Frontdoors, Not Backdoors: Accountable Anonymity for National Digital Identity

Shailesh Mishra
shailesh.mishra@epfl.ch
EPFL
Switzerland

Roxanne Chevalley
roxanne.chevalley@epfl.ch
EPFL
Switzerland

Martin Burkhart
martin.burkhart@armasuisse.ch
armasuisse Science and Technology
Switzerland

Bryan Ford
bryan.ford@epfl.ch
EPFL
Switzerland

***Abstract*—Anonymous credentials (ACs) let users prove identity claims with strong privacy guarantees, and national systems built on them — EUDI and the Swiss e-ID — go live soon. Due to these guarantees, however, AC systems cannot detect *proof proxy attacks*, where malicious users "lend" their identity by presenting their credentials on behalf of other users, exposing a fundamental gap between privacy and accountability. For instance, users could sell anonymous legal age proofs online and get away with it. This lack of accountability will act as an obstacle for AC adoption on grounds of "national safety", as exemplified by the recurring push towards chat control in the EU. Consequently, issuing authorities may well settle for less-than-ideal privacy guarantees of current solutions (e.g., batch issuance of one-show credentials).**

In this paper, we advocate the need for integrated accountability in ACs and introduce the *cryptographic forensic trail* (CFT). A CFT is a randomized encryption of a user's identity that they produce alongside each credential presentation; only when credential misuse, e.g., a proof proxy attack, is detected can the CFT be decrypted for accountability. The design of CFT is based on a legal process requiring probable cause for anonymity revocation. It enforces separation of powers between law enforcement, a judicial body, and a digital privacy advocate (NGO) using privacy-enhancing technologies. The protocol mimics checks and balances of a healthy democracy, in which neither law enforcement nor justice can track people as they will. Even if both branches colluded, the NGO can detect the misuse and block further use.

We implement a prototype of CFT based on both hardware-compatible and zero-knowledge-friendly elliptic curves. Our evaluations show that CFT only adds about 20% computation overhead to credential shows on the user's end, demonstrating the feasibility of deployments in wallets on mobile phones.

I. Introduction

Anonymous credential or AC systems, proposed more than 40 years ago by Chaum [1], provide maximal privacy for users when proving claims about themselves. A seminal AC design was provided by Camenisch and Lysyanskaya [2], which provided concrete desirable privacy properties: (i) *unlinkability* ensures that different shows by a user cannot be linked together as originating from the same user; (ii) *unforgeability* ensures that users cannot manipulate attributes in credentials; (iii) *non-transferability* aims to prevent a user from transferring (part of) their credential to another user, and; (iv) *anonymity revocation* revokes a user's anonymity if they misuse their credential. These properties, if achieved simultaneously, make AC an ideal candidate for a private and accountable identity system.

Recently, ACs have moved ever so close to being deployed: Switzerland and the EU are both planning to roll out privacy-preserving digital-identity ecosystems in 2026 or early 2027 [3], [4]. These national identity projects are AC-inspired, supporting selective disclosure of claims and aiming for unlinkable presentations. However, the actual level of privacy still lags behind what would be expected from an ideal AC system. The Swiss e-ID will support selective disclosure and builds on single-use key pairs (batch issuance) [5]. While this provides unlinkability for verifier-verifier collusion, credential shows are linkable if issuers collude with verifiers. The European Data Protection Supervisor notes that AC technology could enhance privacy, although it has not yet been widely adopted [6]. Both initiatives therefore converge on the goal of secure, interoperable digital identities with selective disclosure, but fall short of providing maximum privacy yet [7], [8]. Deployment of advanced cryptographic primitives, such as BBS signatures [9]–[11] or zero-knowledge proofs (ZKP), is discussed [8], [12]–[14], but currently blocked by a lack of support in secure enclaves on mobile phones and doubts regarding practicality [15]. We expect that standardization and advances in ZKP research [12], [16] will likely weather away these obstacles in a few years.

Yet, there is another obstacle to adopting stronger privacy with national digital identity projects: users could no longer be held accountable for misusing their credentials due to unlinkability. Specifically, a malicious user can execute a *proof proxy attack*, where they show their credential on behalf of another (illegitimate) user, and can go unpunished due to the lack of accountability. For example, a malicious user could sell checks for legal age, which underage citizens could use to access gambling or erotic websites. The general availability of anonymous age checks at large scales would undermine age protection laws in general [17]. Unless addressed, these attacks provide a motivation to governmental authorities to settle with less privacy-preserving solutions planned for roll-out in 2026 or early 2027, which still allow tracking misuse if required. This however, would be a bad trade-off, leaving the vast majority of honest users with suboptimal privacy. Then again, if strong privacy would actually be deployed, governments would turn to undermining users' privacy at a more fundamental level through backdoors [18], e.g., bypassing end-to-end encryption. Recent examples are chat control in the EU [19], [20] and

Switzerland driving away Proton [21].

Hence, we argue that *if we want strong privacy to survive politically, accountability is a necessity*, and cryptographers should design it before politicians do. As mentioned above, this is not an entirely new idea. Camenisch and Lysyanskaya proposed anonymity revocation for holding users accountable already 25 years ago, not as an afterthought, but as one of the important features [2]. Since then, however, accountability has received comparatively little attention in ACs: with some exceptions [22]–[24], the long line of further works [25]–[29] has primarily focused on performance, and stronger privacy. Alternative primitives, such as group signatures and their variants (see Table I), already provide accountability by letting a signature be opened; yet none of them is *standardization compliant*: none fits the technical stack that deployed e-ID systems have already fixed [30], [31]—credentials in the mDoc/mDL formats presented over OID4VP, and signing keys confined to mobile secure enclaves that today support only ECDSA. These are hard external constraints, and they rule out the accountability primitives proposed so far.

Hence, we present the *cryptographic forensic trail* (CFT), an accountability solution for ACs in a governmental identity framework, based on a democratic anonymity revocation committee. The CFT is a randomized encryption of the user's identity that is attached to every credential show. It is used for revoking anonymity if users misuse their credentials, e.g. for proof proxy attacks. Our design adopts a decentralized approach, unlike [2], which uses a central component. We strive towards the notion of *conditional anonymity revocation*, akin to *conditional anonymity* introduced by Kilian and Petrank [32]: a user's anonymity can be revoked *only if* the user is part of an investigation based on probable cause. Our goal is to emulate the legal process of "searching an individual" in the physical world. In particular, our design requires the law enforcement body (the police) to obtain a *digital search warrant* from the judicial body (the judge) before even initiating the anonymity revocation protocol. To make our approach more democratic, we further decentralize our system by including a *non-governmental organization* (NGO). The NGO ensures that the governmental bodies, the police and the judge, do not misuse the system for tracking benign users.

The CFT is designed in Section IV, implemented in Section VI, and evaluated in Section VII. The main bottleneck of CFT is a zero-knowledge proof (ZKP) provided by the user on a smartphone. Using state-of-the art ZKP libraries, we demonstrate its practical feasibility: CFT generation only adds about 20% computation overhead to credential shows and the revocation committee can perform up to 40 anonymity revocations per second. Moreover, we discuss challenges regarding governance, committee composition, operation and legal questions around accountable AC systems, as well as possible future extensions (Section V).

The paper makes the following contributions:

- Advocating accountability as a necessary and integral part of ACs deployed in real systems, leading to the notion of *conditional anonymity revocation*.
- Thorough discussion of *proof proxy attacks* that can pose a major threat to practical digital ID systems.
- The design, implementation, and evaluation of the *cryptographic forensic trail*, an accountable anonymity solution.

II. The Problem: Proof Proxy Attacks

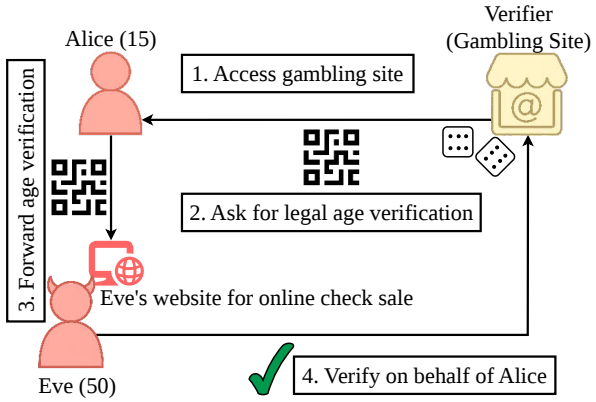

Fig. 1: Proof proxy attacks on electronic IDs: Alice (age 15) is not legally allowed to access a gambling site, but she is able to bypass the age check since Eve (age 50) proves legal age on Alice's behalf.

Before designing the CFT system, we discuss an important underlying problem – proof proxy attacks - and show why existing countermeasures fail to work in general scenarios.

We consider the case where the electronic-ID (e-ID) system is based on AC and ensures strong user privacy. Without additional measures, such a system enables misuse of credentials by proxying proof presentations and getting away with it [41]. For the specific use case of anonymous age verification, researchers have pointed out this potential misuse in an open letter addressed to politicians worldwide [17], calling for a moratorium on age verification legislation until harms are better understood [42] and technical aspects are clarified.

As illustrated in Fig. 1, malicious users could sell anonymous checks on the internet (e.g. for legal age, nationality, personhood [43]) by having people forward the verifiers' requests to them. While device binding of private keys effectively prevents transferability of identities, it does not prevent proof proxying. That is, Eve remains in exclusive possession of her private key (it is not transferred to Alice), but Eve leases (or lends) her identity temporarily for proving a statement on behalf of Alice. Such attacks are not mere speculations or motivated reasoning: in the past, malicious users have abused anonymity of systems, such as narcotics trafficking on Silk Road [44], [45], drug sale on AlphaBay [46], and misuse of cryptocurrencies for illegal transactions [47], [48]. In the e-ID context, such misuse includes illegal sale of age checks, citizenship, personhood, medical certificates, university degrees, etc. A more decoupled variant of this attack is possible, e.g., by users creating age-verified accounts on stock and selling

Scheme	Strong Unlinkability	Standard. compliant	Anonymity Revocation with privacy			
			Multi-party opening	Revoke signature	Trace user ID	Trace user ID with thr.
National e-ID (Swiss e-ID, EUDI)						
Batch issuance-based (current) [30], [31]†	✘	✔	✘	NA	NA	NA
ZKP-based (future) [33]	✔	✔	✘	✘	✘	✘
Existing cryptographic primitives						
Anonymous Credentials [2], [23]	✔	✘	✘	✔	✘	✘
Group Signatures [9], [10], [34]‡	✔	✘	✘	✔	✘	✘
Threshold Group Signatures [35]	✔	✘	✔	✔	✔	✘
Traceable Signatures [36]–[38]	✔	✘	✘	✔	✔	✘
Hidden ID-Based Signatures [39]	✔	✘	✘	✔	✔	✘
Identity Escrow [32], [40]	✔	✘	✔	✔	✘	✘
Cryptographic Forensic Trail (This Work)	✔	✔	✔	✔	✔	✔

TABLE I: Comparison of National e-ID and accountability-related cryptographic schemes. ✔ = supported, ✘ = not supported, ∼ = partial/limited. *Anonymity Revocation with privacy* groups the ways a scheme can revoke anonymity while preserving the privacy of benign users: distributing the opening across several parties (multi-party opening), revoking a single signature, tracing the signatures behind a given identity, and tracing identities that appear more than a threshold. †: batch issuance of single use credentials fails to provide unlinkability if an issuer and a verifier collude and discussing revocability without unlinkability is redundant; ‡: the group signature scheme of [9] introduced the pairing-based construction later adapted into BBS+ [10] and modern BBS [11], which form the basis for unlinkable selective disclosure.

them later on the internet [49]. Legislation aims at protecting minors from adult content but also from adults joining age-restricted chats such as Roblox groups. With emerging markets and services around illegal age verification, such legislations would be rendered obsolete [17].

The Swiss and EU digital identity infrastructure, however, is envisioned to build the foundation for not just a single identity credential and a single use case like age verification, but an entire *ecosystem around digital credentials* [30], [50]. Issuers from various government offices and industry are invited to issue their own domain-specific credentials. Standards for health, financial, and education sectors are currently being developed. Some of these credentials may not require hardware device binding, which simplifies key extraction and misuse even more. If not protected carefully, the next generation of agentic AI systems [51] could obtain credentials proving personhood [43], legal age, citizenship, health insurance coverage, and possession of university degrees. Markets for resold age-verified accounts [49] already demonstrate that the business model is commercially viable at small scale.

How do the identity issuers deal with these threats? A straightforward way of preventing such attacks is to include the verification of holder binding into the design. This is usually done by visually matching a picture stored on the credential with the presenting person [52]. Such an approach is suitable for EUDI and Swiss e-ID because they include a biometric user picture and are device bound, i.e., keys are stored in the secure enclave of the user's smartphone. So whenever the owner presents the e-ID to a human verifier, e.g., when buying a beer in a bar, the verifier can verify holder binding by doing in-person biometric authentication. However, *many online use cases will neglect the verification of holder binding*. Reasons for this are complexity, costs, usability and anonymity. While image- and video-based identification solutions could do the job in theory, they are costly and require biometric data to be processed, which raises data protection issues. Also, users may hesitate to upload their pictures to websites requiring legal age verification. Moreover, cameras might not be available on all devices (e.g., self-checkout systems in shops or restricted office clients). These issues are the very reasons why proponents call for modern e-ID systems not requiring the upload of physical ID card pictures and selfie videos to third-party service providers.

The Federal Office of Information Technology (FOITT) of Switzerland, which operates the Swiss e-ID, acknowledges the threat of proof proxy attacks [53] (chapter 6.9). The impact of the threat is rated as `HIGH`, while the likelihood is considered to be `LOW`, resulting in an overall `MEDIUM` risk level. One option the FOITT discusses for mitigation is to "track down the credential in question and revoke it", which implies that there will be no unlinkability of credential shows. In fact, with the current system using batch issuance [5], FEDPOL, the federal police and issuer of the Swiss e-ID, would actually be able to track down the owner by collaborating with verifiers. Such a design choice is problematic, since it puts all users under the threat of getting tracked, even if there is no apparent cause for doing so. However, if FEDPOL would be progressing to a more privacy-preserving system, e.g., by adopting

ZKPs [12], [54], the ability to "track down" malicious actors would vanish.

A possible countermeasure against QR code forwarding is the FIDO CTAP (Client to Authenticator Protocol) hybrid flow discussed in the EUDI ARF [30]. With CTAP, the browser session with the verifier and the wallet app on the mobile phone are linked through a BLE (Bluetooth Low Energy) tunnel. Hence, the wallet presenting the proof must be physically close (within BLE radius) to the device running the browser session. This raises the bar for malicious users, as simply forwarding the QR code to a remote attacker would no longer work. However, Alice could still share her browser session or establish a Bluetooth tunnel with Eve. Moreover, CTAP could be difficult to roll out consistently for heterogeneous platforms, as it requires integration of browser APIs, wallets, and OS-specific bluetooth drivers. If some consumer setups are not supported (e.g., alternative browsers on Linux), Alice can simply avoid CTAP. Also, CTAP only protects synchronous verifications. If accounts are age-verified on stock and sold later, it does not help. Last but not least, BLE comes with a long list of security vulnerabilities and privacy issues [55]. Mandating all citizens to turn on BLE when using their digital identities may be a bad tradeoff for catching a small set of malicious users.

III. Overview

This section describes the high-level workflow and scenarios for the cryptographic forensic trail (CFT) usage (Sections III-A and III-B), the composition of the anonymity revocation committee (Section III-C), the system model (Section III-D), the threat model (Section III-E), and the desired properties such a system should achieve (Section III-E). Broadly speaking, a CFT is an encrypted identity attached to every credential show, and in case the credential holder is involved in any unlawful activities, such as selling online anonymous checks, this trail is used to revoke the anonymity of the holder. The motivation for introducing the forensic trail[1] parallels that of anonymity revocation in idemix [2], [25], serving a mechanism for enabling accountability in anonymous credentials.

A. Workflow

Figure 2 depicts the high-level workflow of the use of CFT for ensuring accountability in the e-ID system. Our design requires the law enforcement body (the police), the judicial body (the judge), and an NGO to work together for anonymity revocation, thereby decentralizing the accountability framework and making it impossible for a single entity to revoke the anonymity of any user as they will. We discuss the decentralization aspect later in Section III-C.

We discuss the workflow w.r.t. the proof proxy attack (Section II). Firstly, *nothing* related to the buyer's identity (Alice in Figure 1) is stored in a CFT; we only aim to hold (malicious) users that sell their identities accountable (Eve in Figure 1); Our solution requires Eve to create a CFT and prove its correctness to the verifier for each credential show (1). The verifier verifies the correctness of the credential show as well as the CFT, and thereafter, stores the CFT (2). Once Eve's activities come to light, the police obtains a search warrant[2] (3) and use it to retrieve the relevant CFT(s) from the verifier (4). Then, the police, the judge, and the NGO run a protocol together to revoke the anonymity of the user corresponding to the acquired CFT (5).

B. Scenarios for Anonymity Revocation

The number of relevant CFT(s) (4) and the exact semantics of the revocation protocol (5) depend on the amount of information available for investigation, which lead us to three scenarios of anonymity revocation.

1) *Search by request ID.* The police know the specific CFT whose anonymity should be revoked to find out the person behind the anonymous check sale service (Eve). This may be the case when a public service selling age checks is accessible and police perform a sting operation-like procedure [61] by buying an age proof themselves (Section IV-E). This scenario is akin to matching a record in a database and hence, the least leaky setting as only one database record is accessed.
2) *Search by suspect ID.* In this scenario, the police have the ID of the suspect but do not know which requests belong to the suspect. This scenario becomes relevant when a whistleblower (or any other source) informs the police about Eve being potentially involved in suspicious activities or because the police learned Eve's ID from *Search by request ID* (Section IV-F). This scenario is akin to finding database records that match a query with a key condition and hence, is potentially more leaky than the request ID version as multiple database records are accessed.
3) *Search by activity threshold.* In this scenario, the police only have knowledge about someone (not the ID) executing a proof proxy attack. For instance, the police solely learn many underage citizens accessing gambling sites and nothing about Eve. Thus, the police will try to find the ID of a person that appears abnormally often in a time interval (Section IV-G). This scenario is akin to making a predicate query to the database and hence, is potentially more leaky than the previous two scenarios since we are not only touching multiple records but also computing over them (the predicate).

Note that these protocols are useful beyond the investigation of proof proxy attacks within the e-ID system itself. The "search by suspect ID" and "search by activity threshold" protocols can equally inform traditional criminal investigations — for instance, in the context of a court-authorized investigation into organized fraud, where matching e-ID activity to a known suspect provides material evidence. We consider

[1] We use "forensic trail", "trail", and "cryptographic forensic trail" interchangeably.

[2] Search warrants are a prerequisite for any criminal investigation under Swiss law [56, Arts. 269–279], the EU Law Enforcement Directive [57], its national implementations such as Germany's StPO [58, §§ 100a, 100g], and Article 8 of the ECHR [59], [60].

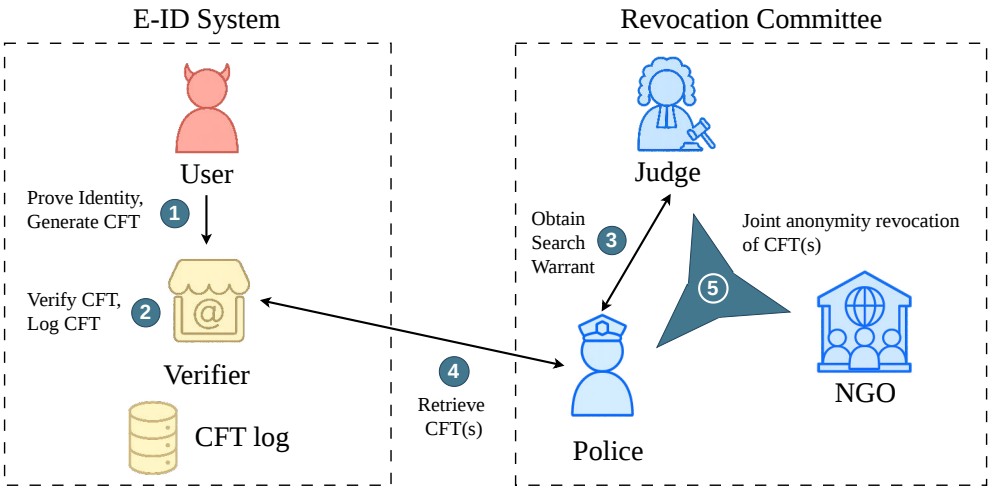

Fig. 2: Cryptographic forensic trail (CFT) workflow: user generates the CFT while showing her credential to the verifier (1), who verifies and logs the CFTs (2); after obtaining a digital search warrant (3), the police retrieve the relevant CFT record(s) (4). Finally, the revocation committee runs a protocol together to revoke the anonymity of the CFT(s) (5).

this a feature, not a bug: a national digital identity system must not stand above the law, granting its users exceptional immunity that does not exist for users of other digital or physical systems. As with banks, telecommunications, and physical identity documents, e-ID should integrate into the governmental framework for legal prosecution — but only under constraints that the governmental framework itself may not always provide. As detailed in the following sections, the CFT prevents misconduct through democratic checks and balances combined with cryptographic protocols that protect the privacy of all citizens, including suspects outside the scope of an active warrant. In the absence of such a transparent and constrained accountability mechanism, political pressure to introduce opaque backdoors becomes harder to resist (see Section III-F).

C. The Decentralized Revocation Committee

Idemix [25] assumes the existence of a *centralized revocation authority* for anonymity revocation, and presents anonymity revocation as an optional feature in anonymous credentials. In contrast, we hold that there must be a *decentralized* unit, which we refer to as the *revocation committee*, for anonymity revocation, and this committee must contain multiple independent entities for running the revocation protocol. Moreover, anonymity revocation should not be optional; it should be universally applicable to all users, but only triggered in cases of legal non-compliance. That is, anonymity revocation should be *conditional*, instead of optional.

But which parties should be involved in anonymity revocation? In this regard, an e-ID system should not re-invent new legal processes; instead it should emulate the law enforcement structure that has existed for centuries in healthy democracies. Specifically, in any democratic country, citizens have privacy rights until they violate some law. Even when the law enforcement body (the police) finds an indication for law violation (probable cause), it cannot simply override the right to privacy of the citizen. The police must convince the judicial authority (the judge) of a violation and get a signed search warrant. Only then, privacy of the citizen may be overridden. This structure has been established and has functioned in democracies for centuries [59], [62]. Hence, we hold that e-ID systems must emulate this process for accountability: the police should be able to search the credential shows of an individual only if the judge issues a (digital) search warrant against that individual.

A concern with only the police and the judge forming the revocation committee in the digital world is the possibility of covert tracking: if the police and the judge collude, then they can track all the credential users, and it is much easier to do so in the digital world than tracking people physically. Hence, the anonymity revocation committee must contain one or more organizations which are independent of government bodies, such as NGOs or international bodies, which can act as deterrents to governments going rogue (we use the term "NGOs" to refer to these independent parties).

a) NGOs for the revocation committee: Genuine independence from the state is the key criterion for selecting NGOs, since most civil society organizations receive at least some public funding. A pragmatic way to identify suitable candidates is to favor organizations that are member- or donor-supported, with a track record of publicly opposing surveillance legislation backed by their own governments. Candidates meeting this bar include noyb [63] and EDRi [64], which have repeatedly co-signed open letters against the EU CSA Regulation [65]; the membership-funded EFF [66]; AlgorithmWatch [67] and Digitale Gesellschaft [68] in Switzerland; and the digital-rights and scientific communities behind recent open letters on age assurance and the EUDI framework [17], [42]. Technical expertise, where needed, could come from organizations such as the Chaos Computer Club [69] or the Tor Project [70]. Combining several such organizations

across jurisdictions further reduces the risk that any single government can coerce the non-governmental party.

D. System Model

Our system, as outlined in Figure 2, consists of the revocation committee and the three entities which are governed by the national ID solution (EUDI, Swiss e-ID, etc.): (i) the issuer who issues credentials (not shown in Fig. 2); (ii) the user who obtains a credential from the issuer, and; (iii) the verifier who verifies the credential of the user before providing some service. The credentials contain attributes of the user, and one of the attributes is a technical identifier, which could be the national identity number (e.g., Social Security Number). The user owns a public key-secret key pair, which is required for generating signatures of knowledge. We assume that secret keys are bound to a Wallet Secure Cryptographic Device (WSCD) — either a local secure enclave on the user's phone or a remote HSM with per-user partitioning — as specified by the EUDI ARF [30] and the Swiss e-ID [31]. In both cases, keys are non-extractable, and an attacker can only generate CFTs that point to her own identity. Resale therefore reduces to selling access to one's own wallet (locally) or one's own remote-signing account (remotely); in both cases CFT correctly identifies the seller. For other credentials in the ecosystem with no hardware binding, leakage or theft would be relatively easy. In that case, CFT allows identification and validity revocation for the corrupted credentials. Whether Eve was a victim, having her keys stolen, or whether she sold the credentials deliberately, would have to be investigated.

In our system, the verifier needs to store the credential shows of (anonymous) users for a certain duration (e.g., six months or a year) (2). This is in line with many national data retention policies requiring phone and internet service providers to store call and connection metadata for some duration [71]–[75].

The anonymity revocation committee members also hold individual public key-secret key pairs. In a CFT deployment, each member organization maintains a computation node running the distributed revocation (5) with the computation nodes of the other members. Each computation node must have access to its owner's secret key. These secret keys must be stored securely, analogous to how certificate authorities protect their private keys, e.g. in a local HSM. As there will be dozens or even hundreds of people - e.g., the police investigators and judges – interacting with the system, we assume the member organizations perform proper access control on the computation nodes and model the required roles for each action in the anonymity revocation protocol. That is, only investigators may start a revocation process by requesting a search warrant, only appointed judges may issue the warrants, and only authorized NGO representatives may object to the revocation. Many government units dealing with sensitive information use smartcards to authenticate and authorize their personnel when accessing buildings, services, and data. Hence, an obvious option would be to leverage these smartcards (with personal private keys and PINs) for logging in and authorizing actions on the computation nodes. The privileges for anonymity revocation could be tied to group properties in the organization's active directory, e.g. implementing the policy that each member of the investigation unit is allowed to trigger an anonymity revocation and upload CFTs as input to the computation. The computation nodes start the actual revocation protocol only when all members have completed their preliminary tasks and provided their consent.

E. Threat Model

Our system aims to defend against a user who attempts to abuse the privacy guarantees of the e-ID by selling her anonymous credential shows. We assume that the issuer issues credentials correctly, i.e., the issuer signs the right set of claims and does not issue duplicate credentials to users.

We assume the identity system is built on anonymous credentials, so credential shows remain unlinkable even under cross-verifier and issuer-verifier collusion. Verifiers are expected to check proofs and log credential shows; to deter deviations, verifier logs can be audited at random, akin to cut-and-choose technique [76]. Furthermore, certification of verifiers and a trust registry - as envisioned by EUDI and Swiss e-ID - could guide users towards responsible verifiers and build incentives for verifiers to comply.

We assume that the wallet application held by users is untrusted. A "trusted wallet" assumption could be enforced via platform attestation (e.g., Google Play Integrity [77], Apple App Attest [78]), which certifies that the wallet app is authentic and runs on a secure device. The required long-term mTLS key, however, would itself act as a linkable identifier across presentations, breaking unlinkability.

We assume that *1-out-of-3* parties among the law enforcement body (the police), the judicial body (the judge), and the non-governmental body (the NGO) is honest, and the protocol (5) is aborted in case an honest party detects tracking attempts or deviation from the protocol by any party. For example, in an authoritarian country, the judge and the police might work together, or a corrupted NGO might collude with the police or the judge. In that case, the remaining honest party could stop the entire process. If collusion of more than two parties is feared, this could be addressed with more decentralization, e.g., involving more NGOs in the anonymity revocation protocol; although involving more parties would lead to more computation / communication costs. Such adaptations can be easily made to our main protocol, and hence, we describe our main solution based on the 1-out-of-3 assumption in Section IV.

For the CFT system, we require the following properties to hold:

Unlinkability: Just as identity presentations, distinct encrypted CFTs must not be linkable to users, not by (multiple) verifiers and not if issuers collude with verifiers.[3]

[3]Linking presentations through network metadata, such as browser attributes or IP addresses, is considered out of scope for this work. To prevent this, countermeasures such as private browsers and TOR could be deployed.

Conditional privacy: Law abiding users must enjoy the privacy guarantees provided by the ID system, whereas users misusing the system can have their anonymity revoked.
Scope restriction: The anonymity revocation must be restricted to the credential shows specified by the search warrant (3)(4).
Transaction binding: A CFT must be bound to a specific identity proof request by a specific verifier via a signature on the request tag and opening randomness, preventing false identity injection and replaying to different contexts.
Oppression resilience: In case a repressive regime is taking over and tries to control multiple (or all) entities, any honest party of the revocation committee must be able to stop the system, e.g., by erasing their secret key.

F. Isn't this a Backdoor?

The cryptographic community has documented the risks of exceptional-access mechanisms, so-called backdoors, in detail [79], [80], including unilateral control, catastrophic failure modes [81], and lawful-access infrastructure being compromised by unintended parties (e.g., Salt Typhoon) [82]. Backdoors are typically associated with several distinct properties listed in Table II, which rightly trigger a defensive reaction among security researchers [20].

However, the *CFT is categorically different from a backdoor*, exhibiting none of the backdoor properties (see Table II). While backdoors correspond to "keys under doormats" [79], the CFT is rather a frontdoor with multiple locks and no accessible keys. With the CFT, we are not trusting the state more to handle exceptional access. We are trusting no single party at all, including the police, the judiciary, and the NGO. This is the same logic by which democratic societies accept that police exist, without concluding that police can enter any home at any time. The answer to "police have too much power" has never been "abolish investigation". It has been separation of powers, checks and balances, probable cause, warrants, and oversight. The CFT is the cryptographic instantiation of that same answer.

Note that ACs without accountability mechanism are not analogous to a private home as we know it. Rather, it is analogous to *diplomatic immunity*, an exemption from law that society only grants in narrow, reciprocal cases. Extending this immunity-level of privacy to every credential show for every user and every transaction is a political proposition that will not survive contact with democratic legislatures. If accountability is not designed into systems, politicians will likely enforce it by weakening the systems — by building *actual* backdoors. Recent example for this are the recurring discussion about chat control in the EU (see Appendix A) but also Switzerland's new proposed surveillance laws that let Proton threaten to leave the country [21].

IV. Cryptographic Forensic Trail

In this section, we detail the core design of the cryptographic forensic trail (CFT). Our solution aims to achieve all the desired properties highlighted above (Section III-E). The main challenges when designing CFT are that first, the user is not trusted. Second, we want to maintain the strong level of unlinkability, resilient against collusion of issuers and verifiers. Thus, the CFT must not introduce data items that give away unlinkability. We hold up unlinkability even for those transactions of malicious users that are not direct subject of investigation. Third, the entities holding users accountable must themselves be held accountable in order to prevent unlawful anonymity revocation.

Figure 2 presents a high-level overview of our solution. We first describe how the CFT is created during credential show (Section IV-B) (1), and then, we discuss how the anonymity of a user can be revoked using the CFT (Section IV-D) (5).

A. Notations

Table III summarizes the notations used. We write group operations additively in a cyclic group $\langle G \rangle$ of prime order q (e.g., an elliptic-curve group) with generator G. Public keys, including the user's accountability public key ID_U (equivalently pk_U), are elements of $\langle G \rangle$, while secret keys and randomness (r_1, r_2) are scalars in $\mathbb{Z}_q$. The user holds a key pair (pk_U, sk_U) with $pk_U = sk_U \cdot G$; ID_U denotes pk_U and is the value encrypted in C_3. The secret key sk_U is hardware-bound (non-extractable), so that the binding signature C_4 also provides device binding. The values $(\mathrm{ts}, \mathrm{RID}, ID_V)$ are treated as public identifiers.

B. Cryptographic Forensic Trail Generation

Figure 3 depicts the steps involved in the generation of the cryptographic forensic trail (CFT) (1). On a high level, a cryptographic forensic trail (CFT) is a randomized encryption of the user's accountability public key (identifier) ID_U under a joint public key of the police, the judge, and the NGO. We instantiate the CFT using an ElGamal-style construction due to three main reasons: (i) since ElGamal is additively homomorphic in the public key, the user can encrypt in one step under $PK_{CFT} = PK_P + PK_J + PK_N$ (rather than nested encryption), keeping the corresponding proof of correct encryption compact; (ii) ElGamal naturally lives on elliptic-curve groups, which enables zk-friendly curve choices [83], [84] and can make the proof of correct encryption cheaper than RSA-based approaches, and; (iii) ElGamal-style operations commute with Pohlig-Hellman-style [85] blinding used in our two-phase conditional opening approach (Section IV-F), which lets the parties transform a probabilistic, unlinkable trail into a deterministic blinded identifier PID within the warrant scope without decrypting benign users.

a) CFT Content: The CFT is formed from four elements:

- $C_1 = r_1 G$: a first ElGamal ephemeral public key.
- $C_2 = r_2 G$: a second ElGamal ephemeral public key, used to derive the binding message at opening time.
- $C_3 = ID_U + r_1 PK_{CFT}$: the ElGamal ciphertext component that encrypts the user's public key ID_U under the joint public key PK_{CFT}.

Property	Backdoor	Cryptographic Forensic Trail
Power to act	One actor, unilaterally	Committee consisting of police, judge, and NGO
Scope of use	Potential for massive scaling: one key, many targets	Clearly defined scope in each warrant
Precondition	Technical capability or knowledge is the only gate; it can be leaked, stolen, or sold	Probable cause, judicial warrant, and NGO consent
Weaker security	Security of the system is deliberately weakened, e.g., patches retained, keys copied, or weak parameters selected	The core identity system is untouched; accountability is an add-on layer
Data confidentiality and integrity	Confidentiality, integrity, and availability properties of all user data are at risk	All transaction data remains secure; only metadata is de-anonymized
Security by obscurity	APIs, even their existence, are hidden from the public	Public governance
Abuse detection	No; use is invisible	Yes; the NGO can detect and halt systemic misuse
Capability revocable	No; once deployed, it persists	Yes; oppression-resilience through key destruction

TABLE II: What makes CFT categorically different from a backdoor.

TABLE III: Notations

Notation	Description
I, U, V	Issuer, User, Verifier
P, J	Police, Judge
N	Non-governmental organization (NGO)
ID_U	User's accountability public key (pk_U), used as an identifier, a group element $ID_U \in \langle G \rangle$
pk_U, sk_U	User key pair with $pk_U = sk_U \cdot G$; $ID_U = pk_U$; sk_U is hardware-bound and used to produce C_4
ID_V	ID of the verifier
RID	Identifier of the presentation request
ts	Timestamp associated with a presentation request
$cred_U$	Credential issued by the issuer to the user
$\vec{attrs}$	Attributes of the user
$\vec{attrs}_V$	Attributes that the user needs to show to the verifier
$\langle G \rangle$	Cyclic group of prime order q (e.g., an elliptic-curve group), written additively
q	(Prime) order of $\langle G \rangle$
G	Generator of $\langle G \rangle$
(PK_i, sk_i)	Public key - secret key pair of party i, where $i \in \{U, V, P, J, N\}$, with $PK_i \in \langle G \rangle$ and $sk_i \in \mathbb{Z}_q$
CFT	Cryptographic forensic trail
t	Event-binding tag derived from the public request transcript: $t = H(\text{ts}, \text{RID}, ID_V)$
SW	Search warrant
ID_{SW}	ID of the search warrant
`reason`	Reason for search warrant
`aux`	Auxiliary information related to the search warrant
`filter`	Predicate which filters which CFTs should be used in conditional unlinkability removal

- $C_4 = \mathsf{Sign}_{sk_U}(H(t, r_2 PK_{CFT}))$: a signature under the user's hardware-held secret key sk_U on the binding message $H(t, r_2 PK_{CFT})$, where the event tag is $t \leftarrow H(\text{ts}, \text{RID}, ID_V)$. Because sk_U never leaves the secure enclave, C_4 simultaneously binds the CFT to the request transcript and provides device binding: only the device that holds sk_U can produce a valid C_4. After opening, the revocation committee verifies $\mathsf{VerifySign}_{ID_U}(C_4, H(t, r_2 PK_{CFT}))$ using the recovered public key ID_U and $r_2 PK_{CFT}$ obtained from C_2 (Theorem 2). Because the binding message depends on $r_2 PK_{CFT}$, which is hidden in C_2 until threshold partial decryption, an adversary cannot verify candidate signatures offline: without knowing $H(t, r_2 PK_{CFT})$, brute-force attempts against C_4 are futile even if t is public.

b) CFT Proof: During the credential show, the user proves in zero-knowledge (to the verifier) that the CFT is well-formed and correctly bound to the public request transcript; we instantiate this proof as a zk-SNARK [86]. The full circuit description of the proof can be found in our anonymous implementation repository [4]. The predicate $\mathsf{VerifyCFT}()$ asserts that $C_1 = r_1 G$, $C_2 = r_2 G$, $C_3 = ID_U + r_1 PK_{CFT}$, and $\mathsf{VerifySign}_{ID_U}(C_4, H(t, r_2 PK_{CFT})) = 1$ for the user's identifier $ID_U = pk_U$.

c) CFT Verification: The user sends the CFT and the proof to the verifier, who must perform following tasks (2): (i) verify the zero-knowledge proof π_U and if the proof fails, it must abort the transaction; (ii) store the CFT along with the ID of the presentation request and a timestamp. In order to prevent verifiers from deviating from the two tasks, we require them to store the proofs and the required randomness for a certain duration, like six months to a year. These proofs do not leak anything about the provers and can be checked at random, akin to the cut-and-choose technique [76].

[4] https://anonymous.4open.science/r/accountability_for_acs-7670/prove_verify/zk-friendly/prove_verify/prove_verify.circom

C. Security of the CFT

We briefly argue event binding and the privacy/security properties of the trail. Formal games and reductions appear in Appendix E. We assume DDH in $\langle G \rangle$, model H as a random oracle, take the signature scheme used for C_4 as EUF-CMA, and take the presentation proof system to be sound and zero-knowledge.

a) Event binding.: Each show derives $t = H(\mathrm{ts}, \mathrm{RID}, ID_V)$ from the public request and forms $C_4 = \mathsf{Sign}_{sk_U}(H(t, r_2 PK_{CFT}))$ under the hardware-held sk_U. At presentation, t is a public input of π_U, so accepting the proof binds $(C_1, \ldots, C_4)$ to *this* verifier, request, and time: replaying the same (CFT, π_U) under another t fails verification, and soundness prevents adapting a valid show for one event into a valid show for another without the credential witness. At opening, the committee recomputes t from the logged metadata, derives $r_2 PK_{CFT}$ from C_2, recovers ID_U, and checks C_4; a trail swapped onto a different logged event fails that check. Producing a trail for a fresh event is covered by unforgeability below.

b) Indistinguishability and unlinkability.: Semantic security is cast as an IND-style game: the adversary picks two candidate identities, receives a challenge trail encrypting one of them, and may sample further honest trails for the same hidden identity.

Theorem 1 (Trail indistinguishability)**.** *Under DDH in $\langle G \rangle$ and modelling H as a random oracle, no PPT adversary wins the IND game with non-negligible advantage.*

Intuition. The reduction places a DDH challenge in (C_1, C_3); in the real case the transcript matches an honest trail, while in the random case C_3 is uniform and hides ID_U. Even knowing both candidate signing keys, verifying C_4 offline still requires $r_2 PK_{CFT}$, which is CDH-hard from public (C_2, PK_{CFT}). Fresh randomness then makes successive trails for the same user indistinguishable from trails of different users, and zero knowledge of π_U adds no linking information beyond the public inputs. Without all three committee keys, deciding cross-show identity equality is hard.

c) Unforgeability.: Anyone can form ElGamal components (C_1, C_2, C_3) for a known ID_U, the binding signature C_4 is the hard part.

Theorem 2 (Trail unforgeability)**.** *Assuming EUF-CMA of the signature scheme used for C_4, and assuming the adversary does not know sk_U, no PPT adversary can produce a valid trail for a fresh tag except with negligible probability.*

Intuition. A forgery for a new tag t' needs a valid signature on $H(t', r_2' PK_{CFT})$ under ID_U. Without the hardware-held sk_U, that reduces to EUF-CMA, even given many honest trails for other events.

Next, we discuss the steps involved in anonymity revocation using CFT in each of the three scenarios discussed in Section III-A.

Algorithm 1: CFTCreate

Parties: User U, Verifier V
Inputs: user's key pair (pk_U, sk_U) with $ID_U \leftarrow pk_U$, verifier's identifier ID_V, set of attributes to be shown $\vec{attrs}_V$, timestamp ts, request ID RID, public keys $\{PK_P, PK_J, PK_N\}$, predicate $f_V(\cdot)$

1: U derives an *event-binding* tag from the public request transcript:

$$t \leftarrow H(\mathrm{ts}, \mathrm{RID}, ID_V) \tag{1}$$

2: U generates the cryptographic forensic trail CFT as an ElGamal-style ciphertext with a signature binding component (where $PK_{CFT} \leftarrow PK_P + PK_J + PK_N$):

$$\begin{aligned} &\text{sample } r_1, r_2 \xleftarrow{\$} \mathbb{Z}_q, \\ &\quad m \leftarrow H(t, r_2 PK_{CFT}), \\ &\quad C_1 \leftarrow r_1 G, \qquad C_2 \leftarrow r_2 G, \\ &\quad C_3 \leftarrow ID_U + r_1 PK_{CFT}, \quad C_4 \leftarrow \mathsf{Sign}_{sk_U}(m), \\ &\quad CFT \leftarrow (C_1, C_2, C_3, C_4) \end{aligned} \tag{2}$$

3: U generates the zero-knowledge proof, π_U:

$$\begin{aligned} \pi_U = \mathsf{SNARK}\Big[\big(\vec{attrs}, ID_U, CFT, r_1, r_2, t, PK_{CFT}\big) : \\ \mathsf{VerifyCred}(\vec{attrs}) = 1 \\ \wedge\ \mathsf{f}_V(\vec{attrs}_V) = 1 \\ \wedge\ \mathsf{VerifyCFT}(PK_{CFT}, CFT, ID_U, t) = 1 \Big] \end{aligned} \tag{3}$$

4: U sends CFT and π_U to V.
5: V verifies the correctness of π_U:

$$\begin{aligned} V_{\pi_U} = \mathsf{VerifySNARK}(vk_{\mathsf{SNARK}}, \pi_U, PK_{CFT} \\ CFT, \vec{attrs}_V, t) \end{aligned} \tag{4}$$

6: **if** $V_{\pi_U} = 1$ **then**
7: $\quad V$ stores CFT and π_U
8: **else**
9: $\quad$ Abort

Fig. 3: Generation of CFT and Zero-Knowledge Proof

D. Anonymity Revocation using CFT

The anonymity revocation is run using three main subroutines: (i) search warrant issuance (3); (ii) obtaining relevant trails (4) and; (iii) partial decryption of relevant trails (5). These subroutines are needed for all three scenarios of Section III-A, but their exact inputs, outputs, and semantics vary. We provide concrete instantiations of the three subroutines in Figure 4 (on the right) with respect to the search by request ID scenario, which we discuss below:

- IssueWarrant: this subroutine enables the police to justify to the judge that a search warrant needs to be issued. After verifying the legitimacy of the need for a search, the judge issues a search warrant, which is a digital signature that ties the search warrant to a search ID, the specific reason why the warrant is being issued (`reason`), and some auxiliary information (`aux`). The auxiliary information specifies information relevant to the search such as the procedure of search, the number of CFTs whose anonymity should be revoked, or the verifiers that should be contacted.

Algorithm 2: CFTRevokeByRequestID

Parties: Police P, Judge J, NGO N, Verifier V
Inputs: `reason`, `aux`, ID_{SW}, key pairs (pk_i, sk_i) for $i \in \{P, J, N\}$

P finds a cause for search warrant issuance.
$\sigma_{SW} \leftarrow \mathsf{IssueWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW})$
$(CFT, ts, \mathrm{RID}, ID_V, \sigma_{SW,V}) \leftarrow \mathsf{ObtainTrail}(P, V, \sigma_{SW})$, where $CFT = (C_1, C_2, C_3, C_4)$
P sends $CFT, (ts, \mathrm{RID}, ID_V), \sigma_{SW}, \sigma_{SW,V}$ to N.
if `reason` not `legitimate` $\vee\, \sigma_{SW}$ is `invalid` $\vee\, \sigma_{SW,V}$ is `invalid` **then**
  N aborts.
$(D_{(1,N)}, D_{(2,N)}) \leftarrow \mathsf{PartialDecrypt}(N, C_1, C_2, sk_N)$
N forwards $(D_{(1,N)}, D_{(2,N)})$ to P.
P sends $CFT, (ts, \mathrm{RID}, ID_V), \sigma_{SW}, \sigma_{SW,V}$ to J.
if σ_{SW} is `invalid` $\vee\, \sigma_{SW,V}$ is `invalid` **then**
  J aborts.
$(D_{(1,J)}, D_{(2,J)}) \leftarrow \mathsf{PartialDecrypt}(J, C_1, C_2, sk_J)$
J forwards $(D_{(1,J)}, D_{(2,J)})$ to P.
$(D_{(1,P)}, D_{(2,P)}) \leftarrow \mathsf{PartialDecrypt}(P, C_1, C_2, sk_P)$
P combines partials and checks the binding:
  $ID_U \leftarrow C_3 - (D_{(1,P)} + D_{(1,N)} + D_{(1,J)})$
  $D_2 \leftarrow D_{(2,P)} + D_{(2,N)} + D_{(2,J)}$
  $t \leftarrow H(ts, \mathrm{RID}, ID_V)$
if $\mathsf{VerifySign}_{ID_U}(C_4, H(t, D_2)) = 0$ **then**
  P contacts V.
P outputs ID_U and the opened trace $(CFT, (ts, \mathrm{RID}, ID_V), \sigma_{SW,V}, ID_U, D_2)$

$\mathsf{IssueWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW})$

P presents `reason` to J.
if `reason` not legitimate **then**
  J aborts.
$\sigma_{SW} \leftarrow \mathsf{Sign}_{sk_J}(ID_{SW} \,\|\, \texttt{reason} \,\|\, \texttt{aux})$
return σ_{SW}

$\mathsf{ObtainTrail}(P, V, \sigma_{SW}, \mathrm{RID})$

if RID is `NULL` **then**
  P buys an anonymous check with id RID.
P sends $\sigma_{SW}, \mathrm{RID}$ to V.
if σ_{SW} is `invalid` **then**
  V aborts.
V retrieves CFT for RID with params (ts, RID, ID_V).
$\sigma_{SW,V} \leftarrow \mathsf{Sign}_{sk_V}(ID_{SW} \,\|\, ts \,\|\, \mathrm{RID} \,\|\, ID_V \,\|\, \mathsf{H}(CFT))$
return $(CFT, ts, \mathrm{RID}, ID_V, \sigma_{SW,V})$

$\mathsf{PartialDecrypt}(X, val_1, val_2, sk_X)$

$D_{(1,X)} \leftarrow sk_X \cdot val_1$
$D_{(2,X)} \leftarrow sk_X \cdot val_2$
return $(D_{(1,X)}, D_{(2,X)})$

Fig. 4: CFT anonymity revocation (left) and its subroutines (right).

- ObtainTrails: the police runs this subroutine with the relevant verifier(s) to obtain the CFT(s) whose anonymity needs to be revoked. In the "search by request ID" scenario, the police get exactly one CFT.
- PartialDecrypt: this subroutine is run by a member of the revocation committee for removing a layer of decryption of CFT.

Next, we detail the whole search procedure for the search by request ID scenario using the subroutines.

E. Searching by Request ID

The whole procedure for search by request ID has been depicted in Figure 4 (left). Before any of these subroutines are called, the police first acquires the reason for search warrant issuance, which is the proof proxy attack in our case, from sources such as investigative journalism, web crawling for suspicious online activities, or voluntary whistleblowing. Thereafter, the police runs the IssueWarrant subroutine with the judge. IssueWarrant outputs the search warrant to proceed with the search.

If the police does not have the request ID (RID) yet, it uses a sting operation-like approach [61] to acquire the relevant trail by approaching the malicious user as a buyer of an anonymous online check for the verifier's service. Specifically, the police works together with a specific verifier (mentioned in the aux) to run the ObtainTrail subroutine. Then, the police uses the request ID (RID) to pinpoint the relevant CFT that the malicious user bought. After obtaining the CFT, the NGO, the judge, and the NGO invoke the PartialDecrypt on the CFT to obtain the user's public key (ID_U). The partial decryption of the CFT can be done in any order but the last PartialDecrypt should perhaps be invoked by the police since the police would continue further investigation. Furthermore, the entire search procedure includes multiple checkpoints, such as parties verifying signatures and the search reason. These checkpoints imply that the search procedure aborts if the party cheats or if the police tries to revoke the anonymity of a benign user.

F. Searching by Suspect ID

We present the complete procedure for Search by Suspect ID in Figure 8 (see appendix). The subroutine IssueCondWarrant is pretty much the same as IssueWarrant, except the police needs to provide more justification for searching multiple records. The ObtainTrails needs minor modifications w.r.t. ObtainTrail since the police cannot use a sting operation-like approach as the RID is unknown; instead the police obtains relevant trails from relevant verifier(s).

PartialDecrypt subroutine, however, needs major changes: if this subroutine is used without any modifications, then the police can see in clear when and how many times a user ID appears. With this leakage, if the police can collude with an issuer, then the police learns the identity of people using a service and their usage patterns. Hence, we propose a paradigm similar to Segal et al. [87]: which we refer to as *two-phase partial decryption*: the key idea is to map each trail to a blinded identifier PID and then, compare each blinded

identifier with the input suspect ID. Our instantiation uses the Pohlig-Hellman trick to convert the randomized ElGamal encrypted CFT to a deterministic blinded identifier PID. Since each PID is generated using a fresh randomness, the blinded identifiers of benign users remain unlinkable as long as one of the members of the revocation committee is honest, while enabling the revocation committee to learn about the usage records of the suspect. Next, we discuss how the two-phase partial decryption protocol fits into the search by activity threshold scenario.

G. Searching by Activity Threshold

A generalization of the *search by suspect ID* scenario is when the police needs to search for IDs that appear more than a threshold number of times. We cannot use Figure 8 as is: using fresh randomness for each CFT would lead to degenerate comparisons of the PID of an unknown ID to the PID of another unknown ID. Thus, we present a new protocol in Figure 9, which we refer to as *conditional opening*, for searching by activity threshold using the two-phase decryption: the parties derive PID values, and check if a PID appears more than a threshold number of times in the records. All PID records are blinded using randomness generated by each party of the revocation committee. Although blinding prevents the identifiers from appearing in plaintext, the access pattern over the encrypted identifiers is still leaked, posing a significant privacy risk [88]–[90].

a) An MPC-based approach: To mitigate the access pattern leakage, we propose another scheme using Plaintext-Equality-Test, derived from [91], and using secure multiparty computation (MPC). In this paradigm, the police, the judge, and the NGO hold shares of the secret key, and the joint public key is used for CFT generation. CFT generation is unchanged; the anonymity-revocation protocol proceeds in three steps:

1) *Plaintext equality test.* For every pair of CFTs the revocation committee computes the homomorphic difference of the two ciphertexts and blinds it with a fresh random scalar contributed jointly by all members. Each member then partially decrypts the blinded difference and provides its partial as a secret-shared input to the next step. No member ever reconstructs the blinded difference in the clear, so no member learns whether any individual pair of CFTs encrypts the same ID.
2) *MPC-based counting.* Inside an MPC, the committee combines the secret-shared partials into the joint blinded difference (still secret-shared), tests whether it is the curve identity (the equality witness), and, for each CFT, counts how many other CFTs encrypt the same ID and tests whether the count meets the threshold τ. Only the per-CFT threshold bits are revealed; the pairwise matching pattern, the counts, and the equality witnesses all remain inside the MPC.
3) *Conditional reveal.* For each CFT whose bit is 1 (i.e. its ID appears at least τ times), the committee threshold-decrypts that CFT and learns the ID. CFTs whose bit is 0 remain encrypted.

We provide more details on the MPC-based approach in Appendix D.

V. Discussion and Future Work

a) Transparency logs for CFT accesses: Decentralizing the revocation committee mitigates the chances of anonymity revocation of benign users. However, the system is still non-transparent to the users: they have no direct visibility into when, or how often, their CFTs are opened. A natural extension for transparency is to maintain a CFT search log — with its warrant identifier, verifier, and timestamp — to an append-only, publicly verifiable log, drawing inspiration from Certificate Transparency [92] and CONIKS [93]. Users could either be notified directly whenever a CFT linked to their identity is opened, or receive periodic aggregate reports summarizing CFT accesses. We leave the precise design of the transparency logs for CFTs to future work.

b) Credential Revocation: Issuers must be able to revoke credentials – for instance when a credential is corrupted, the holder's entitlement is withdrawn, or as part of an investigation enabled by CFT itself. The naïve approach is a status list (as used in the Swiss e-ID [31] and EUDI [30]): the issuer publishes an array indexed by credential, and the holder discloses their index for the verifier to check. This breaks unlinkability when indices are static. Privacy-preserving alternatives – cryptographic accumulators [94], [95] and SNARK proofs over status lists [54] – integrate cleanly with CFT; their implementation is left for future work.

c) MPC at Scale: The search by activity threshold may require the opening of thousands of CFTs under MPC. Running the protocol symmetrically across all three parties per record imposes a sizeable cost on the NGO, typically the least-resourced party. A natural mitigation is *asymmetric MPC* [96], in which parties contribute different amounts of work while jointly realizing the functionality with the same security guarantees. The police and judge, both backed by state infrastructure, could handle the heavy computation involved, while the NGO participates with a much lighter footprint. However, Goyal et al. only address asymmetric network requirements; extending asymmetric MPC to support asymmetric compute and storage remains an open problem.

d) Post-quantum security: As the EUDI / Swiss e-ID systems will not be quantum-safe upon go-live, they will likely face a migration to post-quantum-cryptography [97], [98] within a few years. The dominant prerequisite — and the actual deployment bottleneck — is hardware support for PQ primitives in secure enclaves and their widespread availability across user devices. Yet, digital ID systems are under less store-now-decrypt-later pressure than bulk data encryption: broken signatures can be replaced by PQ signatures and integrity restored, but lost confidentiality cannot be retroactively repaired. The CFT, being an encrypted identifier stored by verifiers, is an exception. However, expected verifier retention periods are short (6–12 months), and well below the timelines projected for cryptographically relevant quantum computers [99]. A non-compliant verifier deliberately

archiving CFTs would re-introduce the risk; CFT-log auditing (Section III-E) thus serves a dual role.

Designing a fully PQ-safe CFT is important future work, but the path is quite clear. The institutional design and protocol architecture are agnostic to cryptographic primitives. ECDSA / EdDSA become Dilithium (ML-DSA). ZKPs migrate to lattice-based or hash-based (STARK) backends. The joint-key ElGamal encryption could be replaced by sequential nested PQ encryption under three independent keys. For the two-phase opening, our MPC-based protocol (Section IV-G) is naturally PQ-safe: under the 1-out-of-3 honesty assumption, dishonest-majority MPC protocols such as maliciously-secure replicated secret sharing are information-theoretically secure and available off-the-shelf in MP-SPDZ [100].

We expect the performance impact of using PQ-safe components to be significant, requiring detailed analysis.

VI. Implementation

This section details the prototype implementations of the CFT generation and anonymity revocation protocols. The implementation source is available in an anonymous repository [5].

A. Proof Generation and Verification

We implemented a zero-knowledge circuit realizing the following credential proof: a zero-knowledge statement that demonstrates possession of a valid, digitally signed credential while proving correct encryption of a CFT and legal age constraints (1)(2)—an important practical application of anonymous credentials [42].

The credential used in our prototype contains 32 attributes: an upper bound on the number of attributes of the Swiss e-ID (24 attributes [101]), and five of these attributes are used in the circuit's logic.

We developed two implementations of this proof: a native C++ version built on the Longfellow stack, and a SNARK-oriented circuit implementation. The circuit proves the following statements in zero-knowledge:

1) **Credential Validity:** Five attributes (identity components validity timestamps, and birthdate) are proven to belong to the credential by showing membership among the full set of 32 attributes committed in a single hash. The issuer signs this attribute hash, which the prover demonstrates possession of without revealing the attributes themselves, nor the signature.
2) **Issuer Signature:** The prover demonstrates possession of a valid signature on the compact attribute hash from the credential issuer (identified by their public key), without revealing the signature itself.
3) **Temporal Constraints:** Three timestamp checks are enforced:
   - Birth date is before a maximum age threshold (≥ 18)
   - Credential validity period has begun (now $\geq$ validFrom)
   - Credential has not expired (now $\leq$ validUntil)
4) **ElGamal Encryption and(Hardware) Binding Signature:** The proof generates two independent random scalars r_1 and r_2 and computes four output values:
   - $c_1 = r_1 \cdot G$ (ephemeral public key for encryption)
   - $c_2 = r_2 \cdot G$ (second ephemeral public key)
   - $c_3 = (ID_x, ID_y) + r_1 \cdot Q$ where Q is the ElGamal public key (the user's public key / identifier encrypted under Q using randomness r_1)
   - $c_4 = \text{EdDSA-Sign}_{sk_u}(\text{Poseidon}(t,\ r_2 \cdot Q))$ (signature under the user's secret key sk_u, which is stored in hardware is therefore non-extractable, the signature binds the request tag and the the public key identifier).

   The public key (ID_x, ID_y) encrypted in c_3 is the same key proven to be a valid attribute in the credential via membership against the signed attribute hash, ensuring the encrypted identity corresponds to the verified credential. Because c_4 is produced with the hardware-held sk_u, device binding is realized by this CFT component rather than by a separate hardware-signature check: only the device that holds sk_u can produce a valid binding signature on $\text{Poseidon}(t,\ r_2 \cdot Q)$. At revocation, the committee recovers (ID_x, ID_y) from c_3 and verifies c_4 against $\text{Poseidon}(t, r_2 \cdot Q)$.

a) Standards-based implementation: This is a native C++ implementation built on the Longfellow stack [12], [102] and standard cryptographic primitives. It uses P-256 for elliptic-curve operations, ECDSA for issuer signatures and for the hardware-held user key used in c_4, and SHA-256 for hashing. These algorithms are supported on most secure enclaves in smartphones, enabling hardware device binding of private keys. Therefore, they are used in verifiable credential standards such as mobile driving license (mDoc/mDL), and the EUDI and Swiss e-ID.

b) ZK-friendly implementation: Longfellow is highly optimized for its specific use case, but the underlying algorithms produce large proof circuits when used in generic scenarios. Therefore, we implement another version using algorithms that are more efficient for generic SNARK circuits: BabyJubjub curve arithmetic, Poseidon hashing, and EdDSA-style signatures. The implementation is expressed as a circuit description (Circom) and produced witness/proof material using rapidsnark as the prover.

c) Credential non-revocation proof.: As an optional extension, both stacks can additionally prove that the presented credential is not revoked against a packed status-list Merkle root: the issuer embeds a revocation index in a signed attribute slot, and the circuit proves a Merkle authentication path to a public root with the corresponding bit unset (i.e., the credential is not revoked). The revocation index stays private in order not to make two presentations linkable. We report the composed cost separately in Section VII-B and Appendix C.

B. Anonymity Revocation

We implemented prototypes for all three scenarios of anonymity revocation: Searching by Request ID (Figure 4),

[5] https://anonymous.4open.science/r/accountability_for_acs-7670/README.md

Searching by Suspect ID (Figure 8), and Searching by Activity Threshold—both without MPC (Figure 9), and with MPC (Appendix D). We use MP-SPDZ [103] for our MPC-based implementation and run it using the MASCOT [104]—a dishonest majority, malicious secure protocol. The simplifications consist of omitting the warrant generation and trail collection steps, and the signatures creation and verification steps.

Our implementations use the same cryptographic primitives as the zk-friendly implementation described above.

C. Integration in Digital Identity Stack

Both, the Swiss and EU solutions use the OID4VP (OpenID for Verifiable Presentations) protocol [105] for handling credential requests and presentations between holders and verifiers [30], [31]. OID4VP provides extension points and profiles to deal with specific formats, trust requirements and transport mechanisms. The basic request object (from verifier to holder wallet) supports optional client metadata that could transport the CFT input parameters (timestamp, RID, etc.) to the holder. The response object also supports additional parameters for transmitting the proof to the verifier. The EU has already drafted a standard for integrating circuit-based ZKPs with EUDI wallets and OID4VP [33]. We suggest to build on this and define an OID4VP profile with CFT support. This would enable seamless integration of CFT into national digital identity stacks.

VII. Practicality and Evaluation

a) Requirements: Digital identity systems such as EUDI and the Swiss e-ID run on mobile devices with limited resources and involve direct user interaction, so any overhead added to a credential presentation impacts the user-perceived latency. Thus, if adding CFT raises the time needed for credential presentation from 500ms to 5s, it will not be accepted by users. Yet, a total latency of about 1-2s for a presentation is often considered acceptable [106], [107].

Hence, in this section, we answer the following questions:

- What is the overhead added by CFTs on users and verifiers and are the proving/verification times practical? (Section VII-A)
- How does composing a non-revocation claim with the CFT presentation scale with population size? (Section VII-B)
- What is the time taken for anonymity revocation when searching by Request ID? (Section VII-D)
- What is the time taken for anonymity revocation when searching by Suspect ID or by Activity Threshold? And how does it scale with the number of trails considered and the percentage of these trails that reach the anonymity revocation stage? (Section VII-D)

A. Proof Generation and Verification

We evaluate the practical performance of the full credential proof described in Section VI-A. We benchmark both the standards-based implementation and the fully zk-friendly implementation.

We run the benchmarks in two distinct environments to capture both constrained client devices and more powerful server hardware, which reflects the intended deployment model: provers run on users' phones, while verifiers are expected to run on server infrastructure.

We also benchmark a no-CFT variant of the same proof. Because device binding is required whether or not CFT is used (via a signature under the hardware-held sk_U), the no-CFT baseline still includes a device-binding claim: the prover signs a random challenge message m under sk_U and proves knowledge of that signature in zero knowledge. The CFT variant replaces this standalone binding with the CFT construction of point (4) in the implementation list (ElGamal encryption plus the binding signature C_4). Comparing the two therefore isolates the overhead of CFT itself, rather than the combined cost of CFT and device binding. The comparison also excludes non-revocation claims (Section VII-B), so that relative CFT overhead doesn't depend on population size.

- **Mobile-like environment:** aarch64 machine with 2 CPU cores and 4 GB of RAM.
- **Server-like environment:** x86_64 machine with 8 CPU cores and 16 GB of RAM.

Each benchmark was repeated 10 times, and we report mean prover and verifier runtimes in Table IV.

Machine	Impl.	Prover		Verifier	
		No CFT	CFT	No CFT	CFT
Mobile	Standard	1959	2213 (+13%)	1472	1686 (+15%)
	zk-friendly	1030	1194 (+16%)	23	23 (+0%)
Server	Standard	1200	1355 (+13%)	854	976 (+14%)
	zk-friendly	314	438 (+39%)	12	13 (+8%)

TABLE IV: Performance of the credential proof (in ms), without a non-revocation claim. Columns show mean runtimes without CFT and with CFT for prover and verifier.

1) Practical Implications: Measured overhead of adding the CFT components depends on the environment and implementation. For the prover it ranges from 13% to 39%, while verifier overheads span 0% to 15%.

Longfellow is highly optimized for SHA-256 and ECDSA (our current primitives in the standards-based implementation). But the zk-friendly implementation is faster for the prover (about 2 to 4 times faster depending on the environment) and the verifier (about 75 times faster). Thus, if hardware supports both implementations, we recommend the zk-friendly implementation: zk-SNARKs give greater flexibility for building new circuits, and verification is much cheaper, leading to a major win for high-volume verifiers.

Overall, CFT introduces modest overhead that is likely acceptable for real-world deployments. The demonstrated performance, together with ongoing framework improvements, suggests practical deployment of CFT is achievable in the near term. The standards-based implementation result remains somewhat above the ideal target of 2 seconds, but further

optimization could make it viable soon. Detailed performance analysis comparing attribute commitment representations (Merkle trees vs. flat hashing) is provided in Appendix B.

B. Proof of Credential Non-Revocation

An additional claim that the credential is not revoked at show time can be added to the credential proof. We time that check as a separate composition on top of the age-check+CFT circuit, varying the population size N on both stacks and machine profiles. Scaling plots and depth-d fits are in Appendix C. On the zk-friendly stack the prover cost grows only gently with Merkle depth ($d = log_2(N) - 7$) and stays interactive even at an EU-scale upper bound $N = 2^{29}$ while the verifier cost is constant (about 1.5 s prove on mobile and ≈13 ms verify on server). Both the prover and the verifier cost grow linearly with the merkle depth for the standards-based stack. At $N = 2^{29}$ prove costs 6.5 s on mobile and verify costs 2.8 s on server. This gap reflects two effects: Poseidon is much cheaper in-circuit than SHA-256 as the Merkle path deepens, and Groth16 verification has fixed cost whereas the Longfellow verifier grows with the SHA-256 circuit.

C. Communication costs of a credential presentation.

A presentation is two messages. The verifier sends a request with ts, RID, ID_V, policy parameters (`now`, `maxBirthDate`), and the current status-list root (192 B total). The wallet derives t locally, so the tag is not sent, and the Merkle path for non-revocation stays private. In the other direction the wallet returns the trail $CFT = (C_1, C_2, C_3, C_4)$ (288 B on both stacks) and the proof π. The sizes below are for our composed toy circuit (age-check+CFT+non-revocation claim). On the zk-friendly stack the Groth16 proof itself is fixed-size ($|\pi| = 128$ B for any N), so this show is 416 B and end-to-end communication about 608 B. A different circuit could still change the total if more public inputs must be exchanged, even though $|\pi|$ stays constant. On the standard stack $|\pi|$ grows with the circuit, and hence with population via the status-list Merkle depth, reaching about 663 KB at Swiss-scale $N = 2^{24}$ and about 716 KB at EU-scale $N = 2^{29}$. That scaling is another reason to prefer the zk-friendly stack when both are available.

D. Anonymity Revocation

We executed the benchmarks for anonymity revocation [5] on the server-like machine described above, as the involved parties are expected to have access to better hardware. Each benchmark was repeated 10 times, and we report mean runtimes in milliseconds.

For the "Searching by Request ID" case, the anonymity revocation benchmark resulted in a mean runtime of 21 ms, i.e., a throughput of roughly 48 revocations per second, which is more than sufficient for the expected frequency of revocations in real-world deployments.

For the "Searching by Suspect ID" and "Searching by Activity Threshold", we varied the size of the set of trails considered (between 10 and 8000 trails), and also varied the size of the set considered for the final decryption step of the link-then-decrypt approach. (between 10 and 50 percent of the full set of trails size). The results are shown in Figure 5.

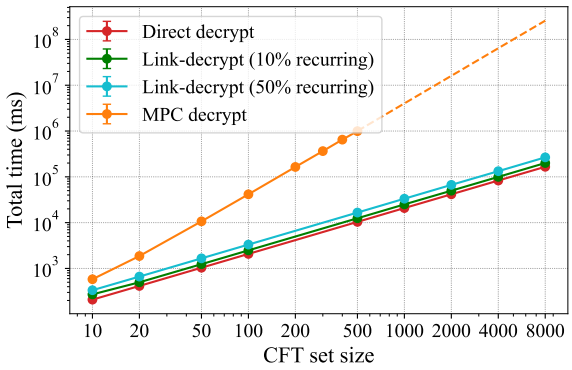

Fig. 5: Total revocation time across the three anonymity revocation techniques as a function of the number of CFTs; dashed lines represent extrapolated values.

As depicted in Figure 5, the time taken for revocation grows linearly with the number of trails to be decrypted for the direct decryption and the link-then-decrypt approach, and quadratically for the MPC-based approach. The standard deviations are too small for the error bars to be visible in the figure. In the link-then-decrypt approach, the time taken also grows linearly with the number of trails that reach the second stage of the algorithm.

We fit the benchmark runtimes of the link-then-decrypt approach to a linear model, and the resulting formula is given by:

$$T(N, M) \approx 23 \cdot N + 21 \cdot M,$$

where N denotes the total number of trails considered and M denotes the number of "recurring ID", i.e., trails remaining after the filtering step, that will go through the anonymity revocation stage. In the "10% of recurring ID" scenario, $M = 0.1N$, and the formula becomes $T(N) \approx 23 \cdot N + 2.1 \cdot N = 25.1 \cdot N$. In the "50% of recurring ID" scenario, $M = 0.5N$, and the formula becomes $T(N) \approx 23 \cdot N + 11.5 \cdot N = 34.5 \cdot N$. The 50% scenario thus takes more time simply because more trails reach the second stage of the link-then-decrypt algorithm.

Opting for the two-stage approach instead of revoking anonymity of all trails adds roughly 2 ms per candidate trail plus an additional 21 ms per trail that reaches the final revocation stage.

For the MPC-based decryption variant, the Search by Activity Threshold case scales as

$$T(N) \approx 8 \cdot (N \cdot (N-1)/2),$$

making the privacy cost explicit: compared to link-then-decrypt, MPC buys stronger privacy at a higher computational price.

VIII. Related Work

a) Anonymity Revocation in AC: ACs were envisioned by Chaum [1] and since then, many works have followed focusing on efficiency, attributes, delegation, and universal composability [2], [25]–[29]. Camenisch and Lysyanskaya [2] introduced *anonymity revocation* via a centralized authority; later work made this authority third-party-auditable [108] and then user-auditable [22]. Yet revocation remained optional in ACs and is absent from today's national identity systems. CFT revives this goal, now made urgent by nationwide deployment. The most direct prior work is PAPR [23], with a concurrent proposal by Daza et al. [24]. CFT differs from PAPR in three ways: (i) PAPR restricts itself to linkable showings; CFT treats unlinkability as a precondition. (ii) PAPR de-anonymizes all of a user's shows; CFT is scoped to the warrant. (iii) PAPR prevents misuse only by detection; CFT prevents it structurally via separation of powers. PAPR's bulletin-board transparency could nonetheless be combined with CFT. Rate-limiting approaches [109] prevent only over-use, offering no way to investigate other misbehavior.

b) Group Signatures and Their Variants: Group signatures [9], [10], [34] let any member sign anonymously to prove membership, while a designated *open* protocol revokes a signer's anonymity—a natural fit for accountable anonymity. Opening, however, is handled typically by the centralized group manager, and the scheme's setup is not ECDSA-compatible, a requirement for national e-IDs. Threshold group signatures [35] decentralize opening across several parties, but do not eliminate the standardization issue. Traceable signatures [36]–[38], along with related bifurcated [38], hidden ID-based signatures [39], and identity escrow [32], [40], add finer-grained privacy properties to group signatures. However, they fall short in standardization compliance and providing generic revocation ability such as search by activity threshold. As highlighted in Table I, none of these primitives simultaneously provides unlinkability, multi-party opening, and standardization compliance—precisely the gap that CFT fills.

c) Accountability for State Processes: Green et al. [110] propose an abuse-resistant law enforcement. CFT differs by targeting ACs rather than encrypted communications, using only standard primitives rather than witness encryption, and integrating institutional separation of powers—a necessity for practicality (Section III-C). A parallel line of work addresses public accountability for "secret" state processes [111]. Frankle et al. [112] make warrant processes auditable in aggregate without revealing individual targets. CFT extends this agenda from accountable issuance to warrant *execution* in an AC system.

d) Cryptographic Pseudonyms: Mayrhofer et al. [113] present cryptographic pseudonyms for European Identity Wallet (EUDIW). For each verifier, the holders will use different but fixed pseudonyms. The verifiers will bind accounts to these pseudonyms. By comparing pseudonyms used across different user accounts, verifiers are therefore capable of detecting potential misuse, e.g., when a user resells his pseudonyms to other users. While this approach demotivates selling numerous proofs, it requires verifiers to track (anonymous) holders using personal accounts [6]. Additionally, the approach does not allow anonymity revocation of misbehaving users or stopping their activity on other sites as pseudonyms are unique to each verifier, it is impossible to build collaborative blocklists.

e) Others: Anonymous tokens have been extended with a hidden metadata bit using algebraic MACs [114], but a single bit is insufficient to encode a forensic trail. Zcash [115] allows opening private transactions via viewing keys, but these keys are user-controlled and must be voluntarily shared with authorities for opening transactions. Tsang et al. [116] blacklist anonymous credentials in a decentralized way, but blacklisting does not de-anonymize misbehaving users and therefore does not enable accountability.

TruAge [117] developed by NACS (National Association of Convenience stores) and Conexxus is a US solution for age verification in stores (points of sales), based on driving license attributes. It uses encrypted, one-time digital tokens and provides admissible proof-of-age verification appended to retailers' transaction logs, which can be unlocked under subpoena and submitted as evidence. While being an industry-driven solution for a specific use case, TruAge demonstrates the importance of privacy-preserving token presentation subjected to anonymity revocation through subpoenas, as intended by our CFT logs and the revocation committee process.

IX. Conclusions

Anonymous credential systems have been researched for decades but have seen limited adoption in practical deployments. With the rise of national electronic identity frameworks in regions such as the EU and Switzerland, privacy-preserving mechanisms inspired by anonymous credentials are gaining renewed relevance. Many existing designs incorporate some notion of unlinkability between credential presentations; however, achieving perfect unlinkability—resilient even against collusion between issuers and verifiers—remains an open challenge in practice, constrained by current performance and standardization limitations.

Yet, once unlinkability approaches this ideal, a new challenge becomes central: accountability. Fully unlinkable systems can be misused, for example, by malicious users reselling identity proofs. We argue that accountability must be an integral part of system design to preempt political pressures to weaken privacy guarantees for all citizens—pressures similar to those seen in current debates around end-to-end encryption and "chat control" initiatives.

To address this, we introduce the *cryptographic forensic trail (CFT)*, a mechanism enabling controlled and transparent revocation of anonymity through a multi-party protocol with democratic checks and balances. We design and implement

[6]When creating these accounts, special care has to be taken by users not to accidentally leak identifying information that was deliberately not disclosed during credential presentation, such as email addresses for account maintenance.

several CFT protocol variants using state-of-the-art cryptographic primitives. Our evaluation demonstrates that practical, privacy-preserving accountability mechanisms for national identity systems are within reach.

Acknowledgments

Ethics Considerations

To every credential presentation, the CFT attaches cryptographic material that can be used to identify the presenting user under specified conditions (targeted warrants). This deliberately weakens the unlinkability property of these presentations and hence affects citizen privacy. The experiments in this paper were performed in the lab and did not involve real user data. In the following, we discuss implications of the CFT if it were deployed on national scale.

We believe that neither the Swiss e-ID Act [118] nor the EUDI Regulation [119] currently provide a legal basis for warrant-bound anonymity revocation of credential presentations, and existing telecommunications surveillance law (BÜPF [71], ePrivacy Directive [120]) may not cover this operation, since credential presentations are not communications in the sense these instruments define. This question requires further investigation. Therefore, deploying CFT could require new legislation, defining the institutional roles, and integrating with judicial-review mechanisms. We view this legislative dependency as a feature, not a bug: such a critical mechanism requires democratic deliberation and should not be available to executive action alone. The CFT proposes what the cryptographic substrate of such regulation could look like if designed with accountability as a first-class requirement.

While the CFT affects citizen privacy, the realistic alternative to CFT is not unconditional anonymity for all users, We have been discussing this throughout the paper (see sections I-III), highlighting that a lack of accountability may lead to an overall worse level of privacy, also for the vast majority of users who never become the subject of a warrant. A balanced discussion of ethical aspects must also address implications of lacking accountability in national digital identity. As we have shown, proof proxy attacks may undermine authenticity of identity proofs, even if credentials are non-transferable and hardware-bound. Legislation depending on these checks (like age verification) may be rendered obsolete, leaving society with the overhead but not the protective effects of digital credentials. If it is possible to cheat with identity proofs and get away, a false sense of security may arise, eventually eroding trust in the overall system.

CFT has some inherent limitations. It always identifies the signing account holder (Eve in our example), which is not necessarily the (only) wrongdoer. Alice, who consumes an illegal service, is not identified by CFT. In case private keys or devices are stolen or people are coerced, the victim is under pressure to proof their innocence. As is the case with misused electronic signatures, this would require police reporting as well as forensic analysis of platform audit trails and session metadata. We recognize that a principled objection remains: any anonymity revocation capability, however constrained, may eventually be abused (e.g., by a future government), and the ethically correct response may be to refuse to design it. We take this view seriously; our disagreement is empirical rather than normative. CFT's institutional guarantees are bounded by

the assumptions on which they rest: a functioning separation of powers and an independent civil-society co-signer with a democratic mandate. We do not recommend CFT deployment where these assumptions fail, and we do not claim to have resolved all ethical questions the design raises; these must eventually be addressed through democratic debate.

Open Science

Implementations of the CFT algorithms and evaluation scripts are publicly available at https://anonymous.4open.science/r/accountability_for_acs-7670.

Use of AI Tools

The authors used generative AI-based tools to improve flow and awkward phrasing. AI-based tools were used to search for additional related work and for pre-submission review.

We have manually verified and are responsible for the accuracy, originality, and integrity of the output of all AI-based tools.

References

[1] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 84–90, 1981.

[2] J. Camenisch and A. Lysyanskaya, "An efficient system for non-transferable anonymous credentials with optional anonymity revocation," in *International conference on the theory and applications of cryptographic techniques*. Springer, 2001, pp. 93–118.

[3] T. F. C. of Switzerland, "e-id act," 2025, on 28 September 2025, Swiss voters approved a state-issued e-ID after rejecting a private scheme in 2021. [Online]. Available: https://www.admin.ch/gov/en/start/documentation/votes/20250928/e-id-act.html

[4] E. Commission, "European digital identity (eudi) regulation," 2024, explains that eIDAS 2.0 requires member states to provide an EU Digital Identity wallet by 2026 and describes the obligations and goals. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/eudi-regulation

[5] F. Weissbaum, A. F. Sang, A. Höss, and J. Niestroj, "Key Management for e-ID - Batch Issuance and Renewal V 1.1," https://github.com/swiyu-admin-ch/community/blob/main/tech-concepts/e-id-key-management-batch-issuance-and-renewal.md, 2025.

[6] E. D. P. Supervisor, "Techdispatch – digital identity wallets," 2025, notes that anonymous credentials can mitigate over-disclosure and linkability and are not yet widely adopted. [Online]. Available: https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2025-12-15-techdispatch-32025-digital-identity-wallets_en

[7] S. G. Popescu, "Swiss e-ID: Use cases, harms, opportunities," 2026. [Online]. Available: https://infoscience.epfl.ch/entities/publication/25a1f247-3f5e-4967-b4d1-3dabb0eb7ed5

[8] C. Baum, O. Blazy, J. Camenisch, J.-H. Hoepman, E. Lee, A. Lehmann, A. Lysyanskaya, R. Mayrhofer, H. Montgomery, N. K. Nguyen *et al.*, "Cryptographers' feedback on the eu digital identity's ARF," *Tech. Rep.*, 2024.

[9] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Advances in Cryptology – CRYPTO 2004*, ser. Lecture Notes in Computer Science, vol. 3152. Springer, 2004, pp. 41–55.

[10] M. H. Au, W. Susilo, and Y. Mu, "Constant-size dynamic k-taa," in *International conference on security and cryptography for networks*. Springer, 2006, pp. 111–125.

[11] S. Tessaro and C. Zhu, "Revisiting bbs signatures," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2023, pp. 691–721.

[12] M. Frigo and abhi shelat, "Anonymous credentials from ECDSA," Cryptology ePrint Archive, Paper 2024/2010, 2024. [Online]. Available: https://eprint.iacr.org/2024/2010

[13] M. Babel and J. Sedlmeir, "Bringing data minimization to digital wallets at scale with general-purpose zero-knowledge proofs," *arXiv preprint arXiv:2301.00823*, 2023.

[14] "EUDI Architecture and Reference Framework - Discussion Paper G - Zero Knowledge Proof," https://github.com/eu-digital-identity-wallet/eudi-doc-architecture-and-reference-framework/discussions/408, 2025.

[15] A. Lehmann, "EU digital identity and anonymous credentials - a happy end?" Talk given at RWC 2025, Video at https://youtu.be/UpQHWObCx4I, 2025.

[16] C. Paquin, G.-V. Policharla, and G. Zaverucha, "Crescent: stronger privacy for existing credentials," *Cryptology ePrint Archive*, 2024.

[17] R. Mayrhofer, B. Preneel, D. Aranha, O. Blazy, A. Lehmann *et al.*, "Joint statement of security and privacy scientists and researchers on Age Assurance," https://csa-scientist-open-letter.org/ageverif-Feb2026, February 2026.

[18] S. E. Landau, *Listening in: Cybersecurity in an insecure age*. Yale University Press, 2017.

[19] "Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse," https://cdn.netzpolitik.org/wp-upload/2025/07/2025-07-24_Council_Presidency_LEWP_CSA-R_Compromise-texts_11596.pdf, 7 2025.

[20] "Joint statement of scientists and researchers on the EU Presidency's new proposal for the Child Sexual Abuse Regulation," https://csa-scientist-open-letter.org/Sep2025, 2025.

[21] B. Hoeisen, "Switzerland's Privacy and Encryption Under Attack: Proposed Surveillance Laws Threaten Fundamental Rights and Cripples Encryption," https://www.isoc.ch/swiss-surveillance-ordinance-encryption-threat-vupf-oscpt/, May 2025.

[22] J. Camenisch and A. Lehmann, "Privacy-preserving user-auditable pseudonym systems," in *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2017, pp. 269–284.

[23] J. Brorsson, B. David, L. Gentile, E. Pagnin, and P. S. Wagner, "Papr: Publicly auditable privacy revocation for anonymous credentials," in *Cryptographers' Track at the RSA Conference*. Springer, 2023, pp. 163–190.

[24] V. Daza, A. Haque, A. Scafuro, A. Zacharakis, and A. Zapico, "Mutual accountability layer: accountable anonymity within accountable trust," in *International Symposium on Cyber Security, Cryptology, and Machine Learning*. Springer, 2022, pp. 318–336.

[25] J. Camenisch and E. Van Herreweghen, "Design and implementation of the idemix anonymous credential system," in *Proceedings of the 9th ACM Conference on Computer and Communications Security*, ser. CCS '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 21–30. [Online]. Available: https://doi.org/10.1145/586110.586114

[26] J. Camenisch and A. Lysyanskaya, "Signature schemes and anonymous credentials from bilinear maps," in *Annual international cryptology conference*. Springer, 2004, pp. 56–72.

[27] M. Belenkiy, M. Chase, M. Kohlweiss, and A. Lysyanskaya, "P-signatures and noninteractive anonymous credentials," in *Theory of cryptography conference*. Springer, 2008, pp. 356–374.

[28] M. Chase, S. Meiklejohn, and G. Zaverucha, "Algebraic macs and keyed-verification anonymous credentials," in *Proceedings of the 2014 acm sigsac conference on computer and communications security*, 2014, pp. 1205–1216.

[29] M. Rosenberg, J. White, C. Garman, and I. Miers, "zk-creds: Flexible anonymous credentials from zksnarks and existing identity infrastructure," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 790–808.

[30] "European Digital Identity Architecture and Reference Framework (ARF)," https://eudi.dev/latest/architecture-and-reference-framework-main/, 2026.

[31] "Swiss e-ID Technology Stack," https://swiyu-admin-ch.github.io/technology-stack/, 2026.

[32] J. Kilian and E. Petrank, "Identity escrow," in *Annual International Cryptology Conference*. Springer, 1998, pp. 169–185.

[33] "Specification for the implementation of Zero-Knowledge Proofs based on arithmetic circuits in the EUDI Wallet," https://github.com/eu-digital-identity-wallet/eudi-doc-standards-and-technical-specifications/blob/main/docs/technical-specifications/ts13-zksnarks.md, EU, January 2026.

[34] M. Abdalla and B. Warinschi, "On the minimal assumptions of group signature schemes," in *Information and Communications Security – ICICS 2004*, ser. Lecture Notes in Computer Science, vol. 3269. Springer, 2004, pp. 1–13.

[35] J. Camenisch, M. Drijvers, A. Lehmann, G. Neven, and P. Towa, "Short threshold dynamic group signatures," in *Security and Cryptography for Networks – SCN 2020*, ser. Lecture Notes in Computer Science, vol. 12238. Springer, 2020, pp. 401–423, iACR ePrint 2020/016.

[36] A. Kiayias, Y. Tsiounis, and M. Yung, "Traceable signatures," in *Advances in Cryptology – EUROCRYPT 2004*, ser. Lecture Notes in Computer Science, vol. 3027. Springer, 2004, pp. 571–589.

[37] M. Abe, S. S. Chow, K. Haralambiev, and M. Ohkubo, "Double-trapdoor anonymous tags for traceable signatures," *International journal of information security*, vol. 12, no. 1, pp. 19–31, 2013.

[38] B. Libert, K. Nguyen, T. Peters, and M. Yung, "Bifurcated signatures: folding the accountability vs. anonymity dilemma into a single private signing scheme," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2021, pp. 521–552.

[39] A. Kiayias and H.-S. Zhou, "Hidden identity-based signatures," in *International Conference on Financial Cryptography and Data Security*. Springer, 2007, pp. 134–147.

[40] J. Camenisch and A. Lysyanskaya, "An identity escrow scheme with appointed verifiers," in *Annual International Cryptology Conference*. Springer, 2001, pp. 388–407.

[41] R. Arnold and D. Longley, "Zero-knowledge proofs do not solve the privacy-trust problem of attribute-based credentials: What if alice is evil?" *IEEE Communications Standards Magazine*, vol. 3, no. 4, pp. 26–31, 2019.

[42] C. Knabenhans, S. Veitch, M. Raynal, T. Stadler, S. Chatel, W. Lueks, and C. Troncoso, "On the (privacy) harms of the european digital identity framework," Cryptology ePrint Archive, Paper 2026/867, 2026. [Online]. Available: https://eprint.iacr.org/2026/867

[43] S. Adler, Z. Hitzig, S. Jain, C. Brewer, W. Chang, R. DiResta, E. Lazzarin, S. McGregor, W. Seltzer, D. Siddarth *et al.*, "Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online," *arXiv preprint arXiv:2408.07892*, 2024.

[44] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," in *24th USENIX security symposium (USENIX security 15)*, 2015, pp. 33–48.

[45] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 213–224.

[46] Federal Bureau of Investigation, "AlphaBay Takedown," https://www.fbi.gov/news/stories/alphabay-takedown, 2017, alphaBay served over 200,000 users and 40,000 vendors, with listings for illegal drugs, stolen identity documents, malware, firearms, and fraudulent services; taken down in Operation Bayonet.

[47] Financial Action Task Force, "Illicit Financial Flows from Cyber-Enabled Fraud," Financial Action Task Force (FATF), Tech. Rep., November 2023. [Online]. Available: https://www.fatf-gafi.org/content/dam/fatf-gafi/reports/Illicit-financial-flows-cyber-enabled-fraud.pdf

[48] U.S. National Counterterrorism Center, "Identifying and Preventing Illicit Use of Cryptocurrency by Terrorists," Joint Counterterrorism Assessment Team (JCAT), Tech. Rep., 2024. [Online]. Available: https://www.dni.gov/files/NCTC/documents/jcat/firstresponderstoolbox/119s_-_First_Responders_Toolbox_-_Identifying_and_Preventing_Illicit_Use_of_Cryptocurrency_by_Terrorists.pdf

[49] F. Jessica, "Age-verified Roblox accounts are already being sold on eBay," https://www.dexerto.com/roblox/age-verified-roblox-accounts-are-already-being-sold-on-ebay-3302680/, January 2026.

[50] "Building a Swiss Digital Trust Ecosystem – Perspectives around an e-ID ecosystem in Switzerland," https://digitalswitzerland.com/building-a-swiss-digital-trust-ecosystem/, 2022.

[51] A. Chan, C. Ezell, M. Kaufmann, K. Wei, L. Hammond, H. Bradley, E. Bluemke, N. Rajkumar, D. Krueger, N. Kolt *et al.*, "Visibility into ai agents," in *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*, 2024, pp. 958–973.

[52] J. Hesse, N. Singh, and A. Sorniotti, "How to bind anonymous credentials to humans," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 3047–3064.

[53] F. Egger, "Security analysis of the swiss e-id and trust infrastructure," https://github.com/user-attachments/files/21157739/Security_Analysis_of_the_Swiss_e_ID___Trust_Infrastructure.pdf, July 2025.

[54] D. Mombelli, "General-Purpose Zero-Knowledge Proofs for Verifiable Credentials," Master's thesis, ETH Zurich, March 2025, available at https://ethz.ch/content/dam/ethz/special-interest/infk/inst-infsec/appliedcrypto/education/theses/masters-thesis_damiano-mombelli.pdf.

[55] A. Barua, M. A. Al Alamin, M. S. Hossain, and E. Hossain, "Security and privacy threats for bluetooth low energy in iot and wearable devices: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 251–281, 2022.

[56] Swiss Confederation, "Swiss Criminal Procedure Code (StPO) of 5 October 2007, SR 312.0," https://www.fedlex.admin.ch/eli/cc/2010/267/en, 2011, arts. 269–279 require a judicial order for surveillance of telecommunications and traffic data.

[57] European Parliament and Council of the European Union, "Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences," https://eur-lex.europa.eu/eli/dir/2016/680/oj, 2016, law Enforcement Directive; requires investigative access to be based on Union or Member State law, including prior judicial authorisation where applicable.

[58] Federal Republic of Germany, "Strafprozessordnung (StPO) – German Code of Criminal Procedure," https://www.gesetze-im-internet.de/englisch_stpo/, 2024, §§ 100a, 100g, 102–110 require a judicial order for telecommunications interception, traffic-data collection, and searches.

[59] Council of Europe, "European Convention on Human Rights, Article 8 – Right to respect for private and family life," https://www.echr.coe.int/documents/d/echr/convention_eng, 1950.

[60] European Court of Human Rights, "Roman Zakharov v. Russia, Application no. 47143/06, Grand Chamber judgment of 4 December 2015," https://hudoc.echr.coe.int/eng?i=001-159324, 2015, establishes that secret-surveillance regimes interfering with Art. 8 ECHR must be subject to prior independent (preferably judicial) authorisation.

[61] G. R. Newman, "Sting operations," U.S. Department of Justice, Office of Justice Programs, Tech. Rep., Oct. 2007.

[62] Congressional Research Service, Library of Congress, "Constitution Annotated: Fourth Amendment – Search and Seizure," https://constitution.congress.gov/constitution/amendment-4/, 2024, official annotated U.S. Constitution; historical background, doctrinal development, and Supreme Court case law.

[63] noyb – European Center for Digital Rights, "About noyb: an independent, non-profit organisation enforcing data protection rights," https://noyb.eu/en/about-us, 2023, accessed: 2026-05-23.

[64] EDRi – European Digital Rights, "About edri: the biggest european network defending rights and freedoms online." https://edri.org/, 2026, accessed: 2026-05-23.

[65] European Digital Rights (EDRi) and 47 civil society organisations, "Open letter: High-risk lists must not become an excuse to abandon fundamental rights – joint civil society statement on the Council's CSA Regulation proposal," https://edri.org/our-work/open-letter-high-risk-lists-must-not-become-an-excuse-to-abandon-fundamental-rights/, 2024.

[66] Electronic Frontier Foundation, "About eff: Member-supported and independent," https://www.eff.org/about, 2024, accessed: 2026-05-23.

[67] AlgorithmWatch, "About algorithmwatch: a non-governmental, non-profit organisation evaluating algorithmic decision-making," https://algorithmwatch.org/en/, 2024, accessed: 2026-05-28.

[68] Digitale Gesellschaft, "Digitale gesellschaft – swiss civil society organisation on digital rights, e-id, and surveillance," https://www.digitale-gesellschaft.ch/, 2024, accessed: 2026-05-28.

[69] Chaos Computer Club, "About the CCC: Europe's largest association of hackers," https://www.ccc.de/en/club, 2024, accessed: 2026-05-28.

[70] The Tor Project, "About the Tor project: independent 501(c)(3) non-profit advancing privacy and anonymity online," https://www.torproject.org/about/history/, 2024, accessed: 2026-05-28.

[71] Swiss Confederation, "Federal Act on the Surveillance of Postal and Telecommunications Traffic (BÜPF/LSCPT), SR 780.1," https://www.fedlex.admin.ch/eli/cc/2018/30/en, 2018, art. 26 obliges telecommunications service providers to retain communications metadata (subscriber, traffic, and location data) for six months.

[72] French Republic, "Code des postes et des communications électroniques, Art. L34-1, and Décret n° 2021-1361 of 20 October 2021," https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044229711, 2021, obliges electronic communications operators to retain civil identity and contract data for the duration of the contract plus one year, and (on Prime Minister's order in case of serious threat) traffic and location data for one year, for the purposes of investigating criminal offences.

[73] Court of Justice of the European Union, "La Quadrature du Net and Others, Case C-470/21, Grand Chamber judgment of 30 April 2024," https://curia.europa.eu/juris/document/document.jsf?docid=285405, 2024, held that general retention of IP addresses for fighting criminal offences is permissible under the ePrivacy Directive, provided strict safeguards are in place.

[74] "Federal law (classified compilation) — 2018/32 (english edition)," https://www.fedlex.admin.ch/eli/cc/2018/32/en, Swiss Federal Chancellery, 2018, published in the Classified Compilation of Swiss Federal Legislation. Accessed: 2026-01-07. [Online]. Available: https://www.fedlex.admin.ch/eli/cc/2018/32/en

[75] Netzpolitik, "Internes dokument: Eu-staaten fordern ein jahr vorratsdatenspeicherung für internet-dienste wie messenger," https://netzpolitik.org/2025/internes-dokument-eu-staaten-fordern-ein-jahr-vorratsdatenspeicherung-fuer-internet-dienste-wie-messenger/, December 2025, accessed: 2026-01-07.

[76] Y. Lindell and B. Pinkas, "An efficient protocol for secure two-party computation in the presence of malicious adversaries," in *Advances in Cryptology – EUROCRYPT 2007*. Springer, 2007, pp. 52–78.

[77] Google, "Play Integrity API," https://developer.android.com/google/play/integrity, 2024, accessed: 2026-05-21.

[78] Apple Inc., "Establishing your app's integrity," https://developer.apple.com/documentation/devicecheck/establishing-your-app-s-integrity, 2024, accessed: 2026-05-21.

[79] H. Abelson, R. Anderson, S. M. Bellovin, J. Benaloh, M. Blaze, W. W. Diffie, J. Gilmore, M. Green, S. Landau, P. G. Neumann *et al.*, "Keys under doormats," *Communications of the ACM*, vol. 58, no. 10, pp. 24–26, 2015.

[80] H. Abelson, R. Anderson, S. M. Bellovin, J. Benaloh, M. Blaze, W. Diffie, J. Gilmore, P. G. Neumann, R. L. Rivest, J. I. Schiller *et al.*, "The risks of key recovery, key escrow, and trusted third-party encryption," 1997.

[81] S. Checkoway, J. Maskiewicz, C. Garman, J. Fried, S. Cohney, M. Green, N. Heninger, R.-P. Weinmann, E. Rescorla, and H. Shacham, "A systematic analysis of the juniper dual ec incident," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 468–479.

[82] CISA, "Countering Chinese State-Sponsored Actors Compromise of Networks Worldwide to Feed Global Espionage System," https://www.cisa.gov/news-events/cybersecurity-advisories/aa25-239a, 2025.

[83] S. Masson, A. Sanso, and Z. Zhang, "Bandersnatch: a fast elliptic curve built over the BLS12-381 scalar field," Cryptology ePrint Archive, Paper 2021/1152, 2021. [Online]. Available: https://eprint.iacr.org/2021/1152

[84] B. WhiteHat, M. Bellés, and J. Baylina, "ERC-2494: Baby Jubjub Elliptic Curve," Ethereum Improvement Proposals, Jan. 2020, accessed: 2026-05-17. [Online]. Available: https://eips.ethereum.org/EIPS/eip-2494

[85] S. Pohlig and M. Hellman, "An improved algorithm for computing logarithms overgf(p)and its cryptographic significance (corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 106–110, 1978.

[86] J. Groth, "On the size of pairing-based non-interactive arguments," in *Annual international conference on the theory and applications of cryptographic techniques*. Springer, 2016, pp. 305–326.

[87] A. Segal, B. Ford, and J. Feigenbaum, "Catching bandits and only bandits: Privacy-Preserving intersection warrants for lawful surveillance," in *4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14)*. San Diego, CA: USENIX Association, Aug. 2014. [Online]. Available: https://www.usenix.org/conference/foci14/workshop-program/presentation/segal

[88] M. S. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation." in *NDSS*, vol. 20, 2012, p. 12.

[89] P. Grubbs, M.-S. Lacharité, B. Minaud, and K. G. Paterson, "Learning to reconstruct: Statistical learning theory and encrypted database attacks," in *IEEE Symposium on Security and Privacy (S&P) 2019*, 2019.

[90] G. Kellaris, G. Kollios, K. Nissim, and A. O'neill, "Generic attacks on secure outsourced databases," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 1329–1340.

[91] M. Jakobsson and A. Juels, "Mix and match: Secure function evaluation via ciphertexts," in *International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT)*. Springer, 2000, pp. 162–177.

[92] B. Laurie, A. Langley, and E. Kasper, "Certificate Transparency," in *RFC 6962, Internet Engineering Task Force (IETF)*, 2013.

[93] M. S. Melara, A. Blankstein, J. Bonneau, E. W. Felten, and M. J. Freedman, "CONIKS: Bringing Key Transparency to End Users," in *24th USENIX Security Symposium*, 2015, pp. 383–398.

[94] S. Jaques, H. Montgomery, and M. Lodder, "Allosaur: Accumulator with low-latency oblivious sublinear anonymous credential updates with revocations," in *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, 2024, pp. 1708–1723.

[95] A. Colombo, "Scalable and Privacy-Preserving Revocation of Verifiable Credentials," https://github.com/alecolo129/eid-revocation-rs, September 2024.

[96] V. Goyal, C.-D. Liu-Zhang, and R. Ostrovsky, "Asymmetric multi-party computation," *Cryptology ePrint Archive*, 2023.

[97] CISA/NSA, "Quantum-Readiness: Migration to Post-Quantum Cryptography," https://www.cisa.gov/resources-tools/resources/quantum-readiness-migration-post-quantum-cryptography, August 2023.

[98] Federal Office for Information Security (BSI), "Quantum-safe cryptography – fundamentals, current developments and recommendations," https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Brochure/quantum-safe-cryptography.html, May 2022.

[99] M. Mosca and M. Piani, "Quantum Threat Timeline Report," https://globalriskinstitute.org/publication/quantum-threat-timeline-report-2025b/, 2025.

[100] M. Keller, "Mp-spdz: A versatile framework for multi-party computation," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2020, pp. 1575–1590.

[101] Swiss Federal Department of Justice and Police. (2026) Beta credential service (bcs) - request a beta-id. Accessed: 2026-05-24. [Online]. Available: https://www.bcs.admin.ch/bcs-web/#/beta-id/create

[102] M. Frigo and abhi shelat, "Longfellow ZK Github Repository," https://github.com/google/longfellow-zk, 2025.

[103] M. Keller, "Mp-spdz: A versatile framework for multi-party computation," in *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 2020, pp. 1575–1590.

[104] M. Keller, E. Orsini, and P. Scholl, "Mascot: faster malicious arithmetic secure computation with oblivious transfer," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 830–842.

[105] O. Terbu, T. Lodderstedt, K. Yasuda, D. Fett, and J. Heenan, "OpenID for Verifiable Presentations 1.0," https://openid.net/specs/openid-4-verifiable-presentations-1_0.html, OpenID, July 2025.

[106] F. F.-H. Nah, "A study on tolerable waiting time: how long are web users willing to wait?" *Behaviour & Information Technology*, vol. 23, no. 3, pp. 153–163, 2004.

[107] J. Nielsen, "Response times: The 3 important limits," https://www.nngroup.com/articles/response-times-3-important-limits/, January 1993, accessed: 2026-05-30.

[108] J. Camenisch and A. Lehmann, "(un) linkable pseudonyms for governmental databases," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1467–1479.

[109] J. Camenisch, S. Hohenberger, M. Kohlweiss, A. Lysyanskaya, and M. Meyerovich, "How to win the clonewars: efficient periodic n-times anonymous authentication," in *Proceedings of the 13th ACM conference on Computer and communications security*, 2006, pp. 201–210.

[110] M. Green, G. Kaptchuk, and G. Van Laer, "Abuse resistant law enforcement access systems," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2021, pp. 553–583.

[111] S. Goldwasser and S. Park, "Public accountability vs. secret laws: can they coexist? a cryptographic proposal," in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017, pp. 99–110.

[112] J. Frankle, S. Park, D. Shaar, S. Goldwasser, and D. Weitzner, "Practical accountability of secret processes," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 657–674.

[113] R. Mayrhofer, A. Lehmann, and a. shelat, "A brief note on cryptographic pseudonyms for anonymous credentials," *arXiv preprint arXiv:2510.05419*, 2025.

[114] M. Chase, F. B. Durak, and S. Vaudenay, "Anonymous tokens with hidden metadata bit from algebraic MACs," Cryptology

ePrint Archive, Paper 2022/1622, 2022. [Online]. Available: https://eprint.iacr.org/2022/1622

[115] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized anonymous payments from bitcoin," in *2014 IEEE symposium on security and privacy*. IEEE, 2014, pp. 459–474.

[116] P. P. Tsang, M. H. Au, A. Kapadia, and S. W. Smith, "Blacklistable anonymous credentials: blocking misbehaving users without ttps," in *Proceedings of the 14th ACM conference on Computer and communications security*, 2007, pp. 72–81.

[117] "Truage," https://www.mytruage.org, NACS, Conexxus, 2026.

[118] "Bundesgesetz über den elektronischen Identitätsnachweis und andere elektronische Nachweise (BGEID)," https://www.fedlex.admin.ch/eli/fga/2025/20/, Swiss Federal Chancellery, 2025.

[119] European Parliament and Council of the European Union, "Regulation (EU) 2024/1183 establishing the European Digital Identity Framework," https://eur-lex.europa.eu/eli/reg/2024/1183/oj/eng, 2024.

[120] ——, "Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (ePrivacy Directive)," https://eur-lex.europa.eu/eli/dir/2002/58/oj, 2002, art. 15(1) permits Member States to adopt legislative measures requiring retention of communications data where necessary, appropriate and proportionate for national security, defence, public security, or criminal investigation.

APPENDIX A
CHAT CONTROL

Chat control is a set of measures planned by the EU to detect CSAM (Child Sex and Abuse Material) in encrypted messages [19]. Proponents of chat control want to work around end-to-end encryption, e.g., using client-side scanning or detection orders for providers. Their goal is to search encrypted content and the proactive detection of suspicious behavior, e.g. CSAM. If suspicious patterns are found, alerts for further investigations are generated. Hence, *probable cause is created* from observing patterns in encrypted data. The politicians' plans have provoked widespread resistance among researchers [20].

In contrast, our idea *does not create probable cause*, but acts on it. The CFT does not search for patterns in encrypted data or workaround end-to-end encryption. It simply revokes anonymity of single malicious users, following a controlled process respecting privacy of benign users.

The recurring push by EU politicians towards chat control demonstrates the necessity of a minimum level of accountability for privacy-preserving protocols. Some politicians would go as far as bypassing end-to-end encryption and surveil all citizens just to generate suspect lists. With the CFT, no encrypted data are mined for potential patterns. It simply balances the individual right to privacy with the societal right for self-protection in case of probable cause, respecting separation of powers in democratic countries.

APPENDIX B
ATTRIBUTE COMMITMENTS: MERKLE TREE VS. FLAT HASH

To prove that claim attributes are indeed part of a valid issuer-signed credential, two commitment representations can be used. In a **Merkle representation**, the issuer signs a Merkle root of all attributes and the prover shows membership proofs for only the used attributes. In a **flat-hash representation**, the issuer signs one flat hash over all attributes and the prover proves consistency of the used attributes with that signed hash. For this benchmark we use a zk-friendly implementation (Circom/rapidsnark, Poseidon hash).

This benchmark isolates only the attribute-to-commitment part: computing/proving consistency up to the Merkle root (Merkle case) or up to the flat hash (flat case). Signature verification is intentionally excluded, because it is the same cost in both variants and would not affect their relative comparison.

This benchmark was executed on a mobile-like aarch64 machine with 2 CPU cores and 4 GB of RAM. Each benchmark was repeated 10 times, and we report mean prover and verifier runtimes.

The table below compares both approaches for 8–64 total attributes and 1–16 used attributes.

TABLE V: Merkle vs. Flat Hash Performance (ms): Prover / Verifier. Fastest approach in bold.

Used Attrs	Total Attrs 8	16	32	64
1	**66 / 16**	**76 / 18**	**84 / 18**	**92 / 16**
	87 / 14	116 / 13	174 / 12	322 / 15
2	88 / 12	**103 / 15**	**110 / 13**	**114 / 14**
	83 / 13	123 / 15	181 / 13	307 / 15
4	111 / 13	142 / 13	**154 / 13**	**165 / 12**
	93 / 14	**112 / 13**	176 / 14	331 / 17
8	183 / 13	205 / 14	243 / 13	**264 / 15**
	84 / 11	**118 / 11**	**185 / 10**	338 / 11
16		366 / 11	408 / 11	445 / 10
	—	**118 / 10**	**185 / 10**	**329 / 10**

Note: Each cell shows Merkle on the first line and flat on the second line. Timings are rounded to the nearest millisecond and averaged over 10 iterations. Bold indicates the faster approach in that case.

Main takeaway: if n denotes the total number of attributes and k the number of attributes used in the proof, then the Merkle-based approach has proving cost $O(k \log n)$, while the flat-hash approach has proving cost $O(n)$.

For small k, the logarithmic factor makes Merkle faster. As k grows, the gap shrinks and flat hashing becomes more competitive. Our implementation uses $n = 32$ and $k = 7$, which is closest to the benchmark row $n = 32$, $k = 8$. In that regime, the table already shows flat hashing performing better, which is why we adopted it.

APPENDIX C
CREDENTIAL NON-REVOCATION PROOF SCALING

We compose a packed status-list non-revocation check with the age-check+CFT presentation of Section VI-A. The issuer embeds a global revocation index in a signed attribute slot. Bits are packed at $B = 253$ per Merkle leaf, a population of N credentials occupies $L = \lceil N/B \rceil$ leaves, padded to the next power of two, so the authentication-path depth is $d = \lceil \log_2 \lceil N/B \rceil \rceil$ (equivalently $d = \ell - 7$ for $N = 2^\ell$). The circuit

proves a Merkle path of length d to a public root and that the indexed bit is unset (not revoked).

We time population sizes $N \in \{2^{12}, 2^{16}, 2^{20}, 2^{24}\}$ on both stacks and both machine profiles (ten runs each). Figures 6 and 7 plot mean prover and verifier latency against $\ell = \log_2 N$, solid markers are measured, dashed segments extrapolate a depth-d fit out to $N = 2^{29}$ (EU-population upper bound, $d = 22$). At $d = 0$ the status-list path vanishes and intercepts should recover the CFT-only baseline of Table IV.

a) Standard stack.: Prove and verify grow roughly linearly in d:

$$T^{\text{mobile}}_{\text{prove}}(d) \approx 197\,d + 2119, \quad T^{\text{mobile}}_{\text{verify}}(d) \approx 156\,d + 1537,$$
$$T^{\text{server}}_{\text{prove}}(d) \approx 115\,d + 1296, \quad T^{\text{server}}_{\text{verify}}(d) \approx 88\,d + 898.$$

Extrapolated to $N = 2^{29}$ ($d = 22$) this is about 6.5 s / 5.0 s on mobile and 3.8 s / 2.8 s on server environment.

b) Zk-friendly stack.: Prover time grows gently with d, verifier time is independent of N (fixed-size Groth16 proofs):

$$T^{\text{mobile}}_{\text{prove}}(d) \approx 14.6\,d + 1197, \qquad T^{\text{mobile}}_{\text{verify}} \approx 34,$$
$$T^{\text{server}}_{\text{prove}}(d) \approx 2.6\,d + 480, \qquad T^{\text{server}}_{\text{verify}} \approx 13.$$

Extrapolated to $N = 2^{29}$ proving is about 1.52 s on mobile and 0.54 s on server. Verification stays at about 34 ms (mobile) and 13 ms (server).

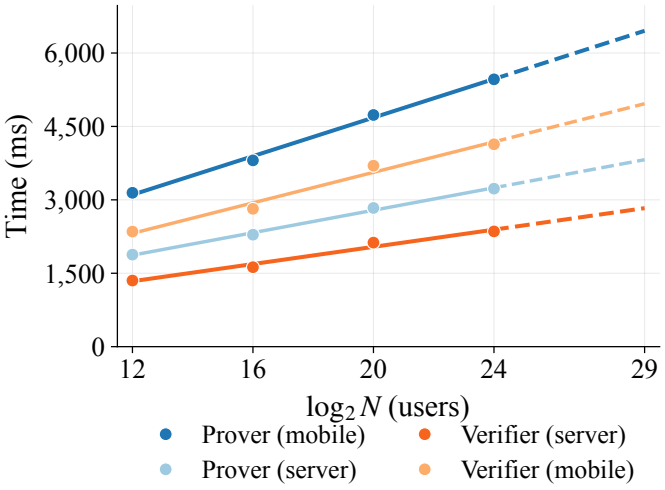

Fig. 6: Packed status-list non-revocation composed with age-check+CFT (standard stack): mean prover and verifier time versus population size N ($\ell = \log_2 N$).

Appendix D
Fully Private Search by Activity Threshold

A. Detailed Algorithms for Suspect-ID and Activity-Threshold Searches

B. Setting and Notation

Let $\mathbb{G}$ be the prime-order subgroup of BabyJubjub generated by the basepoint $G = \texttt{Base8}$, with scalar field $\mathbb{Z}_q$ where $q = \texttt{BABYJUB_ORDER}$. A holder secret key $sk \in \mathbb{Z}_q$ has public key $Q = sk \cdot G \in \mathbb{G}$, and sk is additively-shared among

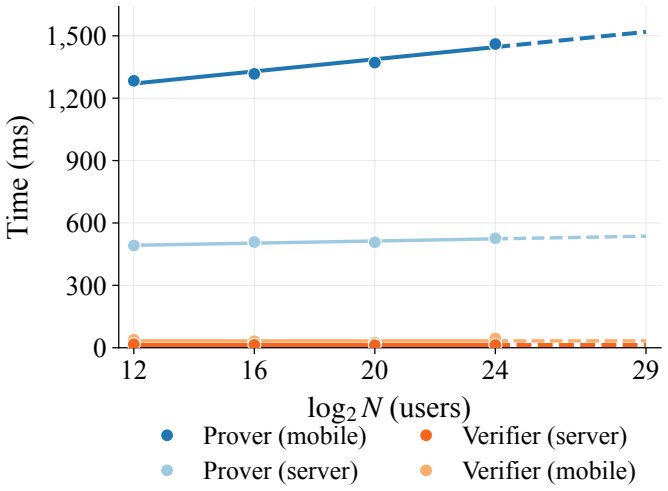

Fig. 7: Packed status-list non-revocation composed with age-check+CFT (zk-friendly stack): mean prover and verifier time versus population size N ($\ell = \log_2 N$).

n parties. ElGamal on $\mathbb{G}$ encrypts a BabyJubjub point identifier $ID \in \mathbb{G}$ as

$$\mathsf{Enc}_Q(ID; r) \;=\; (c_1, c_3) \;=\; \big(r \cdot G,\; ID + r \cdot Q\big),$$

for $ID \in \mathbb{G}$ and $r \in \mathbb{Z}_q$. We write $ct = (c_1, c_3)$. Decryption requires all the parties to each contribute a partial $D_i = s_i \cdot c_1$, which are then combined to obtain $sk \cdot c_1$, recovering $ID = c_3 - sk \cdot c_1$ without ever reconstructing sk.

The corpus is a list $C = (ct_1, \ldots, ct_N)$ of ciphertexts encrypting unknown identities $ID_1, \ldots, ID_N$ under Q. For a threshold τ, we want to reveal exactly those ID_i whose identity appears at least τ times in C:

$$P_\tau(C, i) \;=\; \big[\, |\{j : ID_j = ID_i\}| \;\geq\; \tau \,\big].$$

The protocol has two phases: a pairwise *Plaintext Equality Test* (PET) phase that yields a blinded equality witness for every pair, and an MPC phase that consumes those witnesses to compute and reveal the per-row predicate bits $P_\tau(C, i)$. Rows whose bit is 1 are then opened.

C. The Plaintext Equality Test

The PET, due to Jakobsson and Juels [91], is the workhorse primitive. Given $ct_a = (c_{1,a}, c_{3,a})$ and $ct_b = (c_{1,b}, c_{3,b})$ encrypting ID_a, ID_b under the same key Q, define the homomorphic difference

$$\Delta(ct_a, ct_b) \;\coloneqq\; \big(c_{1,a} - c_{1,b},\; c_{3,a} - c_{3,b}\big),$$

which, by linearity, is an encryption of $ID_a - ID_b$. Each of the t cooperating parties samples a fresh $\alpha_i \in \mathbb{Z}_q$ uniformly at random and contributes $\alpha_i \cdot \Delta$. The summed contribution gives the *blinded difference*

$$\Delta'(ct_a, ct_b) \;\coloneqq\; \big(\alpha \cdot (c_{1,a} - c_{1,b}),\; \alpha \cdot (c_{3,a} - c_{3,b})\big),$$

where $\alpha = \sum_i \alpha_i \bmod q$; this is an encryption of $\alpha(ID_a - ID_b)$.

Were the parties to threshold-decrypt Δ' here, they would recover the joint point $M' = \alpha(ID_a - ID_b) \in \mathbb{G}$ in the clear, and the bit $[M' = O]$ would tell them whether the two IDs are equal — exposing the entire pairwise matching pattern of the corpus across all PETs. To prevent this leakage, our protocol does *not* reconstruct M'. Instead, each party i computes its own partial decryption solely. The joint point M' then lives only inside the MPC, in shared form, and is never opened. The cooperating parties therefore learn nothing about whether $ID_a = ID_b$ from the PET itself.

Lemma 3 (PET output distribution)**.** *Inside the MPC, the (still-shared) joint point recovered from the D_i and $\Delta'.c_3$ equals $M' = \alpha(ID_a - ID_b)$. If $ID_a = ID_b$, then $M' = O$ (the identity element of $\mathbb{G}$) with probability* 1*. If $ID_a \neq ID_b$ and at least one honest party contributed a uniform α_i, then M' is uniformly distributed over $\mathbb{G} \setminus \{O\}$.*

Proof sketch. $M' = \alpha(ID_a - ID_b)$. When $ID_a = ID_b$ this is O for any α. Otherwise $ID_a - ID_b$ generates $\mathbb{G}$ (since $\mathbb{G}$ has prime order), the map $\alpha \mapsto \alpha(ID_a - ID_b)$ is a bijection $\mathbb{Z}_q \to \mathbb{G}$, and a single honest α_i suffices to make α uniform; the probability of hitting O is $1/q$ and is treated as negligible. □

D. Computing the Predicate inside MPC

Running a PET on every unordered pair (i, j) with $i < j$ yields, for each pair, t partial decryptions $D_{ij,1}, \ldots, D_{ij,t} \in \mathbb{G}$ (one from each cooperating party) plus the public ciphertext component $\Delta'_{ij}.c_3$. The MPC takes the partials as secret-shared inputs and performs the remaining work: reconstruct M'_{ij} in shared form, test it against O, aggregate the equality matrix, and reveal only the per-CFT predicate bits.

We use a OT-based dishonest majority protocol (MP-SPDZ's `mascot-party.x`). The MPC program executed by the parties is:

Step 1. Secret-share the partials. For each pair (i, j), each cooperating party ℓ feeds its additive share $D_{ij,\ell}$ into the MPC. The ciphertext component $\Delta'_{ij}.c_3$ is provided as a public constant.

Step 2. Reconstruct the joint point as shares. Inside the MPC, sum the t shared partials and subtract from $\Delta'_{ij}.c_3$ to obtain a shared representation of M'_{ij}. This is $t - 1$ twisted-Edwards point additions per pair, performed on secret-shared coordinates.

Step 3. Per-pair zero test. For each pair, the MPC computes a shared bit $\widetilde{b}_{ij} = [\, M'_{ij} = O \,]$. This bit is never opened; it lives only as a Shamir share.

Step 4. Build the matrix and sum rows. Place the $\widetilde{b}_{ij}$ into a shared symmetric $N \times N$ matrix $\widetilde{B}$ with $\widetilde{b}_{ii} = 1$ for the self-match. Compute per-row sums $\widetilde{\mathsf{count}}_i = \sum_j \widetilde{b}_{ij}$. Summation is free in Shamir.

Step 5. Threshold comparison. For each i, evaluate the shared predicate $\widetilde{P}_i = [\, \widetilde{\mathsf{count}}_i \geq \tau \,]$ using the standard MPC comparison subprotocol.

Step 6. Reveal predicate bits. Open exactly the N bits $\widetilde{P}_1, \ldots, \widetilde{P}_N$. Nothing else from inside the MPC is revealed: not M'_{ij}, not the equality matrix $\widetilde{B}$, not the row counts $\widetilde{\mathsf{count}}_i$, and not which specific pairs caused $\widetilde{P}_i = 1$.

E. Conditional Opening

Once the parties hold the public bit vector $P_1, \ldots, P_N$, they threshold-decrypt ct_i exactly for each i with $P_i = 1$, recovering ID_i. Duplicates among the revealed identities are deduplicated client-side (the same ID may have $P_i = 1$ for several i). Records with $P_i = 0$ remain encrypted.

F. Implementation

Implementing Step 2 of Section D-D on MP-SPDZ [100] is non-trivial: MP-SPDZ exposes secret-shared field arithmetic over its own native prime field, not over BabyJubjub. Realising a twisted-Edwards point addition over shared inputs therefore requires encoding the curve formula as a circuit of MP-SPDZ multiplications and additions (the projective `add-2008-bbjlp` formula uses ten multiplications per addition), and re-encoding shares into BabyJubjub's coordinate field at every input/output boundary.

For the prototype we evaluate in this paper, we factor the protocol along the curve/MPC boundary and run each side in its native field. Concretely: the cooperating parties broadcast their PET partials and combine them into M'_{ij} *outside* MPC — in our JavaScript coordinator process, where BabyJubjub arithmetic is cheap — and feed only the coordinates of M'_{ij} to the MPC, which then runs Steps 3–6 unchanged. We refer to this variant as *twisted-Edwards on-disk* [7]. Under this variant, the cooperating parties learn the per-pair equality bits during the PET phase — a leakage that does *not* occur in the ideal protocol above. The variant therefore measures the cost of every other step accurately but does not implement the full privacy guarantee.

APPENDIX E
PRIVACY AND SECURITY OF THE CRYPTOGRAPHIC FORENSIC TRAIL (CFT)

We prove the statements of Theorems 1 and 2 from Section IV-C. Throughout, let E be an elliptic curve over $\mathbb{F}_p$, let $G \in E(\mathbb{F}_p)$ be a generator of a cyclic subgroup of prime order q, and model the hash function used in the construction as a random oracle H. The user holds a key pair (ID_u, sk_u) with $ID_u = sk_u G$; the secret key sk_u is hardware-bound (WSCD) and never leaves the device. We write $PK := sk \cdot G$ for the joint revocation public key ($sk = sk_P + sk_J + sk_N$ in the distributed setting), and write $m := \mathsf{H}(t, r_2 PK)$ for the binding message signed in C_4.

Background. We now present some background to better understand the following proofs:

[7] "On-disk" is a slight misnomer: the curve sums happen in the coordinator process, not in storage. We use the term to emphasise that the points cross the MPC boundary in plaintext.

- **The DDH Assumption:** the decisional Diffie-Hellman problem is hard in $\langle G \rangle$.
- **The Random Oracle Model:** The hash function is modeled as a random oracle H: $\{0,1\}^* \to \{0,1\}^\ell$, accessible to all parties including the adversary.
- **EUF-CMA:** the signature scheme $(\mathsf{Sign}, \mathsf{Verify})$ used for C_4 is existentially unforgeable under chosen-message attack.

A. IND-Style Indistinguishability

The adversary chooses two candidate key pairs (ID_0, sk_0) and (ID_1, sk_1)—including the secret keys—and is given a challenge CFT that encrypts one of them, while also having access to a sampling oracle that returns fresh valid CFTs for the *same* hidden key pair, with fresh randomness each time. The adversary may query this oracle an arbitrary number of times, as long as the total number of queries is bounded by some polynomial $q(\lambda)$ in the security parameter.

Giving $\mathcal{A}$ both secret keys does not break the game: to check C_4 under either key, $\mathcal{A}$ needs the binding message $m = \mathsf{H}(t, r_2 PK)$, and computing $r_2 PK$ from the public values $(C_2, PK) = (r_2 G, skG)$ is equivalent to CDH. Thus $\mathcal{A}$ cannot run Verify offline.

1) Sample $sk \xleftarrow{\$} \mathbb{Z}_q \setminus \{0,1\}$. Set $PK := skG$.
2) The adversary outputs two candidate key pairs (ID_0, sk_0), (ID_1, sk_1) with $ID_i = sk_i G$, and a tag $t^\star \in \mathbb{Z}_q \setminus \{0\}$. Sample a challenge bit $b \xleftarrow{\$} \{0,1\}$ and a challenge triple $(r_1^\star, r_2^\star) \xleftarrow{\$} (\mathbb{Z}_q \setminus \{0,1\})^2$. Set
$$(C_1^\star, C_2^\star, C_3^\star, C_4^\star) := \big(r_1^\star G,\ r_2^\star G,\ ID_b + r_1^\star PK,\ \mathsf{Sign}(sk_b, \mathsf{H}(t^\star, r_2^\star PK))\big).$$
Give $\mathcal{A}$ the public parameters (G, PK) and the challenge transcript $(C_1^\star, C_2^\star, C_3^\star, C_4^\star, t^\star)$.
3) Oracle Samp: on each query, sample fresh
$$(r_1, r_2, t) \xleftarrow{\$} (\mathbb{Z}_q \setminus \{0,1\})^2 \times (\mathbb{Z}_q \setminus \{0\})$$
and return
$$(C_1, C_2, C_3, C_4, t) := (r_1 G,\ r_2 G,\ ID_b + r_1 PK,\ \mathsf{Sign}(sk_b, \mathsf{H}(t, r_2 PK)),\ t).$$
4) The adversary outputs a bit b' and wins if $b' = b$.

If an adversary $\mathcal{A}$ wins this game with probability $\frac{1}{2} + \varepsilon$ while making at most q_λ queries to Samp and H, then there exists a PPT algorithm $\mathcal{B}$ solving DDH with advantage $\mathsf{Adv}^{DDH}(\mathcal{B}) \geq \varepsilon - q_\lambda/q$, which yields Theorem 1.

a) Proof of Theorem 1: We construct an explicit PPT algorithm $\mathcal{B}$ that uses $\mathcal{A}$ as a subroutine to distinguish DDH tuples.

Input to $\mathcal{B}$. $\mathcal{B}$ receives a DDH challenge (G, aG, bG, Z) where either $Z = abG$ (real case) or $Z \xleftarrow{\$} \langle G \rangle$ (random case), and must output a bit.

Simulation by $\mathcal{B}$. $\mathcal{B}$ simulates the indistinguishability game for $\mathcal{A}$ as follows:

1) Set the public key $PK := bG$ (so the implicit secret key is $sk = b$, unknown to $\mathcal{B}$).
2) Receive (ID_0, sk_0), (ID_1, sk_1) and $t^\star$ from $\mathcal{A}$ (abort if $ID_i \neq sk_i G$). Sample a challenge bit $\beta \xleftarrow{\$} \{0,1\}$ and let (ID_β, sk_β) denote the hidden key pair. Sample $r_2^\star \xleftarrow{\$} \mathbb{Z}_q \setminus \{0,1\}$ and set
$$C_1^\star := aG,\ C_2^\star := r_2^\star G,\ C_3^\star := ID_\beta + Z.$$
(Implicitly, define $r_1^\star := a$ so that $C_1^\star = r_1^\star G$.) Since $\mathcal{B}$ knows the scalar $r_2^\star$ and the point PK, it can compute $r_2^\star PK$ by scalar multiplication, query $m^\star \leftarrow \mathsf{H}(t^\star, r_2^\star PK)$, and set
$$C_4^\star := \mathsf{Sign}(sk_\beta, m^\star).$$
Answer all other H-queries uniformly at random, consistently.
3) Send $\mathcal{A}$ the challenge transcript $(C_1^\star, C_2^\star, C_3^\star, C_4^\star, t^\star)$ along with the public parameters (G, PK).
4) Answer each Samp query from $\mathcal{A}$ by sampling fresh:
$$(r_i, r_{2,i}, t_i) \xleftarrow{\$} (\mathbb{Z}_q \setminus \{0,1\})^2 \times (\mathbb{Z}_q \setminus \{0\})$$
Define $r_{1,i} := r_1^\star + r_i$ and set
$$C_1 := C_1^\star + r_i G,\ C_2 := r_{2,i} G,\ C_3 := C_3^\star + r_i PK.$$
(If $Z = abG$ then $C_3^\star = ID_\beta + r_1^\star PK$, hence $C_3 = ID_\beta + r_{1,i} PK$ and $C_1 = r_{1,i} G$.)
Compute $m_i \leftarrow \mathsf{H}(t_i, r_{2,i} PK)$ and $C_4 := \mathsf{Sign}(sk_\beta, m_i)$.
Return $(C_1, C_2, C_3, C_4, t_i)$.

$\mathcal{B}$ outputs 1 iff $\mathcal{A}$ outputs the correct bit β, otherwise $\mathcal{B}$ outputs 0.

Case 1: $Z = abG$ (real case).
Then $C_1^\star = aG$ and

$$C_3^\star = ID_\beta + Z = ID_\beta + abG = ID_\beta + a \cdot PK = ID_\beta + r_1^\star PK.$$

Moreover, for each Samp reply we pick fresh $r_i \xleftarrow{\$} \mathbb{Z}_q$ and define $r_{1,i} := r_1^\star + r_i$; then $C_1 = C_1^\star + r_i G = r_{1,i} G$ and $C_3 = C_3^\star + r_i PK = ID_\beta + r_{1,i} PK$. The signatures C_4 are honest signatures under sk_β on the correct binding messages. Therefore the simulated transcript (and likewise each Samp reply) is *identically distributed* to a real game transcript. Therefore:

$$\Pr[\mathcal{A} \text{ wins} \mid Z = abG] = \frac{1}{2} + \varepsilon.$$

This means that when $Z = abG$, $\mathcal{B}$ outputs 1 with probability

$$\Pr[\mathcal{B} = 1 \mid Z = abG] = \frac{1}{2} + \varepsilon.$$

Case 2: $Z \xleftarrow{\$} \langle G \rangle$ (random case).
Z is a uniformly random point in $\langle G \rangle$, independent of ID_β, PK, $C_1^\star$ and $C_2^\star$. Consequently, the C_3 components are distributed as $ID_\beta + (Z + r_i PK)$ for fresh $r_i \xleftarrow{\$} \mathbb{Z}_q$. Since $r_i PK$ is uniform over $\langle G \rangle$, every C_3 is uniform over $\langle G \rangle$ and

independent of ID_β, thus even polynomially many sampling queries leak no information about ID_β through C_3.

It remains to argue that C_4 does not help either, even though $\mathcal{A}$ knows both sk_0 and sk_1. To test whether C_4 verifies under ID_0 or ID_1, $\mathcal{A}$ must evaluate $\mathsf{Verify}(ID_i, C_4, m)$ for $m = \mathsf{H}(t, r_2 PK)$. By the random oracle assumption, $\mathcal{A}$ obtains this m only by querying H at the exact point $(t, r_2 PK)$. Computing $r_2 PK$ from the public pair $(C_2, PK) = (r_2 G, skG)$ is CDH-hard; equivalently, the probability that any of $\mathcal{A}$'s q_λ random-oracle queries coincidentally hits the correct point is at most q_λ/q. Without m, $\mathcal{A}$ cannot run verification under either key, so C_4 yields no distinguishing advantage beyond this negligible loss. Therefore the challenge bit remains hidden except with this negligible loss, and we get

$$\Pr[\mathcal{B} = 1 \mid Z \text{ random}] \;\le\; \frac{1}{2} + \frac{q_\lambda}{q}.$$

Distinguishing Advantage. $\mathcal{B}$'s DDH advantage is:

$$\begin{aligned} \mathsf{Adv}^{DDH}(\mathcal{B}) \;&=\; \big|\Pr[\mathcal{B} = 1 \mid Z = abG] \\ &\qquad -\; \Pr[\mathcal{B} = 1 \mid Z \text{ random}]\big| \\ &\ge\; \varepsilon \;-\; \frac{q_\lambda}{q}. \end{aligned}$$

If ε is non-negligible, so is $\mathcal{B}$'s advantage, contradicting the DDH assumption.

B. Unforgeability

We now prove that the CFT is also unforgeable, that is, a CFT cannot be repurposed to a fresh request tag t'. Throughout this subsection we assume that the adversary $\mathcal{A}$ *does not know* the user's hardware-held signing key sk_u but may know the corresponding public key $ID_u = sk_u G$, the joint public key PK, and previously issued CFT transcripts.

A purported forgery for a new tag t' is a tuple $(C_1', C_2', C_3', C_4', t')$ that opens correctly under ID_u:

$$C_3' = ID_u + sk\, C_1' \;\text{ and }\; \mathsf{Verify}\big(ID_u,\, C_4',\, \mathsf{H}(t', sk\, C_2')\big) = 1.$$

Knowing ID_u, $\mathcal{A}$ can freely sample r_1', r_2' and set $C_1' = r_1' G$, $C_2' = r_2' G$, $C_3' = ID_u + r_1' PK$, so forging consistent (C_1', C_2', C_3') is easy. The binding check, however, requires a valid signature

$$C_4' = \mathsf{Sign}\big(sk_u,\, \mathsf{H}(t', r_2' PK)\big).$$

Under EUF-CMA, no PPT adversary that does not hold sk_u can produce such a C_4' on a fresh message $\mathsf{H}(t', r_2' PK)$. Therefore forging a complete valid CFT for a new tag is infeasible, which yields Theorem 2.

Algorithm 3: CFTRevokeBySuspectID

Parties: Police P, Judge J, NGO N, verifiers V.
Inputs: `reason`, `aux`, ID_{SW}, `filter`, $\mathrm{ID}_{\mathrm{suspect}}$, key pairs (pk_i, sk_i) for $i \in \{P, J, N\}$.
Convention: $\#(x, Y)$ counts occurrences of x in multiset Y.

$\sigma_{SW} \leftarrow \mathsf{IssueCondWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW}, \texttt{filter}, \mathrm{ID}_{\mathrm{suspect}})$
$\mathcal{C}_V \leftarrow \mathsf{ObtainTrails}(P, \{V\}, \sigma_{SW}, \texttt{filter})$
P prepends $\mathrm{ID}_{\mathrm{suspect}}$ to each entry in $\mathcal{C}_V$.
$\mathcal{C}_P \leftarrow \mathsf{BatchLink}(P, \mathcal{C}_V, \texttt{filter}, sk_P)$

P sends $\mathcal{C}_P$, σ_{SW} to J.
if σ_{SW} invalid **then**
  J aborts.
$\mathcal{C}_J \leftarrow \mathsf{BatchLink}(J, \mathcal{C}_P, \texttt{filter}, sk_J)$
$\sigma_{SW,J} \leftarrow \mathsf{Sign}_{sk_J}(ID_{SW} \,\|\, \mathsf{H}(\mathcal{C}_J))$
J sends $\mathcal{C}_J$, $\sigma_{SW,J}$ to P.

P sends $\mathcal{C}_J$, σ_{SW}, $\sigma_{SW,J}$ to N.
if σ_{SW} or $\sigma_{SW,J}$ invalid **then**
  N aborts.
$\mathcal{C}_N \leftarrow \mathsf{BatchLink}(N, \mathcal{C}_J, \texttt{filter}, sk_N)$ (entries carry $\mathrm{PID} = D_{(3,N)}$)
$\mathcal{C}_{N,2} \leftarrow \mathsf{BatchDeanonymize}(N, \{\text{entries in } \mathcal{C}_N \text{ with PID} = \mathrm{PID}_{\text{suspect}}\}, sk_N)$
$\sigma_{SW,N} \leftarrow \mathsf{Sign}_{sk_N}(ID_{SW} \,\|\, \mathsf{H}(\mathcal{C}_{N,2}))$
N sends $\mathcal{C}_{N,2}$, σ_{SW}, $\sigma_{SW,N}$ to P.

P forwards $\mathcal{C}_{N,2}$, $\mathrm{PID}_{\text{suspect},N}$, σ_{SW}, $\sigma_{SW,N}$ to J.
if σ_{SW} or $\sigma_{SW,N}$ invalid **then**
  J aborts.
$\mathcal{C}_{J,2} \leftarrow \mathsf{BatchDeanonymize}(J, \{\text{entries in } \mathcal{C}_{N,2} \text{ with PID} = \mathrm{PID}_{\text{suspect}}\}, sk_J)$
P receives $\mathcal{C}_{J,2}$, $\mathrm{PID}_{\text{suspect},J}$ from J.

$\mathcal{C}_{P,2} \leftarrow \mathsf{BatchDeanonymize}(P, \{\text{entries in } \mathcal{C}_{judge,2} \text{ with PID} = \mathrm{PID}_{\text{suspect}}\}, sk_P)$
P combines partial decryptions and checks bindings:
for all entries in $\mathcal{C}_{P,2}$ **do**
  $D_2 \leftarrow D_{(2,P)} + D_{(2,N)} + D_{(2,J)}$
  $t \leftarrow H(ts, \mathrm{RID}, ID_V)$
  if $\mathsf{VerifySign}_{\mathrm{ID}_{\mathrm{suspect}}}(C_4, H(t, D_2)) = 0$ **then**
    discard entry from $\mathcal{C}_{P,2}$.
P outputs $\mathrm{ID}_{\mathrm{suspect}}$ and the set of opened traces $\mathcal{C}_{P,2}$

$\mathsf{IssueCondWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW}, \texttt{filter}, \mathrm{ID}_{\mathrm{suspect}})$

P presents $(\texttt{reason}, \mathrm{ID}_{\mathrm{suspect}}, \texttt{filter})$ to J.
if `reason`, $\mathrm{ID}_{\mathrm{suspect}}$, or `filter` not legitimate **then**
  J aborts.
$\sigma_{SW} \leftarrow \mathsf{Sign}_{sk_J}(ID_{SW} \,\|\, \texttt{reason} \,\|\, \texttt{aux} \,\|\, \texttt{filter} \,\|\, \mathrm{ID}_{\mathrm{suspect}})$
return σ_{SW}

$\mathsf{ObtainTrails}(P, \{V\}, \sigma_{SW}, \texttt{filter})$

$\mathcal{C}_V \leftarrow \emptyset$
for all verifiers V in scope **do**
  P sends σ_{SW} to V.
  if σ_{SW} invalid **then**
    V aborts.
  for all $(CFT, ts, \mathrm{RID}, ID_V)$ satisfying `filter` **do**
    $\sigma_{SW,V} \leftarrow \mathsf{Sign}_{sk_V}(ID_{SW} || ts || \mathrm{RID} || ID_V || H(CFT))$
    $D_{(1,V)} \leftarrow C_1$
    $D_{(3,V)} \leftarrow C_3$
    add $(CFT, (ts, \mathrm{RID}, ID_V), D_{(1,V)}, D_{(3,V)}, \sigma_{SW,V})$ to $\mathcal{C}_V$.
return $\mathcal{C}_V$

$\mathsf{BatchLink}(X, \mathcal{C}_{\mathrm{in}}, \texttt{filter}, sk_X)$

$\mathcal{C}_{\mathrm{out}} \leftarrow \emptyset$.
for all e in $\mathcal{C}_{\mathrm{in}}$ with valid $\sigma_{SW,V}$ satisfying `filter` **do**
  generate fresh k_X;
  Parse e as: $(\mathrm{PID}_{(\mathrm{suspect,prev})}, CFT, (ts, \mathrm{RID}, ID_V), D_{(1,\mathrm{prev})}, D_{(3,\mathrm{prev})}, \sigma_{SW,V})$
  $(D_{(1,X)}) \leftarrow k_X \cdot D_{(1,\mathrm{prev})}$
  $(D_{(3,X)}) \leftarrow k_X \cdot (D_{(3,\mathrm{prev})} - sk_X \cdot D_{(1,\mathrm{prev})})$
  $\mathrm{PID}_{(\mathrm{suspect},X)} \leftarrow k_X \cdot \mathrm{PID}_{(\mathrm{suspect,prev})}$
  add the following tuple to $\mathcal{C}_{\mathrm{out}}$:
  $(\mathrm{PID}_{(\mathrm{suspect},X)}, CFT, (ts, \mathrm{RID}, ID_V), D_{(1,X)}, D_{(3,X)}, \sigma_{SW,V})$
return $\mathcal{C}_{\mathrm{out}}$

$\mathsf{BatchDeanonymize}(X, \mathcal{C}_{\mathrm{in}}, sk_X)$

$\mathcal{C}_{\mathrm{out}} \leftarrow \emptyset$.
for all e in $\mathcal{C}_{\mathrm{in}}$ with valid $\sigma_{SW,V}$ satisfying `filter` **do**
  get $CFT = (C_1, C_2, C_3, C_4)$ from e
  $D_{(2,X)} \leftarrow sk_X \cdot C_2$
  append $D_{(2,X)}$ to the entry e
  add e to $\mathcal{C}_{\mathrm{out}}$.
return $\mathcal{C}_{\mathrm{out}}$

Fig. 8: Searching by suspect ID: CFTRevokeBySuspectID (left) and its subroutines (right).

$\mathsf{CFTCondOpen}(P, J, N, \{V\}, \texttt{reason}, \texttt{aux}, ID_{SW}, \texttt{filter}, \tau)$

Parties: Police P, Judge J, NGO N, verifiers V.
Inputs: `reason`, `aux`, ID_{SW}, `filter`, τ, key pairs (pk_i, sk_i) and warrant-scoped symmetric keys k_i for $i \in \{P, J, N\}$ (fresh per warrant instance).
Convention: $\#(x, Y)$ counts occurrences of x in multiset Y.

$\sigma_{SW} \leftarrow \mathsf{IssueCondWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW}, \texttt{filter}, \tau)$
$\{CFT\}_V \leftarrow \mathsf{ObtainTrails}(P, \{V\}, \sigma_{SW}, \texttt{filter})$
$\{CFT\}_P, k_P \leftarrow \mathsf{BatchLink}(P, \{CFT\}_V, \texttt{filter}, sk_P)$

P sends $\{CFT\}_P$, σ_{SW} to J.
if σ_{SW} invalid **then**
  J aborts.
$\{CFT\}_J, k_J \leftarrow \mathsf{BatchLink}(J, \{CFT\}_P, \texttt{filter}, sk_J)$
$\sigma_{SW,J} \leftarrow \mathsf{Sign}_{sk_J}(ID_{SW} \,\|\, \mathsf{H}(\{CFT\}_J))$
J sends $\{CFT\}_J$, $\sigma_{SW,J}$ to P.

P sends $\{CFT\}_J$, σ_{SW}, $\sigma_{SW,J}$ to N.
if σ_{SW} or $\sigma_{SW,J}$ invalid **then**
  N aborts.
$\{CFT\}_N, k_N \leftarrow \mathsf{BatchLink}(N, \{CFT\}_J, \texttt{filter}, sk_N)$
  (entries carry $\mathrm{PID} = D_{(3,N)}$)
$\{\mathrm{PID}\}_{\text{suspect}} = \{\mathrm{PID} : \mathrm{PID} \in \{CFT\}_N \text{ s.t. } \#(\mathrm{PID}, \{CFT\}_N)\} > \tau$
if $\{\mathrm{PID}\}_{\text{suspect}}$ is empty **then**
  N aborts.
$\{CFT\}_{N,2} \leftarrow \mathsf{BatchDeanonymize}(N,$
  $\{\text{entries in } \{CFT\}_N \text{ with PID} \in \{\mathrm{PID}\}_{\text{suspect}}\}, sk_N, k_N)$
$\sigma_{SW,N} \leftarrow \mathsf{Sign}_{sk_N}(ID_{SW} \,\|\, \mathsf{H}(\{CFT\}_{N,2}))$
N sends $\{CFT\}_{N,2}$, σ_{SW}, $\sigma_{SW,N}$ to P.

P forwards $\{CFT\}_{N,2}$, σ_{SW}, $\sigma_{SW,N}$ to J.
if σ_{SW} or $\sigma_{SW,N}$ invalid **then**
  J aborts.
$\{\mathrm{PID}\}_{\text{suspect},N} = \{\mathrm{PID} : \mathrm{PID} \in \{CFT\}_{N,2} \text{ s.t. } \#(\mathrm{PID}, \{CFT\}_{N,2})\} > \tau$
$\{CFT\}_{J,2} \leftarrow \mathsf{BatchDeanonymize}(J,$
  $\{\text{entries in } \{CFT\}_{N,2} \text{ with PID} \in \{\mathrm{PID}\}_{\text{suspect},N}\}, sk_J, k_J)$
P receives $\{CFT\}_{J,2}$, $\{\mathrm{PID}\}_{\text{suspect},J}$ from J.

$\{\mathrm{PID}\}_{\text{suspect},J} = \{\mathrm{PID} : \mathrm{PID} \in \{CFT\}_{J,2} \text{ s.t. } \#(\mathrm{PID}, \{CFT\}_{J,2})\} > \tau$
$\{CFT\}_{P,2} \leftarrow \mathsf{BatchDeanonymize}(P,$
  $\{\text{entries in } \{CFT\}_{J,2} \text{ with PID} \in \{\mathrm{PID}\}_{\text{suspect},J}\}, sk_P, k_P)$
  (entries carry $ID_U = \mathrm{PID}_P$)
P combines partial decryptions and checks bindings:
for all entries in $\{CFT\}_{P,2}$ **do**
  $D_2 \leftarrow D_{(2,P)} + D_{(2,N)} + D_{(2,J)}$
  $t \leftarrow H(ts, \mathrm{RID}, ID_V)$
  if $\mathsf{VerifySign}_{ID_U}(C_4, H(t, D_2)) = 0$ **then**
    discard entry from $\{CFT\}_{P,2}$.
$\{ID_U\}_{\text{suspect}} = \{ID_U : ID_U \in \{CFT\}_{P,2} \text{ s.t. } \#(ID_U, \{CFT\}_{P,2})\} > \tau$
$\{CFT\}_{\text{suspect}} = \{\text{entries in } \{CFT\}_{P,2} \text{ with } ID_U \in \{ID_U\}_{\text{suspect}}\}$
P outputs $\{ID_U\}_{\text{suspect}}$ and the set of opened traces $\{CFT\}_{\text{suspect}}$

$\mathsf{IssueCondWarrant}(P, J, \texttt{reason}, \texttt{aux}, ID_{SW}, \texttt{filter}, \tau)$

P presents $(\texttt{reason}, \tau, \texttt{filter})$ to J.
if `reason`, τ, or `filter` not legitimate **then**
  J aborts.
$\sigma_{SW} \leftarrow \mathsf{Sign}_{sk_J}(ID_{SW} \,\|\, \texttt{reason} \,\|\, \texttt{aux} \,\|\, \texttt{filter} \,\|\, \tau)$
return σ_{SW}

$\mathsf{BatchLink}(X, \{CFT\}_{\text{in}}, \texttt{filter}, sk_X)$

generate fresh k_X; $\{CFT\}_{\text{out}} \leftarrow \emptyset$.
for all e in $\{CFT\}_{\text{in}}$ with valid $\sigma_{SW,V}$ satisfying `filter` **do**
  Parse e as: $(CFT, (ts, \mathrm{RID}, ID_V), D_{(1,\text{prev})}, D_{(3,\text{prev})}, \sigma_{SW,V})$
  $(D_{(1,X)}) \leftarrow k_X \cdot D_{(1,\text{prev})}$
  $(D_{(3,X)}) \leftarrow k_X \cdot (D_{(3,\text{prev})} - sk_X \cdot D_{(1,\text{prev})})$
  add $(CFT, (ts, \mathrm{RID}, ID_V), D_{(1,X)}, D_{(3,X)}, \sigma_{SW,V})$ to $\{CFT\}_{\text{out}}$.
return $\{CFT\}_{\text{out}}, k_X$

$\mathsf{BatchDeanonymize}(X, \{CFT\}_{\text{in}}, sk_X, k_X)$

$\{CFT\}_{\text{out}} \leftarrow \emptyset$.
for all e in $\{CFT\}_{\text{in}}$ with valid $\sigma_{SW,V}$ satisfying `filter` **do**
  get $\mathrm{PID}_{\text{prev}}$ from e
  $\mathrm{PID}_X = k_X^{-1} \cdot \mathrm{PID}_{\text{prev}}$
  replace $\mathrm{PID}_{\text{prev}}$ with PID_X in e
  $D_{(2,X)} \leftarrow sk_X \cdot C_2$
  append $D_{(2,X)}$ to the entry e
  add e to $\{CFT\}_{\text{out}}$.
return $\{CFT\}_{\text{out}}$

Fig. 9: Searching by activity threshold (without MPC): conditional CFT opening $\mathsf{CFTCondOpen}$ (left) and its subroutines (right).