When Age Checks Need Not Reveal Identity: A Construction and Distribution Criterion for Privacy-Preserving Attribute Verification

Greg Magarshak
IE University (NYC) and Safebots, Inc.
greg@magarshak.com

Abstract

A wave of 2026 statutes conditions online access on proving an attribute such as being over 18; the deployed method, uploading an ID and selfie to a third-party verifier, de-anonymizes adults and builds identity honeypots already breached. The prevailing view, held by civil-liberties groups, security researchers, and regulators alike, is that this is unavoidable: that age verification is unsafe *by design* because it must link a real identity to an online action, and that link, stored somewhere even briefly, becomes a target. We refute the premise. The unlinkable-token primitive proving a predicate without revealing the value exists, but the field treats it as insufficient because the surrounding system, who issues the token, what its issuance era leaks, how a live person is bound to it, still forces linkage somewhere; we close those gaps, eliminating the linkage rather than relocating it. Four constructions: an M-of-N threshold committee, so no party, the state included, can deny, tag, or deanonymize below threshold; closure of a vintage-leakage channel surviving unlinkable redemption, where the issuance epoch bounds age from below, by proactive resharing of a stable key or by mixing among current holders, protection provably monotone in the mixes; a band generalization with a computable non-individuating width; and a per-proof binding to a live person via a key sealed in the device's secure element behind an on-device liveness gate, non-replayable and resistant to the remote-deepfake attack that defeats verifier-side selfie checks. The consequence: a jurisdiction enforces an age mandate and preserves privacy at once, leaving no honeypot, with no point in the system where identity is linked to action. The result: a person proves what an institution requires rather than surrendering identity to it, with a computable boundary for where the substitution is safe.

1 Introduction

1.1 A construction, not an impossibility proof

This is an engineering paper. It does not refute a theorem; it builds a system. The distinction matters because the "impossibility" this paper answers is a worst-case statement, and worst-case statements are routinely mistaken for typical-case requirements. Shannon's source-coding bound says no scheme compresses every input below its entropy, yet zip compresses the overwhelming majority of real files, because real files carry structure the worst case lacks. The CAP theorem says no distributed store is consistent and available during a partition, yet eventual consistency runs most of the internet, because partitions are rare. In each case a true worst-case limit was read as a ban on the typical case, and the engineering win came from serving the typical case cheaply while ceding only the sparse hard corner. Privacy-preserving age verification has been trapped in exactly this confusion. The claim that one "cannot prove you are 18 without proving who you are"

is, read precisely, a statement about deployed systems that all happen to link identity to action; it has been heard as a statement about the verification task itself. This paper builds the system the confusion said could not exist, and works for the deployments that make up essentially all of real-world age checking. A narrow residual remains, and we state it in one place (Section 9) rather than scattering it, because naming the rare hard corner is not the same as conceding the typical case, and the two should not be allowed to blur.

1.2 The false binary

The privacy community, civil-liberties organizations, and the regulators charged with implementing these mandates have converged on a pessimistic consensus: that age verification is *unsafe by design*. The argument is structural, not incidental. To check an attribute, the reasoning goes, the system must at some moment connect a real identity to an online action; that connection has to be computed and stored somewhere, even if only briefly; and wherever it lives, it is a target for breach, subpoena, or abuse. On this view the honeypot is not an implementation flaw to be engineered away but an inherent feature of the task, so the most one can hope for is to minimize and protect the linkage, never to remove it. The position is held widely and in good faith: hundreds of security and privacy researchers have asked governments to pause deployment until the underlying science settles, and the deployed systems keep proving the pessimists right, including a national digital-identity wallet marketed as privacy-preserving and zero-knowledge that researchers defeated within minutes of its release by flipping a configuration flag.

This paper's central claim is that the structural premise is false. There need be no moment, and no place, where identity is linked to action. The linkage that the consensus treats as irreducible is an artifact of *how the surrounding system is built*, not of the verification task itself, and we give constructions that eliminate it rather than relocate or shrink it. The novelty is the construction, not a new primitive, and that is the point worth stating clearly rather than apologizing for. The *primitive* that proves a predicate without revealing its value, an unlinkable attribute token, is standard and deployed; engineering is the act of composing standard parts into a system that does what the parts alone do not. What the field has treated as unsolved, and treated the lack of a solution as evidence of impossibility, is the *system around* that primitive: who issues the token without becoming a single point of surveillance, what the token's issuance era silently leaks even when redemption is unlinkable, how a live person is bound to a proof without a reusable identifier, and how a coarsened predicate avoids becoming an identifier in a thin population. Close those four gaps and the linkage disappears. That is the contribution: not a new way to prove age, but a working demonstration that the honeypot everyone has accepted as the price of compliance was never necessary.

Legislation conditioning online access on a verified attribute presents operators with what looks like a forced choice: to know which users are minors, collect identifying information from everyone. The choice is false in the same way the worst-case framings in the two companion papers are false. A verifier almost never needs the underlying value; it needs the truth of a predicate over that value, "age ≥ 18" or "age $\in [30, 39]$". A token that carries exactly the predicate and nothing else satisfies the verifier and discloses no identity. The cryptography for such a token, blind issuance and unlinkable redemption, is standard and deployed; the Privacy-Pass lineage and its browser instantiation as Private State Tokens give issuer-unlinkable redemption at scale, and attribute-based anonymous credentials give predicate proofs over committed values.

What is not solved, and what this paper addresses, is everything around the proof. Two problems matter and neither is cryptographic in the narrow sense.

The first is *who issues*. Every deployed design trusts a single issuer. In the age-verification

setting the issuer is a government or a government-licensed party, and the threat model of the very statutes driving deployment is precisely a state that would like to know who reads, watches, or says what. A single issuer can refuse classes of applicants, tag cohorts through key choices, or, if it learns redemption contexts, correlate a person's sites. Moving the proof to zero knowledge does nothing about an issuer that is itself the adversary.

The second is *distribution*. A right of access conditioned on a credential creates an obligation to issue that credential universally, including to people without smartphones, without passports, without stable addresses, the same population whose privacy the scheme is nominally protecting. The credential literature writes "assume the issuer distributes credentials" and proceeds. The assumption is the hard part.

1.3 The mandates are already here

This is not a hypothetical design exercise. As of 2026 a wave of statutes compels operators to verify users' ages and explicitly rejects self-declaration, so the only question left is *how* verification is done, by identity disclosure or by a privacy-preserving proof.

- **United Kingdom.** The Online Safety Act, in force from July 2025, requires "highly effective" age assurance for services exposing minors to harmful content; the regulator's accepted methods are photo-ID matching, facial age estimation, open banking, mobile-operator and digital-identity checks, and self-declaration is explicitly insufficient. Enforcement is active: the Information Commissioner fined Reddit £14.5 million in early 2026 for over-reliance on self-declaration, and penalties reach the greater of £18 million or 10% of global turnover.

- **European Union.** The Digital Services Act obliges platforms to protect minors, and the Commission's age-verification blueprint, feature-ready 15 April 2026, lets a user prove they are over 18 (or another band such as 13+ or 65+) without disclosing any other personal data, built on the same specifications as the eIDAS 2.0 EU Digital Identity Wallet and slated for member-state rollout by end of 2026. The released application was promoted as zero-knowledge and privacy-preserving; security researchers nonetheless bypassed its protections within minutes, defeating its biometric gate by toggling a configuration flag and reaching credentials stored outside secure hardware, an illustration of the gap between invoking the right primitive and building a system in which the linkage is genuinely absent.

- **France.** Act No. 2024-449 (SREN) mandates robust age verification for pornographic services, and the regulator ARCOM requires a "double anonymity" property: neither the website nor the verification provider may link a user's identity to the content accessed.

- **United States.** With no federal statute, roughly half the states have enacted age-verification requirements for adult content or social media, nine taking effect in 2025 and more in 2026, several under First Amendment challenge but the direction unmistakable. A second front targets the device layer directly: app-store statutes in Texas (SB 2420), Utah, and Louisiana require store operators to verify a user's age and bind minor accounts to a parent, and California's Digital Age Assurance Act (AB 1043) requires operating-system providers to collect age at device setup and broadcast it to apps from 2027, placing the verification checkpoint at exactly the device-and-OS layer where a device-sealed proof would live (Section 8).

- **Australia.** Age-Restricted Material Codes apply from March 2026 across social media, app stores, search, and other services, on a technology-neutral standard that demands demonstrated effectiveness.

Two consequences frame this paper. First, the privacy-preserving direction is now the *official* European position, not a fringe proposal, which means our contribution cannot be the bare idea of proving an attribute without identity; it must be the trust model, the distribution, and the binding of proof to a present person (Section 1.5). Second, operators face a genuine bind: the same statutes that compel age checks coexist with data-protection law, the GDPR's data-minimization and purpose-limitation principles, that penalizes building the identity honeypot the naive "upload your ID" approach creates. A construction that satisfies the age mandate while disclosing no identity resolves the bind in both directions at once; we return to this in Section 7.

1.4 The impossibility consensus, in its own words

The claim this paper refutes is not a strawman we have constructed for rhetorical convenience. It is the explicit, repeated, on-the-record position of the civil-liberties organizations, security researchers, and journalists who have studied age verification most closely, and it is stated as a matter of structural necessity rather than of present engineering immaturity. We quote it directly, because a reader should weigh the strength of the consensus before weighing our answer to it.

The Electronic Frontier Foundation, the most prominent digital-rights organization in the United States, states the position without qualification. In a December 2025 explainer surveying every deployed method, it concludes that no available technology is "entirely privacy-protective" [1], and that every system of age verification or estimation demands personal information that, in its words, "links their offline identity to their online activity" [1]. Elsewhere the organization puts it more bluntly still: online age verification is "incompatible with privacy" [2]. Crucially for a cryptographic audience, the EFF has addressed the obvious technical rejoinder head-on, arguing in a dedicated 2025 analysis that zero-knowledge proofs "alone" do not solve the problem [3], on the grounds that the surrounding identity system still collects and links. The same structural argument recurs across the field in nearly identical language. The recurring impossibility formula appears in two interchangeable forms: that "you cannot prove you are 18 without proving who you are," and, attributed to journalist Taylor Lorenz and repeated widely, that there is "no way to reliably verify someone's age without verifying who they are" [6]. Security author Cory Doctorow, relaying his EFF colleague Jason Kelley, calls age verification flatly "an impossibility" [4], in an essay whose title dismisses the very notion of privacy-preserving age verification as nonsense. An EFF representative told a national broadcaster that the practice "fundamentally undermines" the ability to engage in anonymous speech [5], and the Center for Democracy and Technology observes that even methods marketed as privacy-preserving still "risk exposing users' identities" [7] because all of them require collecting and retaining more data. The conviction is strong enough that courts have acted on it: reviewing a year of litigation, the EFF reports that judges have repeatedly held such mandates unconstitutional, confirming it is "nearly impossible" to impose online age-verification requirements without violating users' First Amendment rights [8].

Read together, these statements make a single, precise structural claim, the one our constructions are built to falsify. It is not merely that deployed systems are sloppy, though they are; it is that the identity-to-action linkage is held to be *irreducible*, an inherent feature of the verification task rather than an artifact of how a given system is assembled. That formula, "you cannot prove you are 18 without proving who you are," is, if true, a statement about the task itself, and it is on that understanding that hundreds of researchers asked the European Commission to pause deployment until the science settles. The deployed evidence has repeatedly vindicated the pessimists: a national digital-identity wallet promoted as zero-knowledge and privacy-preserving was defeated by security researchers within minutes of its release, its biometric gate bypassed by toggling a configuration flag and its credentials reached outside secure hardware [9].

This paper's disagreement is narrow and exact. We accept that every system the consensus examined does link identity to action, and we accept that the bare predicate-proof primitive, which we did not invent and for which we claim no novelty, does not by itself remove that linkage, which is precisely why the EFF is correct that zero-knowledge proofs "alone" are insufficient. What we deny is the universal quantifier. The linkage is a property of *where the token comes from, what its issuance reveals over time, and how a live person is bound to it*, and once those three system-level channels are closed, the constructions of Sections 4 through 7 exhibit a working age check in which no party holds, and no protocol step computes, a record tying a person to the action they took. The impossibility is real for the systems that prompted it and false in general, and the gap between those two statements is the contribution.

1.5 Contributions and relation to the companion papers

This is the third of three papers built on one method: deliver exactly the predicate a system needs, withhold everything else, and make the boundary of what is safe a computable object. The companion election paper [10] delivers eligibility without identity; the companion reporting paper [11] delivers group-level results without individual exposure; this paper delivers an attribute without identity. The three share machinery, not just a slogan: the threshold blind-credential issuance and the provable multiparty shuffle used here are the same primitives the election paper uses for ballots, and the coarseness analysis here imports the reporting paper's bounds directly.

We contribute:

1. **Issuer-distrust by threshold issuance** (Section 4). Tokens are issued by an M-of-N committee; no coalition smaller than M can selectively deny, deanonymize, or covertly tag. We state the guarantee and its boundary, an independence assumption on the N that is the exact analogue of the dual-root independence the election paper [10] flags as its own soft spot.

2. **Vintage-privacy** (Section 5). We identify the epoch-key leakage channel, then close it unconditionally by proactive resharing of a stable predicate key, and retrofittably by asynchronous mixing among current holders. We prove monotonicity of the mixing guarantee and locate its dependence on the anonymity set.

3. **Coarsened predicates with a computable privacy floor** (Section 6). The token proves an arbitrary band, not only a bit, via a committed value and a Bulletproofs range proof [12]; the minimum band width that keeps the token from individuating a holder is read off the population using the reporting paper's identifiability machinery [11].

4. **Device-bound presence** (Section 7). The holder proves, from a key sealed in the device's secure element, a fresh liveness-bound presence at the moment of each confirmation, so the proof is non-replayable and is resistant to the remote-deepfake attack that defeats verifier-side selfie checks. This narrows the credential-versus-person gap and keeps the biometric on the device, never at a bureau.

Delta from the EU blueprint and eIDAS wallet. The European age-verification solution (Section 1.3) already proves "over 18" without disclosing a birthdate, so privacy-preserving *proof* is not our claim. Our construction differs in three load-bearing ways. The wallet trusts a *single* member-state issuer; we distribute issuance across an M-of-N committee so no single authority, including the state, can deny, tag, or deanonymize (Section 4). The wallet binds to a *long-lived* identity credential, a reusable identifier whose every use risks linkage, the precise tracking concern

privacy advocates raise about it; we bind instead to a *per-action, device-sealed presence proof* that carries no persistent identifier and cannot be replayed (Section 7). And the wallet performs its biometric enrolment *once at a government bureau*, minting a credential then used as one key for every lock; we keep biometric presence *on the user's own device at each moment of use*, which is more sovereign and removes the central identity-and-biometric store as a target.

The construction has two genuine boundaries and one legal precondition. We do not restate them at every section; they are collected in Section 9, where each is also placed in context, because each turns out to be either narrow, shared by every alternative including the paper ID it replaces, or a one-line matter of statute rather than of cryptography.

2 Related work and what is new

Every component primitive we use is prior art; the contribution is their composition into a trust model and distribution layer for attribute verification, and the vintage-privacy, coarsening, and presence results that composition requires. We state the delta against the closest lines of work.

Anonymous credentials (CL, BBS+, Idemix, U-Prove). Attribute-based anonymous credentials [15, 17, 21] let a holder prove a predicate over certified attributes without revealing them, and selective disclosure with range proofs over a committed birthdate is exactly how one proves "over 18" privately. We use this machinery rather than improve it. What these systems assume and we do not is a *single trusted issuer*: the credential is signed by one authority that can refuse, tag, or, if consulted at use time, observe. Our issuance is an M-of-N threshold (Section 4), removing that single point, and we address a leakage these systems do not model, the issuance *vintage* encoded by credential timestamps or key epochs (Section 5).

Threshold-issuance credentials (Coconut). Coconut [14] is the closest prior system: it provides threshold-issued, selectively-disclosable, unlinkable credentials, and our issuance layer is Coconut-like. The delta is not threshold issuance, which Coconut already gives, but three things built on top of it and specific to attribute verification. First, vintage-privacy: Coconut does not consider that an attribute credential's issuance epoch leaks a tighter bound than the predicate (Prop. 2), and we close that channel both unconditionally (stable predicate key via proactive resharing, Thm. 2) and retrofittably (asynchronous mixing with the monotonicity guarantee of Thms. 3–4). Second, the computable coarsening floor (Thm. 5) that ties band width to a population non-individuation criterion. Third, device-bound presence (Section 7). Coconut is a credential scheme; we build an attribute-verification system with a stated distribution and legal-discharge story around it.

Anonymous tokens and Privacy Pass (PST, private metadata). Privacy Pass and its browser instantiation Private State Tokens [13], and anonymous tokens with a private metadata bit [19] and keyed-verification credentials [18], give unlinkable redemption at scale and a small amount of issuer-controlled metadata. We adopt PST as the redemption rail. These systems are single-issuer and rotate keys per epoch; that rotation is exactly the vintage channel of Prop. 1(b). Our contribution over this line is to make the issuer a threshold committee and to remove the epoch signal, either by a non-rotating predicate key under proactive resharing or by mixing over the existing rotating-key rail, so a deployment that must build on PST as it exists can still obtain vintage-privacy.

The EU wallet and age-verification blueprint (eIDAS 2.0). The deployed system closest to our goal is the EU Digital Identity Wallet and its age-verification blueprint [20], which already proves "over 18" without disclosing a birthdate and is privacy-preserving in that narrow sense, so we do not claim private proof of age as novel. The delta is threefold and explicit (Section 1.5): the wallet trusts a *single* member-state issuer where we distribute issuance across a committee; it binds a *long-lived* identity credential, a reusable identifier whose repeated use invites linkage, the tracking concern raised about it, where we bind a *per-action* presence proof carrying no persistent identifier; and it performs biometric enrolment once at a *central authority* that holds an identity-and-biometric store, where we keep biometric presence on the holder's own device with no central store to compromise or compel.

Hardware-backed presence (FIDO2/WebAuthn). WebAuthn and FIDO2 [22] provide non-exportable device keys gated by local user presence and verification, and our presence construction (Section 7) uses exactly this class of authenticator. The novelty is not the authenticator but its use to bind an *anonymous attribute proof* to a live present person, yielding a non-replayable, deepfake-resistant presentation (Thm. 6) whose security we reduce to the authenticator's non-exportability and gate parameters, and the observation that this resolves the data-minimization bind the deployed selfie-upload methods create.

Differential privacy and disclosure limitation. The coarsening floor (Thm. 5) draws on the disclosure-limitation analysis of the companion paper [11] rather than on local differential privacy; the distinction, and why a one-account-per-person token gives an inflation resistance that randomized-response and local-DP self-report do not, is treated there. We import its identifiability machinery and do not restate it.

3 Model and primitives

Parties. A set of *issuers* $I_1, \ldots, I_N$ jointly form an issuance committee with threshold M. A *holder* is a person who has been enrolled and possesses one or more tokens. A *verifier* is an online service that must learn the truth of a predicate π over the holder's attribute. An *adversary* may corrupt up to $M-1$ issuers, observe network traffic, and operate verifiers, and seeks either to learn a holder's identity or attribute beyond π, to link a holder's actions across verifiers, or to obtain a token for a holder who does not satisfy π.

Attribute and predicate. Each enrolled holder has an attribute $a \in \mathcal{A}$ (for age, a birthdate, equivalently an integer year). A predicate $\pi : \mathcal{A} \to \{0, 1\}$ is a coarsening: a threshold $\pi(a) = [a \geq \tau]$, or a band $\pi(a) = [a \in [\ell, h]]$. The *coarseness* of a band is $h - \ell$; the threshold is the limiting case of an unbounded band.

Redemption rail. We assume a Privacy-Pass-style rail: a holder redeems a token at a verifier such that the issuer, even in collusion with the verifier up to the corruption bound, cannot link the redemption to the issuance. Private State Tokens [13] is the concrete instantiation we target; the constructions use only the abstract unlinkable-redemption interface, so any rail with that interface suffices.

Shuffle. We assume a provable multiparty shuffle: a set of mix servers permutes and re-randomizes a batch of committed tokens, producing a proof that the output multiset equals the input multiset under some secret permutation, sound unless all mix servers collude. This is the same primitive the election paper [10] uses to anonymize ballots; we reuse its guarantees rather than re-deriving them.

Nullifier. Each token carries a deterministic, unlinkable nullifier so that a token consumed in redemption or in a mix cannot be reused, as in the election paper's double-vote prevention. "Spend" below means: nullifier recorded, token no longer valid.

4 Threshold issuance and issuer-distrust

Construction 1 (Threshold attribute issuance)**.** Enrollment binds a holder to a holder secret hk (a key the holder controls; the binding factor, device or biometric-gated, is a deployment choice discussed in Section 9). To issue a token for predicate π:

1. The holder presents enrollment evidence and a blinded request to the committee.
2. Each issuer I_j verifies the evidence establishes $\pi(a) = 1$ and returns a partial blind signature on the request bound to π and to a commitment to hk.
3. Any M valid partial signatures combine into a token: an unlinkable credential asserting "the holder of hk satisfies π," redeemable on the rail.

A holder may obtain as many tokens as needed; each is independently unlinkable.

Theorem 1 (Issuer-distrust)**.** *Let the issuance scheme of Construction 1 instantiate an (M, N) threshold blind signature that is (a)* existentially unforgeable under chosen message attack *against any coalition of at most $M-1$ issuers, and (b)* blind*, i.e. the issuer's view of an issuance session is independent of the resulting token, and let the redemption rail satisfy* redemption unlinkability*: the joint view of all corrupted issuers and the verifier is independent of which issuance session produced a redeemed token. Then against any probabilistic polynomial-time adversary $\mathcal{A}$ corrupting a set C of issuers with $|C| \leq M - 1$ and operating any number of verifiers:*

1. (Soundness.) *The probability that $\mathcal{A}$ produces a token accepted for predicate π on behalf of a holder with $\pi(a) = 0$ is negligible.*
2. (Unlinkability.) *For any two issuance sessions s_0, s_1 completed by honest holders and any token t redeemed from one of them, $\mathcal{A}$'s advantage in deciding which session produced t is negligible.*
3. (Availability.) *Every holder with $\pi(a) = 1$ obtains a valid token whenever at least M issuers are honest; denial requires at least $N - M + 1$ issuers to refuse.*

Proof. (1) Soundness. A token is by construction a threshold blind signature carrying M partial signatures under the committee key, bound to π and to the commitment to hk. Suppose $\mathcal{A}$, corrupting C with $|C| \leq M - 1$, outputs a token t^* accepted for π on behalf of a holder with $\pi(a) = 0$. Every honest issuer $I_j \notin C$ evaluates the enrollment evidence and, by the construction's issuing rule, emits a partial signature for π only when its local check certifies $\pi(a) = 1$; since $\pi(a) = 0$, no honest issuer emits a partial signature for this (π, hk) pair. Thus at most the $|C| \leq M - 1$ corrupted issuers contribute partial signatures, which is strictly fewer than the M required to combine. For t^* to verify, $\mathcal{A}$ must therefore have produced a valid threshold signature

with fewer than M legitimately obtained shares, i.e. a forgery. We build a reduction $\mathcal{B}$ that runs $\mathcal{A}$, simulating the honest issuers using the threshold-signature unforgeability game's signing oracle for the uncorrupted shares (never querying it on (π, hk) with $\pi(a) = 0$): a successful t^* is a valid signature on a message never signed by the honest shares, hence an EUF-CMA forgery against assumption (a). Soundness follows, with the holder's binding to hk preventing a different holder's valid token from being repurposed, since the signed message includes the hk-commitment.

(2) Unlinkability. Fix sessions s_0, s_1 and a redeemed token t. The adversary's total view comprises the issuance transcripts of the corrupted issuers and the verifier's redemption transcript. By blindness (b), each issuer's issuance view is independent of the produced token, so the corrupted issuers' joint issuance view is independent of the bit b identifying which session yielded t. By redemption unlinkability, the verifier's view together with the corrupted issuers' view is independent of that session identity. A standard hybrid over the two assumptions bounds $\mathcal{A}$'s distinguishing advantage by the sum of the blindness and redemption-unlinkability advantages, each negligible; hence $\mathcal{A}$'s advantage is negligible. When the holder additionally mixes (Construction 3), the unlinkability of the mix composes by the same hybrid argument, and is unconditional in the assumed honest-shuffle model except that it fails if all mix servers collude.

(3) Availability. Issuance requires M valid partial signatures. Each honest issuer presented with evidence establishing $\pi(a) = 1$ emits its partial signature by the construction's rule, so if at least M issuers are honest the holder collects M valid shares and combines a token. Conversely a token is withheld only if fewer than M issuers sign, i.e. at least $N - M + 1$ refuse. This is exactly the stated bound, and it is tight: a coalition of $N - M + 1$ refusing issuers blocks issuance, and no smaller coalition can. □

Remark 1 (Where the trust now sits)*.* Theorem 1 reduces issuer trust from one party to a quorum: the privacy guarantee holds unless M of the N issuers collude. Choosing the N to be genuinely independent, different jurisdictions, an NGO, a university, makes that collusion a broad conspiracy rather than a single point of failure. We return to the strength and the limits of this assumption in Section 9.

5 Vintage-privacy

5.1 The leakage channel

A token proving π should reveal π and not a tighter fact. But issuance occurs in time, and naive constructions leak issuance time through two channels.

Definition 1 (Vintage)**.** The *vintage* of a token is its issuance epoch. A construction *leaks vintage* if a verifier can infer the issuance epoch from a valid presentation.

Proposition 1 (Two leakage channels)**.** *A token may leak vintage (a) explicitly, if the credential binds an issuance timestamp presented to the verifier, or (b) implicitly, if issuers rotate keys per epoch and the token verifies against an epoch-specific public key, so that "verifies under the 2026 key" identifies the epoch even when no timestamp is present.*

Channel (a) is closed by binding the credential to the predicate evaluated at issuance, "$\pi(a) = 1$," and never to the issuance time; a presentation proves the predicate and carries no timestamp. Channel (b) is the substantive one: key rotation, which security and revocation require, reintroduces vintage through the verifying key. Why vintage matters for age is immediate.

Proposition 2 (Vintage is an age lower bound)**.** *If a threshold token for $\pi = [a \geq \tau]$ is known to have vintage year y, then at any later year y' a verifier learns $a \geq \tau + (y' - y)$, strictly stronger than π whenever $y' > y$.*

Proof. At issuance in year y the committee certified $\pi(a) = 1$ at that time, i.e. the holder's age $\text{age}_y \geq \tau$. Age advances with calendar time, so by year $y' \geq y$ the holder's age is $\text{age}_{y'} = \text{age}_y + (y' - y) \geq \tau + (y' - y)$. The predicate π asserts only $\text{age}_{y'} \geq \tau$; the vintage-augmented inference asserts $\text{age}_{y'} \geq \tau + (y' - y)$, which is strictly stronger for $y' > y$. Hence revealing the vintage discloses a tighter lower bound on age than the predicate alone. □

5.2 Unconditional closure: a stable predicate key

Construction 2 (Stable predicate key via proactive resharing)**.** The committee maintains, for each predicate π, a single public verification key that does *not* rotate across epochs. Committee *membership* rotates for security through proactive secret resharing of the same key: at each epoch the current share holders reshare the secret to the next committee, leaving the public key fixed. Every token for π, of any vintage, verifies under the one stable key.

Theorem 2 (Vintage-freedom)**.** *Under Construction 2, the verifying key for predicate π is identical across all epochs, so neither leakage channel of Proposition 1 is present: a valid presentation reveals π and is statistically independent of the issuance epoch. Security against issuer compromise is preserved at every epoch by the proactive resharing, which tolerates a mobile adversary corrupting fewer than M shares per epoch.*

Proof. Write a presentation as the pair $(\mathsf{tok}, \mathsf{vk}_\pi)$ where tok is the token shown and vk_π is the key against which the verifier checks it. We show the joint distribution of a verifier's view is independent of the issuance epoch e.

Channel (a) is absent. By the predicate-only binding of Construction 1, the signed message is $(\pi, \text{commitment to } \mathsf{hk})$ and contains no function of e; hence tok carries no timestamp, and nothing in the presented token is a function of e.

Channel (b) is absent. By Construction 2 the verification key vk_π is a single fixed value, identical for all epochs, so the random variable vk_π is constant in e. The two components of the view are thus each independent of e: tok because its message is epoch-free and the signature is rerandomized at redemption (so its distribution depends only on π and the fixed key), and vk_π because it is constant. Therefore the verifier's view is statistically independent of e, which is the claimed vintage-freedom; formally $I(E; \text{view}) = 0$ for the epoch random variable E.

Security across epochs. Let the per-epoch threshold be M. Proactive resharing (Construction 2) implements a proactive secret-sharing scheme for the fixed secret key: at each epoch boundary the current shareholders run a resharing subprotocol that produces a fresh sharing of the *same* secret and erases old shares. We invoke the standard guarantee of such schemes [16]: against a mobile adversary that corrupts at most $M - 1$ shares within any single epoch (and may move between epochs), (i) the secret is never reconstructible by the adversary, since fewer than M shares reveal nothing about it by the perfect privacy of the underlying (M, N) secret sharing, and (ii) shares learned in one epoch are independent of the sharing in any later epoch, since resharing re-randomizes the share polynomial subject only to fixing its constant term. Consequently the unforgeability assumption of Theorem 1 holds at every epoch under the same $M - 1$ bound, now against a mobile rather than static adversary, and vintage-freedom above holds jointly with it. □

This is the unconditional answer: privacy by construction, no dependence on holder behavior. Its cost is operational, the committee must run proactive resharing of a stable key, which not every

deployed rail supports.

5.3 Retrofittable closure: asynchronous mixing

When the deployment must build on a rail that rotates keys per epoch (the default for Private State Tokens as deployed), vintage cannot be removed at issuance and must be laundered after the fact. We do so by mixing, restricted to holders who already possess a valid token, hence are already eligible.

Construction 3 (Asynchronous closed-pool vintage mix)**.** A *pool* accepts, in a round, tokens from holders who each prove possession of a valid token for π (of any vintage). The mix servers run the provable shuffle on the batch, *spending* each input nullifier and *issuing* a fresh token of the current epoch to each holder, re-randomized and permuted so that output tokens are unlinkable to inputs. A holder may enter the pool at any time, in any epoch, any number of times; each entry consumes the holder's current token and returns exactly one fresh token, so token count is invariant.

The spend-and-reissue discipline is what makes repeated mixing safe: because each mix consumes the old token, a holder cannot multiply tokens by cycling the pool.

Definition 2 (Round anonymity set)**.** For a mix round r, let $\mathcal{V}_r$ be the multiset of vintages of the tokens entering r, and let $H(\mathcal{V}_r)$ be its Shannon entropy.

Theorem 3 (Mixing protection equals round-set entropy)**.** *Consider a round r with input tokens $x_1, \ldots, x_n$ of vintages $v_1, \ldots, v_n$ (the multiset $\mathcal{V}_r$), processed by a shuffle that is* sound *in the sense that, conditioned on the output batch, the permutation matching inputs to outputs is uniform over all $n!$ permutations to any adversary not controlling all mix servers. Fix a target holder h whose input is x_{i_h}, and let V_h denote h's vintage. For an adversary who observes the output batch and knows $\mathcal{V}_r$ but not the permutation:*

1. *the adversary's posterior on V_h given its view equals the empirical vintage distribution $\hat{p}_r(v) = \frac{1}{n}|\{j : v_j = v\}|$, which is also its prior; consequently*
2. *the round leaks* no *information about V_h beyond what $\mathcal{V}_r$ already determines: $I(V_h; \text{view} \mid \mathcal{V}_r) = 0$;*
3. *the adversary's residual uncertainty about V_h is $H(\hat{p}_r)$, which falls short of the maximum achievable uncertainty on the same support, $\log|\mathrm{supp}(\hat{p}_r)|$, by the non-uniformity gap $\log|\mathrm{supp}(\hat{p}_r)| - H(\hat{p}_r) \geq 0$; this gap is 0 exactly when $\hat{p}_r$ is uniform on its support, the case of maximal protection.*

Proof. Let σ be the (secret) permutation the shuffle applied and $y_1, \ldots, y_n$ the outputs, with $y_{\sigma(i)}$ the re-randomized image of x_i. The adversary's view is the output batch together with $\mathcal{V}_r$.

(1) Posterior equals $\hat{p}_r$. By soundness, the conditional law of σ given the view is uniform on S_n. The holder h's output is $y_{\sigma(i_h)}$; to attribute a vintage to h the adversary must invert, i.e. guess the input index $i_h = \sigma^{-1}(h\text{'s output position})$. Since σ is uniform, i_h is uniform on $\{1, \ldots, n\}$ conditioned on the view, so the input the adversary associates with h is a uniformly random one of the n inputs, and its vintage is distributed as $\hat{p}_r$. Before observing the round, h is an exchangeable member of a batch with vintage multiset $\mathcal{V}_r$, so the prior on V_h is likewise $\hat{p}_r$. Posterior and prior coincide.

(2) Zero vintage leakage. Mutual information is the expected log-ratio of posterior to prior; since the two distributions are identically $\hat{p}_r$ for every realization of the view (the argument in (1) holds conditional on any fixed output batch), the ratio is 1 and $I(V_h; \text{view} \mid \mathcal{V}_r) = 0$. The round reveals

which *outputs* exist but, by the uniform-σ premise, nothing that updates the vintage attribution of h beyond the batch composition $\mathcal{V}_r$ itself. (The view does carry $\log n - \sum_v \hat{p}_r(v) \log |\{j : v_j = v\}|$ bits about the *index* i_h, but index information that does not change the vintage distribution is not vintage leakage; this is the distinction the bound makes precise.)

(3) Residual uncertainty and the uniformity gap. The adversary's uncertainty about V_h is the entropy of its posterior, $H(\hat{p}_r)$. Among all distributions supported on $\mathrm{supp}(\hat{p}_r)$, the entropy is maximized by the uniform distribution, with value $\log |\mathrm{supp}(\hat{p}_r)|$ (Gibbs' inequality), so

$$0 \;\le\; \log |\mathrm{supp}(\hat{p}_r)| - H(\hat{p}_r),$$

with equality iff $\hat{p}_r$ is uniform on its support. Thus a round delivers maximal protection (full posterior entropy on its support) exactly when its vintages are balanced, and the gap above quantifies the shortfall when they are not. The soundness of the uniform-σ premise is the shuffle's, holding unless all mix servers collude [10]. □

Theorem 4 (Monotonicity: more mixing never hurts)**.** *Let a holder pass through rounds $r_1, \ldots, r_k$ in order, each round a sound shuffle in the sense of Theorem 3, with input-vintage multisets $\mathcal{V}_{r_1}, \ldots, \mathcal{V}_{r_k}$. Define the holder's* anonymity set *after round r_j as the set A_j of vintages the adversary cannot rule out as the holder's true vintage given its entire view through round r_j. Then $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k$; each additional sound round satisfies $A_j \supseteq A_{j-1}$, so the anonymity set is monotone non-decreasing in the number of mixes and never shrinks.*

Proof. We argue by induction on j that $A_j \supseteq A_{j-1}$, with A_0 the singleton (or small set) consistent with the holder's pre-mix vintage.

Base. After r_1, by Theorem 3 the adversary's posterior on the holder's vintage is supported on $\mathrm{supp}(\hat{p}_{r_1})$, the vintages present in round r_1; thus $A_1 = \mathrm{supp}(\hat{p}_{r_1}) \supseteq A_0$, since the holder's own (pre-mix) vintage is present in r_1 as the holder's input.

Step. Assume A_{j-1} is the adversary's consistent-vintage set for the holder entering round r_j; that is, conditioned on the adversary's view through round r_{j-1}, the holder's current token has vintage in A_{j-1} and every element of A_{j-1} has positive posterior probability. The holder submits this token to r_j. By soundness of r_j, conditioned on its output batch the matching permutation is uniform, so the holder's output token is equiprobably the image of any of r_j's inputs. Let B_j be the set of vintages the adversary assigns positive probability among r_j's *inputs*; the holder's own incoming token contributes a vintage ranging over all of A_{j-1} (with the posterior weights carried from round r_{j-1}), and the co-inputs contribute their vintages. Because the output-to-input attribution is uniform, the holder's post-r_j posterior support is the union of the support contributed by the holder's own token, A_{j-1}, and the support contributed by the co-inputs; hence $A_j = A_{j-1} \cup S_j \supseteq A_{j-1}$, where S_j is the co-input vintage support. The inclusion $A_j \supseteq A_{j-1}$ is the only fact we need and holds regardless of whether r_j is vintage-rich or homogeneous: even a homogeneous round (S_j a single vintage already in A_{j-1}) gives $A_j = A_{j-1}$, never smaller.

The non-removal is what makes the inclusion hold rather than merely the cardinality bound: the spend-and-reissue discipline of Construction 3 leaves the holder with exactly one live token, whose nullifier history is erased (each prior nullifier spent, the new token freshly re-randomized), so the adversary's view after r_j provides no predicate that excludes a vintage previously in A_{j-1}; formally, every $v \in A_{j-1}$ retains positive posterior because the uniform attribution keeps the holder's token confusable with at least one input of incoming vintage v. Therefore $A_{j-1} \subseteq A_j$ for every j, and by induction $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k$, the claimed monotonicity. □

Corollary 1 (A single mix is a floor)**.** *One mix leaves the adversary with posterior entropy $H(\hat{p}_r)$ on the holder's vintage, which is positive whenever the round contains more than one vintage and*

reaches its maximum $\log|\mathrm{supp}(\hat{p}_r)|$ *when the round's vintages are balanced; a vintage-homogeneous round gives no protection. By Theorem 4 repetition across rounds never decreases the anonymity set and drives the posterior toward the full over-*τ *vintage distribution. A holder who misses epochs loses nothing: an old token of any vintage, mixed once whenever the holder appears, acquires the posterior entropy of that round's set.*

Proof. The protection (residual posterior entropy on the vintage) is $H(\hat{p}_r)$ by Theorem 3(3); $H(\hat{p}_r) > 0$ iff $\hat{p}_r$ is non-degenerate, i.e. the round has at least two distinct vintages, and $H(\hat{p}_r) = \log|\mathrm{supp}(\hat{p}_r)|$ iff $\hat{p}_r$ is uniform. The non-decrease under repetition and the miss-years-lose-nothing statement are Theorem 4 applied to the holder's sequence of rounds, which makes no assumption that the rounds are consecutive in time. □

Remark 2 (What mixing relies on). Theorem 3 concerns vintage-as-attribute leakage. The unlinkability it relies on is the shuffle's, so the guarantee inherits the shuffle's anonymity-set model against a global passive adversary; we do not re-derive mixnet anonymity but import it, as the election paper does.

5.4 Which closure to use

Proposition 3 (Selection criterion)**.** *Use the stable-predicate-key construction (Construction 2) when the deployment controls the issuance committee and can run proactive resharing; it gives vintage-freedom unconditionally, independent of holder participation. Use the asynchronous mix (Construction 3) when the deployment must build on an existing rotating-key rail; it retrofits vintage-privacy without changing issuance, at the cost that protection depends on the round anonymity set (Theorem 3) and accrues with use (Theorem 4).*

6 Coarsened predicates and the computable privacy floor

6.1 From a bit to a band

A threshold token answers one verifier's need (a site that must exclude minors). Other verifiers need a coarser fact than a birthdate but finer than a single threshold: a dating service filters by an age *band*, a decade, not a bit and not a birthdate. We support arbitrary bands.

Construction 4 (Banded attribute token)**.** Two variants:

1. *Issuer menu.* The committee issues, per holder, a fixed menu of band tokens (over-18, over-21, $[30, 39]$, ...) chosen by policy; the holder presents the band the verifier requires. No range proof at presentation; the available bands are fixed by the issuer.

2. *Committed value with range proof.* The credential carries a Pedersen commitment to the attribute a; at presentation the holder proves $a \in [\ell, h]$ in zero knowledge via a Bulletproofs range proof [12], for any band the verifier requests, revealing nothing of a beyond membership.

The menu variant is cheaper and fixes disclosure at issuance; the committed-value variant permits any band at presentation at the cost of a range proof. The vintage analysis of Section 5 applies to both, with the committed-value variant requiring that the commitment and its mixing preserve the ability to range-prove while erasing the issuance epoch.

6.2 Coarseness is the privacy knob, and its floor is computable

A band that is too narrow in a thin population is itself an identifier: "born 1947 in this small locale" may single out a person as surely as a name. The companion reporting paper [11] answers exactly the analogous question for demographic cells, how coarse a release must be before an individual is identifiable, and we import its machinery.

Definition 3 (Individuating band)**.** Relative to a population P and auxiliary structure available to the verifier, a band $[\ell, h]$ is *individuating* for a holder if presenting membership in $[\ell, h]$, combined with what the verifier already holds, isolates the holder to within a set smaller than a stated anonymity parameter k.

Theorem 5 (Computable minimum band width)**.** *Let the population P have known attribute distribution, and for an attribute value a and band $B = [\ell, h] \ni a$ let $N_P(B) = |\{p \in P : a_p \in B\}|$ be the count of members whose attribute lies in B. Call B* k-individuating *for a holder if, combined with the auxiliary partition the verifier already holds, the holder is isolated to a set of size $< k$. Then:*

1. *(Monotonicity.) For fixed left endpoint, $N_P([\ell, h])$ is non-decreasing in h, and for fixed a, widening B to $B' \supseteq B$ gives $N_P(B') \geq N_P(B)$.*
2. *(Threshold.) There is a minimum width w^*, computable in closed form from the distribution, such that every band of width $\geq w^*$ containing a has anonymity set of size $\geq k$ within every auxiliary cell, hence is not k-individuating; and the admissibility test is exactly the no-noise safety rule of the reporting paper [11] applied with the band as a released one-way margin.*

Proof. (1) Monotonicity. $N_P(\cdot)$ is a counting measure on the attribute axis: $N_P(B) = \sum_{p \in P} \mathbf{1}[a_p \in B]$. If $B \subseteq B'$ then $\mathbf{1}[a_p \in B] \leq \mathbf{1}[a_p \in B']$ pointwise, so $N_P(B) \leq N_P(B')$; the special case of fixed ℓ and increasing h is immediate. The same holds within any auxiliary cell c by restricting the sum to $p \in c$, giving $N_{P \cap c}(B) \leq N_{P \cap c}(B')$.

(2) Threshold and closed form. Disclosing membership in band B reveals exactly the event $a \in B$ and nothing finer; this is informationally identical to releasing a one-way margin that bins the attribute axis at the cut points ℓ, h. The holder's anonymity set within auxiliary cell c is then precisely $N_{P \cap c}(B)$, the number of population members sharing both the cell and the band. By the reporting paper's identifiability analysis [11], the induced identifiability interval for any individual in cell c has width determined in closed form by the released marginal counts, and the no-noise safety rule (its Cor. for the safe threshold) certifies k-anonymity exactly when every such count is $\geq k$. Define

$$w^* \;=\; \min\Big\{\, w :\; \forall a,\; \forall \text{ auxiliary cells } c,\; \min_{B \ni a,\, |B| = w} N_{P \cap c}(B) \;\geq\; k \,\Big\}.$$

By part (1) the inner quantity $\min_{B \ni a, |B| = w} N_{P \cap c}(B)$ is non-decreasing in w, so the set of admissible widths is an up-set $[w^*, \infty)$ and w^* is well-defined as its least element; it is computable by scanning widths against the known distribution (or directly by inverting the cumulative counts). Any band of width $\geq w^*$ containing a then has anonymity set $\geq k$ in every auxiliary cell by monotonicity, hence is not k-individuating. The admissibility test coincides with the reporting paper's no-noise rule by the margin identification above. □

Remark 3 (The trilogy's shared coarseness). Theorem 5 is the point at which the three papers share not only method but mathematics: the rule that decides how coarsely a demographic result may be

reported is the same rule that decides how coarse an attribute band must be to avoid individuating its holder. Eligibility-without-identity, report-without-exposure, and attribute-without-identity are three predicates over one coarseness calculus.

6.3 Governance of coarseness

Who fixes the available predicates is a policy parameter, not a cryptographic one. In the menu variant the issuer, hence the governing authority, selects the bands; in the committed-value variant the verifier requests a band and the floor of Theorem 5 bounds how fine a request may be honored. How that authority is exercised, by administrative choice or by a participatory mechanism such as the continuous standing-quantity vote of the companion election paper [10], is itself a governance-design question that this paper does not adjudicate; we note only that the coarseness parameter is a legitimate object of such a mechanism and leave its design to that line of work.

7 Device-bound presence and the regulatory bind

The constructions so far prove that a credential satisfies a predicate and that the presenter holds the credential secret. They do not, on their own, bind the proof to a *live person present at the moment of use*, and that binding is what both the strongest attacks and the strongest regulatory framing turn on.

7.1 Construction

Construction 5 (Device-sealed presence proof)**.** The holder secret hk is generated in and never leaves the device's secure element (a Secure-Enclave / TPM / FIDO-authenticator-class component); the device gates each use of hk behind a local liveness check (on-device biometric or equivalent presence test) whose template also never leaves the device. To confirm an action act (a token redemption, a ballot in the companion construction [10], a statistic contribution in the companion construction [11]), the verifier issues a fresh challenge nonce c; the device performs the liveness check and, only on success, signs (act, c, t) under hk, where t is a timestamp. The presentation carries this signature together with the attribute token; the verifier accepts only if the signature is valid, c matches the challenge it issued, and t is fresh.

The biometric and its template are inputs to a *local* gate on a device-resident key; what crosses the wire is a signature over an action and a verifier-chosen nonce, never the biometric, never a reusable identifier.

7.2 What it guarantees

Theorem 6 (Non-replayable, deepfake-resistant presence)**.** *Under Construction 5, with a secure element whose key is non-exportable and a liveness gate with false-accept probability ε:*

1. (No replay.) *A presentation for action* act *under challenge c is not accepted for any other action* $\mathsf{act}' \neq \mathsf{act}$ *or any other challenge $c' \neq c$; a captured presentation cannot be reused, since acceptance binds to the verifier's fresh nonce and the specific action.*

2. (No remote deepfake.) *An adversary who can synthesize the holder's face or voice at a remote verifier cannot produce a valid presentation without also defeating the* on-device *liveness gate and exercising the non-exportable key; a verifier-side facial match, by contrast, is defeated by exactly such synthesis.*

3. (Bounded impersonation.) *The probability that a non-holder in physical possession of the device produces a valid presence proof is at most ε per attempt.*

Proof. We model the secure element as an oracle that, on a successful local liveness gate, returns a signature under a key sk whose corresponding vk is bound to the holder; sk is non-exportable, meaning the adversary may obtain signatures only by passing the gate, never the key itself. The signature scheme is EUF-CMA. The verifier accepts a presentation $(\mathsf{act}, c, t, \sigma)$ iff σ verifies under vk on message (act, c, t), c equals the nonce the verifier just issued, and t is within the freshness window.

(1) No replay. Suppose an adversary, having observed a transcript $(\mathsf{act}, c, t, \sigma)$, gets a presentation accepted for a different action $\mathsf{act}' \neq \mathsf{act}$ or under a different challenge $c' \neq c$. Acceptance requires a valid σ' on (act', c', t'). If the adversary never queried the device oracle on this message (it cannot, without passing the gate, and by hypothesis it is replaying rather than present), then σ' is a signature on a message not previously signed, i.e. an EUF-CMA forgery; we reduce directly, using the observed transcript as the adversary's auxiliary input. The verifier's issuance of a fresh, unpredictable c' for each session guarantees $c' \neq c$ except with the negligible probability of nonce collision, so the replayed $(\mathsf{act}, c, \cdot)$ fails the c'-match check. Hence replay succeeds only with negligible probability.

(2) No remote deepfake. An adversary who synthesizes the holder's biometric at the *verifier* provides input to the verifier's channel, not to the device's local gate. Producing an accepting presentation still requires a signature under sk on the session message, obtainable only through the device oracle, which releases it only on a successful *on-device* gate. Synthesis aimed at the verifier never invokes that oracle, so by part (1)'s unforgeability the adversary cannot produce the signature; its advantage is negligible. The contrast with verifier-side matching is structural: a scheme that accepts a biometric match computed at the verifier has its accept predicate satisfied by the synthesized input directly, with no unforgeable token interposed.

(3) Bounded impersonation. A non-holder in physical possession of the device must pass the local gate to invoke the oracle. By assumption each gate attempt admits a non-holder with probability at most ε (the false-accept rate), independent across attempts; so the probability of at least one success in q attempts is at most $1 - (1 - \varepsilon)^q \leq q\varepsilon$, and per-attempt at most ε. Conditioned on not passing the gate, by (1) no signature is obtainable, so impersonation reduces exactly to defeating the gate, bounding success by ε per attempt. This ε is the irreducible residual recorded in Section 9; it does not vanish, and voluntary transfer of an unlocked device sets it to 1, which no cryptographic property prevents. □

7.3 Sovereignty and the data-protection bind

Theorem 6 also reframes the policy question. The mandates of Section 1.3 push operators toward collecting identity; data-protection law, the GDPR's data-minimization and purpose-limitation principles, penalizes exactly the identity honeypot that collection creates. Construction 5 discloses to the verifier only a predicate proof and a per-action presence signature, no identity, no biometric, no reusable identifier, so a single mechanism satisfies the age mandate and the data-minimization duty at once, the bind dissolved rather than traded off.

The contrast with the bureau-issued wallet model is the substantive one. That model performs biometric enrolment once at a central authority and mints a durable credential; the authority holds a biometric-and-identity store, and the credential is a long-lived identifier whose reuse invites linkage. Construction 5 keeps the biometric on the holder's device, proves presence per action with no persistent identifier, and leaves no central store to compromise or compel. This is what we

mean by *sovereign*: the proof of a present, live, entitled person is produced at the edge, under the holder's control, and is fresh each time.

Remark 4 (A way for the law to thread the needle)*.* The construction offers regulators a path that satisfies the stated aim of the statutes, reliable exclusion of underage users, without the dragnet the same statutes' critics decry. A safe-harbor rule (Section 9) that deems a committee-issued token presented under a device-sealed presence proof to discharge the operator's duty would let a jurisdiction keep its age mandate and its privacy commitments simultaneously, rather than choosing between them.

8 Why the deployed alternative is worse: deepfakes and honeypots

The presence construction of Section 7 is not merely an improvement over a hypothetical baseline. The baseline being deployed under the mandates of Section 1.3, verifier-side document-and-selfie checks routed through third-party providers, is actively failing on two fronts that the device-bound approach removes by construction.

8.1 Visual document checks lose to real-time deepfakes

The dominant deployed method asks a user to photograph a government ID and a selfie, and matches them at the verifier or a contracted provider. Two facts defeat it. First, the visual security of the document is irrelevant over a camera: a REAL-ID-compliant license and an ordinary one look the same to a webcam, and only a cryptographic challenge to an embedded chip, NFC or BLE, distinguishes a genuine credential from a printed or screen image. Second, real-time face synthesis now lets an attacker hold up a victim's document and present a live, matching, animated face to the verifier's liveness check; the same synthesis that makes "hold your ID next to your face" a weak ceremony makes impersonation easier, not harder, as the models improve.

Device-bound presence (Construction 5) is immune to both because the liveness gate is *on the holder's own device*, against a template that never leaves it, releasing a *non-exportable* key; the synthesized face presented to a remote verifier never reaches that gate (Thm. 6, parts 1–2). The attack that defeats verifier-side matching has nothing to act on when the proof is a device-sealed, per-action signature rather than a transmitted face.

8.2 Collection creates honeypots, and the honeypots leak

The mandates have pushed identity collection to a layer that has repeatedly failed to hold it. In October 2025 attackers reached roughly 70,000 government-ID images that Discord's users had supplied for age verification, through a compromised third-party support vendor; the prior year a major age-verification provider was breached as well, and a dating-safety app collecting IDs leaked them the same year. The pattern is structural, not incidental: a mandate to verify age induces a store of identity documents, the store is high-value, and "we delete it immediately" neither survives interception in transit nor audits that platforms rarely permit. Every collection point is a target, and the number of targets grows with every jurisdiction that mandates collection.

A construction in which the verifier receives a predicate proof and a presence signature, and no identity document ever leaves the device, has *no honeypot to leak*. This is the same property that resolves the data-protection bind of Section 7: what is never collected cannot be breached, and cannot be demanded.

8.3 The trajectory: an attribute layer, not only age

The mandates are converging on the device and operating-system layer, where this construction lives. App-store statutes in Texas (SB 2420, in force in 2026 after the Fifth Circuit stayed an injunction), Utah, and Louisiana require the store operator to verify a user's age category and bind minor accounts to a parent; California's Digital Age Assurance Act (AB 1043) requires operating-system providers to collect age at device setup and broadcast an age signal to apps from 2027. Platform leaders have publicly urged that this verification be centralized at the operating-system and app-store layer, and the OS makers, while objecting that the laws force collection of sensitive data "even to download a weather app," are complying by building age signals into the platform.

Two things follow. First, the law is already placing the checkpoint exactly where a device-sealed presence proof would live, so the construction is deployable into the emerging architecture rather than against it. Second, and this is why the paper is no longer only about age: once the platform can attest an age band privately, the same mechanism attests *any* coarsened attribute, residency, eligibility, membership, a protected-class self-report (the companion reporting paper's applications [11]). Observers have named the destination an identity-mediated internet, where access is configured around verified attributes before any interaction; age is merely the first attribute deployed at that layer. That layer can be built as a per-attribute, device-sealed, unlinkable, threshold-issued proof, or as a centralized identity broker. This paper is the case that it can be the former, and the construction is general over the predicate, not specific to age.

8.4 Head-to-head with the deployed methods

A deployer choosing how to meet an age mandate today picks from two fielded options: the dominant one, uploading a government ID and a selfie to a third-party verifier, and the emerging official one, a bureau-issued identity wallet (the eIDAS 2.0 line). Against both, on the properties that decide whether a mandate can be met without creating a surveillance liability, this construction wins on every axis but one, which we name.

Property	**ID+selfie upload**	**Bureau wallet**	**This work**
No identity honeypot created	no	no	**yes**
No central biometric store	no	no	**yes**
Issuer cannot deny / tag / deanonymize	no	no	**yes (M-of-N)**
No reusable linkable identifier	no	no	**yes (per-action)**
Resists remote deepfake	no	partial	**yes**
Discloses only the predicate	no	yes	yes
Meets data-minimization duty	no	partial	**yes**
Integrates with no extra setup	**yes**	no	no

The wins are not incremental. The deployed ID-upload method builds exactly the honeypot the mandates' critics fear, and those honeypots have already leaked at scale (Section 8); it is also defeated by real-time face synthesis, which the on-device liveness gate is not (Theorem 6). The bureau wallet removes the per-site honeypot but keeps a single member-state issuer that can deny or tag, a central biometric enrolment store, and a long-lived identifier whose reuse invites linkage, the three things threshold issuance, on-device presence, and per-action proofs each remove here. On the property the whole debate is about, whether meeting the mandate forces an identity honeypot into existence, this is the only one of the three that answers no.

The single axis where an incumbent leads is the last row: ID-and-selfie upload integrates with no special setup, a photo and an API call, whereas this construction asks a deployment to stand

up an M-of-N committee and rely on a device secure element. That is a real operational cost, and it is the price of removing the honeypot rather than protecting it. We leave the fuller weighing of that tradeoff to the reader; the point of the table is that on privacy and security, the properties the statutes and their critics actually argue over, the construction leads.

9 Boundaries

A construction is judged by what survives once every assumption is made explicit, so we collect here the three things this one depends on. None is hidden elsewhere in the paper, and each is narrower than it first sounds. We state the limit, then state its limit, because a residual that requires an implausible conjunction, or that burdens every alternative equally, does not bear on whether the construction works in the deployments that actually occur.

Threshold independence, and why collusion is hard to arrange. The privacy guarantee of Theorem 1 holds unless M of the N issuers collude. This is a real assumption, and it is also the weakest form a trust assumption can take: it replaces the single trusted issuer that every deployed system has, including the EU wallet, with a quorum, and it fails only when a majority-sized coalition of the issuers acts as one. A deployment chooses its issuers, and the natural choice is bodies with divergent incentives and jurisdictions, a civil-liberties NGO, a university, a foreign registrar, a standards body, for which silent coordinated collusion is the kind of broad conspiracy that is expensive to organize and hard to keep quiet. The assumption degrades gracefully: with one honest issuer among the N above threshold the holder still gets a token, and below-threshold corruption leaks nothing. The failure case is not a subtle cryptographic edge; it is "every independent institution you chose was secretly the same actor," which a deployer controls by not choosing that way.

Credential versus person, a limit every age check shares. Every method here proves that a credential satisfies a predicate and that the presenter holds the credential's secret; a holder who hands a minor their unlocked device, or their key, defeats the predicate without breaking any cryptography. This sounds like a gap until one notices that *it is the same gap in every age-verification method ever fielded, including the physical ID it replaces.* A bouncer who checks a driver's license cannot stop the adult from handing the same license to a friend around the corner; an ID-and-selfie upload is defeated by borrowing a parent's face for thirty seconds. Against this universal residual the construction does strictly better than the alternatives, not worse: the device-sealed presence proof of Section 7 binds each use to a live, on-device liveness check, so casual sharing is replaced by deliberate transfer of an unlocked personal device at each moment of use, and replay and remote deepfakes, which defeat verifier-side selfie checks outright, are removed entirely (Theorem 6). The remaining false-accept rate ε of the local gate is a hardware parameter, the same one guarding the device's banking apps, and voluntary transfer of one's own unlocked device is a behavior no age-verification technology can prevent and none claims to. The honest statement is that this construction narrows the credential-versus-person gap further than any deployed alternative and closes the parts, replay and synthesis, that are actually technological.

Legal discharge, a one-line matter of statute. A privacy-preserving token proves age without proving identity; whether presenting one *discharges* an operator's legal duty under a "knew or should have known" standard is a question for the statute, not the cryptography. This is not a flaw in the construction; it is a gap in law that the construction makes it possible to close cleanly. A safe-harbor provision deeming presentation of a committee-issued token under a device-sealed

presence proof to satisfy the operator's duty is a single clause, and it is one legislators are far more likely to grant for a mechanism that demonstrably protects citizens' privacy than for the identity-dragnet alternative. The construction supplies the private mechanism; the statute supplies one sentence; together they let a jurisdiction keep its age mandate and its privacy commitments at once.

9.1 A bound on the failure region

The three boundaries above are not independent risks that accumulate; they are necessary conditions that must *co-occur* for a guarantee to break, and necessary conditions multiply. We make this precise, because a worst-case residual stated without its probability invites a reader to imagine it larger than it is, which is exactly the move that lets a true worst-case limit masquerade as a typical-case barrier.

Definition 4 (Failure events)**.** For a single verification event, say the *privacy guarantee fails* if some party links the holder's identity to the action, and the *soundness guarantee fails* if a holder with $\pi(a) = 0$ is accepted. Define:

- Col: the infrastructure-independence assumption fails, i.e. at least M of the N issuers collude covertly, or all mix servers of a round collude;
- Xfer: the legitimate holder voluntarily transfers an unlocked, enrolled device (or its secret) to an ineligible person;
- Gate: the on-device liveness gate false-accepts, probability $\leq \varepsilon$ per attempt (Thm. 6).

Theorem 7 (The failure region is a product of independent small factors)**.** *Under the assumptions of Theorems 1–6:*

1. (Privacy.) *The privacy guarantee fails only on* Col*. If each of the N issuers is independently compromised with probability at most q, then*

$$\Pr[\textit{privacy failure}] \;\leq\; \sum_{k=M}^{N} \binom{N}{k} q^k \;\leq\; \binom{N}{M} q^M,$$

and an n_S-server mix contributes at most $q_S^{n_S}$. Both decay geometrically in the threshold, and the deployer sets M, N, n_S, and issuer independence, hence sets this bound.

2. (Soundness.) *The soundness guarantee fails only on* $\mathsf{Xfer} \cup \mathsf{Gate}$*, so* $\Pr[\textit{soundness failure}] \leq \Pr[\mathsf{Xfer}] + \varepsilon$*. The term* Xfer *is a deliberate act of the legitimate holder, present identically in every age-verification method including the physical identity document this replaces, and disjoint from the privacy claim the consensus calls impossible; ε is the hardware false-accept rate already trusted to guard the same device's payment and banking applications.*

Proof. (1) By Theorem 1(2) and Theorems 3–4, an adversary corrupting fewer than M issuers and not all mix servers has a view statistically independent of which holder acted; linkage therefore requires Col. The displayed inequality is the binomial tail probability of $\geq M$ independent compromises, bounded above by its largest term times the binomial coefficient; the mix term is the probability that all n_S independent servers collude. *(2)* By Theorem 1(1), no token issues to a holder with $\pi(a) = 0$ except by a forgery of negligible probability, so an accepted ineligible holder must present a validly issued token, which reaches them only via Xfer, and must clear the presence gate, which a non-holder passes only on Gate (probability $\leq \varepsilon$, Thm. 6). The union bound gives the stated sum. □

Proposition 4 (Rational incentives shrink the privacy-failure region further)**.** Col *is not a random event but an organized act, and a rational adversary undertakes it only if the value V of deanonymizing the targeted population exceeds the expected cost $C = p_{\text{exp}}\, D + K$, where p_{exp} is the probability the covert multi-party collusion is eventually exposed (conspiracies among independent parties leak over time), D is the legal-and-reputational penalty on exposure, and K is the operational cost of co-opting M independent bodies across independent jurisdictions. For any named institution D is the loss of its standing and its exposure to the same data-protection and safe-harbor law this construction invokes; consequently C exceeds V for the overwhelming majority of real deployments, and a deployer raises C without bound by adding issuers in further-separated jurisdictions, which simultaneously lowers the $\binom{N}{M}q^M$ ceiling of Theorem 7.*

To make the multiplicative structure concrete, with the explicit caveat that the following are illustrative figures chosen to be pessimistic rather than measured rates: take an $(M, N) = (3, 5)$ committee of deliberately independent bodies, an NGO, a university, a foreign registrar, a standards body, a regulator, and suppose, generously, that each would covertly collude with probability $q = 0.1$. The privacy-failure ceiling is then $\binom{5}{3}(0.1)^3 = 10 \times 10^{-3} = 10^{-2}$ before the rational-incentive filter of Proposition 4 removes the deployments where $V < C$; raising the committee to $(M, N) = (5, 7)$ drives the ceiling to $\binom{7}{5}(0.1)^5 \approx 2 \times 10^{-4}$. The specific number is whatever a deployer computes from its own issuer roster; the structural fact is that the residual is a product of independent small factors and shrinks geometrically as the deployer chooses to harden it, which is the opposite of an irreducible barrier.

The realistic conclusion. Assemble the pieces. The privacy property the consensus calls impossible holds outright except on Col, a deployer-controlled product of independent small factors that rational incentives shrink further; the only soundness residuals are a behavior every age check shares and a hardware rate already trusted elsewhere; and the legal precondition is a single sentence. There is no regime in which a generic, resourceful adversary defeats the construction at will. There is only a rare, deliberately arranged, high-exposure conjunction that a deployer prices out by adding independent issuers. In the vast majority of scenarios that arise in practice, a website meeting an age mandate, a platform attesting a band, an organization verifying membership or residency or a protected-class self-report, the construction works, privately, with no honeypot to breach or subpoena. And because it requires neither a national election day's apparatus nor a central identity bureau, it puts private, verifiable attribute checking within reach of ordinary websites and small organizations, not only of states with the resources of a census. The impossibility was a worst-case theorem; the worst case is a corner a deployer can see coming and stay out of.

10 Conclusion

The age-verification debate is conducted as a tradeoff between protecting minors and preserving everyone's privacy, as though the two were in necessary tension. They are not in tension at the level of the proof, where the predicate-without-identity primitive is old and deployed. They are in tension only at the two layers the literature assumes away, the trust model of the issuer and the distribution of credentials, and at the legal layer where a predicate bit meets a negligence standard. We have given a construction that removes the single-issuer assumption by threshold issuance, closes the vintage channel that survives unlinkable redemption either unconditionally or retrofittably, generalizes the bit to a coarsened band whose minimum safe width is computable from the same calculus that governs demographic reporting, and binds each proof to a live present person through a device-sealed, per-action presence signature that defeats the replay and remote-deepfake

attacks the deployed document-and-selfie methods fall to, while collecting no identity to breach. Age is the lead instance, not the limit: the construction is general over the predicate, so the same device-sealed, unlinkable, threshold-issued proof attests any coarsened attribute, and the law is already building the checkpoint at the device and operating-system layer where it would live. The consensus that age verification cannot be made private was reasoning from the systems that exist, all of which link identity to action, to a claim about the task, which does not. The three things the construction depends on (Section 9) are an independence assumption the deployer controls, a residual every age check shares and this one handles best, and a one-sentence safe harbor. Outside that narrow corner, which is the worst case and not the common one, a person proves a fact about themselves without revealing who they are, and the honeypot the debate treats as inevitable is simply not built.

References

[1] Electronic Frontier Foundation, *Age Assurance Methods Explained*, Deeplinks, December 2025. https://www.eff.org/deeplinks/2025/12/age-assurance-methods-explained

[2] Electronic Frontier Foundation, *Age Verification is a Privacy Nightmare*, Deeplinks, 2026. https://www.eff.org/deeplinks/2026/05/age-verification-privacy-nightmare

[3] Electronic Frontier Foundation, *Zero Knowledge Proofs Alone Are Not a Digital ID Solution to Protecting User Privacy*, Deeplinks, July 2025. https://www.eff.org/deeplinks/2025/07/zero-knowledge-proofs-alone-are-not-digital-id-solution-protecting-user-privacy

[4] C. Doctorow, *"Privacy preserving age verification" is bullshit*, August 2025. https://doctorow.medium.com/privacy-preserving-age-verification-is-bullshit-0aefd53019e0

[5] NBC News, *Age verification laws prompt privacy and free-speech concerns* (remarks of J. Mackey, Electronic Frontier Foundation), April 2026. https://www.nbcnews.com/tech/tech-news/age-verification-laws-advocates-express-concerns-rcna331835

[6] The Intercept, *Online Age Verification Could Kill Whistleblowing*, June 2026. https://theintercept.com/2026/06/28/age-verification-privacy-surveillance-journalists-whistleblowers/

[7] Center for Democracy and Technology, *Mitigating Risk to Rights with Age Verification*, October 2025. https://cdt.org/insights/mitigating-risk-to-rights-with-age-verification-privacy-preserving-guardrails-that-should-a

[8] Electronic Frontier Foundation, *Fighting Online ID Mandates: 2024 in Review*, Deeplinks, December 2024. https://www.eff.org/deeplinks/2024/12/effs-2024-battle-against-online-age-verification-defending-youth-privacy-and-free

[9] Proton, *The EU's age verification app was hacked in two minutes*, April 2026. https://proton.me/blog/eu-age-verification-app-hacked

[10] Companion paper, *When Remote Voting Beats Paper: A Construction and Quantitative Criterion for Making Elections More Trustworthy and Accessible.*

[11] Companion paper, *When Exact Reporting Beats Adding Noise: A Construction and Quantitative Criterion for Reporting How Groups Voted Without Exposing Anyone.*

[12] B. Bünz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille, G. Maxwell, *Bulletproofs: Short Proofs for Confidential Transactions and More*, IEEE S&P, 2018.

[13] Private State Tokens (formerly Trust Tokens), W3C / Privacy Sandbox; building on *Privacy Pass* (Davidson, Goldberg, Sullivan, Tankersley, Valsorda), PETS 2018.

[14] A. Sonnino, M. Al-Bassam, S. Bano, S. Meiklejohn, G. Danezis, *Coconut: Threshold Issuance Selective Disclosure Credentials with Applications to Distributed Ledgers*, NDSS 2019.

[15] J. Camenisch, A. Lysyanskaya, *Signature Schemes and Anonymous Credentials from Bilinear Maps*, CRYPTO 2004.

[16] A. Herzberg, S. Jarecki, H. Krawczyk, M. Yung, *Proactive Secret Sharing Or: How to Cope With Perpetual Leakage*, CRYPTO 1995.

[17] D. Boneh, X. Boyen, H. Shacham, *Short Group Signatures*, CRYPTO 2004; and *BBS+* as formalized by Au, Susilo, Mu (SCN 2006) and Tessaro–Zhu (EUROCRYPT 2023), the basis of current selective-disclosure credential standardization.

[18] M. Chase, S. Meiklejohn, G. Zaverucha, *Algebraic MACs and Keyed-Verification Anonymous Credentials*, ACM CCS 2014.

[19] B. Kreuter, T. Lepoint, M. Orrù, M. Raykova, *Anonymous Tokens with Private Metadata Bit*, CRYPTO 2020.

[20] European Commission, *European Digital Identity (eIDAS 2.0) and the EU Digital Identity Wallet*; and the *Age Verification Solution Blueprint*, 2025–2026.

[21] J. Camenisch, E. Van Herreweghen, *Design and Implementation of the Idemix Anonymous Credential System*, ACM CCS 2002; and Microsoft U-Prove (Paquin–Zaverucha, 2013).

[22] W3C, *Web Authentication (WebAuthn) Level 2*, and FIDO Alliance, *FIDO2/CTAP*; hardware-backed, non-exportable credential keys with local user-presence and user-verification gates.