Genuine Personal Identifiers and Mutual Sureties for Sybil-Resilient Community Formation

Genuine Personal Identifiers and Mutual Sureties for Sybil-Resilient Community Formation

Gal Shahaf[1], **Ehud Shapiro**[1] **& Nimrod Talmon**[2]
[1]**Weizmann Institute of Science**
[2]**Ben-Gurion University**
{**gal.shahaf, ehud.shapiro**}**@weizmann.ac.il, talmonn@bgu.ac.il**

Abstract. While most of humanity is suddenly on the net, the value of this singularity is hampered by the lack of credible digital identities: Social networking, person-to-person transactions, democratic conduct, cooperation and philanthropy are all hampered by the profound presence of fake identities, as illustrated by Facebook's removal of 5.4Bn fake accounts since the beginning of 2019.

Here, we introduce the fundamental notion of a *genuine personal identifier*—a globally unique and singular identifier of a person—and present a foundation for a decentralized, grassroots, bottom-up process in which every human being may create, own, and protect the privacy of a genuine personal identifier. The solution employs mutual sureties among owners of personal identifiers, resulting in a mutual-surety graph reminiscent of a web-of-trust. Importantly, this approach is designed for a distributed realization, possibly using distributed ledger technology, and does not depend on the use or storage of biometric properties. For the solution to be complete, additional components are needed, notably a mechanism that encourages honest behavior [29] and a sybil-resilient governance system [30].

Introduction

Providing credible identities to all by 2030 is a UN Sustainable Development Goal. Yet, current top-down digital identity-granting solutions are unlikely to close the 1Bn-people gap [6] in time, as they are not working for citizens of failed states nor for people fleeing physical or political harshness [17, 5]. Concurrently, humanity is going online at an astonishing rate, with more than half the world population now being connected. Still, online accounts do not provide a solution for credible digital identity either, as they may easily be fake, resulting in lack of accountability and trust. For example, Facebook reports the removal of 5.4Bn (!) fake accounts since the beginning of 2019 [16, 21].

The profound penetration of fake accounts on the net greatly hampers its utility for credible human discourse and any ensuing deliberations and democratic decision making; it makes the net unsuitable for vulnerable populations, including children and the elderly; it makes the use of the net for person-to-person transactions, notably direct philanthropy, precarious; and in general it turns the net into an inhuman, even dangerous, ecosystem. As an aside, we note that the panacea of cryptocurrencies for the lack of credible personal identities on the net is the reckless employment of the environmentally-harmful proof-of-work protocol.

Our aim is a conceptual and mathematical foundation for allowing every person to create, own, and protect the privacy of a globally-unique and singular identifier, henceforth referred to as *genuine personal identifier*. We believe that successful deployment and broad adoption of genuine personal identifiers will afford solutions to these problems and more, providing for: egalitarian digital democratic governance in local communities and in global movements; digital cooperatives and digital credit unions; direct philanthropy; child-safe digital communities; preventing unwanted digital solicitation; banishing deep-fake (by marking as spam videos not signed by a genuine personal identifier); credible and durable digital identities for people fleeing political or economic harshness; accountability for criminal activities on the net; and egalitarian cryptocurrencies employing an environmentally-friendly Byzantine-agreement consensus protocol among owners of genuine personal identifiers. Furthermore, genuine personal identifiers may provide the necessary digital foundation for a notion of *global citizenship* and, subsequently, for democratic global governance [31].

Granting an identity document by a state is a complex process as it requires careful verification of the person's credentials. The process culminates in granting the applicant a state-wide identifier that is unique (no two people have the same identifier) and singular (no person has two identifiers). Granting a genuine personal identifier might seem even more daunting, as it needs to be globally-unique, not only state-wide unique, except for following fundamental premise:
Every *person deserves a genuine personal identifier*. Thus, there are no specific credentials to be checked, except for the existence of the person. As a result, a solution for all people to create and own genuine personal identifiers may focus solely on ensuring the one-to-one correspondence between people and their personal identifiers.

A solution that is workable for all must be decentralized, distributed, grassroots, and bottom-up. Solid foundations are being laid out by the notion of self-sovereign identities [22] and the W3C Decentralized Identifiers [13] and Verifiable Claims [34] emerging standards, which aim to let people freely create and own identifiers and associated credentials. We augment this freedom with the goal that each person declares exactly one identifier as her *genuine personal identifier*. We note that besides the genuine personal identifier, one may create, own and use any number of identifiers of other types.

Becoming the owner of a genuine personal identifier is simple. With a suitable app, this could be done with a click of a button:

1. Choose a new cryptographic key-pair (v, v^{-1}), with v being the public key and v^{-1} the private key.

2. Claim v to be your genuine personal identifier by publicly posting a declaration that v is a genuine personal identifier, signed with v^{-1}.

Lo and behold! You have become the proud rightful owner of a genuine personal identifier.[1] Note that a declaration of a genuine personal identifier, by itself, does not reveal the person making the declaration; it only reveals to all that someone who knows the secret key for the public key v claims v as her genuine personal identifier. Depending on personality and habit, the person may or may not publicly associate oneself with v. E.g., a person with truthful social media accounts may wish to associate these accounts with its newly-minted genuine personal identifier.

If becoming the rightful owner of a genuine personal identifier is so simple, what could go wrong? In fact, so many things can go wrong, that this paper is but an initial investigation into describing, analyzing, and preventing them. Some of them are enumerated below; h denotes an agent:

1. The key-pair (v, v^{-1}) is not new, or else someone got hold of it between Step 1 and Step 2 above. Either way, someone else has declared v to be a genuine personal identifier prior to the declaration by h. In which case h cannot declare v.
2. Agent h failed to keep v^{-1} secret so that other people, e.g. h', know v^{-1}, in which case h' is also an *owner* of v and, thus, v is *compromised*. Figure 1 (left) illustrates a compromised personal identifier.
3. The agent h intended to divulge his association with his public key v only on a need-to-know basis, but the association of v and h has become public knowledge, prompting agent h to replace his genuine personal identifier v with a new one.
4. Agent h declared v as his genuine personal identifier, but later also declared another personal identifier v'. Then, v and v' are *duplicates*, v' is a *sybil*, and agent h is *corrupt*. An *honest* agent does not declare sybils. Figure 1 illustrates honest and corrupt agents.

We aim to develop the foundation for genuine personal identifiers utilizing basic concepts of public-key cryptography, graph theory, social choice theory and game theory. Here, we focus on utilizing the first two disciplines; see [30] for sybil-resilient social choice, which provides a foundation for democratic governance of digital communities with bounded sybil penetration; incorporating sybils in cooperative game theory, with the goal of forming a sybil-free grand coalition, is work in progress.

Related Work. Sybil-resilience has received considerable attention in AI research [2, 3, 7, 8, 10, 11, 12, 30, 35, 36] as elaborated below.

Philosophically, genuine personal identifiers aim to bridge the gap between agents and their corresponding identifiers. This distinction, acknowledged in semiotics as the difference between *signified* and *signifier*, strongly relates to the study of sense and reference initiated by Frege [15] followed by vast literature in analytic philosophy and the philosophy of language. Conceptually, the formal framework suggested here may be viewed as an attempt to computationally realize unique and singular signifiers of agents in a distributed setting.

Practically, digital identities are a subject of extensive study, with many organizations aiming at providing solutions, notably Self-Sovereign Identifiers [22]. Business initiatives include the Decentralized Identity Foundation (identity.foundation), the Global Identity Foundation [14] and Sovrin [33]. None of these projects are concerned with the uniqueness or singularity of personal identities.

[1] As the public key may be quite long, one may also associate oneself with a shorter "nickname", a hash of the public key, e.g. a 128-bit hash (as a UUID) or a 256-bit hash (as common in the crypto world).

Other high-profile projects to provide nationwide digital identities include India's Aadhaar system [23], Sierra Leone's Kiva identity protocol [32] and the World Food Programmes cash aid distribution program in refugee camps [19]. Here we are concerned with self-sovereign and global personal identities, not bound to any national boundaries or entities, and argue that top-down approaches fail to provide such a solution. In this context, we mention the concept of Proof of Personhood [7], aiming at providing unique and singular identities by means of conducting face-to-face encounters, an approach suitable only for small communities.

Our solution is based upon the notion of trust, thus we mention Andersen et al. [3], studying axiomatizations of trust systems. They are not concerned, however, with sybils, but with quality of recommendations. We mention work on sybil-resilient community growth [26], describing algorithms for the growth of an online community that keep the fraction of sybils in it small; and work on sybil-resilient social choice [30], describing aggregation methods to be applied in situations where sybils have infiltrated the electorate. In these two papers, a notion of genuine identities and sybils is used without specifying what they are; here, we define a concrete notion of genuine personal identifiers, and derive from it a formal definition of sybils and related notions of honest and corrupt agents and byzantine identifiers. Finally, we mention the work of Conitzer [12, 2] regarding computerized tests for sybil fighting: A test that is hard for a person to pass more than once [12] and a test that is hard to pass simultaneously by one person [2].

Genuine Personal Identifiers

Ingredients. The ingredients needed for a realization of genuine personal identifiers are:

1. A set of agents. It is important to note that, mathematically, the agents form a set (of unique entities) not a multiset (with duplicates). Intuitively, it is best to think of agents as people (or other physical beings with unique personal characteristics, unique personal history, and agency, such as intelligent aliens), which cannot be duplicated, but not as software agents, which can be.
2. A way for agents to create cryptographic key-pairs. This can be realized, e.g., using the RSA standard [27]. Our solution does not require a global standard or a uniform implementation for public key encryption: Different agents can use different technologies for creating and using such key-pairs, as long as the signatures-verification methods are declared.
3. A way for agents to sign strings using their key-pairs. As we assume cryptographic hardness, an agent that does not know a certain key-pair cannot sign strings with this key-pair.
4. A bulletin board or public ledger, to which agents may post and observe signed messages, where all agents observe the same order of messages. The weaker requirement that the same order is observed only eventually, as is standard with distributed ledger protocols, could also be accommodated. Considering partial orders is a subject of future work.

Agents and their Personal Identifiers. We assume a set of agents $\mathcal{H}$ that is fixed over time.[2] Agents can create new key-pairs (v, v^{-1}). We assume that an agent that has a key-pair can sign a string, and denote by $v(x)$ the string resulting from signing the string x with v^{-1}. Intuitively, each agent corresponds to a human being. Importantly,

[2] Birth and death of agents will be addressed in future work.

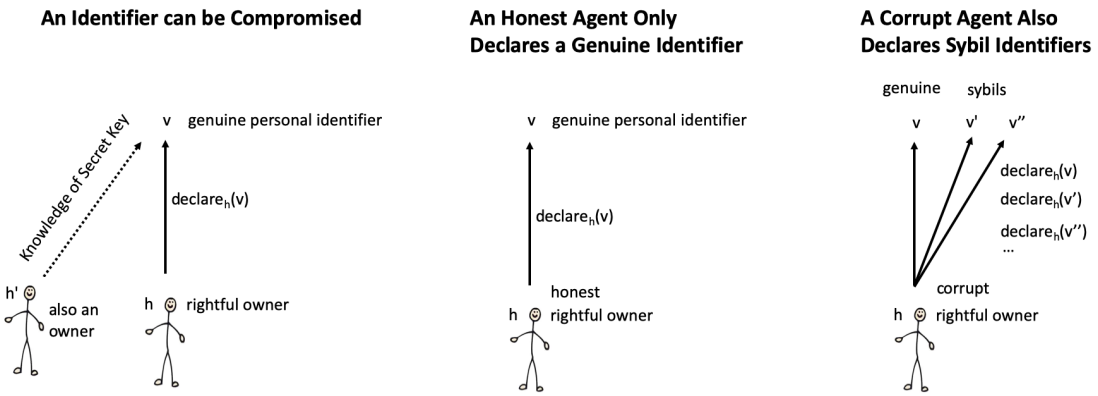

Figure 1. A compromised identifier (left), honest agent (middle), and corrupt agent (right).

members of the set $\mathcal{H}$ of agents (e.g., containing all human beings) cannot be referenced explicitly and, in particular, posted signed messages never refer directly to agents $h \in \mathcal{H}$. A key motivation for our work is providing people with digital genuine personal identifiers without accessing any of their intrinsic (e.g. biometric) private properties and without depending upon such properties as identifiers.

As we aim personal identifiers to be self-sovereign identities that conform to the W3C Decentralized Identifiers emerging standards, we let agents create and own their personal identifiers. An agent h can publicly declare a personal identifier v for which it knows the private key v^{-1}. A *personal identifier declaration* has the form $v(gid(v))$ and can be effected by agent h posting $v(gid(v))$ to a public ledger. We denote this action by $declare_h(gid(v)) \in \mathcal{C}$. Recall that all agents have the same view of the sequence of all declarations made; subsequent work may relax this assumption.

Definition 1 (Personal Identifier) *Let $\mathcal{C}$ be a sequence of personal identifier declaration events and $declare_h(gid(v)) \in \mathcal{C}$ the first declaration event in which v occurs. Then v is a* personal identifier *and h is the* rightful owner *of v, given $\mathcal{C}$.*

Definition 2 (Genuine Personal Identifier, Sybil, Honest, and Corrupt Agents) *Let $\mathcal{C}$ be a sequence of personal identifier declaration events and h be the rightful owner of personal identifier v in $\mathcal{C}$. Then v is* genuine *if it is the first personal identifier declared in $\mathcal{C}$ by h, else v is a* sybil*. An agent h is* corrupt *if it declares any sybils, else h is* honest*. (All notions are relative to $\mathcal{C}$.)*

See Figure 1 and some remarks: (1) An agent is the rightful owner of its genuine personal identifier as well as of any subsequent sybils that it declares. (2) If h, the rightful owner of v, is corrupt, then its first declared identifier is genuine and the rest of its declared identifiers are all sybils. (3) An honest agent may create and use many key-pairs for various purposes, yet remain honest as long as it has declared at most one public key as a personal identifier.

Mutual Sureties and Their Graphs

A key element of our approach is the pledging of mutual sureties by agents. Intuitively, mutual surety pledges provide a notion of trust between the owners of personal identifiers. They allow two agents that know each other and know the personal identifiers declared by each other to vouch for the other regarding the good standing of the personal identifiers. This notion is key to help honest agents fend-off sybils. We consider several types of mutual sureties, of increasing strength, and illustrate the corresponding sybil-resilience each type of mutual sureties can obtain.

For an agent, say h, to provide surety regarding the personal identifier of another agent, say h', h first has to know h'. How this knowledge is established is not specified in our formal framework, but this is quite an onerous requirement that cannot be taken lightly or satisfied casually. E.g., we may assume that one knows one's family, friends and colleagues, and may diligently get to know new people if one so chooses. We consider several types of sureties of increasing strength, in which an agent h with personal identifier v makes a pledge regarding the personal identifier v' of another agent h'; all assume that the agent h knows the agent h'. We describe four *Surety Types*, which are cumulative as each includes all previous ones, explain on what basis one may choose to pledge each of them, and present results that utilize them,

Surety of Type 1: *Ownership* of a personal identifier. Agent h pledges that agent h' owns personal identifier v'.

Agent h' can prove to agent h that she owns v' without disclosing v'^{-1} to h. This can be done, for example, by h asking h' to sign a novel string x and verifying that $v'(x)$ is signed using v'^{-1}. This surety type is the weakest of all four, it is the one given in "key signing parties", and is implicitly assumed by applications such as PGP and web-of-trust [1]. For a given surety type, we say that the surety is *violated* if its assertion does not hold; in particular, a surety of Type 1 is violated if h' in fact does not know the secret key v'^{-1}.

In general, mutual surety between two agents with two personal identifiers is pledged by each of the two agents pledging a surety to the personal identifier of the other agent.[3] We define below three additional surety types, where the format of a surety pledge of Type X by the owner of v to the owner of v' is $v(suretyX(v'))$, $X \in \{1, 2, 3, 4\}$. The corresponding surety event is $pledge_h v(suretyX(v'))$, and the surety enters into effect once both parties have made the mutual pledges. We now take $\mathcal{C}$ to be a record of both declaration events and pledge events.

[3] We consider undirected graphs, as we require surety to be symmetric. Indeed, one may consider directed sureties.

Definition 3 (Mutual Surety) *The personal identifiers v, v' have* mutual surety of type X, *$X \in \{1,2,3,4\}$, if there are $h, h' \in \mathcal{H}$ for which $pledge_h v(suretyX(v')) \in \mathcal{C}$ & $pledge_{h'} v'(suretyX(v)) \in \mathcal{C}$, in which case h and h' are the* witnesses *for the mutual surety between v and v'.*

A sequence of events induces a sequence of surety graphs in which the vertices are personal identifiers that correspond to personal identifier declarations and the edges correspond to mutual surety pledges.

Definition 4 (Surety Graph) *Let $\mathcal{C} = c_1, c_2, \ldots$ be a sequence of events and let $\mathcal{C}_k$ denote its first $k \geq 0$ events. Then, for each $k \geq 0$, $\mathcal{C}_k$ induces a* surety graph of type X, *$GX_k = (V_k, EX_k)$, $X \in \{1,2,3,4\}$, as follows:*[4]

$$V_k = \{v \mid declare_h(gid(v)) \in \mathcal{C}_k \text{ for some } h \in \mathcal{H}\}$$

$$EX_k = \{(v, v') \mid pledge_h v(suretyX(v')) \in \mathcal{C}_k, \\ pledge_{h'} v'(suretyX(v)) \in \mathcal{C}_k, \\ \text{for some } h, h' \in \mathcal{H}, v, v' \in V_k\}$$

Remark 1 *Observe that mutual sureties can be easily pledged by agents, technically. However, we wish agents to be prudent and sincere in their mutual surety pledges. Thus, we expect a mechanism that, on one hand, rewards the pledging of sureties but, on the other hand, punishes for surety violations, for example based on the approach of [29]. While the specifics of such a mechanism is beyond the scope of the current paper, note that with such a mechanism in place, the commissive illocutionary force [28] of a surety pledge will come to bear.*

Updating a Personal Identifier with Mutual Sureties

Once creating a genuine personal identifier is provided for, one must also consider the many circumstances under which a person may wish to update their personal identifier:

1. **Identifier loss:** The private key was lost.
2. **Identifier theft:** The private key was stolen, robbed, extorted, or otherwise compromised.
3. **Identifier breach of privacy:** The association between the personal identifier and the person was accidentally or maliciously disclosed with unwarranted consequences.
4. **Identifier refresh:** Proactive identifier update to protect against all the above.

Update in Case of Loss or Theft. The personal identifier declaration event $declare_h(gid(v))$ establishes v as a personal identifier. To support updating a personal identifier, we add the personal identifier update event $declare_h(gid(v, v'))$, which declares that v' is a new personal identifier that replaces v. A public declaration of identifier update has the form $v(gid(v, v'))$, i.e., it is signed with the new identifier. We refer to declarations of both types as *personal identifier declarations*, and extend the assumption that a new identifier can be declared at most once to this broader definition of identifier declaration. The *validity* of an identifier update declaration is defined inductively, as follows.

[4] We allow surety pledges to be made before the corresponding personal identifier declarations, as we do not see a reason to enforce order.

Definition 5 (Valid identifier Update declaration) *Let $\mathcal{C}$ be a sequence of declarations, V the set of global identities declared in $\mathcal{C}$, and $h \in \mathcal{H}$. A* personal identifier update event *over V has the form $declare_h(gid(v, v'))$, $v, v' \in V$.*

A personal identifier update event $declare_h(gid(v, v')) \in \mathcal{C}$ is valid *and h is the* rightful owner *of v if it is the first identifier declaration event of v and h is the rightful owner of v'.* □

Valid personal identifier declarations should form linear chains, one for each agent, each starting from $gid(v)$ and ending with the currently valid personal identifier of the agent:

Definition 6 (Identifier Provenance Chain) *Let $\mathcal{C}$ be a sequence of declarations and V the declared set of global identities. An* identifier provenance chain *(*provenance chain *for short) is a subsequence of $\mathcal{C}$ of the form (starting from the bottom):*

$$declare_{h_k}(gid(v_k, v_{k-1})),$$
$$declare_{h_{k-1}}(gid(v_{k-1}, v_{k-2})),$$
$$\ldots$$
$$declare_{h_1}(gid(v_1)).$$

Such a provenance chain is valid *if the declarations in it are valid. Such a provenance chain is* maximal *if there is no declaration*

$$declare_h(gid(v, v_k)) \in \mathcal{C}$$

for any $v \in V$ and $h \in \mathcal{H}$. A personal identifier v is current *in $\mathcal{C}$ if it is the last identifier $v = v_k$ in a maximal provenance chain in $\mathcal{C}$.* □

Note that it is very easy for an agent to make an update declaration for its identifier. However, it is just as easy for an adversarial agent wishing to steal the identifier to make such a declaration. Hence, this ability must be coupled with a mechanism that protects the rightful owner of an identifier from identifier theft through invalid identifier update declarations. Here we propose to use a stronger type of mutual sureties to support valid identifier update declarations and help distinguish between them and invalid declarations.

Surety of Type 2: *Rightful* ownership of a personal identifier. Agent h pledges that h' is the rightful owner of personal identifier v'.

In addition to proving to h that it owns v', h' must provide evidence that h' itself, and not some other agent, has declared v'. A selfie video of h' pressing the *declare* button with v', signed with a certified timestamp promptly after the video was taken, and then signed by v', may constitute such evidence. A suitable app may record, timestamp, and sign such a selfie video automatically during the creation of a genuine personal identifier. In particular, this surety is violated if h' in fact did not declared v' as a personal identifier.

Note that immediately following an identifier update declaration, the new identifier may not have any surety edges incident to it. Thus, as a crude measure, we may require that the identifier update would come to bear only after all the Type 2 surety neighbors of the old identifier, or a sufficiently large majority of them, would update their mutual sureties to be with the new identifier. To achieve that, an agent wishing to update its identifier would have to approach its neighbors and to create such updated Type 2 mutual surety pledges.

Example 1 *Consider two friends, agent h and agent h' having a mutual surety pledge between them. If h' would lose her identifier, she would create a new key-pair, make an identifier update declaration, and ask h for a new mutual surety pledge between h's identifier and h''s new identifier.*

The following observation follows from: (1) a valid provenance chain has a single owner; and (2) whether a Type 2 surety between two identifiers is violated depends on their rightful owners.

Observation 1 *Let $\mathcal{C}$ be a sequence of update declarations and C_1, C_2 be two valid provenance chains in $\mathcal{C}$. If a Type 2 surety pledge between two global identities $v_1 \in C_1, v_2 \in C_2$ is valid, then any Type 2 surety pledge between two personal identifiers in these provenance chains, $u_1 \in C_1, u_2 \in C_2$ is valid.*

The import of Observation 1 is that a Type 2 mutual surety can be "moved along" valid provenance chains as they grow, without being violated, as it should be. Below we argue that invalid identifier update declarations are quite easy to catch, thus the risk of stealing identities can be managed. In effect, we show the value of Type 2 surety pledges in defending an identifier against theft via invalid update declarations.

Let $\mathcal{C}$ be a sequence of declarations, C_1, C_2 be two provenance chains in $\mathcal{C}$, and assume that there is a valid Type 2 surety pledge between the two current global identifiers $v_1 \in C_1, v_2 \in C_2$, made by $h_1 =$ *Marry* and $h_2 =$ *John*. Now assume that the identifier update declaration $c = declare_h(gid(v, v_1))$ is made, namely, some agent *Sue* $\neq$ *Marry* has declared to replace v_1 by v. Then, it will be hard for *Marry* to secure surety from *John* and, if she attempts to do so, then *John* will know that c is not valid and thus (if *John* is honest) a Type 2 mutual surety between v and v_2 will not be established. Consider the following case analysis:

- Assume *Marry* notices c. Then she would inform *John* that she did not declare c, and thus *John* will know that c is not valid.
- Assume that *John* notices c. He would approach *Marry* to update the Type 2 mutual surety between them accordingly; *Marry* would deny owning v, and thus *John* will know that c is invalid.
- Alternatively, *Sue* would approach *John* to update the Type 2 mutual surety of v_2 with v_1 to be with v instead; *John* will see (or suspect, if *Sue* did not reveal herself) that *Sue* is not *Marry*, will double check with *Marry* and deem the declaration c invalid.

Reset in Case of Breach of Privacy. Note that provenance chains address identifier update in case of loss of theft, but do not address breach of privacy, as the updated new identifier is publicly tied to the previous one. To address breach of privacy, a person first has to invalidate his existing identifier, then create a new genuine identifier that is not linked to the previous one; this may result in loss of any public credit and goodwill associated with the old genuine personal identifier, but there may be circumstances in which a person would need to protect his privacy even at the expense of such loss. To facilitate that, an identifier reset declaration uses the special value **null** and has the form $declare_h(gid(v, \textbf{null}))$, which nullifies v as a personal identifier, and enters into effect if supported by those who initially provided the surety to v. Agent h would then be free to declare a new personal identifier v', which is not tied to the now-defunct v, and at the same time without v' being a sybil and without h becoming corrupt as a result of this declaration.

We note that using **null**, an initial genuine identifier declaration could be of the form $declare_h(gid(\textbf{null}, v))$, thus dispensing with the unary declaration format $declare_h(gid(v))$ altogether.

GDPR. In general, our approach does not imply any "data controllers" or "data processors" [24] other than people and their trusted friends, but if realized on a large scale it might require such, possibly democratically-appointed by the large community that needs them. Furthermore, our approach and does not record or store any "personally identifiable information", and as such we believe is compliant with GDPR [24]. One exception is perhaps the association of a genuine personal identifier with the person that owns it, which could be made public on purpose by the owner, inadvertently by the owner or by another person, or maliciously by another person. Closely tied to this exception is GDPR's "right to be forgotten" [24, Article 17], which is notoriously difficult to realize in a distributed setting and hence resulted in sweeping legal opinions regarding the inapplicability of distributed ledger/blockchain technology for the storage of personally identifiable information [25].

Our method of personal identifier reset first invalidates an existing identifier (which could be coupled with a demand of erasure of every reference to this identifier from any public data controller, e.g. Facebook) and then creates an independent and unlinked new personal identifier. This may be the first proposal on how to realize the "right to be forgotten" in a decentralized setting.

Sybil- and Byzantine-Resilient Community Growth

Ideally, we would like to attain sybil-free communities, but acknowledge that one cannot prevent sybils from being declared and, furthermore, perfect detection and eradication of sybils is out of reach. Thus, our aim is to provide the foundation for a digital community of genuine personal identifiers to grow by admitting new identifiers indefinitely, while retaining a bounded sybil penetration. As noted above, democratic governance can be achieved even with bounded sybils penetration [30].

Community history. For simplicity, we assume a single global community A and consider elementary transitions obtained by either adding a single member to the community or removing a single community member:

Definition 7 (Elementary Community Transition) *Let A, A' denote two communities in V. We say that A' is obtained from A by an* elementary community transition*, and we denote it by $A \to A'$, if:*

- *$A' = A$, or*
- *$A' = A \cup \{v\}$ for some $v \in V \setminus A$, or*
- *$A' = A \setminus \{v\}$ for some $v \in A$.*

Definition 8 (Community History) *Let $\mathcal{C} = c_1, c_2, ...$ be a sequence of events. A* community history *wrt. $\mathcal{C}$ is a sequence of communities $\mathcal{A} = A_1, A_2, \ldots$ such that $A_i \subseteq V_i$ and $A_i \to A_{i+1}$ holds for every $i \geq 1$.*

We do not consider community governance in this paper, only the effects of community decisions to add or remove members. Hence, we assume that the sequence of events $\mathcal{C}$ includes the events $add_A(v)$ and $remove_A(v)$. With this addition, $\mathcal{C} = c_1, c_2, \ldots$ induces a community history $A_1, A_2, ...$, where $A_{i+1} = A_i \cup \{v\}$ if $c_i = add_A(v)$ for $v \in V \setminus A_i$; $A_{i+1} = A_i \setminus \{v\}$ if $c_i = remove_A(v)$ for $v \in A_i$; else $A_{i+1} = A_i$.

Definition 9 (Community, Sybil penetration rate) *Let $\mathcal{C}$ be a sequence of events and let $S \subseteq V$ denote the sybils in V wrt. $\mathcal{C}$. A community in V is a subset of identifiers $A \subseteq V$. The* sybil penetration *$\sigma(A)$ of the community A is given by $\sigma(A) = \frac{|A \cap S|}{|A|}$.*

The following observation is immediate.

Observation 2 *Let $A_1, A_2, ...$ be the community history wrt. a sequence of declarations $\mathcal{C}$. Assume that $A_1 \subseteq H$, and that whenever $A_{i+1} = A_i \cup \{v\}$ for some $v \notin A_i$, it holds that $Pr(v \in S) \leq \sigma$ for some fixed $0 \leq \sigma \leq 1$. Then, the expected sybil penetration rate for every A_i is at most σ.*

That is, the observation above states that a sybil-free community can keep its sybil penetration rate below σ, as long as the probability of admitting a sybil to it is at most σ. While the simplicity of Observation 2 might seem promising, its premise is naively optimistic. Due to the ease in which sybils can be created and to the benefits of owning sybils in a democratic community, the realistic scenario is of a hoard of sybils and a modest number of genuine personal identifiers hoping to join the community. Furthermore, once a fraction of sybils has already been admitted, it is reasonable to assume that all of them (together with their perpetrators of course) would support the admission of further sybils. Thus, there is no reason to assume neither the independence of candidates being sybils, nor a constant upper bound on the probability of sybil admission to the community. Hence, in the following we explore sybil-resilient community growth under more realistic assumptions.

Sybil-Resilient Community Growth

A far more conservative assumption includes a process employed by the community with the aim of detecting sybils. We shall use the abstract notion of *sybil detector* in order to capture such process, that may take the form of a query, a data-based comparison to other identifiers, or a personal investigation by some other agent. To leverage this detector to sybil-resilient community growth regardless of the sybil distribution among the candidates, we shall utilize a stronger surety type, defined as follows:

Surety of Type 3: Rightful ownership of a *genuine* personal identifier. Agent h pledges Surety Type 2 and that v' is the genuine personal identifier of h'.

Providing this surety requires a leap of faith. In addition to h obtaining from h' a proof of rightful ownership of v', h must also trust h' not to have declared any other personal identifier prior to declaring v'. There is no reasonable way for h' to prove this to h, hence the leap of faith.

Since Type 3 sureties inherently aim to distinguish between genuine identifiers and sybils, sybil-resilient community growth is established upon the underlying $G3$ surety graph. Specifically, we consider a setting where potential candidates to join the community are identifiers with a surety obtained from current community members. Conversely, we consider a violation of a surety in one direction as a strong indication that the surety in the other direction is violated. That is, if $(v, v') \in E$ and v' was shown to be sybil, (i.e., h' has declared some other v'' as a personal identifier before declaring v' as a personal identifier), then v should undergo a thorough investigation in order to determine whether it is sybil as well.

Next, we formalize this intuition in a simple stochastic model where admissions of new members are interleaved with random sybil detection among community members:

(1) An identifier is admitted to the community via an elementary community transition $A \rightarrow A \cup \{v\}$ only if there is some $a \in A$ with $(a, v) \in E$.
(2) Every admittance of a candidate is followed by a random sybil detection within the community: An identifier $a \in A$ is chosen uniformly at random. If a is genuine it is declared as such. If a is sybil, it is successfully detected with probability $0 < p \leq 1$.
(3) The detection of a sybil implies the successful detection of its entire connected sybil component (with probability 1). That is, if a is detected as sybil, then the entire connected component of a in the sybil subgraph $G|_S$ is detected and expelled from the community.
(4) The sybils are operated from at most $j \leq k$ disjoint sybil components in A. Furthermore, we assume that sybils join sybil components uniformly at random, i.e., a new sybil member has a surety to a given sybil component with probability $\frac{1}{k}$, else, it forms a new sybil component with probability $\frac{k-j}{k}$.

Note that assumption (2) is far weaker than the premise in Observation 2 as it presumes nothing on the sybil penetration among the candidates, but rather on the proactive ability to detect a sybil, once examined. Assumption (3) exploits the natural cooperation among sybils, especially if owned by the same agent, and assumes that if a sybil is detected by a shallow random check with probability p, then all its neighbours will be thoroughly investigated and will be detected if sybil with probability 1, continuing the investigation iteratively until the entire connected sybil component of the initially-detected sybil is identified. In assumption (4), the parameter k and the locations of the components are adversarial – the attacker may choose how to operate. While realistic attackers may also choose to which component shall the new (sybil) member join, uniformity is assumed to simplify the analysis. Possible relaxations of this model are future work.

For this setting we show an upper bound on the expected sybil penetration, assuming bounded computational resources of the attacker.

Theorem 1 *In the stochastic model described above, obtaining an expected sybil penetration $E[\sigma(A)] \geq \epsilon$ is NP-hard for every $\epsilon > 0$.*

Proof. Let $X_i \subseteq A$ denote a sybil component within the community at time i. In the stochastic model described above, X_i is detected and immediately expelled with probability $p \cdot \frac{|X_i|}{|A_i|}$. The expected size of the component in this model is obtained in a steady state, i.e., in a state i in time where $E[|X_{i+1}|] = |X_i|$, that is:

$$(1 - px/n) \cdot \frac{1}{k} \cdot (x + 1) + (1 - px/n) \cdot (1 - \frac{1}{k}) \cdot x = x,$$

where $n := |A_i|$ and $x = |X_i|$. It follows that $x^2 + \frac{1}{k}x - \frac{n}{pk} = 0$. Solving this quadratic equation implies that the size of a single sybil component in the steady state is $x \leq \sqrt{n/pk}$. It follows that the number of sybils in the community in a steady state is $xk \leq \sqrt{nk/p}$.

The crucial observation now is that operating from k nonempty sybil components corresponds to obtaining an independent set of size k (at least, choosing a single vertex in each component). The theorem follows from the fact that approximating independent set within a constant factor is NP-hard (see, e.g., [4]). □

The following corollary establishes an upper bound on the sybil penetration rate regardless of the attacker's computational power. The result is formulated in terms of the second eigenvalue of the graph restricted to A_i. ($\lambda(G|_{A_i})$ is defined in the supplementary materials section.

Corollary 1 *Let $\mathcal{A} = A_1, A_2, \ldots$ be a community history wrt. a sequence of events $\mathcal{C}$. If every community $A_i \in \mathcal{A}$ with $A_i = A_{i-1} \uplus \{v\}$ satisfies $\lambda(G|_{A_i}) < \lambda$, then the expected sybil penetration in every $A_i \in \mathcal{A}$ under the stochastic model depicted above, is at most $\sqrt{\lambda/p}$.*

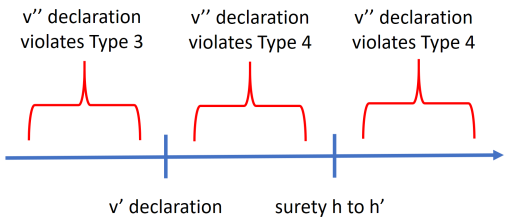

Figure 2. Violations of Sureties of Type 3 and Type 4. The axis of time goes from left to right, with the points in time in which v' was declared by h' and the surety from h to h' was declared. Then, the braces describe the time regions in which, if another identifier v'' was to be declared by h', would correspond to a violation of a surety of Type 3 or 4.

Proof. Recall that the size of the maximal independent set is a trivial upper bound on k. The cardinality of an independent set in a λ-expander is at most λn [18]; thus, $k \leq \lambda n$. It follows that the number of sybils in the community in a steady state is $xk \leq \sqrt{nk/p} \leq n\sqrt{\lambda/p}$. □

Byzantine-Resilient Community Growth

Here we consider the challenge of byzantine-resilient community growth. Intuitively, the term *byzantines* aims to capture identifiers owned by agents that are acting maliciously, possibly in collaboration with other malicious agents. Formally, we define byzantines as follows.

Definition 10 (Byzantine and harmless identifiers, Byzantine penetration) *An identifier is said to be* byzantine *if it is either a sybil or the personal identifier of a corrupt agent. Non-byzantine identifiers are referred to as* harmless*. We denote the byzantine and harmless identifiers in V by $B, H \subseteq V$, respectively. The* byzantine penetration *$\beta(A)$ of a community $A \subseteq V$ is given by $\beta(A) = \frac{|A \cap B|}{|A|}$.*

Since $S \subseteq B$, it holds that $\sigma(A) \leq \beta(A)$ for every community A, hence an upper bound on the byzantine penetration also provides an upper bound on the sybil penetration.

As byzantine identifiers include genuine identifiers, they are inherently harder to detect, and thus the detection-based model described in Section is no longer applicable in this setting. Rather, to achieve byzantine-resilient community growth, we rely on a stronger surety type, defined as follows:

Surety of Type 4: Rightful ownership of a genuine personal identifier by an *honest* agent. Agent h pledges Surety Type 3 and, furthermore, that v' is a genuine personal identifier of an honest agent h'.

Here h has to put even greater trust in h': Not only does h has to trust that the past actions of h' resulted in v' being her genuine personal identifier, but she also has to take on faith that h' has not declared any sybils since and, furthermore, that h' will not do so in the future. Note that a Type 4 surety is violated if after h' declares v' it ever declares some other v'' as a personal identifier. See Figure 2 for illustrations of violations of sureties of Types 3 and 4.

In the following, we provide sufficient conditions for Type 4 sureties to be used for byzantine-resilient community growth. We utilize the notation $A \twoheadrightarrow A'$ to indicate that A' was obtained from A via a finite sequence of elementary community transitions of incremental growth. Formally, $A \twoheadrightarrow A'$ if there exists $k \in \mathbb{N}$ with $A = A_0 \to A_1 \to ...A_k = A'$, with $|A_i \setminus A_{i-1}| = 1$ for all $i \in [k]$.

Theorem 2 *Let $\mathcal{A} = A_1 \twoheadrightarrow A_2 \twoheadrightarrow A_3 \twoheadrightarrow \ldots$ be a community history wrt. a sequence of events $\mathcal{C}$. Set a sequence of degrees $d_1, d_2, \ldots$ and parameters $\alpha, \beta, \gamma, \delta \in [0, 1]$. Assume:*

1. *$deg(v) \leq d_i$ for all $v \in A_i$, $i \in \mathbb{N}$.*
2. *Every $a \in A_i$ satisfies $\frac{|\{x \in A_i \mid (a,x) \in E\}|}{d_i} \geq \alpha$.*
3. *$\frac{|A_1 \cap B|}{|A_1|} \leq \beta$.*
4. *$\frac{e(A_i \cap H, A_i \cap B)}{vol_{A_i}(A_i \cap H)} \leq \gamma$.*
5. *$|A_i \setminus A_{i-1}| \leq \delta |A_{i-1}|$, with $\beta + \delta \leq \frac{1}{2}$.*
6. *$\Phi(G|_{A_i}) > \frac{\gamma}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$.*

Then, every community $A_i \in \mathcal{A}$ has Byzantine penetration $\beta(A_i) \leq \beta$.

Roughly speaking, Theorem 2 suggests that whenever: (1) Each graph $G|_{A_i}$ has a bounded degree d_i; (2) Sufficiently many edges are within A_i; (3) Byzantine penetration to A_1 is bounded; (4) Edges between harmless and byzantine identifiers are scarce; (5) Community growth in each step is bounded; (6) The conductance within $G|_{A_i}$ is sufficiently high; Then, the community may grow indefinitely with bounded byzantine penetration.

As Theorem 2 allows byzantine resilient growth at a fixed rate δ, it may be interpreted as an extension of a related result by Poupko et al. [26], where new members were added one at a time. Moreover, [26] assumed a notion of honest (what we refer to as genuine) and byzantine identities without defining what they are; here we formalize these notions and provide more sound definitions of harmless and byzantine identifiers. A formal definition of graph conductance Φ and the proof of Theorem 2 may be found in the supplementary material.

Remark 2 *A potential application of Theorem 2 is a byzantine-resilient union of two communities. Let $A, A' \subseteq V$ denote two communities that have some overlap (non-empty intersection) and wish to unite into $A_2 := A \cup A'$. Then, if Theorem 2 holds for (A_1, A_2) in case $A_1 := A$ and also in case $A_1 := A'$, this would provide both A and A' the necessary guarantee that the union would not result in an increase of the sybil penetration rate for either community.*

Outlook

Digital identity systems face the "Decentralized Identity Trilemma", of being (*i*) privacy-preserving, (*ii*) sybil-resilient and (*iii*) self-sovereign, all at the same time [20]. It has been claimed that no existing identity system satisfies all three corners of the Trilemma [20]; our approach may be the first to do so.

While this paper provides a formal mathematical framework, we aimed the constructions to be readily amenable to implementation. Realizing the proposed solution entails developing additional components, notably sybil-resilient governance mechanisms, e.g. along the lines of [30]; a mechanism for encouraging honest behavior and discouraging corrupt behavior, e.g. along the lines of [29]; and a cryptocurrency to fuel such a mechanism and the system in general. Once all components have been designed, we aim to implement, simulate, test, deploy, and evaluate the proposed framework, hopefully realizing the potential of genuine personal identifiers.

References

[1] Alfarez Abdul-Rahman, 'The pgp trust model', in *EDI-Forum: the Journal of Electronic Commerce*, volume 10 (3), pp. 27–31, (1997).

[2] Garrett Andersen and Vincent Conitzer, 'Atucapts: Automated tests that a user cannot pass twice simultaneously.', in *Proceedings of IJCAI '16*, pp. 3662–3669, (2016).

[3] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz, 'Trust-based recommendation systems: an axiomatic approach', in *Proceedings of WWW '08*, pp. 199–208, (2008).

[4] Sanjeev Arora and Boaz Barak, *Computational complexity: a modern approach*, Cambridge University Press, 2009.

[5] R. Baird, K. Migiro, D. Nutt, A. Kwatra, S. Wilson, J. Melby, A. Pendleton, M. Rodgers, and J. Davison. Human tide: the real migration crisis, 2007.

[6] World Bank. Identification for development (ID4D) global dataset, 2018.

[7] M. Borge, E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, and B. Ford, 'Proof-of-personhood: Redemocratizing permissionless cryptocurrencies', in *Proceedings of EuroS&PW '17*, pp. 23–26, (2017).

[8] Maria Borge, Eleftherios Kokoris-Kogias, Philipp Jovanovic, Linus Gasser, Nicolas Gailly, and Bryan Ford, 'Proof-of-personhood: Redemocratizing permissionless cryptocurrencies', *Proceedings of EuroS&PW '17*, 23–26, (2017).

[9] Jeff Cheeger, 'A lower bound for the smallest eigenvalue of the laplacian', in *Proceedings of the Princeton conference in honor of Professor S. Bochner*, (1969).

[10] V. Conitzer, N. Immorlica, J. Letchford, K. Munagala, and L. Wagman, 'False-name-proofness in social networks', in *Proceedings of the 6th International Workshop on Internet and Network Economics (WINE '10)*, pp. 209–221, (2010).

[11] V. Conitzer and M. Yokoo, 'Using mechanism design to prevent false-name manipulations', *AI magazine*, **31**(4), 65–78, (2010).

[12] Vincent Conitzer, 'Using a memory test to limit a user to one account', in *Proceedings of AMEC '08*, pp. 60–72, (2008).

[13] D. Longley C. Allen R. Grant M. Sabadello D. Reed, M. Sporny, 'Decentralized identifiers (DIDs) v0. 12–data model and syntaxes for decentralized identifiers (DIDs)', *Draft Community Group Report*, **29**, (2018).

[14] The Global Identity Foundation. Global identity challenges, pitfalls and solutions, 2014.

[15] Gottlob Frege, 'On sense and reference', *oversatt av Max Black, i J. Guitérrez-Rexach (red.): Semantics: Crictical concepts in linguistics*, **1**, 7–25, (2003).

[16] Brian Fung and Ahiza Garcia, 'Facebook has shut down 5.4 billion fake accounts this year', *CNN*, (November 2019).

[17] Charles Geisler and Ben Currens, 'Impediments to inland resettlement under conditions of accelerated sea level rise', *Land Use Policy*, **66**, 322–330, (2017).

[18] Shlomo Hoory, Nathan Linial, and Avi Wigderson, 'Expander graphs and their applications', *Bulletin of the American Mathematical Society*, **43**(4), 439–561, (2006).

[19] Nir Kshetri and Jeffrey Voas, 'Blockchain in developing countries', *It Professional*, **20**(2), 11–14, (2018).

[20] Maciek Laskus. Decentralized identity trilemma, 2018. Available at http://maciek.blog/dit/?cookie-state-change=1574327093444.

[21] Niall McCarthy, 'Facebook deleted more than 2 billion fake accounts in the first quarter of the year', *Forbes*, (May 2019).

[22] A. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, 'A survey on essential components of a self-sovereign identity', *Computer Science Review*, **30**, 80–86, (2018).

[23] Government of India. Home - unique identification authority of india, 2018. Available at https://uidai.gov.in.

[24] European Parliament. General data protection regulation (gdpr).

[25] European Parliament. Blockchain and the general data protection regulation: Can distributed ledgers be squared with european data protection law?, 2019.

[26] Ouri Poupko, Gal Shahaf, Ehud Shapiro, and Nimrod Talmon, 'Sybil-resilient conductance-based community growth', in *Proceedings of CSR '19*, (2019). A preliminary version appears in https://arxiv.org/abs/1901.00752.

[27] R. L. Rivest, A. Shamir, and L. Adleman, 'A method for obtaining digital signatures and public-key cryptosystems', *Communications of the ACM*, **21**(2), 120–126, (1978).

[28] J. R. Searle, S. Willis, and D. Vanderveken, *Foundations of illocutionary logic*, CUP Archive, 1985.

[29] Sven Seuken and David C Parkes, 'Sybil-proof accounting mechanisms with transitive trust', in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 205–212. International Foundation for Autonomous Agents and Multiagent Systems, (2014).

[30] G. Shahaf, E. Shapiro, and N. Talmon, 'Sybil-resilient reality-aware social choice', in *IJCAI '19*, pp. 572–579, (2019).

[31] E. Shapiro, 'Point: foundations of e-democracy', *Communications of the ACM*, **61**(8), 31–34, (2018).

[32] Susan Staats, Alfonso Sintjago, and Renata Fitzpatrick, 'Kiva microloans in a learning community: An assignment for interdisciplinary synthesis', *Innovative Higher Education*, **38**(3), 173–187, (2013).

[33] A. Tobin and D. Reed, 'The inevitable rise of self-sovereign identity', *The Sovrin Foundation*, **29**, (2016).

[34] W3C. Verifiable credentials data model 1.0, 2019.

[35] B. Waggoner, L. Xia, and V. Conitzer, 'Evaluating resistance to false-name manipulations in elections.', in *Proceedings of the 26rd AAAI Conference on Artificial Intelligence (AAAI '12)*, (2012).

[36] L. Wagman and V. Conitzer, 'Optimal false-name-proof voting rules with costly voting.', in *Proceedings of the 22st AAAI Conference on Artificial Intelligence (AAAI '08)*, pp. 190–195, (2008).

Supplementary Material

Graph notations and terminology

We provide some more detailed notations regarding graphs and conductance.

Let $G = (V, E)$ be an undirected graph. The *degree* of a vertex $x \in V$ is $\deg(x) := |\{y \in V \mid (x, y) \in E\}|$. The *volume* of a given subset $A \subseteq V$ is the sum of degrees of its vertices, $vol(A) := \sum_{x \in A} \deg(x)$. Additionally, we denote the subgraph induced on the set of vertices A as $G|_A$, by $\deg_A(x)$ the degree of vertex $x \in A$ in $G|_A$, and by $vol_A(B) := \sum_{x \in B} \deg_A(x)$ the volume of a set $B \subseteq A$ in $G|_A$. Given two disjoint subsets $A, B \subseteq V$, the size of the cut between A and B is denoted by

$$e(A, B) = |\{(x, y) \in E \mid x \in A, y \in B\}| .$$

Connectivity measures. In the following, we define two fundamental notions of graph connectivity that play a substantial role in safe community growth.

Definition 11 (Combinatorial Conductance) *Let $G = (V, E)$ be a graph. The* conductance *of G is defined by:*

$$\Phi(G) = \min_{\emptyset \neq A \subset V} \frac{e(A, A^c)}{\min\{vol(A), vol(A^c)\}} .$$

($A^c := V \setminus A$ is the complement of A.)

The following definition is an algebraic measure for connectivity.

Definition 12 (Algebraic conductance) *Let G be a graph, and let $\lambda_n \leq \lambda_{n-1} \leq ... \leq \lambda_2 \leq \lambda_1$ be the eigenvalues of its random walk matrix. Then, G is a said to be a λ-expander if its generalized second eigenvalue $\lambda(G) := \max_{i \neq 1} |\lambda_i|$ satisfies $\lambda(G) \leq \lambda$.*

We note that the notions of conductance and algebraic conductance are tightly related via the celebrated Cheeger inequality [9]. We refer the reader to the text of Hoory et al. [18] for through exposition and elaborate discussion regarding conductance and graph expansion.

Proof of Theorem 2

Theorem 2 follows by induction from the following Lemma:

Lemma 1 *Let $G = (V, E)$ be a surety graph with $A \subseteq A' \subseteq V$, and set $\alpha, \beta, \gamma, \delta \in [0, 1]$ and $d > 0$.*

Assume:

1. $deg(v) \leq d$ *for all* $v \in A'$.
   [The graph has a bounded degree].
2. *Every* $a \in A'$ *satisfies* $\frac{|\{x \in A' \mid (a,x) \in E\}|}{d} \geq \alpha$.
   [Sufficiently many edges are within members of A']
3. $\frac{|A \cap B|}{|A|} \leq \beta$.
   [Byzantine penetration to the initial community is bounded]
4. $\frac{e(A' \cap H, A' \cap B)}{vol_{A'}(A' \cap H)} \leq \gamma$.
   [the edges between harmless and byzantine identifiers are relatively scarce]
5. $|A' \setminus A| \leq \delta |A|$, *with* $\beta + \delta \leq \frac{1}{2}$.
   [Community growth is bounded]
6. $\Phi(G|_{A'}) > \frac{\gamma}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$.
   [the conductance within A' is sufficiently high]

Then, $\frac{|A' \cap B|}{|A'|} \leq \beta$.

Proof. We first note that due to $A \subseteq A'$, and assumptions (3), (5), we have

$$\begin{aligned} |A' \cap B| &\leq |A \cap B| + |A' \setminus A| \\ &\leq \beta|A| + \delta|A| \\ &\leq \frac{|A|}{2} < \frac{|A'|}{2}. \end{aligned}$$

As $V = B \uplus H$, it follows that

$$|A' \cap B| < |A' \cap H|. \tag{1}$$

We now utilize assumption (1):

$$\begin{aligned} vol_{A'}(A' \cap B) &:= \sum_{a \in A' \cap B} |\{x \in A' \mid (a, x) \in E\}| \\ &\geq \sum_{a \in A' \cap B} \alpha d = \alpha d |A' \cap B|. \end{aligned} \tag{2}$$

Similarly, we have

$$vol_{A'}(A' \cap H) \geq \alpha d |A' \cap H|. \tag{3}$$

Inequalities 3 and 1 imply that

$$vol_{A'}(A' \cap H) \geq \alpha d |A' \cap B|,$$

and together with Inequality 2, we have:

$$\min\{vol(A' \cap H), vol(A' \cap B)\} \geq \alpha d |A' \cap B|. \tag{4}$$

Now, Inequality 4 and assumption (6) imply that:

$$\begin{aligned} \frac{e(A' \cap H, A' \cap B)}{\alpha d |A' \cap B|} &\geq \frac{e(A' \cap H, A' \cap B)}{\min\{vol(A' \cap H), vol(A' \cap B)\}} \\ &> \frac{\gamma}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right), \end{aligned}$$

or equivalently

$$\frac{e(A' \cap H, A' \cap B)}{d\gamma |A' \cap B|} \geq \frac{1-\beta}{\beta}. \tag{5}$$

Assumptions (1) and (4) imply

$$\frac{e(A' \cap H, A' \cap B)}{d|A' \cap H|} \leq \frac{e(A' \cap H, A' \cap B)}{vol_{A'}(A' \cap H)} \leq \gamma,$$

or equivalently

$$|A' \cap H| \geq \frac{e(A' \cap H, A' \cap B)}{d\gamma}. \tag{6}$$

Combining Inequalities 5, 6 we get:

$$\begin{aligned} \frac{|A'|}{|A' \cap B|} &= \frac{|A' \cap H| + |A' \cap B|}{|A' \cap B|} \\ &\geq \frac{e(A' \cap H, A' \cap B)}{d\gamma |A' \cap B|} + 1 \\ &> \left(\frac{1-\beta}{\beta}\right) + 1 = \frac{1}{\beta}, \end{aligned}$$

where the first equality holds as $A = (A \cap H) \uplus (A \cap B)$, the second inequality stems from Equation 6 and the third inequality stems from Equation 5. Flipping the nominator and the denominator then gives $\beta(A') := \frac{|A' \cap B|}{|A'|} < \beta$. □