resistant law enforcement access systems. In: EUROCRYPT (3). Lecture Notes in Computer Science, vol. 12698, pp. 553–583. Springer (2021). https://doi.org/10.1007/978-3-030-77883-5_19
23. Groth, J.: Fully anonymous group signatures without random oracles. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 164–180. Springer, Heidelberg (Dec 2007). https://doi.org/10.1007/978-3-540-76900-2_10
24. Groth, J., Sahai, A.: Efficient non-interactive proof systems for bilinear groups. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 415–432. Springer, Heidelberg (Apr 2008). https://doi.org/10.1007/978-3-540-78967-3_24
25. Grubbs, P., Arun, A., Zhang, Y., Bonneau, J., Walfish, M.: Zero-knowledge middleboxes. In: USENIX Security Symposium. pp. 4255–4272. USENIX Association (2022)
26. Grubbs, P., Lu, J., Ristenpart, T.: Message franking via committing authenticated encryption. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part III. LNCS, vol. 10403, pp. 66–97. Springer, Heidelberg (Aug 2017). https://doi.org/10.1007/978-3-319-63697-9_3

27. Issa, R., Alhaddad, N., Varia, M.: Hecate: abuse reporting in secure messengers with sealed sender. In: 31st USENIX Security Symposium (USENIX Security 22). pp. 2335–2352 (2022)
28. U.S. Senate Committee on the Judiciary: Graham, Blumenthal, Hawley, Feinstein Introduce EARN IT Act to Encourage Tech Industry to Take Online Child Sexual Exploitation Seriously (2020)
29. U.S. Department of Justice: Attorney General Barr Signs Letter to Facebook From US, UK, and Australian Leaders Regarding Use of End-To-End Encryption (2019)
30. U.S. Department of Justice: International Statement: End-To-End Encryption and Public Safety (2020)
31. Kamara, S., Knodel, M., Llansó, E., Nojeim, G., Qin, L., Thakur, D., Vogus, C.: Outside looking in (2021)
32. Kulshrestha, A., Mayer, J.R.: Identifying harmful media in end-to-end encrypted communication: Efficient private membership computation. In: USENIX Security Symposium. pp. 893–910. USENIX Association (2021)
33. Libert, B., Joye, M.: Group signatures with message-dependent opening in the standard model. In: Benaloh, J. (ed.) Topics in Cryptology - CT-RSA 2014 - The Cryptographer's Track at the RSA Conference 2014, San Francisco, CA, USA, February 25-28, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8366, pp. 286–306. Springer (2014). https://doi.org/10.1007/978-3-319-04852-9_15, http://dx.doi.org/10.1007/978-3-319-04852-9_15
34. Liu, L., Roche, D.S., Theriault, A., Yerukhimovich, A.: Fighting fake news in encrypted messaging with the fuzzy anonymous complaint tally system (facts). Cryptology ePrint Archive, Report 2021/1148 (2021)
35. Matthew Green: An evaluation of the risks of client-side scanning (2022)
36. Naor, M., Reingold, O.: Number-theoretic constructions of efficient pseudo-random functions. J. ACM **51**, 231–262 (2004)
37. Newman, L.: The EARN IT Act Is a Sneak Attack on Encryption. Wired (2020)
38. Peale, C., Eskandarian, S., Boneh, D.: Secure complaint-enabled source-tracking for encrypted messaging. In: Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. pp. 1484–1506 (2021)
39. Pinkas, B.: The private set intersection (psi) protocol of the apple csam detection system (2021), https://decentralizedthoughts.github.io/2021-08-29-the-private-set-intersection-psi-protocol-of-the-apple-csam-detection-system/
40. Prokos, J., Jois, T.M., Fendley, N., Schuster, R., Green, M., Tromer, E., Cao, Y.: Squint hard enough: Evaluating perceptual hashing with machine learning. Cryptology ePrint Archive, Report 2021/1531 (2021)
41. Schnorr, C.P.: Efficient identification and signatures for smart cards. In: Brassard, G. (ed.) CRYPTO'89. LNCS, vol. 435, pp. 239–252. Springer, Heidelberg (Aug 1990). https://doi.org/10.1007/0-387-34805-0_22
42. Tyagi, N., Grubbs, P., Len, J., Miers, I., Ristenpart, T.: Asymmetric message franking: Content moderation for metadata-private end-to-end encryption. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part III. LNCS, vol. 11694, pp. 222–250. Springer, Heidelberg (Aug 2019). https://doi.org/10.1007/978-3-030-26954-8_8
43. Tyagi, N., Miers, I., Ristenpart, T.: Traceback for end-to-end encrypted messaging. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. pp. 413–430 (2019). https://doi.org/10.1145/3319535.3354243