

Binding Humans to Keys: A Secure Decentralized Proof of Personhood Protocol

Bilel Zaghdoudi

Sorbonne University
CNRS, LIP6, Paris, France
bilel.zaghdoudi@lip6.fr

Gewu BU

Université Clermont Auvergne, Clermont Auvergne INP
CNRS, LIMOS, Clermont-Ferrand, France
gewu.bu@uca.fr

Frederic Hayek

Antoine University TICKET Lab
Baabda, Lebanon
frederic.hayek@ext.uca.fr

Pascal Lafourcade

Université Clermont Auvergne, Clermont Auvergne INP
CNRS, LIMOS, Clermont-Ferrand, France
ASTEROIDE, Trust4Sign, La Monnerie-le-Montel, France
pascal.lafourcade@uca.fr

Maria Potop-Butucaru

Sorbonne University
CNRS, LIP6, Paris, France
maria.potop-butucaru@lip6.fr

Abstract—Decentralised proof of personhood, cryptographically binding a digital identity to a unique living human without a trusted central authority, is a foundational problem in permissionless networks. We present a novel protocol, Secure-GYID (S-GYID), resilient to Identity Theft Attack (i.e., adversarial authorities substitute a victim’s biometric fingerprint under a malicious public key) and Ghost Key Attack (i.e., colluding authorities fabricate a registration for a non-existent user). Our protocol closes both vulnerabilities by replacing raw biometric transmission with an encryption-based digital fingerprint derivation computed under a one-way, chosen-plaintext-secure, deterministic scheme, so that no authority can substitute or forge a binding pair without knowledge of the user’s secret key. S-GYID further introduces a three-stratum generational authority hierarchy with bounded tenure and VRF-based session selection to limit long-term collusion structurally. We prove Identity Theft Resistance, Biometric Detectability, and Ghost Key Resistance under standard cryptographic assumptions, and provide a hypergeometric session-capture analysis that establishes concrete parameter design rules for both the key activation and authority promotion thresholds.

Index Terms—Distributed Identity, Proof of Personhood, Security Analysis

I. INTRODUCTION

Decentralized identity management has emerged as a foundational challenge in open, permissionless networks, where the absence of a trusted central authority makes identity verification both necessary and difficult. Self-sovereign identity (SSI) standards such as Decentralized Identifiers (DIDs) [1] provide a data model for user-controlled identity, yet they do not address the harder question of *proof of personhood*, cryptographically binding a digital identity to a unique, living human being.

In [2] the authors bring to the attention of the academic community this gap and propose a protocol, GYID (*Get-Your-ID*), to solve it. A user u obtains a certified binding between their biometric fingerprint F_u and their public key pk_u by interacting with a randomly selected set of K authorities forming a *signing party*. If a quorum of authorities confirms

the biometric scan and countersigns the binding pair (F_u, pk_u) , the binding is appended to a public ledger. Despite its elegant design, GYID is vulnerable to two classes of collusion attacks among authorities: *Identity Theft Attack* and *Ghost Key Attack*.

In *Identity Theft Attack*, a coalition of adversarial authorities intercepts the fingerprint F_u of a honest user u during the signing session and substitutes the public key pk_z of a malicious user z for pk_u . Because the original GYID endorsement message carries only the raw binding pair (F_u, pk_u) with no cryptographic commitment from the user, colluding authorities can freely replace pk_u with pk_z and produce a quorum of endorsements for the forged binding pair (F_u, pk_z) , resulting in identity theft recorded on the ledger.

In *Ghost Key Attack* a coalition of adversarial authorities fabricates a signing-party session for a fictitious user g . Because GYID does not require the user’s active participation to be cryptographically proven, colluding authorities can jointly construct a synthetic fingerprint F_g and register it under an adversarial key pk_g without any genuine user being present.

Our fix: *encryption-based fingerprint binding*. We propose **S-GYID**, which closes both attack vectors through two complementary mechanisms. First, the digital fingerprint is computed as the asymmetric encryption of the user’s physical biometric under the user’s own public key, cryptographically binding biometric and key at the point of acquisition. Second, and critically, each endorsement record published on the ledger exposes the binding pair in plaintext: any observer, including the legitimate user, can immediately verify whether their biometric fingerprint has been misused.

Contributions. This paper makes the following contributions:

- (1) **Formal attack analysis.** We formally define and prove the Identity Theft Attack and the Ghost Key Attack against the state of art, GYID [2], as cryptographic games (Appendix B).
- (2) **Encryption-based fingerprint.** We replace raw fingerprint transmission with the asymmetric encryption of

the biometric under the user's public key, and expose the binding pair in plaintext within each endorsement record, enabling immediate fraud detection and privacy-preserving ownership proof without biometric disclosure (Section VI).

- (3) **Generational authority stratification.** We introduce a three-stratum authority hierarchy with 3-cycle tenure and mandatory cooling periods, limiting long-term collusion (Section VI).
- (4) **VRF-based session selection.** Signing parties are assembled using EC-VRF seeded from the previous block header, ensuring per-session unpredictability and public verifiability (Section VI).
- (5) **Formal security proofs.** We prove Identity Theft Resistance (Lemma 1), Biometric Detectability (Lemma 2), and Ghost Key Resistance (Lemma 3) under standard cryptographic assumptions (Section VII).
- (6) **Session-capture probability.** We derive a closed-form hypergeometric bound on the probability that an adversary controlling a fraction f of each stratum can corrupt an entire signing party of size K (Section VIII).

Paper organization. Section II surveys related work. Section III presents the system model. Section IV formalizes the two attacks. Section V defines the adversary model. Section VI specifies S-GYID. Section VII provides the security analysis. Section VIII analyses session-capture probability.

II. RELATED WORK

Decentralised identity has been extensively studied through self-sovereign identity (SSI) and decentralised identifier (DID) frameworks, which enable users to manage verifiable credentials without relying on central registries (e.g. [1], [3]). Blockchain-backed public-key infrastructures such as PB-PKI [4] and Web-of-Trust systems such as Dunitier [5] demonstrate how decentralised credential management and revocation can be achieved in practice. More recent systems, including CanDID [6], strengthen Sybil resistance through commitment-based duplicate detection within the DID model. However, these approaches do not solve the fundamental problem of ensuring that each registered identity corresponds to exactly one unique physical human. Proof-of-personhood schemes address this challenge by introducing physical-world uniqueness assumptions. Early pseudonym-party constructions [7]–[9] established physical co-location as a trust anchor for Sybil-resistant identity, while Borge *et al.* [10] formalised proof-of-personhood in the blockchain setting through attendance-based attestations. Other systems such as Proof of Humanity [11] combine social attestation with human verification. Although effective in some deployment settings, these systems either face scalability limitations or rely on trust assumptions that do not formally bound collusion.

The work most closely related to ours is Get-Your-ID (GYID) [2], which distributes biometric verification across randomly selected signing parties rather than delegating identity issuance to a single trusted authority. By requiring users to obtain multiple biometric endorsements before activation,

GYID provides a decentralised mechanism for biometric proof of personhood. Our work builds directly on this design. We identify two previously unformalised collusion attacks against GYID, the Identity Theft Attack and the Ghost Key Attack, which arise from the lack of an explicit cryptographic binding between biometric data and the claimed public key during endorsement. S-GYID addresses this vulnerability by introducing encryption-based biometric–key binding together with structural anti-collusion authority rotation.

Our proposition also draws on established cryptographic primitives. Biometric cryptosystems such as fuzzy commitments and fuzzy extractors [12]–[14] provide the standard abstraction for handling noisy biometric measurements, enabling S-GYID to define its biometric acquisition function as an abstract interface independent of the underlying biometric modality. The round-based authority rotation adopted in S-GYID is consistent with standard Byzantine fault-tolerant models of partial synchrony and committee reconfiguration [15]–[18], while unpredictable authority selection leverages verifiable random functions (VRF) [19], [20].

S-GYID provides a proof-of-personhood substrate for decentralised applications requiring one-person-one-vote guarantees, including anonymous voting and personhood-based registration systems [21]–[23]. By preventing duplicate identity registration, it mitigates the classical Sybil attack identified by Douceur [24], in which an adversary gains disproportionate influence through low-cost identity multiplication.

III. SYSTEM MODEL

We consider a network composed of a set of mobile nodes. When two nodes u and v are within each other's communication range, we say that there exists a bidirectional communication link between them, also referred to as an edge, $e = \{u, v\}$. Otherwise, no communication link exists between them. Therefore, we model the system as an undirected time-varying graph. We denote the network at time i as $VarG_i = (V_i, E_i)$, where V_i is the set of nodes present at time i , and E_i is the set of edges at time i . Two nodes u and v are said to be neighbors at time i if there exists an edge between them, i.e., u and v are neighbors if $\exists e = \{u, v\} \in E_i$.

The nodes in the system are divided into **user** nodes and a special **ledger** node, ldg .

We define the set of users, at a given time i , as $U_i = V_i \setminus ldg$. Each user $u \in U_i$ is associated with:

- a public–private key pair (pk_u, sk_u) , which is distinct from those of other users.
- a unique physical biometric data P_u , which is distinct from those of other users, and serves as the biometric identity of user u ,
- a biometric data acquisition device D , which is the same for all users, and is the only means to access a user's physical biometric identity.

We model the biometric data acquisition process via the device as a function $Acq : \mathcal{P} \times \mathcal{X} \rightarrow \mathcal{F}$, where, $P \in \mathcal{P}$ denotes a physical biometric data, $aux \in \mathcal{X}$ denotes auxiliary input parameters required during the acquisition process, and

$F = \text{Acq}(P, \text{aux}) \in \mathcal{F}$ denotes the digital representation of the biometric data P .

For a user u , we denote $F_u = \text{Acq}(P_u, \text{aux})$ the digital representation of P_u (or the digital fingerprint of u), acquired through the biometric data acquisition device with function Acq .

We consider the existence of a special ledger node, ldg , which is a neighbor to all users at any time i . Formally, $\forall i, \{\{\text{ldg}, u\} \mid u \in U_i\} \subseteq E_i$.

The ledger is considered a public append-only registry. It receives messages from users and stores them as records. Meanwhile, the ledger allows users to query the records stored on it. The ledger can store different types of records. Records of type x are denoted as r^x . The records in the ledger remain valid unless they are explicitly marked as invalid by the ledger. Let R_i denote the set of all valid records stored in the ledger starting from time i . Ledger satisfies the following properties:

- **Append-only:** Records in the ledger cannot be removed or modified by anyone, but can be marked as invalid by the ledger itself.
- **Consistency:** Any record in the ledger is visible and consistent across all users.
- **Public verifiability:** All records in the ledger are publicly accessible and can be independently verified by any user.

Additionally, both users and the ledger are assumed to have access to a globally synchronized clock.

The edges in the system capture two categories of communication, namely User–User edges and User–Ledger edges.

For each User–User edge, it consists of two types of interaction channels.

1) *Network Communication Channel:*

Communication between users is performed via message passing over a network with bounded communication delay, Δd .

2) *Biometric Acquisition Channel:*

Users do not have direct access to their physical biometric data P ; instead, P can only be obtained via biometric data acquisition performed by a dedicated acquisition device D with acquisition function Acq under close-range physical contact between the acquiring user and the target user. Upon successful acquisition, the acquiring user obtains a digital representation F of the biometric data P , i.e., $F = \text{Acq}(P, \text{aux})$, while the raw biometric data P remains inaccessible. The acquisition time is assumed to be negligible.

For User–Ledger edges, it also consists of two types of interaction channels.

1) *Write Channel:*

Similar to the network communication channel of the User–User edge, when a user writes to the ledger, messages are transmitted via the same network with bounded communication delay, Δd . Upon successful write, the message is stored in the ledger as a record.

2) *Read Channel:*

When a user queries the records in the ledger, the dedicated read channel is used. Upon successful read, the requesting user obtains the corresponding records information. The read operation time is assumed to be instantaneous without loss of generality.

IV. PROOF-OF- IDENTITY HYPOTHESIS, ARCHITECTURE AND PROPERTIES

We aim to design a proof-of-identity protocol, where each user binds the digital representation of his physical biometric information to his public key, and stores this binding on the ledger as a record. By successfully executing this protocol, the user activates their public key as their unique identity in the system, thereby completing user registration.

Let $r_{F_u, pk_u}^{\text{reg}}$ denote a registration record belonging to user u , which binds the binding pair (F_u, pk_u) between the digital representation of physical biometric information F_u with its public key pk_u . A user u is said to be successfully registered since time i if there exists a registration record $r_{F_u, pk_u}^{\text{reg}} \in R_i$, and for any other registration record $r_{F_v, pk_v}^{\text{reg}} \in R_i$, it holds that $F_u \neq F_v \wedge pk_u \neq pk_v$. A user u remains successfully registered until the corresponding registration record is marked as invalid by the ledger.

An *authority* is a special type of user in the system responsible for user registration. A registered user who satisfies certain proof-of-identity conditions may become an authority. Let $Au_i \subseteq U_i$ denote the set of all authorities at time i .

During system initialization, a subset of users is pre-registered, i.e., their registration records are already stored in the ledger at the initialization of the system. These users are referred to as *genesis users*. Moreover, all genesis users are also designated as genesis authorities.

To complete registration, a user is required to participate in a signing party $\mathcal{SP}^{(sid, t, T)}$, with a unique identifier sid , organized by a set of authorities during a specific time interval $[t, t+T]$, where t denotes the starting time of the signing party and T denotes its duration. The set of organizing authorities remains their authority status during the entire interval $[t, t+T]$. We denote by $Au_i^{(sid, t, T)}$ the set of all authorities that organize *signing party* $\mathcal{SP}^{(sid, t, T)}$. It is a subset of the set of all authorities at time i , i.e., $Au_i^{(sid, t, T)} \subseteq Au_i$, $\forall i \in [t, t+T]$.

Since the set of organizing authorities remains unchanged throughout the signing party execution, we have $Au_x^{(sid, t, T)} = Au_y^{(sid, t, T)}$, $\forall x, y \in [t, t+T]$. And thus we denote this time-invariant set as $Au^{(sid, t, T)}$ for the interval $[t, t+T]$.

Moreover, since all organizing authorities must already be authorities at the beginning of the signing party, it holds that $Au^{(sid, t, T)} \subseteq Au_t$.

Let $Un_i^{(sid, t, T)}$ denote the set of all non-organizing users participating in signing party $\mathcal{SP}^{(sid, t, T)}$. It is a subset of all non-organizing users at time i , i.e., $Un_i^{(sid, t, T)} \subseteq U_i \setminus Au^{(sid, t, T)}$, $\forall i \in [t, t+T]$.

We define the full participating user set for the signing party $\mathcal{SP}^{(sid, t, T)}$ at time i as $U_i^{(sid, t, T)} = Au^{(sid, t, T)} \cup Un_i^{(sid, t, T)}$, $\forall i \in [t, t+T]$.

Finally, including the ledger node, ldg , we define the set of all participating nodes for the signing party $\mathcal{SP}^{(sid,t,T)}$ at time i as $V_i^{(sid,t,T)} = U_i^{(sid,t,T)} \cup \{ldg\}$, $\forall i \in [t, t+T]$.

During a signing party, all organizing authorities move to a specific region in the system at the beginning time t and remain in this region throughout the interval $[t, t+T]$. All non-organizing users may move into this region during $[t, t+T]$ to participate in the signing party, or leave the region to stop participating.

The region where the signing party takes place is assumed to be significantly smaller than the communication range of user nodes. Therefore, all users participating in the signing party at time i (including both organizing authorities and non-organizing users) are neighbors, i.e., $\{\{u, v\} \mid u, v \in V_i^{(sid,t,T)}, u \neq v\} \subseteq E_i$, $\forall i \in [t, t+T]$.

Let $E_i^{(sid,t,T)}$ denote the set of all edges among participants of signing party $\mathcal{SP}^{(sid,t,T)}$. All communications occurring during the execution of signing party, including those over the Network Communication Channel of User–User edges and the Write Channel of User–Ledger edges, are assumed to be completed by the end of the signing party at time $t+T$. Formally, let d denote the starting time of a communication. Then, it holds that $d + \Delta d \leq t+T$, where Δd is the bounded communication delay.

A signing party $\mathcal{SP}^{(sid,t,T)}$ can thus be viewed as a subgraph, $VarG_i^{(sid,t,T)} = (V_i^{(sid,t,T)}, E_i^{(sid,t,T)})$, of the global graph $VarG_i = (V_i, E_i)$, for all $i \in [t, t+T]$.

During the execution of a signing party $\mathcal{SP}^{(sid,t,T)}$, users, authorities, and the ledger in $VarG_i^{(sid,t,T)}$ collaboratively execute the proof-of-identity protocol to complete user registration. The registration process consists of two steps.

a) *Step 1: Endorsement.*: At any time i , a non-organizing user $u \in Un_i^{(sid,t,T)}$ can interact with any organizing authority $au_x \in Au^{(sid,t,T)}$ to request an endorsement. The authority au_x collects the biometric information of u with device D , then sends an endorsement message msg^{eds} to the ledger. The endorsement message contains the public key pk_{au_x} of the authority, together with a pair (F_u, pk_u) , where F_u is the digital representation of user u 's biometric information and pk_u is the public key of u .

Upon receiving such an endorsement message, the ledger stores it as an endorsement record r^{eds} . If an endorsement record r^{eds} contains the pair (F_u, pk_u) of user u , then user u is said to be the endorsement target of this endorsement record.

We model the endorsement verification process via a function $Reveal : \mathcal{R}^{eds} \times \mathcal{X} \rightarrow \{true, false\}$, where $r^{eds} \in \mathcal{R}^{eds}$ denotes an endorsement record in the ledger, $arg_u \in \mathcal{X}$ denotes auxiliary input provided by user u , and $Reveal(r^{eds}, arg_u)$ outputs *true* if and only if u is the endorsement target of r^{eds} , and *false* otherwise.

b) *Step 2: Key Activation.*: At any time i , a user $u \in Un_i^{(sid,t,T)}$ who is the endorsement target of *strictly more than* $\alpha_{act}K$ endorsement records from different authorities $au_x \in Au^{(sid,t,T)}$ can initiate key activation, where $\alpha_{act} \in (0, 1]$ is

the activation threshold (baseline: $\alpha_{act} = \frac{1}{2}$). Formally, there exist strictly more than $\alpha_{act}K$ records $r_{au_x}^{eds} \in R_i$, with $au_x \in Au^{(sid,t,T)}$ pairwise distinct, such that $Reveal(r_{au_x}^{eds}, arg_u) = true$.

The user u then sends a key activation message msg^{aky} to the ledger, which includes the public keys of the endorsing authorities and the parameter arg_u .

Upon receiving msg^{aky} , the ledger stores it as a key activation record r^{aky} . The ledger then verifies, using arg_u and the function $Reveal$, that u is indeed the endorsement target of strictly more than $\alpha_{act}K$ specified endorsement records from distinct authorities. If this condition holds, the ledger further checks that there is no conflicting registration record in R_i , i.e., for any $r_{F_v, pk_v}^{reg} \in R_i$, it holds that $F_u \neq F_v \wedge pk_u \neq pk_v$, where F_u and pk_u are the digital representation of the biometric information and the public key of the registering user u , respectively. If no such conflict exists, the ledger adds a new registration record r_{F_u, pk_u}^{reg} to R_i , thereby completing the registration of user u by guaranteeing the uniqueness of each registration record in the ledger.

A proof-of-identity protocol should satisfy the following properties:

- **Liveness.** Eventually, the proof-of-identity protocol allows the recurrent initiation of new signing parties.
- **Availability.** For any signing party $\mathcal{SP}^{(sid,t,T)}$, any unregistered user $u \in Un_i^{(sid,t,T)}$ participating within $[t, t+T]$, following the registration process under the proof-of-identity protocol completes the user registration. More precisely, there exists a registration record of u in the ledger by the end of the signing party at time $t+T$, i.e., $\exists r_{F_u, pk_u}^{reg} \in R_{t+T}$, where F_u is the digital representation of the biometric identity of u , and pk_u is the public key of u .
- **Authenticity.** The binding pair (F_u, pk_u) in a registration record must corresponds to exactly one user in the system. Formally, for any $r_{F_u, pk_u}^{reg} \in R_i$, there exists a unique user $v \in V_i$ possessing F_v and pk_v such that $F_v = F_u \wedge pk_v = pk_u$, where F_u is the digital representation of the biometric identity of u , and pk_u is the public key of u .

V. SECURITY & ADVERSARY MODEL

A. Cryptographic Primitives and Security Assumptions

In our system, we assume the following cryptographic primitives:

a) **Hash Function:** $H(x)$ denotes a cryptographic hash function that outputs the hash value of input x . We assume a cryptographic hash function that is collision-resistant.

b) **Digital Signatures Function:** $Sign(k, x)$ denotes a digital signature algorithm that, given a signing key k and a message x , outputs the signed message (x, σ) , where σ is the signature on x . For a key pair (pk, sk) , messages may be signed using the secret key sk and verified using the corresponding public key pk . We assume the signature scheme is secure against existential forgery under chosen-message attacks (EUF-CMA). That is, no probabilistic polynomial-time (PPT) adversary can produce a valid signature on a new

message for an uncorrupted user with more than negligible probability.

c) **Encryption Function:** $Ep(k, x)$ denotes an encryption function that, given a key k and a message x , outputs the ciphertext of x under key k . We assume the encryption scheme with key pair (pk, sk) is secure against one-wayness under chosen-plaintext attacks (OW-CPA). That is, a PPT adversary cannot recover the plaintext from a given ciphertext with more than negligible probability.

B. Adversary Model

We consider a PPT adversary that controls a subset of non-genesis users, including unregistered users, registered users, and authorities. Users under the control of the adversary are referred to as adversarial users, or adversarial authorities if the users under the control of the adversary are authorities. The adversary has full access to these users, including their key pairs (pk, sk) , physical biometric data P , and access to the biometric acquisition device D . The adversary can use this information to generate and send messages, as well as interact with the ledger and other users.

We consider the adversary capable of initiating Adaptive Chosen Message Attacks (A-CMA). In this scenario, an adversary user can not only choose a set of messages and obtain their corresponding ciphertexts or plaintexts, but also select subsequent messages based on the outcomes of the previously chosen messages.

The adversary can attempt to forge signatures or break encryption; however, signatures and encryption schemes are assumed secure (EUF-CMA and OW-CPA, respectively), so the success probability is negligible. Moreover, we assume that the probability that the adversary, given a fixed fingerprint F_u , generates a key pair (pk_z, sk_z) for a biometric data P_z such that $Ep(pk_z, P_z) = F_u$ is negligible.

The adversary, however, cannot control the following entities: the ledger, ldg ; the genesis users or the genesis authorities.

Additionally, the adversary cannot tamper with the following entities: the physical biometric data P of a user, the biometric data acquisition device D and the corresponding acquisition process function Acq .

Finally, adversary users do not introduce additional delays or refuse to process communications when interacting with other users over the user-user edges and the user-ledger edges. And for adversarial authorities, they respond to the organization of signing parties if required by the proof-of-identity protocol.

VI. FORMAL PROTOCOL SPECIFICATION: S-GYID

In the following we propose the detail presentation of S-GYID, our proof-of-identity protocol.

A. Temporal System State Abstraction

In the S-GYID scheme, we impose a discrete, strictly hierarchical temporal structure $\mathcal{T}$ to coordinate network-wide state transitions, leveraging the globally accessible clock available

to all nodes in the system. This structure is defined by three nested granularities:

- Slot (s): The atomic unit of temporal synchronization. It constrains that a signing party $\mathcal{SP}^{(sid, t, t+T)}$ must be completed within a single slot, i.e., the start time t and the end time $t+T$ must both lie within the same slot.
- Epoch (eps): A contiguous, ordered sequence of n slots constituting an intermediate scheduling granularity.
- Cycle (C): A macro-temporal unit comprising m consecutive epochs. The cycle serves as the canonical interval for evaluating authority tenure eligibility and enforcing generational transitions within the authority set.

1) **Authority Lifecycle and Tenure Governance:** To preclude the entrenchment of permanent power monopolies and to mitigate the risk of adversarial cluster consolidation, the authority role is explicitly transient. Let $Auth : \mathcal{AU} \times \mathcal{C} \rightarrow \{0, 1\}$ be a function that determines whether an authority $au \in \mathcal{AU}$ is active during cycle $C \in \mathcal{C}$. The function outputs 1 if au is active during cycle C , and 0 otherwise. Its lifecycle is governed by the following formally defined predicates:

- Activation (C_{start}): The cycle in which a user u publicly reveals valid endorsements and is admitted to the authority role.
- Active Tenure: An authority au is considered active for exactly three cycles immediately following its activation cycle: $Auth(au, C) = 1$ for $C \in \{C_{start} + 1, C_{start} + 2, C_{start} + 3\}$. The one-cycle delayed activation ensures that a newly admitted authority cannot influence any state transition within the same cycle in which it was activated, thereby preventing last-minute manipulation of the admission process.
- Expiration: Authority status is automatically and irrevocably revoked at the commencement of the fourth cycle following activation: $Auth(au, C_{start} + 4) = 0$. This mandatory expiration mechanism is essential to prevent the entrenchment of permanent authority coalitions and to preserve
- Mandatory Cooling Period: To prevent identity recycling and ensure the non-overlap of consecutive tenure windows for any given unique identity F_U , re-activation requires satisfaction of a mandatory inter-tenure condition: $Auth(au, C - 1) = 0$ must hold prior to any new activation attempt.

B. Generational Stratification and Authority Set Composition

For any time i with a cycle $C > 4$, wherein all generational strata have been fully populated, the active authorities in Au_i is partitioned into three mutually disjoint generational subsets, each defined by its historical activation cycle relative to the current cycle C :

- Parents (S_P): $\{a \in Au_i \mid Auth(a, C - 1) = 1\}$
- Grandparents (S_{GP}): $\{a \in Au_i \mid Auth(a, C - 2) = 1\}$
- Great-Grandparents (S_{GGP}): $\{a \in Au_i \mid Auth(a, C - 3) = 1\}$

1) *Stochastic Session Selection via Verifiable Random Function*: To hold a signing party $\mathcal{SP}^{(sid,t,T)}$, the organizing authorities are invited by selection. The invited authority set $Inv = Au^{(sid,t,T)}$ is constructed by means of a Verifiable Random Function (VRF), which provides cryptographically guaranteed unpredictability and public verifiability of the selection outcome. This mechanism ensures multi-generational oversight and provides formal resistance against targeted adversarial pre-selection.

The selection process enforces strict *cross-generational parity*: each active stratum contributes an equal number of authorities to the invited set.

a) *Parametric Extension to Arbitrary Generational Depth*: Although the baseline instantiation of the protocol employs three generational strata, Parents (S_P), Grandparents (S_{GP}), and Great-Grandparents (S_{GGP}), corresponding to $G = 3$, the number of active strata G is a *configurable protocol parameter*. The protocol generalizes to any $G \geq 2$ by incorporating additional historical cohorts (denoted *Ancestors*), and both the authority selection procedure and the associated security guarantees scale accordingly.

The present analysis fixes the *per-stratum* cardinality N as the primary parameter; the total network size therefore grows as $G \cdot N$. An alternative parameterisation fixes the *total* node count M and distributes them as $N = \lfloor M/G \rfloor$ per stratum; in that model, increasing G reduces N and can diminish the diversity benefit analysed in Section VIII. All numerical evaluations in this paper use the fixed- N model.

The VRF-based selection mechanism relies on the following infrastructure preconditions: (i) *Public-key infrastructure*: each authority holds a long-term key-pair registered on ldg ; VRF proofs are verified against these public keys. (ii) *Agreed VRF input*: the unpredictable seed is derived from the hash of the previous record in the ledger, $H(record_{prev})$, whose availability is guaranteed by the existing Slot/Epoch/Cycle synchronisation structure. (iii) *Cryptographic primitive*: the recommended instantiation is the Elliptic-Curve VRF, which provides the required uniqueness and pseudorandomness properties.

Let $N = |S_i|$ denote the uniform stratum cardinality, assumed equal across all strata, and let G be the number of active generational strata. For each stratum S_i , $i = 1, \dots, G$, the invited subset Inv_i is defined as:

$$Inv_i = \text{VRF}\left(S_i, H(record_{prev}), \left\lfloor \frac{K}{G} \right\rfloor\right)$$

The total invited set is $Inv = \bigcup_{i=1}^G Inv_i$, satisfying the equal-representation constraint:

$$|Inv \cap S_i| = \left\lfloor \frac{K}{G} \right\rfloor \quad \forall i \in \{1, \dots, G\}$$

where K denotes the total number of authorities invited per session and $\lfloor K/G \rfloor$ is the per-stratum invited count. In the canonical three-generation instantiation ($G = 3$), each stratum contributes exactly $\lfloor K/3 \rfloor$ authorities.

When K is not divisible by G , the floor allocation $\lfloor K/G \rfloor$ yields $G \cdot \lfloor K/G \rfloor \leq K$, leaving up to $G - 1$ invitation slots unused. Three remedies are available: (i) accept the minor shortfall; (ii) assign $\lfloor K/G \rfloor$ or $\lceil K/G \rceil$ invitations per stratum in round-robin order to achieve exactly K total; or (iii) require $K \equiv 0 \pmod{G}$ by protocol definition. The security analysis below uses the floor-allocation model throughout; the distinction is negligible for the K values of practical interest.

b) *Notational Convention*: In the sequel, K consistently denotes the *total* invited count across all strata. The per-stratum invited count is $k = \lfloor K/G \rfloor$. The adversarial fraction f is always measured relative to the stratum cardinality N , not relative to the total session size K .

C. Key Activation and Escalation Criteria

1) *Key Activation (User Level)*: The key activation is similar to that in the original GYID scheme [2], where a user needs to obtain endorsements from invited authorities in Inv across all three generational strata. Once a user has obtained a sufficient number of endorsements, it can send a key activation message to the ledger to activate its public key and complete the registration process.

Let $Active : \mathcal{PK} \rightarrow \{0, 1\}$ be a function that determines whether a public key $pk \in \mathcal{PK}$ is active, i.e., whether there exists a registration record $r_{F, pk}^{reg}$ in the ledger that binds the public key pk to a user's digital fingerprint F . The function outputs 1 if pk is active, and 0 otherwise. Let $E_u \subseteq Inv$ denote the set of authorities that endorse user u during a signing party, and let $e_i = |E_u \cap S_i|$ for $i \in \{P, GP, GGP\}$. Key activation is granted if and only if the following conditions are jointly satisfied:

- 1) *Unicity*: $\nexists pk' \in ldg$ such that the binding (F_u, pk') is currently active. This condition enforces resistance to Sybil registration by ensuring each biometric fingerprint maps to at most one active key.
- 2) *Global Quorum*: $|E_u| > \frac{1}{2}|Inv|$. A strict majority of the total invited set must endorse the candidate.
- 3) *Cross-Generational Representation*: For every non-empty stratum S_i , the candidate must satisfy a minimum per-stratum endorsement threshold: $e_i \geq \lceil \alpha \cdot |S_i| \rceil$, where $\alpha \in (0, 1]$ is a protocol-defined representation parameter. This requirement ensures that no single generational cohort can unilaterally block or unilaterally grant activation.

2) *Authority Promotion*: Promotion to the authority registry is subject to a strictly elevated consensus threshold relative to standard key activation, providing formal resistance against Sybil escalation attacks. A user u with $Active(pk_u) = \text{true}$ is promoted to the authority set if and only if:

- 1) *Supermajority Quorum*: $|E_u| > \beta \cdot |Inv|$, where $\beta \in (\frac{1}{2}, 1)$ is a protocol-defined promotion threshold strictly greater than the activation majority $\frac{1}{2}$. The baseline instantiation uses $\beta = \frac{3}{4}$ (a three-quarters supermajority), chosen to align with classical Byzantine fault-tolerance conventions. A three-quarters supermajority of the invited set is required to endorse promotion.

- 2) Cross-Generational Veto Threshold: For every non-empty stratum S_i , the per-stratum endorsement must satisfy $e_i \geq \lceil \beta \cdot |S_i| \rceil$, where $\beta > \alpha$. This stricter per-stratum requirement effectively endows the *Great-Grandparent* cohort, the most temporally senior generation, with a structural veto power over all new promotion events, thereby anchoring the chain of trust in well-established, long-tenured authorities.

D. Bootstrapping Phase

Prior to the full population of all generational strata, the protocol operates in a bootstrapping mode utilizing a foundational set of pre-registered authorities, $Conf_Auth$. Two distinct mechanisms govern this phase:

- Genesis-Based Stochastic Selection: Session authorities are selected via VRF directly from the genesis-defined set: $Inv = \text{VRF}(Conf_Auth, H(record_{prev}), K)$.
- Dual-Session Endorsement Requirement: To obtain authority promotion during bootstrapping, a candidate must submit two independent endorsement sets $\{E_{u,s_n}, E_{u,s_m}\}$, each derived from distinct, non-overlapping slots $s_n \neq s_m$, and each individually satisfying the $\frac{3}{4}$ supermajority quorum. This dual-session requirement ensures that promotion cannot be obtained by compromising a single session during the network's most vulnerable initialization phase, substantially reducing the attack surface in the absence of full generational coverage.

E. Updated Endorsement Step

Once a signing party is organized under S-GYID, users wishing to register can participate and execute the procedure for requiring endorsements from the organizing authorities. We follow the main idea of the Original-GYID [2] presented in Appendix A, with two main improvements:

- 1) Enhanced biometric data acquisition function Acq : When an authority au collects the physical biometric identity from a legitimate user u , au receives a concatenation of the hash of P_u and the encryption of P_u under pk_u as the digital fingerprint F_u , rather than only the hash of P_u . This encryption operation can be implemented by requiring the user to provide a public key during the biometric data collection process. Specifically, the acquisition function Acq is defined as $F = H(P) \parallel Ep(pk, P)$, where the hash of a physical biometric identity P is concatenated with its ciphertext encrypted under the public key pk to form the digital fingerprint F .
- 2) New format of endorsement message is applied to submitted to the ledger: An authority au sends $\text{Sign}(sk_{au}, \text{sid} \parallel \text{ENDORSE} \parallel F_u \parallel pk_u \parallel H(F_u \parallel pk_u \parallel r))$ to the ledger.

That is, au directly includes in the endorsement message the value F_u , pk_u , and additionally appends the hash of F_u , pk_u , and the random value r previously chosen by u . So that once an endorsement record is visible on the

ledger, anyone can know which is the binding pair (F, pk) inside this endorsement record.

VII. SECURITY ANALYSIS ON S-GYID

In S-GYID, authorities are continuously invited to form signing parties through a VRF-based selection mechanism. Users can freely participate in any signing party to interact with a selected set of K authorities for endorsement, key activation, and registration. Upon successful completion of these procedures, the user's registration record is stored on the ledger. Therefore, similar to the Original-GYID scheme, S-GYID satisfies the Liveness, Availability and Authenticity properties when there is no adversary in the system.

In this section, we focus on proving the resistance to Identity Theft Attacks and Ghost Key Attacks defined in Appendix B.

We consider the adversary model described in Section V-B. Specifically, the adversary $\mathcal{Z}$ controls an arbitrary number of users and a fixed fraction f of authorities within each generational stratum, each of which contains N active authorities.

Lemma 1 (Identity Theft Resistance). *The probability that an Identity Theft Attack succeeds by the adversary $\mathcal{Z}$ during a signing party $\mathcal{SP}^{(sid,t,T)}$ is negligible, provided that the number of authorities in $Au^{(sid,t,T)}$ controlled by $\mathcal{Z}$ is at most αK , i.e., the adversary does not reach the key-activation quorum $|E_u| > \alpha K$, where $K = |Au^{(sid,t,T)}|$ is the number of total invited authorities, and $\alpha \in (0, 1]$ is the activation threshold.*

Proof. Consider an Identity Theft Attack in which a subset of adversarial authorities colludes with a malicious user z . The goal of the adversary is to obtain the biometric-derived fingerprint F_u of an honest user u during an interaction with adversarial authorities in $Au^{(sid,t,T)}$, and to bind F_u with the public key pk_z of z , such that the forged binding pair (F_u, pk_z) is accepted and recorded on the ledger.

Assume, for contradiction, that the attack succeeds. Then, at time $t + T$, there exists a registration record $r_{F_u, pk_z}^{reg} \in R_{t+T}$, which implies that the ledger has received strictly more than αK endorsement messages on the binding pair (F_u, pk_z) from distinct authorities in $Au^{(sid,t,T)}$ (i.e., the key-activation quorum has been met).

By the hypothesis of the lemma, at most αK of the invited authorities are adversarial. Consequently, at least one honest authority $au_h \in Au^{(sid,t,T)}$ must be among the endorsers of (F_u, pk_z) .

However, this leads to a contradiction. When the malicious user z interacts with the honest authority au_h , the authority collects the biometric data P_z and computes $F_z = Ep(pk_z, P_z)$ though the secure biometric acquisition device D . Since the adversary cannot forge the biometric data of user u , it holds that $P_z \neq P_u$ as the biometric data of all users are distinct. Therefore, the honest authority au_h will only endorse the binding pair (F_z, pk_z) , rather than (F_u, pk_z) .

Furthermore, as declared in Section V-B, the probability that a PPT adversary $\mathcal{Z}$ can generate a key pair (pk_z, sk_z) such

that $Ep(pk_z, P_z) = F_u$ during the biometric data acquisition is negligible.

Therefore, the adversary user z cannot obtain more than αK endorsements on (F_u, pk_z) whenever at most αK of the invited authorities are adversarial, so the activation quorum is not met and the attack succeeds only with negligible probability. $\square$

Lemma 2 (Detectability of Identity Theft Attacks). *A successful Identity Theft Attack, under the condition that the adversary $\mathcal{Z}$ controls strictly more than αK authorities in $Au^{(sid,t,T)}$, where $K = |Au^{(sid,t,T)}|$ is the number of total invited authorities, and $\alpha \in (0, 1]$ is the activation threshold (i.e., the activation quorum is reached entirely through adversarial endorsements), can be detected. The attack is said to be detected if any genesis user can identify the public keys of any adversarial authorities involved in the attack.*

Proof. According to Lemma 1, the adversary $\mathcal{Z}$ cannot successfully launch an Identity Theft Attack as long as at most αK of the invited authorities in $Au^{(sid,t,T)}$ are adversarial. However, if strictly more than αK authorities in $Au^{(sid,t,T)}$ are controlled by $\mathcal{Z}$, the adversary can obtain the activation quorum and successfully perform the attack, as also observed in the Original-GYID scheme.

The advantage of the S-GYID scheme is that the endorsed digital fingerprint F is recorded in plaintext within the endorsement record on the ledger. Therefore, once a forged endorsement record involving a binding pair (F_u, pk_z) appears on the ledger, any user can inspect the record and verify whether their own fingerprint has been misused.

In particular, when an honest user u observes a registration record binding its digital fingerprint F_u to a public key $pk_z \neq pk_u$, it can initiate a claim of ownership. Specifically, u can prove to a genesis authority that the value F_u can be correctly derived from its own public key pk_u and biometric data P_u though the secure biometric acquisition device D , i.e., $F_u = Ep(pk_u, P_u)$.

Since only the legitimate user u possesses the biometric data P_u , this proof establishes u as the rightful owner of F_u . Consequently, the forged endorsement record can be identified as malicious, and all authorities that endorsed the binding pair (F_u, pk_z) can be exposed by their public keys recorded in the ledger, as they are registered user. Once an attack is detected, additional mechanisms can be employed to assist the victim in revoking the fraudulent endorsement and registration records on the ledger.

Therefore, even when the adversary controls enough authorities in a signing party to meet the activation quorum and succeeds in launching an Identity Theft Attack, the attack can be detected. $\square$

Lemma 3 (Ghost Key Resistance). *The Ghost Key Attack cannot succeed by the adversary $\mathcal{Z}$ during a signing party $\mathcal{SP}^{(sid,t,T)}$, provided that at most αK of the invited authorities in $Au^{(sid,t,T)}$ are controlled by $\mathcal{Z}$, where $K = |Au^{(sid,t,T)}|$ is the number of total invited authorities, and $\alpha \in (0, 1]$ is*

the activation threshold, i.e., the adversary does not reach the key-activation quorum.

Proof. Similar to Lemma 1, assume that the adversary $\mathcal{Z}$ successfully launches a Ghost Key Attack. That is, at the end of the signing party $t + T$, there exists a registration record $r_{F_g, pk_g}^{reg} \in R_{t+T}$.

This implies that strictly more than αK authorities in $Au^{(sid,t,T)}$ have endorsed the binding pair (F_g, pk_g) . By the hypothesis of the lemma, at most αK of the invited authorities are adversarial, hence at least one honest authority $au_h \in Au^{(sid,t,T)}$ must be among the endorsers.

However, this leads to a contradiction. An honest authority cannot perform biometric acquisition for a non-existent user, nor will it endorse a binding pair corresponding to a fake identity. Moreover, since the adversary $\mathcal{Z}$ cannot forge or modify the biometric data of a user, it also cannot coerce or induce other users to use the falsified biometric data P_g to obtain endorsement from au_h .

Therefore, the honest authority au_h would not endorse the binding pair (F_g, pk_g) , which contradicts the requirement that strictly more than αK authorities have endorsed it while at most αK of the invited authorities are adversarial. $\square$

The resistance against Identity Theft attacks (Lemma 1) and Ghost Key attacks (Lemma 3) relies on the difficulty for the adversary $\mathcal{Z}$ to control enough invited authorities in a signing party to reach the key-activation quorum, i.e., strictly more than αK of the K invited authorities, where $\alpha \in (0, 1]$ is the activation threshold. Equivalently, the attacks fail whenever at most αK of the invited authorities are adversarial.

Lemma 2 provides a complementary guarantee for the residual case: even when $\mathcal{Z}$ does succeed in controlling strictly more than αK invited authorities and a fraudulent registration reaches the ledger, the attack is *detectable*. Every endorsing authority's public key is recorded in plaintext on the ledger, so any honest user can identify the corrupt signatories and initiate revocation. This detectability property is essential to the overall security argument: it ensures that a non-negligible $P_{capture}^{(\alpha)}$ does not silently compromise the system but instead produces auditable evidence of the attack.

We denote by $P_{capture}^{(\alpha)}$ the probability that the adversary $\mathcal{Z}$ controls strictly more than αK of the K authorities in a signing party. Consequently, $1 - P_{capture}^{(\alpha)}$ is the probability that the adversary falls short of the activation quorum, in which case both Lemma 1 and Lemma 3 guarantee attack failure.

In the following section, we derive a closed-form hypergeometric bound on $P_{capture}^{(\alpha)}$ and show that, with appropriate parameter settings, it can be made negligible.

VIII. SECURITY ANALYSIS: SESSION CAPTURE PROBABILITY

Lemmas 1 and 3 reduce resistance against Identity Theft and Ghost Key attacks to a single combinatorial event: the adversary controls strictly more than an α -fraction of the invited authorities in each generational stratum, where $\alpha \in (0, 1]$ is the per-stratum activation threshold inherited from

the Key Activation rules of Section VII. We denote this event *session capture* and its probability $P_{\text{capture}}^{(\alpha)}$. The two operational settings are $\alpha_{\text{act}} = 1/2$ (strict majority, used for key activation) and $\alpha_{\text{prom}} = 3/4$ (strict supermajority, used for authority promotion); the limit $\alpha = 1$ (per-stratum unanimity) is retained as a conservative reference point. This section bounds $P_{\text{capture}}^{(\alpha)}$ under the VRF-based stratified invitation mechanism of S-GYID and derives parameter regimes in which the bound is negligible.

A. Formal Model of Session Capture

1) *Notation and system parameters*: The following notation is used throughout this section.

- N : *stratum cardinality*. All strata are assumed to have the same size N . Numerical evaluations fix $N = 100$ unless stated otherwise.
- G : number of generational strata. The baseline instantiation uses $G = 3$; the protocol admits any $G \geq 2$.
- K : total invited authorities per signing party, aggregated across all strata.
- $k = \lfloor K/G \rfloor$: per-stratum invited count.
- f : adversarial fraction per stratum, defined relative to N (not relative to K).
- $m = \lfloor f \cdot N \rfloor$: adversarial node count per stratum.
- $\alpha \in (0, 1]$: per-stratum activation threshold. Capture in a stratum requires strictly more than αk adversarial endorsements. We denote the corresponding integer cut-off $t_\alpha(k) := \min(k, \lfloor \alpha k \rfloor + 1)$, so that the attack needs at least $t_\alpha(k)$ of the k invited authorities to be adversarial. The canonical settings are $\alpha = 1/2$ (activation), $\alpha = 3/4$ (promotion), and $\alpha = 1$ (unanimity; $t_1(k) = k$).

Running example ($G = 3$, $K = 60$, $f = 40\%$, $N = 100$, $\alpha = 1/2$): Each stratum contains $N = 100$ nodes, of which $m = 40$ are adversarial. The session invites $K = 60$ authorities, partitioned as $k = 20$ per stratum. Under the activation threshold $\alpha = 1/2$, session capture requires at least $t_{1/2}(20) = 11$ of the 20 invited slots in every stratum to be filled by the adversary's pool of 40 compromised nodes. Under the promotion threshold $\alpha = 3/4$ the corresponding cut-off rises to $t_{3/4}(20) = 16$; under the unanimous reference $\alpha = 1$ it rises further to $t_1(20) = 20$ (all 20 slots).

2) *Hypergeometric tail model*: The VRF selects k authorities uniformly at random without replacement from the N nodes of each stratum, independently across strata. Let $P_{\text{single}}^{(\alpha)}(k, m, N)$ denote the probability that at least $t_\alpha(k)$ of the k selected nodes in a single stratum are adversarial:

$$P_{\text{single}}^{(\alpha)}(k, m, N) = \sum_{j=t_\alpha(k)}^{\min(k, m)} \frac{\binom{m}{j} \binom{N-m}{k-j}}{\binom{N}{k}}. \quad (1)$$

Equation (1) is the upper tail of the hypergeometric distribution $\text{Hypergeom}(N, m, k)$ beyond the activation cut-off $t_\alpha(k)$. The unanimous limit $\alpha = 1$ collapses the tail to a single term and recovers the closed form $P_{\text{single}}^{(1)}(k, m, N) = \binom{m}{k} / \binom{N}{k}$, with the standard convention $\binom{m}{k} = 0$ whenever $k > m$.

The VRF draws are independent across strata because the G strata $S_1, \dots, S_G$ are disjoint node sets: the k nodes drawn from S_i are selected uniformly at random from the N nodes of S_i alone, and since $S_i \cap S_j = \emptyset$ for $i \neq j$, the outcome in one stratum provides no information about the draw in any other. Under this stochastic independence and assuming a uniform adversarial fraction f across all G strata, the overall session-capture probability is

$$P_{\text{capture}}^{(\alpha)}(K, G, f, N) = [P_{\text{single}}^{(\alpha)}(k, m, N)]^G. \quad (2)$$

Feasibility and the safe regime: The tail in (1) is empty whenever $t_\alpha(k) > m$, i.e. the adversary cannot reach the activation cut-off from only m compromised nodes. Writing $m_\alpha^* := \lceil m/\alpha \rceil$ for $\alpha < 1$ and $m_1^* := m + 1$, this condition collapses to the uniform rule

$$k \geq m_\alpha^* \iff P_{\text{single}}^{(\alpha)}(k, m, N) = 0, \quad (3)$$

with consequently $P_{\text{capture}}^{(\alpha)} = 0$. We refer to this outcome as the *safe regime*; the corresponding capture probability is exactly 0 in all tables below, reflecting provable impossibility of capture rather than missing data. Under the canonical activation threshold $\alpha = 1/2$ the safe regime is $k \geq 2m$, twice as stringent as the unanimous reference $k > m$.

3) *Worst-case adversarial allocation*: Equation (2) assumes a uniform adversarial fraction $f_i = f$ across strata. We show that this is the worst case for the defender among all heterogeneous allocations that satisfy the per-stratum feasibility constraint $m_i \geq m_\alpha^*$, first in the unanimous limit $\alpha = 1$ (closed-form discrete concavity) and then via numerical verification for the operational thresholds $\alpha \in \{1/2, 3/4\}$.

Lemma 4 (Worst-case uniform allocation, unanimous limit). *Let $\alpha = 1$. Fix G , N , k , and a total adversarial budget $M = \sum_{i=1}^G m_i$ with $M \geq Gk$. Among all allocations $(m_1, \dots, m_G)$ with $m_i \in [k, N]$ and $\sum_i m_i = M$, the product $\prod_{i=1}^G \binom{m_i}{k} / \binom{N}{k}$ is maximised at the uniform allocation $m_i = M/G$ (or its floor-ceiling balanced counterpart when $G \nmid M$).*

Proof sketch. The map $m \mapsto \log \binom{m}{k}$ is concave on $\{k, k+1, \dots, N\}$ for fixed k , since its discrete second difference is $\log \binom{m+1}{k} - 2 \log \binom{m}{k} + \log \binom{m-1}{k} = \log(1 - \frac{k}{(m+1)(m-k+1)}) < 0$. By Schur-concavity of the sum of a symmetric concave function on a simplex with identical bounds, $\sum_i \log \binom{m_i}{k}$ is maximised at the balanced allocation. Equivalently, $\prod_i \binom{m_i}{k}$ is maximal there, and dividing by the stratum-independent constant $\binom{N}{k}^G$ preserves the extremum. $\square$

For $\alpha < 1$ the per-stratum factor is the hypergeometric tail sum in (1), for which a closed-form discrete-concavity argument is not available. Exhaustive numerical evaluation over the parameter ranges tabulated in this section ($N = 100$, $G \leq 10$, $\alpha \in \{1/2, 3/4\}$, all feasible (K, m)) confirms that the uniform allocation $m_i = M/G$ remains the maximiser in every case examined; a general proof is left to follow-up work.

The concentration argument is subtle: an adversary that *concentrates* its budget in a single stratum increases the per-stratum capture probability there, but simultaneously drives other strata to infeasibility ($m_j < m_\alpha^*$), sending their factor to 0 and annihilating the product. This is why the per-stratum feasibility constraint $m_i \geq m_\alpha^*$ is essential in Lemma 4: without it, a concentrated allocation is trivially the *best* case for the defender. The uniform allocation is therefore the worst case *within the adversary's feasible strategy set*.

Equation (2) is therefore a conservative (upper-bound) estimate of $P_{\text{capture}}^{(\alpha)}$ under any adversarial budget Gm that admits session capture at all.

B. Safe Regime, Vulnerable Regime, and the Combinatorial Cliff

The monotonicity of $P_{\text{capture}}^{(\alpha)}$ in the parameter G is *not* global: for fixed (K, α) , $P_{\text{capture}}^{(\alpha)}(G)$ switches between two qualitatively distinct regimes.

1) *The two regimes*: Recall the safe-regime threshold $m_\alpha^* = \lceil m/\alpha \rceil$ for $\alpha < 1$ and $m_1^* = m + 1$.

Safe regime ($k \geq m_\alpha^*$). The tail in (1) is empty, so $P_{\text{capture}}^{(\alpha)} = 0$ identically. The adversary is combinatorially incapable of filling the per-stratum activation quorum from only m compromised nodes, regardless of strategy.

Vulnerable regime ($k < m_\alpha^*$). At least one tail term is positive, so $P_{\text{capture}}^{(\alpha)} > 0$. Within this regime, $P_{\text{capture}}^{(\alpha)}$ is minimised at the boundary and increases as k decreases further below m_α^* , due to the combinatorial cliff described below.

2) *The combinatorial cliff*: At the boundary $k = m_\alpha^* - 1$, the number of combinatorially distinct adversarial-majority configurations is at its minimum within the vulnerable regime. Each unit decrease of k below $m_\alpha^* - 1$ enlarges the tail sum sharply. For fixed K , decreasing k corresponds to increasing G (since $k = \lfloor K/G \rfloor$), so in the vulnerable regime *increasing* G further raises $P_{\text{capture}}^{(\alpha)}$. We term this phenomenon the *combinatorial cliff*. In the unanimous limit $\alpha = 1$ it reduces to the classical cliff at $k = m$: the numerator collapses to $\binom{m}{m} = 1$ at the edge, and each unit decrease below raises the numerator via $\binom{m}{m-1} = m$, $\binom{m}{m-2} = \binom{m}{2}$, etc.

3) *Critical generation count*: Let $G_{\text{cliff}}^{(\alpha)}$ denote the smallest $G \geq 2$ that places the protocol in the vulnerable regime, i.e. the smallest $G \geq 2$ for which $\lfloor K/G \rfloor < m_\alpha^*$. Solving this inequality for the smallest integer $G \geq 2$ gives

$$G_{\text{cliff}}^{(\alpha)}(K, m) = \max\left(2, \left\lfloor \frac{K}{m_\alpha^*} \right\rfloor + 1\right). \quad (4)$$

Proposition 1. *Formula (4) is the unique smallest integer $G \geq 2$ such that $\lfloor K/G \rfloor < m_\alpha^*$.*

Proof. The condition $\lfloor K/G \rfloor < m_\alpha^*$ is equivalent to $K/G < m_\alpha^*$, i.e. $G > K/m_\alpha^*$. The smallest integer strictly greater than K/m_α^* is $\lfloor K/m_\alpha^* \rfloor + 1$ (regardless of whether $m_\alpha^* \mid K$, since the inequality is strict and the +1 accounts for the integer-boundary case). Taking the maximum with 2 enforces $G \geq 2$. $\square$

Substituting $m_1^* = m + 1$ into (4) recovers the classical cliff formula $G_{\text{cliff}}^{(1)} = \max(2, \lfloor K/(m + 1) \rfloor + 1)$. At $\alpha = 1/2$, $m_{1/2}^* = 2m$ for positive integer m , so $G_{\text{cliff}}^{(1/2)} = \max(2, \lfloor K/(2m) \rfloor + 1)$: the cliff shifts to roughly *double* the unanimous threshold in K . At $\alpha = 3/4$, $m_{3/4}^* = \lceil 4m/3 \rceil$, giving a $4m/3$ threshold.

An alternative expression $\tilde{G}_{\text{cliff}} = \max(\lfloor K/m \rfloor, 2)$ sometimes encountered in the literature is *not* equivalent to (4) at $\alpha = 1$. At $K = 142$, $m = 70$, this expression yields $\tilde{G}_{\text{cliff}} = 2$, yet $\lfloor 142/2 \rfloor = 71 > 70$, so $G = 2$ lies in the safe regime and $G_{\text{cliff}}^{(1)} = 3$ per (4). The two expressions coincide only when $m \mid K$; in general the alternative is off by one. Equation (4) is used throughout.

The role of $G_{\text{cliff}}^{(\alpha)}$ is summarised as follows.

- **Safe zone** ($G < G_{\text{cliff}}^{(\alpha)}$): $k \geq m_\alpha^*$, $P_{\text{capture}}^{(\alpha)} = 0$.
- **Cliff edge** ($G = G_{\text{cliff}}^{(\alpha)}$): $k < m_\alpha^*$ with $k = m_\alpha^* - 1$ whenever $m_\alpha^* \mid K$; $P_{\text{capture}}^{(\alpha)}$ attains its minimum non-zero value within the vulnerable regime.
- **Vulnerable zone** ($G > G_{\text{cliff}}^{(\alpha)}$): $k \ll m_\alpha^*$, and $P_{\text{capture}}^{(\alpha)}$ rises (often sharply) with each additional generation.

Figure 2 plots $G_{\text{cliff}}^{(1)}$ as a function of K for four adversarial fractions, illustrating this staircase structure under the unanimous reference.

IX. CONCLUSION

This paper identified and formalised two collusion attacks against the state of the art proof-of-identity protocol GYID [2]: the Identity Theft Attack, in which adversarial authorities bind a victim's biometric fingerprint to a malicious public key, and the Ghost Key Attack, in which colluding authorities fabricate a registration for a non-existent user. Both attacks exploit the absence of a cryptographic link between the biometric and the claimed public key in the endorsement message of Original-GYID, and the inability of a victim to detect or counter the fraud without disclosing their raw biometric data.

S-GYID closes both vulnerabilities through two complementary mechanisms: an encryption-based fingerprint that binds biometric and public key at the point of acquisition, and an endorsement format that exposes the binding pair in plaintext on the ledger, enabling immediate fraud detection and privacy-preserving ownership proof. Under standard cryptographic assumptions (OW-CPA encryption, EUF-CMA signatures, and collision-resistant hashing), Lemma 1 proves that Identity Theft Attacks fail whenever the adversary controls at most $\alpha_{\text{act}}K$ of the invited authorities, Lemma 2 guarantees that any attack that does succeed leaves auditable evidence on the ledger, and Lemma 3 proves Ghost Key Resistance under the same quorum condition. The probabilistic analysis bounds the probability that the adversary reaches the quorum: under the operational activation threshold $\alpha_{\text{act}} = 1/2$, a session of $K = 120$ authorities drawn from $G = 3$ generational strata over a network where at most 30% of authorities are adversarial achieves $P_{\text{capture}}^{(1/2)} < 10^{-12}$. The generalised cliff formula $G_{\text{cliff}}^{(\alpha)}$ characterises the boundary between a safe regime, where capture is combinatorially impossible,

and a vulnerable regime, providing operators with concrete parameter design rules.

Several directions remain open. The proof of Lemma 1 relies on the OW-CPA property of a *deterministic* encryption scheme; a full reduction under standard assumptions is left to follow-up work. The system model assumes loosely synchronised clocks, which is in tension with the strict slot/epoch/cycle boundaries used in the protocol; a formal treatment of partial synchrony in the spirit of Dwork, Lynch, and Stockmeyer [18] would strengthen the liveness analysis.

ACKNOWLEDGMENT

This work has been partially supported by the European Atrial Project ERASMUS-EDU-2024:101184483 and the French National Research Agency through the PRIVA-SIQ: Privacy-preserving secure communications despite subversions, interceptions, and quantum adversaries (PRIVASIQ) project (ANR-23-CE39-0008).

REFERENCES

- [1] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen, “Decentralized identifiers (DIDs) v1.0: Core architecture, data model, and representations,” World Wide Web Consortium (W3C), W3C Recommendation, Jul. 2022, available at <https://www.w3.org/TR/did-core/>. [Online]. Available: <https://www.w3.org/TR/did-core/>
- [2] P. Lafourcade and M. Lombard-Platet, “Get-your-ID: Decentralized proof of identity,” in *Proceedings of the 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2019, pp. 1–9.
- [3] M. Gestin, “Privacy preserving and fully-distributed identity management systems. (systèmes de gestion de l’identité totalement distribués et respectant la vie privée),” Ph.D. dissertation, University of Rennes 1, France, 2024. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-04808780>
- [4] L. Axon and M. Goldsmith, “PB-PKI: A privacy-aware blockchain-based PKI,” in *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (SECURITY)*. SCITEPRESS, 2017, pp. 311–318.
- [5] Duniter Consortium, “Duniter: A Web of Trust cryptocurrency,” Online, 2016, available at <https://duniter.org/en/>. [Online]. Available: <https://duniter.org/en/>
- [6] D. Maram, H. Malvai, F. Zhang, N. Jean-Louis, A. Frolov, T. Kell, and A. Miller, “CanDID: Can-do decentralized identity with legacy compatibility, Sybil-resistance, and accountability,” in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 1348–1366.
- [7] B. Ford and J. Strauss, “An offline foundation for online accountable pseudonyms,” in *Proceedings of the 1st Workshop on Social Network Systems (SocialNets)*. ACM, 2008, pp. 31–36.
- [8] B. Ford, “Pseudonym parties: An offline foundation for online accountability,” École Polytechnique Fédérale de Lausanne (EPFL), Tech. Rep., 2020, technical report; preliminary draft first circulated March 2012. Available at <https://bford.info/pub/net/pseudonym/>.
- [9] —, “Identity and personhood in digital democracy: Evaluating inclusion, equality, security, and privacy in pseudonym parties and other proofs of personhood,” arXiv preprint arXiv:2011.02412, 2020. [Online]. Available: <https://arxiv.org/abs/2011.02412>
- [10] M. Borge, E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, and B. Ford, “Proof-of-personhood: Redemocratizing permissionless cryptocurrencies,” in *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2017, pp. 23–26.
- [11] Kleros and Democracy Earth, “Proof of humanity: A protocol to verify humans on chain,” Online whitepaper, 2021, available at <https://www.proofofhumanity.id/>. [Online]. Available: <https://www.proofofhumanity.id/>
- [12] A. Juels and M. Wattenberg, “A fuzzy commitment scheme,” in *Proceedings of the 6th ACM Conference on Computer and Communications Security (CCS)*. ACM, 1999, pp. 28–36.
- [13] Y. Dodis, L. Reyzin, and A. Smith, “Fuzzy extractors: How to generate strong keys from biometrics and other noisy data,” in *Advances in Cryptology – EUROCRYPT 2004*, ser. Lecture Notes in Computer Science, vol. 3027. Berlin, Heidelberg: Springer, 2004, pp. 523–540.
- [14] K. Delac and M. Grgic, “A survey of biometric recognition methods,” in *Proceedings of the 46th International Symposium on Electronics in Marine (ELMAR)*. IEEE, 2004, pp. 184–193.
- [15] M. Castro and B. Liskov, “Practical byzantine fault tolerance,” in *Proceedings of the 3rd USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. USENIX Association, 1999, pp. 173–186.
- [16] P. McCorry, S. F. Shahandashti, and F. Hao, “A smart contract for boardroom voting with maximum voter privacy,” in *Proceedings of the 21st International Conference on Financial Cryptography and Data Security (FC)*, ser. Lecture Notes in Computer Science, vol. 10322. Berlin, Heidelberg: Springer, 2017, pp. 357–375.
- [17] M. Yin, D. Malkhi, M. K. Reiter, G. Golan Gueta, and I. Abraham, “Hot-Stuff: BFT consensus with linearity and responsiveness,” in *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC)*. ACM, 2019, pp. 347–356.
- [18] C. Dwork, N. Lynch, and L. Stockmeyer, “Consensus in the presence of partial synchrony,” *Journal of the ACM*, vol. 35, no. 2, pp. 288–323, 1988.
- [19] S. Micali, M. O. Rabin, and S. P. Vadhan, “Verifiable random functions,” in *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 1999, pp. 120–130.
- [20] S. Goldberg, L. Reyzin, D. Papadopoulos, and J. Vcelák, “Verifiable random functions (VRFs),” Internet Engineering Task Force (IETF), RFC 9381, Aug. 2023. [Online]. Available: <https://www.rfc-editor.org/rfc/rfc9381>
- [21] A. Ide and T. Sharma, “Personhood credentials: Human-centered design recommendation balancing security,” arXiv preprint arXiv:2502.16375, 2025. [Online]. Available: <https://arxiv.org/abs/2502.16375>
- [22] L. H. Merino, S. Colombo, R. Reyes, A. Azhir, S. Mishra, P. Tennage, and B. Ford, “TRIP: Coercion-resistant registration for E-Voting with verifiability and usability in Votegral,” in *Proceedings of the ACM SIGOPS 31st Symposium on Operating Systems Principles (SOSP ’25)*. ACM, 2025, pp. 834–874.
- [23] N. Swearingen, X. Zou, and N. Li, “Fully transparent, privacy-preserving yet verifiable, attack-resistant, and practical remote electronic voting rendering assured and fair elections,” *IEEE Transactions on Privacy*, 2025.
- [24] J. R. Douceur, “The Sybil attack,” in *Revised Papers from the First International Workshop on Peer-to-Peer Systems (IPTPS)*, ser. Lecture Notes in Computer Science, vol. 2429. Berlin, Heidelberg: Springer, 2002, pp. 251–260.

A. Original-GYID scheme

The Original-GYID presented in [2] proposes a Proof-of-Identity protocol that aims to guarantee the three properties described in Section IV: Liveness, Availability, and Authenticity. In the following, we provide a high-level overview of the GYID scheme.

The Original-GYID allows authorities to spontaneously organize signing parties by sending an announcement message to the ledger as a announcement record, that contains the identifier of the planned signing party, its scheduled time, geographic region, and the public keys of all authorities participating in the organization of this signing party. When the signing party takes place, any user wishing to register may physically attend and participate within the specified time and region. Users can voluntarily organize and participate in signing parties, such that the Original-GYID guarantees the Liveness property.

During the signing party $\mathcal{SP}^{(sid,t,T)}$, for the endorsement step, a user $u \in \mathcal{U}_i^{(sid,t,T)}$ wishing to register can require multiple endorsements from different authorities organizing the signing party. The procedure for requiring an endorsement with an authority $au \in \mathcal{AU}^{(sid,t,T)}$ is as follows:

- 1) u generates a random number r and sends $\text{Sign}(sk_u, pk_u \parallel r)$ to au .
- 2) au receives $\text{Sign}(sk_u, pk_u \parallel r)$ and verifies the validity of the signature. If the signature is invalid, au aborts the procedure. Otherwise, au collects the digital fingerprint F_u of u through the acquisition function Acq , defined as a simple hash function, $F_u = H(P_u)$.
- 3) au sends F_u to u .
- 4) au sends the endorsement message to the ledger with the following format: $\text{Sign}(sk_{au}, sid \parallel \text{ENDORSE} \parallel H(F_u \parallel pk_u \parallel r))$. Ledger then store the whole message as the endorsement record.

The Original-GYID defines the endorsement verification function *Reveal* as $\text{Reveal}(r^{eds}, (F, pk, r))$. It returns *True* if $H(F \parallel pk \parallel r) = H(F'_u \parallel pk'_u \parallel r')$, where the latter value is contained in r^{eds} ; otherwise, it returns *False*. That is, if a user u can provide (F_u, pk_u, r) such that $H(F_u \parallel pk_u \parallel r)$ matches the value $H(F'_u \parallel pk'_u \parallel r')$ contained in the endorsement record r^{eds} , then u is proven to be the endorsement target of r^{eds} . This is because, due to the properties of the hash function, the probability that anyone who does not simultaneously know F_u, pk_u , and r can construct a value $X = F'_u \parallel pk'_u \parallel r'$ such that $H(X) = H(F_u \parallel pk_u \parallel r)$ is negligible.

For the key activation step, once a user u has obtained a sufficient number of endorsement messages recorded on the ledger from different authorities within the signing party, may reveal F_u, pk_u, r , and the corresponding endorsement records in the ledger by sending a key activation message to the ledger for the endorsement verification function. Anyone at that moment can verify whether u is the endorsement target of at least $K = \beta \cdot |\mathcal{AU}^{(sid,t,T)}|$ endorsement records from distinct authorities, where $\beta \in (0, 1]$. This confirms that u is

the legitimate owner of pk_u and F_u . Ledger then check if there exists a conflicting registration record with u , if yes, ledger aborts the procedure; Otherwise, ledger adds a registration record of u , meaning the public key pk_u is bound to its physical biometric data P_u , and pk_u becomes activated in the system. The registration of u therefore complete.

B. Original-GYID vulnerability

We identify two attack scenarios targeting the proof-of-identity proposed by Original-GYID, in which an adversary, as defined in Section V, can successfully violate the Availability and Authenticity properties.

- **Identity Theft Attack:** In this attack, adversarial authorities may collude with an adversarial user z to obtain the digital representation F_u of biometric data P_u of a victim user u , and let z register its public key pk_z with the F_u .
 - 1) The adversary prepares this attack by initiating a signing party organized by at least $K = \beta \cdot |\mathcal{AU}^{(sid,t,T)}|$ adversarial authorities in $\mathcal{SP}^{(sid,t,T)}$, where $\beta \in (0, 1]$ given by Original-GYID.
 - 2) Adversary authorities can obtain the F_u of u at the phase (2) of endorsement step of Original-GYID, once u requires an adversary authority for the endorsement.
 - 3) The adversary authorities then subsequently allow z to complete the endorsement step using F_u , by directly constructing endorsement messages containing a random value r_z chosen by z , the public key of z , pk_z and F_u , instead of using z 's own digital fingerprint F_z .
 - 4) As z colludes with multiple adversarial authorities, z may therefore become the endorsement target of at least K endorsement records in the ledger faster than u . Once z has obtained K corresponding endorsement records sooner than u , z can send a key activation message including F_u, pk_z and r_z to the ledger to let the ledger verify the K endorsement records and then complete the registration with F_u and pk_z .
 - 5) When u finally obtains K endorsements, upon sending a key activation message to the ledger, its request will be rejected, because z has already successfully registered the digital fingerprint of u . Therefore, there already exists a registration record r_{F_u, pk_z}^{reg} in the ledger that conflicts with the binding pair (F_u, pk_u) . As a result, u cannot complete the registration.

Since F_u contained in the endorsement messages for z is hashed together with pk_z and r_z according to the phase (4) of endorsement step of the Original-GYID, user u does not notice that its digital fingerprint F_u has been stolen until z 's registration is successfully completed, at which point the F_u becomes publicly visible in the registration record for z .

This violates the Availability property, as a user u that follows the protocol honestly cannot complete the registration. It also violates the Authenticity property, as the Original-GYID protocol allows a successful registration r_{F_u, pk_z}^{reg} , while there exists no user $u' \in \mathcal{U}_i$ who simultaneously possesses both F_u and pk_z , since F_u belongs to

user u and pk_z belongs to user z , and each user's digital biometric data $F = H(P)$ and key pair (pk, sk) are both distinct.

- **Ghost Key Attack:** In this attack, adversary authorities collude with a sufficient number of other adversary authorities to fabricate ghost registered users. The ghost user g does not correspond to any user in the system, i.e., $g \notin V_i, \forall i$, but possesses a key pair (pk_g, sk_g) and a fictitious digital fingerprint F_g , which can be directly accessed by any colluding adversary authority. Both the key pair and the digital fingerprint of a ghost user g are distinct from those of other users.

- 1) The adversary prepares this attack by initiating a signing party organized by at least $K = \beta \cdot |Au^{(sid,t,T)}|$ adversarial authorities in $\mathcal{SP}^{(sid,t,T)}$, where $\beta \in (0, 1]$ given by Original-GYID.
- 2) The adversary authorities then collude to create and send to the ledger multiple endorsement messages binding F_g to pk_g without performing the actual biometric data acquisition process.
- 3) Once at least K corresponding endorsement records targeting g are on the ledger, the adversary authorities claim these endorsements on behalf of g , and require the activation of pk_g . As a result, the public key pk_g of the ghost user g becomes active and is bound to the fictitious digital fingerprint F_g . The registration of this ghost user thus completes.

Once a ghost user is successfully registered, the system is unable to detect or expose them. This violates the Authenticity property, since there exists a registration record r_{F_g, pk_g}^{reg} in the ledger, while no user $v \in V_i$ possessing the pair (P_v, pk_v) such that $H(P_v) = F_g \wedge pk_v = pk_g$, as F_g and pk_g are distinct from those of other users in the system.

C. Numerical Tables and Figures for Session Capture Probability

We report $P_{\text{capture}}^{(\alpha)}$ for $G \in \{3, 10\}$, $N = 100$, $f \in \{10\%, \dots, 80\%\}$, and the three canonical thresholds $\alpha \in \{1/2, 3/4, 1\}$. Exact rational evaluation uses Python's `Fractions.Fraction` with `math.comb`; entries equal to 0 satisfy the safe regime condition (3) and denote $P_{\text{capture}}^{(\alpha)} = 0$ exactly. Intermediate values $G \in \{4, \dots, 9\}$ exhibit the same trend and are omitted.

We present the three thresholds in order of operational relevance. Tables I and II ($\alpha = 1/2$) correspond to the *key activation* rule and are the primary security target of the protocol. Tables III and IV ($\alpha = 3/4$) correspond to the *authority promotion* rule, and quantify the additional margin bought by the elevated promotion threshold. Tables V and VI ($\alpha = 1$, per-stratum unanimity) are retained as an optimistic reference: they collapse to the classical bound $P_{\text{capture}}^{(1)} = \left[\binom{m}{k} / \binom{N}{k}\right]^G$ and illustrate the best attainable security under the strongest possible threshold.

The tables exhibit three complementary security mechanisms; Figure 1 plots the $\alpha = 1$ curves across $G \in \{3, 10\}$ for visual reference.

Threshold amplification. For fixed (K, G, f) , lowering the activation threshold α dramatically raises $P_{\text{capture}}^{(\alpha)}$. At $G = 3$, $K = 60$, $f = 30\%$ the probability rises from $P_{\text{capture}}^{(1)} = 1.76 \times 10^{-40}$ (unanimity reference) to $P_{\text{capture}}^{(3/4)} = 1.79 \times 10^{-20}$ (promotion) and $P_{\text{capture}}^{(1/2)} = 6.05 \times 10^{-7}$ (activation, a gap of roughly 33 orders of magnitude between activation and unanimity. This gap is the quantitative cost of replacing unanimous quorum with strict majority quorum and is the principal security trade-off made by the protocol.

Session size amplification. For fixed G and α , increasing K decreases $P_{\text{capture}}^{(\alpha)}$ super-exponentially as long as the system remains in the vulnerable regime $k < m_\alpha^*$. At $\alpha = 1$, $G = 3$, $f = 80\%$ the probability falls from 1.60×10^{-2} at $K = 20$ to 1.75×10^{-35} at $K = 200$, a reduction of more than 33 orders of magnitude.

Generational amplification. For fixed (K, f, α) , increasing G within the safe regime keeps $P_{\text{capture}}^{(\alpha)} = 0$, but once the cliff is crossed further generations *raise* the probability. Within the vulnerable regime the sign of the effect depends on both α and the specific (K, m) ; see §VIII-B.

A moderately sized session ($K = 60$, $G = 3$) renders session capture under the activation threshold negligible only for low-density adversaries ($f \leq 30\%$, $P_{\text{capture}}^{(1/2)} < 10^{-6}$). The regime $f \geq 50\%$ at the activation threshold is no longer cryptographically negligible: $P_{\text{capture}}^{(1/2)} = 0.065$ at $K = 60$, $G = 3$, $f = 50\%$. Operators targeting resilience against majority-dense adversaries must therefore compensate with either larger K , larger G (so long as the cliff is not crossed), or the elevated promotion threshold $\alpha_{\text{prom}} = 3/4$.

1) Tolerance-Based Design: In practice, the actual adversarial fraction f is unknown. The standard engineering approach is to fix a tolerance f_{tol} below which the protocol is required to remain secure, set $m_{\text{tol}} = \lfloor f_{\text{tol}} N \rfloor$ and $m_{\text{tol}, \alpha}^* = \lceil m_{\text{tol}} / \alpha \rceil$ (or $m_{\text{tol}} + 1$ for $\alpha = 1$), and design G relative to $m_{\text{tol}, \alpha}^*$:

$$G_{\text{cliff}}^{(\alpha)}(K, f_{\text{tol}}) = \max\left(2, \left\lfloor \frac{K}{m_{\text{tol}, \alpha}^*} \right\rfloor + 1\right). \quad (5)$$

Any $G < G_{\text{cliff}}^{(\alpha)}(K, f_{\text{tol}})$ guarantees $P_{\text{capture}}^{(\alpha)} = 0$ whenever the actual adversarial fraction $f \leq f_{\text{tol}}$; concretely, $G = 2$ suffices whenever $K \geq 2 m_{\text{tol}, \alpha}^*$.

For the remainder of this subsection we report tolerance-design figures under the *unanimous reference* $\alpha = 1$, which corresponds to the case analysed in prior work and provides a conservative optimistic bound. Operational design under $\alpha_{\text{act}} = 1/2$ follows the same pattern with $m_{\text{tol}, 1/2}^* = 2m_{\text{tol}}$ in place of $m_{\text{tol}} + 1$: the cliff locations shift right by roughly a factor of two in K , and the absolute values of $\log_{10} P_{\text{capture}}^{(1/2)}$ at the cliff edge are smaller in magnitude by several tens of orders compared to the unanimous reference.

We consider three canonical tolerance regimes.

Classical BFT ($f_{\text{tol}} = 1/3$, $m_{\text{tol}} = 33$): The first transition to $G_{\text{cliff}} = 3$ occurs at $K = 68$; subsequent cliff transitions advance in steps of 34 invited authorities.

Majority adversary ($f_{\text{tol}} = 1/2$, $m_{\text{tol}} = 50$): The first transition to $G_{\text{cliff}} = 3$ occurs at $K = 102$; steps advance in intervals of 51.

Supermajority adversary ($f_{\text{tol}} = 0.70$, $m_{\text{tol}} = 70$): The first transition to $G_{\text{cliff}} = 3$ occurs at $K = 142$; steps advance in intervals of 71. This regime is strictly stronger than either of the above. Figure 3 plots $G_{\text{cliff}}^{(1)}(K, f_{\text{tol}})$ across all three tolerance regimes; Figure 4 details the staircase for the supermajority case.

Table VII summarises $G_{\text{cliff}}^{(1)}$ and $\log_{10} P_{\text{capture}}^{(1)}$ (evaluated at the cliff edge $G = G_{\text{cliff}}^{(1)}$) under each tolerance, for several actual adversarial fractions f and for the *unanimous reference* $\alpha = 1$. Entries equal to 0 satisfy $k = \lfloor K/G_{\text{cliff}}^{(1)} \rfloor > m_{\text{act}}$, i.e. the *actual* adversary is weaker than the tolerance bound and capture is combinatorially impossible.

Key observations.: Guaranteed safety below the threshold. Any $G < G_{\text{cliff}}^{(\alpha)}$ satisfies $\lfloor K/G \rfloor \geq m_{\text{tol}, \alpha}^*$ and places the protocol in the safe regime regardless of $G_{\text{cliff}}^{(\alpha)}$.

Security headroom under conservative design. When the actual adversarial fraction f is strictly below f_{tol} , a protocol calibrated to the tolerance over-provisions against a more demanding adversary than is actually present. The per-stratum invited count k then exceeds the actual adversarial count m_{act} , driving the corresponding entries to 0 (safe by construction). This headroom scales with the gap $f_{\text{tol}} - f$.

Tolerance cost. For fixed N and α , tolerating a larger f_{tol} forces K to grow proportionally to obtain the same $G_{\text{cliff}}^{(\alpha)}$. At $\alpha = 1$, $G_{\text{cliff}}^{(1)} = 3$ is first attained at $K = 68$ for $f_{\text{tol}} = 1/3$, $K = 102$ for $f_{\text{tol}} = 1/2$, and $K = 142$ for $f_{\text{tol}} = 0.70$, roughly doubling the session size to move from the classical BFT bound to the supermajority bound. At $\alpha = 1/2$ the same targets require roughly twice as many invited authorities again, because $m_{\text{tol}, 1/2}^* = 2m_{\text{tol}}$.

Practical recommendation. Operators without a precise estimate of f should default to the classical BFT tolerance $f_{\text{tol}} = 1/3$ and select G in the safe zone, $G < G_{\text{cliff}}^{(\alpha)}(K, 1/3) = \max(2, \lfloor K/m_{\text{tol}, \alpha}^* \rfloor + 1)$, using the activation threshold α actually enforced by the protocol (typically $\alpha = 1/2$). Environments facing stronger adversaries can switch to $f_{\text{tol}} = 1/2$ or $f_{\text{tol}} = 0.70$, at a proportional cost in K .

2) *Scaling Across Network Sizes:* We finally examine how session-capture security scales with the network size N , under the *unanimous reference* $\alpha = 1$. For each tolerance f_{tol} , we adopt the canonical sizing rule $K = 3 \cdot m_{\text{tol}}$, which by (5) yields $G_{\text{cliff}}^{(1)} = 3$ across all values of N . Equivalently: $K \approx N$ for $f_{\text{tol}} = 1/3$, $K \approx 1.5N$ for $f_{\text{tol}} = 1/2$, and $K \approx 2.1N$ for $f_{\text{tol}} = 0.70$. Each stratum then receives $k = m_{\text{tol}}$ invitations. Operational deployment at $\alpha = 1/2$ requires doubling K to keep $G_{\text{cliff}}^{(1/2)} = 3$; the scaling exponent with N remains qualitatively identical.

Table IX reports $\log_{10} P_{\text{capture}}^{(1)}$ at the resulting cliff-edge configuration across four network scales.

Observations.: Universal sizing rule. The parameterisation $K = 3m_{\text{tol}}$ uniformly yields $G_{\text{cliff}}^{(1)} = 3$ across all scales under the unanimous reference, confirming three generational strata as a practical sweet spot between security and operational overhead.

Super-linear scaling with N . Comparing Small-70% and XLarge-70% at $f = 70\%$: $\log_{10} P_{\text{capture}}^{(1)}$ goes from -76.4 to -3973.7 , a factor of roughly $52\times$ in the exponent for a $50\times$ increase in N . In absolute terms $P_{\text{capture}}^{(1)}$ improves by approximately 3900 orders of magnitude with no increase in algorithmic complexity: each stratum's hypergeometric denominator $\binom{N}{k}$ grows combinatorially, and the exponent G compounds that growth.

Tolerance buys absolute immunity against weaker adversaries. A zero entry indicates that the actual adversary cannot fill a single per-stratum quota, rendering capture combinatorially impossible. Larger f_{tol} yields a wider band of absolute immunity, at the cost of a proportionally larger session size.

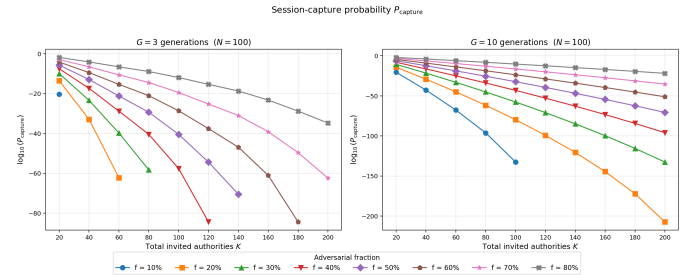


Fig. 1. Session-capture probability $P_{\text{capture}}^{(1)}$ (per-stratum unanimity reference) as a function of total invited authorities K , for $G = 3$ and $G = 10$, $N = 100$, and adversarial fractions $f \in \{10\%, \dots, 80\%\}$. Curves terminate where the system enters the safe regime ($k > m$, $P_{\text{capture}}^{(1)} = 0$). The operational thresholds $\alpha \in \{1/2, 3/4\}$ yield uniformly higher curves; see Tables I–IV.

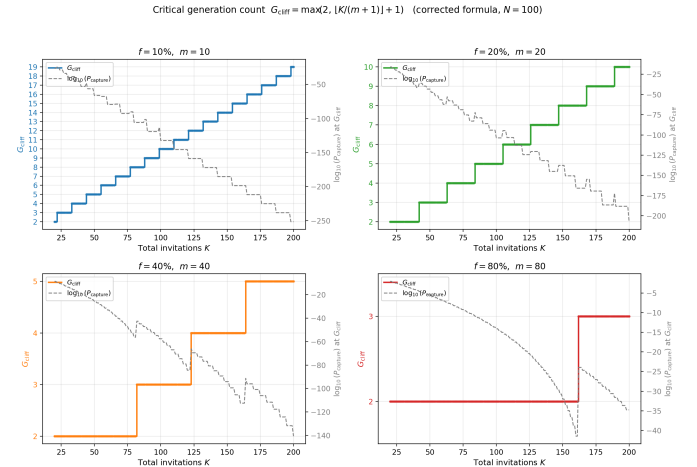


Fig. 2. Critical generation count $G_{\text{cliff}}^{(1)}$ (unanimous reference) as a function of total invited authorities K , for four adversarial fractions $f \in \{10\%, \dots, 80\%\}$ (stratum cardinality $N = 100$). The cliff location for the operational thresholds $\alpha \in \{1/2, 3/4\}$ shifts right by the factor $m_{\alpha}^*/(m+1) \approx 1/\alpha$; e.g. for $\alpha = 1/2$, the first transition to $G_{\text{cliff}}^{(1/2)} = 3$ at $f = 30\%$ is at $K = 120$ rather than at $K = 60$ for $\alpha = 1$.

TABLE I
 $P_{\text{CAPTURE}}^{(1/2)}$ FOR $G = 3$ (STRICT MAJORITY, ACTIVATION THRESHOLD). $N = 100$; $k = \lfloor K/3 \rfloor$; $m = \lfloor fN \rfloor$; $t_{1/2}(k)$ IS SHOWN PER COLUMN. 0: SAFE REGIME ($k \geq 2m$, $P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($t=4$)	$K=40$ ($t=7$)	$K=60$ ($t=11$)	$K=80$ ($t=14$)	$K=100$ ($t=17$)	$K=120$ ($t=21$)	$K=140$ ($t=24$)	$K=160$ ($t=27$)	$K=180$ ($t=31$)	$K=200$ ($t=34$)
10%	10	3.81×10^{-10}	1.26×10^{-15}	0	0	0	0	0	0	0	0
20%	20	2.70×10^{-6}	5.24×10^{-8}	5.11×10^{-13}	4.60×10^{-17}	1.28×10^{-21}	0	0	0	0	0
30%	30	2.67×10^{-4}	1.22×10^{-4}	6.05×10^{-7}	2.16×10^{-8}	2.00×10^{-9}	4.61×10^{-13}	4.65×10^{-16}	1.87×10^{-19}	0	0
40%	40	5.09×10^{-3}	9.77×10^{-3}	1.05×10^{-3}	4.26×10^{-4}	4.47×10^{-4}	2.85×10^{-5}	6.05×10^{-6}	3.16×10^{-6}	2.70×10^{-8}	5.84×10^{-10}
50%	50	3.89×10^{-2}	0.125	6.48×10^{-2}	6.89×10^{-2}	0.125	7.37×10^{-2}	7.44×10^{-2}	0.125	7.37×10^{-2}	7.23×10^{-2}
60%	60	0.161	0.486	0.473	0.584	0.788	0.798	0.867	0.957	0.969	0.989
70%	70	0.420	0.858	0.911	0.968	0.996	0.999	≈ 1	≈ 1	≈ 1	≈ 1
80%	80	0.747	0.989	0.998	≈ 1	≈ 1	≈ 1	≈ 1	≈ 1	≈ 1	≈ 1

TABLE II
 $P_{\text{CAPTURE}}^{(1/2)}$ FOR $G = 10$ (STRICT MAJORITY, ACTIVATION THRESHOLD). $N = 100$; $k = \lfloor K/10 \rfloor$; $m = \lfloor fN \rfloor$. 0: SAFE REGIME ($P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($t=2$)	$K=40$ ($t=3$)	$K=60$ ($t=4$)	$K=80$ ($t=5$)	$K=100$ ($t=6$)	$K=120$ ($t=7$)	$K=140$ ($t=8$)	$K=160$ ($t=9$)	$K=180$ ($t=10$)	$K=200$ ($t=11$)
10%	10	3.86×10^{-21}	3.05×10^{-26}	4.00×10^{-32}	1.38×10^{-38}	1.07×10^{-45}	1.27×10^{-53}	1.23×10^{-62}	3.02×10^{-73}	1.07×10^{-86}	0
20%	20	6.94×10^{-15}	7.77×10^{-17}	2.72×10^{-19}	5.92×10^{-22}	8.86×10^{-25}	9.14×10^{-28}	6.31×10^{-31}	2.77×10^{-34}	7.30×10^{-38}	1.07×10^{-41}
30%	30	2.75×10^{-11}	1.00×10^{-11}	1.22×10^{-12}	1.06×10^{-13}	7.66×10^{-15}	4.80×10^{-16}	2.63×10^{-17}	1.26×10^{-18}	5.26×10^{-20}	1.88×10^{-21}
40%	40	9.44×10^{-9}	2.62×10^{-8}	2.27×10^{-8}	1.40×10^{-8}	7.43×10^{-9}	3.60×10^{-9}	1.64×10^{-9}	7.15×10^{-10}	2.98×10^{-10}	1.19×10^{-10}
50%	50	8.62×10^{-7}	7.85×10^{-6}	2.00×10^{-5}	3.41×10^{-5}	4.85×10^{-5}	6.25×10^{-5}	7.56×10^{-5}	8.77×10^{-5}	9.89×10^{-5}	1.09×10^{-4}
60%	60	3.42×10^{-5}	5.66×10^{-4}	2.28×10^{-3}	5.71×10^{-3}	1.13×10^{-2}	1.93×10^{-2}	3.02×10^{-2}	4.42×10^{-2}	6.15×10^{-2}	8.23×10^{-2}
70%	70	7.64×10^{-4}	1.40×10^{-2}	5.57×10^{-2}	0.129	0.225	0.335	0.448	0.554	0.650	0.732
80%	80	1.12×10^{-2}	0.142	0.379	0.607	0.773	0.877	0.937	0.969	0.985	0.994

TABLE III
 $P_{\text{CAPTURE}}^{(3/4)}$ FOR $G = 3$ (STRICT SUPERMAJORITY, PROMOTION THRESHOLD). $N = 100$; $k = \lfloor K/3 \rfloor$; $m = \lfloor fN \rfloor$; $t_{3/4}(k)$ IS SHOWN PER COLUMN. 0: SAFE REGIME ($P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($t=5$)	$K=40$ ($t=10$)	$K=60$ ($t=16$)	$K=80$ ($t=20$)	$K=100$ ($t=25$)	$K=120$ ($t=31$)	$K=140$ ($t=35$)	$K=160$ ($t=40$)	$K=180$ ($t=46$)	$K=200$ ($t=50$)
10%	10	7.08×10^{-15}	4.51×10^{-33}	0	0	0	0	0	0	0	0
20%	20	1.24×10^{-9}	1.08×10^{-17}	3.03×10^{-33}	7.93×10^{-47}	0	0	0	0	0	0
30%	30	6.97×10^{-7}	1.58×10^{-11}	1.79×10^{-20}	2.04×10^{-25}	1.02×10^{-34}	0	0	0	0	0
40%	40	4.80×10^{-5}	1.04×10^{-7}	2.53×10^{-13}	1.32×10^{-15}	5.57×10^{-20}	2.95×10^{-29}	3.17×10^{-35}	2.29×10^{-49}	0	0
50%	50	1.07×10^{-3}	4.53×10^{-5}	1.60×10^{-8}	2.23×10^{-9}	2.15×10^{-11}	2.53×10^{-16}	2.14×10^{-18}	1.04×10^{-22}	4.27×10^{-33}	6.10×10^{-43}
60%	60	1.14×10^{-2}	3.44×10^{-3}	3.98×10^{-5}	3.51×10^{-5}	7.29×10^{-6}	2.70×10^{-8}	9.99×10^{-9}	4.44×10^{-10}	4.08×10^{-14}	8.03×10^{-16}
70%	70	7.13×10^{-2}	6.91×10^{-2}	9.15×10^{-3}	1.81×10^{-2}	1.76×10^{-2}	2.32×10^{-3}	3.87×10^{-3}	3.18×10^{-3}	2.17×10^{-4}	2.79×10^{-4}
80%	80	0.282	0.435	0.256	0.464	0.601	0.473	0.671	0.799	0.730	0.895

3) Worked Example and Design Recommendations:

A worked example.: For $K = 60$, $f = 30\%$, $m = 30$, $N = 100$, and $\alpha = 1/2$: $m_{1/2}^* = 60$, hence

$$G_{\text{cliff}}^{(1/2)} = \max(2, \lfloor \frac{60}{60} \rfloor + 1) = 2.$$

This means even $G = 2$ already loses the safe regime: the protocol is already vulnerable at two generations. Table VIII verifies the non-monotonic behaviour across $G = 2, \dots, 7$: the probability oscillates as both k and the tail cut-off $t_{1/2}(k)$ change discretely. Under the unanimous reference $\alpha = 1$ ($m_1^* = 31$), the cliff for the same configuration is $G_{\text{cliff}}^{(1)} = \max(2, \lfloor 60/31 \rfloor + 1) = 2$, likewise at $G = 2$, but $P_{\text{capture}}^{(1)}$ is many orders of magnitude smaller owing to the narrower tail.

Design recommendations.: Given K , an estimate or tolerance bound on m , and an activation threshold α (typically

$\alpha_{\text{act}} = 1/2$), the operator should prefer configurations *strictly below* the cliff:

- 1) **Provable safety** ($P_{\text{capture}}^{(\alpha)} = 0$). Choose any $G < G_{\text{cliff}}^{(\alpha)}$, equivalently any G satisfying $\lfloor K/G \rfloor \geq m_{\alpha}^*$. The minimal such choice is $G = 2$, valid whenever $K \geq 2m_{\alpha}^*$ (i.e. $K \geq 4m$ at $\alpha = 1/2$).
- 2) **Cliff edge** ($G = G_{\text{cliff}}^{(\alpha)}$). Use only when operational constraints impose $G \geq G_{\text{cliff}}^{(\alpha)}$. Yields the minimum non-zero capture probability among vulnerable-regime configurations.
- 3) **Never set** $G > G_{\text{cliff}}^{(\alpha)}$ **without verification**. In the vulnerable regime the trend with G is non-monotonic (cf. Table VIII), so operators must consult the per-configuration table before committing to a setting beyond the cliff.

We emphasise that $G_{\text{cliff}}^{(\alpha)}$ is a *threshold*, not an optimum.

TABLE IV
 $P_{\text{CAPTURE}}^{(3/4)}$ FOR $G = 10$ (STRICT SUPERMAJORITY, PROMOTION THRESHOLD). $N = 100$; $k = \lfloor K/10 \rfloor$; $m = \lfloor fN \rfloor$. 0: SAFE REGIME ($P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($t=2$)	$K=40$ ($t=4$)	$K=60$ ($t=5$)	$K=80$ ($t=7$)	$K=100$ ($t=8$)	$K=120$ ($t=10$)	$K=140$ ($t=11$)	$K=160$ ($t=13$)	$K=180$ ($t=14$)	$K=200$ ($t=16$)
10%	10	3.86×10^{-21}	1.94×10^{-43}	6.82×10^{-48}	4.52×10^{-73}	1.57×10^{-80}	6.49×10^{-115}	0	0	0	0
20%	20	6.94×10^{-15}	8.29×10^{-30}	2.02×10^{-30}	2.07×10^{-45}	5.79×10^{-47}	3.54×10^{-63}	1.18×10^{-65}	6.83×10^{-84}	1.26×10^{-87}	4.03×10^{-109}
30%	30	2.75×10^{-11}	2.78×10^{-22}	3.00×10^{-21}	1.04×10^{-31}	2.67×10^{-31}	4.18×10^{-42}	4.06×10^{-42}	1.29×10^{-53}	4.59×10^{-54}	1.49×10^{-66}
40%	40	9.44×10^{-9}	4.73×10^{-17}	4.02×10^{-15}	1.20×10^{-22}	2.98×10^{-21}	8.85×10^{-29}	1.19×10^{-27}	1.96×10^{-35}	1.62×10^{-34}	1.02×10^{-42}
50%	50	8.62×10^{-7}	4.88×10^{-13}	1.24×10^{-10}	5.38×10^{-16}	4.08×10^{-14}	2.64×10^{-19}	1.22×10^{-17}	7.24×10^{-23}	2.47×10^{-21}	1.03×10^{-26}
60%	60	3.42×10^{-5}	8.85×10^{-10}	3.36×10^{-7}	7.20×10^{-11}	7.41×10^{-9}	2.98×10^{-12}	1.86×10^{-10}	9.15×10^{-14}	4.40×10^{-12}	2.16×10^{-15}
70%	70	7.64×10^{-4}	4.89×10^{-7}	1.50×10^{-4}	7.48×10^{-7}	5.19×10^{-5}	5.41×10^{-7}	2.07×10^{-5}	3.12×10^{-7}	8.71×10^{-6}	1.60×10^{-7}
80%	80	1.12×10^{-2}	1.14×10^{-4}	1.46×10^{-2}	9.23×10^{-4}	2.15×10^{-2}	2.76×10^{-3}	3.08×10^{-2}	5.90×10^{-3}	4.29×10^{-2}	1.07×10^{-2}

TABLE V
 $P_{\text{CAPTURE}}^{(1)}$ FOR $G = 3$ (PER-STRATUM UNANIMITY, OPTIMISTIC REFERENCE). $N = 100$; $k = \lfloor K/3 \rfloor$; $m = \lfloor fN \rfloor$. 0: SAFE REGIME ($k > m$, $P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($k=6$)	$K=40$ ($k=13$)	$K=60$ ($k=20$)	$K=80$ ($k=26$)	$K=100$ ($k=33$)	$K=120$ ($k=40$)	$K=140$ ($k=46$)	$K=160$ ($k=53$)	$K=180$ ($k=60$)	$K=200$ ($k=66$)
10%	10	5.47×10^{-21}	0	0	0	0	0	0	0	0	0
20%	20	3.44×10^{-14}	1.30×10^{-33}	6.49×10^{-63}	0	0	0	0	0	0	0
30%	30	1.24×10^{-10}	4.78×10^{-24}	1.76×10^{-40}	6.01×10^{-59}	0	0	0	0	0	0
40%	40	3.34×10^{-8}	4.85×10^{-18}	1.70×10^{-29}	3.65×10^{-41}	2.53×10^{-58}	3.85×10^{-85}	0	0	0	0
50%	50	2.37×10^{-6}	1.24×10^{-13}	6.80×10^{-22}	5.25×10^{-30}	3.73×10^{-41}	4.17×10^{-55}	3.08×10^{-71}	0	0	0
60%	60	7.41×10^{-5}	3.84×10^{-10}	4.78×10^{-16}	9.97×10^{-22}	2.66×10^{-29}	2.84×10^{-38}	1.32×10^{-47}	9.58×10^{-62}	3.85×10^{-85}	0
70%	70	1.33×10^{-3}	2.97×10^{-7}	2.76×10^{-11}	4.07×10^{-15}	3.93×10^{-20}	6.53×10^{-26}	1.09×10^{-31}	8.13×10^{-40}	2.40×10^{-50}	3.94×10^{-63}
80%	80	0.0160	8.71×10^{-5}	2.87×10^{-7}	1.33×10^{-9}	1.26×10^{-12}	4.78×10^{-16}	2.16×10^{-19}	6.04×10^{-24}	1.70×10^{-29}	1.75×10^{-35}

TABLE VI
 $P_{\text{CAPTURE}}^{(1)}$ FOR $G = 10$ (PER-STRATUM UNANIMITY, OPTIMISTIC REFERENCE). $N = 100$; $k = \lfloor K/10 \rfloor$; $m = \lfloor fN \rfloor$. 0: SAFE REGIME ($P_{\text{CAPTURE}} = 0$ EXACTLY).

f	m	$K=20$ ($k=2$)	$K=40$ ($k=4$)	$K=60$ ($k=6$)	$K=80$ ($k=8$)	$K=100$ ($k=10$)	$K=120$ ($k=12$)	$K=140$ ($k=14$)	$K=160$ ($k=16$)	$K=180$ ($k=18$)	$K=200$ ($k=20$)
10%	10	3.86×10^{-21}	1.94×10^{-43}	2.88×10^{-68}	6.84×10^{-97}	4.14×10^{-133}	0	0	0	0	0
20%	20	6.94×10^{-15}	8.29×10^{-30}	1.32×10^{-45}	2.02×10^{-62}	1.92×10^{-80}	6.15×10^{-100}	2.70×10^{-121}	3.66×10^{-145}	8.34×10^{-173}	5.11×10^{-208}
30%	30	2.75×10^{-11}	2.78×10^{-22}	9.40×10^{-34}	9.47×10^{-46}	2.48×10^{-58}	1.43×10^{-71}	1.49×10^{-85}	2.17×10^{-100}	3.19×10^{-116}	3.06×10^{-133}
40%	40	9.44×10^{-9}	4.73×10^{-17}	1.20×10^{-25}	1.45×10^{-34}	7.93×10^{-44}	1.81×10^{-53}	1.60×10^{-63}	4.93×10^{-74}	4.78×10^{-85}	1.27×10^{-96}
50%	50	8.62×10^{-7}	4.88×10^{-13}	1.77×10^{-19}	4.00×10^{-26}	5.42×10^{-33}	4.25×10^{-40}	1.86×10^{-47}	4.29×10^{-55}	5.00×10^{-63}	2.76×10^{-71}
60%	60	3.42×10^{-5}	8.85×10^{-10}	1.71×10^{-14}	2.41×10^{-19}	2.46×10^{-24}	1.76×10^{-29}	8.69×10^{-35}	2.88×10^{-40}	6.24×10^{-46}	8.56×10^{-52}
70%	70	7.64×10^{-4}	4.89×10^{-7}	2.59×10^{-10}	1.13×10^{-13}	4.00×10^{-17}	1.14×10^{-20}	2.56×10^{-24}	4.52×10^{-28}	6.13×10^{-32}	6.32×10^{-36}
80%	80	0.0112	1.14×10^{-4}	1.04×10^{-6}	8.41×10^{-9}	6.06×10^{-11}	3.85×10^{-13}	2.15×10^{-15}	1.04×10^{-17}	4.35×10^{-20}	1.56×10^{-22}

Strict safety is attained by any G below it, and $G = 2$ suffices whenever $K \geq 2m_\alpha^*$.

4) *Summary*: Resistance to Identity Theft and Ghost Key attacks reduces to bounding $P_{\text{capture}}^{(\alpha)}$, the probability that a VRF-selected signing party contains enough adversarial authorities in every stratum to reach the per-stratum activation quorum threshold α . Under the hypergeometric tail model of S-GYID, $P_{\text{capture}}^{(\alpha)}$ admits the closed-form expression (1)–(2). Two qualitatively distinct regimes emerge, separated by the generalised threshold

$$G_{\text{cliff}}^{(\alpha)} = \max(2, \lfloor K/m_\alpha^* \rfloor + 1), \quad m_\alpha^* = \begin{cases} \lceil m/\alpha \rceil & \alpha < 1, \\ m + 1 & \alpha = 1. \end{cases}$$

Below the threshold, session capture is combinatorially impossible ($P_{\text{capture}}^{(\alpha)} = 0$); above it, $P_{\text{capture}}^{(\alpha)}$ is positive but remains

negligible for moderately sized K and G provided α is not too small. For the protocol's operational activation threshold $\alpha_{\text{act}} = 1/2$, the safe regime requires $k \geq 2m$, i.e. $G = 2$ suffices whenever $K \geq 4\lfloor f_{\text{tol}}N \rfloor$; the stricter promotion threshold $\alpha_{\text{prom}} = 3/4$ relaxes this to $K \geq \lceil 8f_{\text{tol}}N/3 \rceil$. The corresponding design rule for the operator is: given a tolerance f_{tol} and the threshold α actually enforced, choose K and G so that $\lfloor K/G \rfloor \geq m_{\text{tol},\alpha}^*$. In particular $G = 2$ always suffices once K is large enough to clear the safe-regime inequality at the targeted threshold.

TABLE VII

$G_{\text{CLIFF}}^{(1)}$ AND $\log_{10} P_{\text{CAPTURE}}^{(1)}$ AT THE CLIFF EDGE UNDER THREE TOLERANCE REGIMES ($\alpha = 1$ UNANIMOUS REFERENCE); $N = 100$. COLUMNS REPORT THE SESSION-CAPTURE PROBABILITY FOR THE LISTED *actual* ADVERSARIAL FRACTION f . —: $P_{\text{CAPTURE}}^{(1)} = 0$ (SAFE REGIME, $\log_{10}$ UNDEFINED).

K	$f_{\text{tol}} = 1/3$				$f_{\text{tol}} = 1/2$			$f_{\text{tol}} = 0.70$			
	G_{cliff}	$f=33\%$	$f=40\%$	$f=50\%$	G_{cliff}	$f=40\%$	$f=50\%$	G_{cliff}	$f=50\%$	$f=60\%$	$f=70\%$
40	2	-23.9	-19.2	-14.1	2	-19.2	-14.1	2	-14.1	-10.2	-7.04
80	3	-51.6	-40.4	-29.3	2	-56.3	-36.3	2	-36.3	-25.0	-16.8
100	3	-79.4	-57.6	-40.4	2	—	-58.0	2	-58.0	-36.3	-23.6
120	4	-86.9	-66.2	-47.2	3	-84.4	-54.4	2	—	-56.3	-33.1
140	5	-96.6	-74.8	-53.8	3	—	-70.5	2	—	—	-50.9
160	5	-123.2	-91.4	-64.5	4	-112.6	-72.5	3	—	-61.0	-39.1
200	6	-158.8	-115.2	-80.9	4	—	-116.0	3	—	—	-62.4

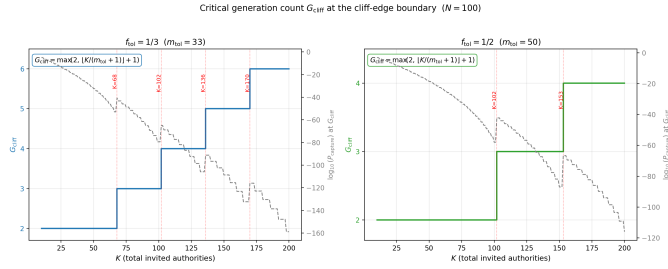


Fig. 3. $G_{\text{cliff}}^{(1)}(K, f_{\text{tol}})$ under the three canonical tolerance regimes $f_{\text{tol}} \in \{1/3, 1/2, 0.70\}$, $N = 100$, for the unanimous reference $\alpha = 1$. Staircase transitions occur at $K = m_{\text{tol}} + 1, 2(m_{\text{tol}} + 1), 3(m_{\text{tol}} + 1), \dots$. Under the operational activation threshold $\alpha = 1/2$ the transitions shift to $K = 2m_{\text{tol}}, 4m_{\text{tol}}, \dots$.

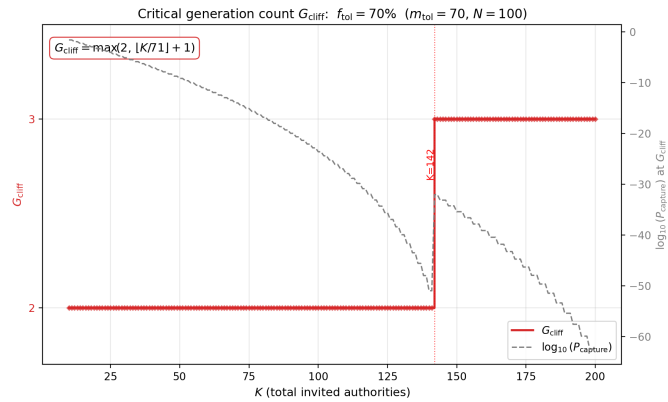


Fig. 4. Detail of the supermajority-tolerance regime ($f_{\text{tol}} = 0.70$, $m_{\text{tol}} = 70$, $N = 100$, $\alpha = 1$): $G_{\text{cliff}}^{(1)}(K)$ and the associated cliff-edge $\log_{10} P_{\text{capture}}^{(1)}$ at $f = 70\%$. The staircase advances in steps of 71, mirroring Figure 3 at the coarsest of the three regimes.

TABLE VIII

WORKED EXAMPLE, ACTIVATION THRESHOLD $\alpha = 1/2$: $K = 60$, $f = 30\%$, $m = 30$, $N = 100$. THE CLIFF $G_{\text{CLIFF}}^{(1/2)} = 2$ IS ALREADY CROSSED AT THE MINIMUM ALLOWED GENERATION COUNT, AND $P_{\text{CAPTURE}}^{(1/2)}$ SUBSEQUENTLY OSCILLATES WITH G .

G	$k = \lfloor 60/G \rfloor$	$t_{1/2}(k)$	$P_{\text{single}}^{(1/2)}$	$P_{\text{capture}}^{(1/2)}$
2	30	16	1.19×10^{-3}	1.41×10^{-6}
3	20	11	8.46×10^{-3}	6.05×10^{-7}
4	15	8	3.68×10^{-2}	1.82×10^{-6}
5	12	7	2.94×10^{-2}	2.19×10^{-8}
6	10	6	3.88×10^{-2}	3.39×10^{-9}
7	8	5	5.04×10^{-2}	8.29×10^{-10}

TABLE IX

RECOMMENDED ($K, G_{\text{CLIFF}}^{(1)}$) CONFIGURATIONS WITH $K = 3m_{\text{tol}}$ AND $G_{\text{CLIFF}}^{(1)} = 3$ ($\alpha = 1$ UNANIMOUS REFERENCE). ENTRIES: $\log_{10} P_{\text{CAPTURE}}^{(1)}$. —: $f < f_{\text{tol}}$, SAFE REGIME ($\log_{10}$ UNDEFINED).

Scenario	f_{tol}	N	K	m_{tol}	$f=33\%$	$f=50\%$	$f=70\%$
Small-1/3	1/3	100	99	33	-79.4	-40.4	-19.4
Small-1/2	1/2	100	150	50	—	-87.0	-35.4
Small-70%	0.70	100	210	70	—	—	-76.4
Medium-1/3	1/3	500	498	166	—	-205.7	-98.4
Medium-1/2	1/2	500	750	250	—	-447.2	-178.4
Medium-70%	0.70	500	1050	350	—	—	-393.7
Large-1/3	1/3	1000	999	333	—	-413.6	-197.8
Large-1/2	1/2	1000	1500	500	—	-898.3	-357.1
Large-70%	0.70	1000	2100	700	—	—	-791.2
XLarge-1/3	1/3	5000	4998	1666	—	-2071.6	-990.2
XLarge-1/2	1/2	5000	7500	2500	—	-4509.6	-1786.9
XLarge-70%	0.70	5000	10500	3500	—	—	-3973.7