This is an a candidate anti-Sybil mechanism for digital identity sytem Upala. It withstands phone camera hacks, bot-net spamming and realistic dolls spamming. The goal is to make sure that nobody is able to register multiple accounts within the system. In other words, one person - one ID.

In short

Every user has a global and a local reputation. Local means geographically local.

Users earn local reputation when they meet each other in person in random pairs defined by the system. Users confirm they see a real person with the right ID photo. Face recognition algorithm detects sibyls. Twins have to prove that they are both real.

Global reputation is calculated by “routing” local user reputation from a small city to the global level. Two cities establish a connection when their citizens travel. The more travelers the more the bandwidth of the connection. When reputation is transferred through a connection it loses some points, depending on the connection bandwidth. Thus global reputation is smaller than the local.

Local reputation

Users meet each other in the real world in pseudo-random pairs defined by the system. If both users confirm personhood, they both gain reputation score. Otherwise, they lose points. Their previous contacts also lose or gain points depending on the result of the current handshake.

Registration

To sign up a new user:

- Make a small registration deposit (it is returned after personhood is confirmed — explained further).
- Take a selfie with the Upala app. This is your ID photo.

Random local handshake

Every user has a reputation balance. It can be positive or negative. Reputation may be earned only from other users. Users have to meet each other in person

to gain a reputation. Pairs of users are assigned pseudo-randomly by the system.

The protocol:

- Announce that you are available for a meeting. The system finds a random user nearby (IDs are not visible).
- Agree on system-proposed time (last chance to disagree with no punishment). After the time is chosen the system selects a random nearby location for a meeting. It is now your responsibility to show up.
- Meet the other person and confirm: there is a real person, the person is acting at their own will, the person’s ID photo is correct. Alternatively: take a picture with your phone and let face-recognition nodes deal with the matching.

Successful handshake brings points to both users. Rejection subtracts points from both.

Reputation propagation

Once the reputation score is received by a user the system selects a number of their contacts (pseudorandomly) and propagates reputation to them.

If the user gets an approval, those who approved them earlier receive points and those who rejected lose. And vice versa. If a user gets a rejection, those who approved them get a rejection, those who rejected get approval.

Accepting a handshake is like having a share of that user in your reputation portfolio. You don’t want a bot in your portfolio, because you’ll be receiving negative scores from it. But you’d better not reject real people because those who approve them will bring you negative scores.

The system stimulates to approve as many non-malicious people as possible. It also stimulates a user to register early (the earlier — the more reputation you can get).

Achieving a certain local reputation score a user may be considered a human locally. After surpassing this threshold the user may choose to continue handshakes and gain even more reputation, thus healing the system. The global reputation score is different and will be explained further.

[

1050×402

](https://miro.medium.com/max/1050/1*5N3BYpHWYMAzKfk7VMO8jg.jpeg)

Reputation scoring in random handshakes.

Example

Alice has already met two people and received 2 points from them.

Now she is on her third meeting but nobody is showing up. Probably her vis-a-vis has a reason not to show up, but maybe it is a bot. She rejects the handshake. She gets a rejection as well and loses 1 point. Her previous contacts (those who approved her) also lose points.

She keeps meeting real people. And eventually, she earns a good score.

With every successful handshake those who approved Alice receive points and those who rejected lose. Alice also gains points when a real human meets and rejects “her” bot.

But whenever she gets rejected, those who approved her lose and those who rejected (bots) gain points. And if Alice’s bot “meets” another cooperating bot, Alice loses points.

Humans vs bots

Location matters. Outlines of real cities and villages are represented in Upala. It is within these outlines that users earn local reputation.

Under normal conditions, all localities are densely “populated” with humans. They can easily outnumber and downvote a small botnet. But they cannot withstand a massive botnet attack. The bots could then prevent normal users to even meet each other. And real humans will only get negative scores.

There are two measures to withstand bots: the invitation deposit

and the ability to create a new cluster

in the area.

Invitation deposit

A candidate must pay a registration deposit when signing up. It is returned as soon as the candidate local person-hood is confirmed.

But just like there is a threshold to be considered a human within an area, there is a negative threshold to be considered a bot.

When a bot is confirmed its deposit is paid to those who had “met” and rejected it.

When FOAM is delivered invitation deposit can be (partly) replaced with dynamic location proof. Then it would make it harder for bot-farms to maintain their armies. They would have to move around a lot in the physical world.

Clusters

Invitation deposit makes it expensive for bots to bother human clusters. But what if bots populated an area first? Real people will then lose their stake when they try to register in their own city.

In situations like this humans can create a new cluster. Then pairs for handshakes are selected only within this cluster. Two instances of the locality emerge: humans and bots. And they don’t trust each other. This is what we want. We will then distinguish them from the outside (explained further).

There is a very high incentive for a human cluster to separate itself from bot-nets. But there is no incentive to create too many clusters in the same area. A single densely interconnected human cluster is easier to reach from the outside than a set of small ones. Two human clusters may choose to merge.

Sybil detection

The mechanism above allows only humans with a correct ID photo to enter the system. It confirms person-hood and location. But it cannot catch the same human entering multiple times. It cannot deal with uniqueness. Uniqueness is guarded by face recognition algorithm.

There are roughly 0,5% of people in the world that have an identical twin. When a user takes a selfie with the Upala app, the app guides them to take several photos from different angles. It does help to increase accuracy, but we can only try to predict how many people will be suspected as a twin. I would bet it won't be higher than 5%. But let's assume everybody has at least one twin somewhere (or the algorithm is giving us false positives giving the same result). It means that everyone should expect to do additional work sometime after they registered.

Uniqueness proof procedure

Once a new candidate is registered, the algorithm tries to find a twin for them. If there are no twins in the system nothing happens.

If there is a twin, they get notified. When the candidate proves global reputation, the candidate and the twin need to undergo uniqueness proof procedure.

Related twins or similar people from the same area may choose to appear together at a handshake. The third person confirms there are really two similar looking people.

Twins from different parts of the world need to appear in different places of the world within an interval too short to travel from one place to another.

There are penalties for refusing or delaying the procedure. So twins are incentivized to cooperate.

On privacy:

- Photos are stored as a search index only.
- There is no information attached to a photo except location (we will try to detach it too).
- The contacts are random so there is no way to deanonymize anyone by deanonymizing their friends.
- Additionally, for every uploaded photo, the system may generate a number of fakes.

Global reputation

There are local and global user reputation scores. The local is the user's reputation in a local community (city, village, etc). The global reputation represents how much of a user's local reputation is "visible" to the global community.

The visibility depends on how well the user's locality is connected to the global community.

Any two localities can have a connection. These connections are established by traveling users. Every connection has a bandwidth.

To calculate a user's global reputation their local reputation is "routed" through the best bandwidth to the global level. Reputation points are lost along the way on low bandwidth connections. Thus global reputation is always lower or equal to the local one.

Connections between cities

The network is hierarchical. Developers (global community) are at the top level.

There might be a different definition of what the global community is depending on the incentive model we chose. But for now, let's use "developers" as the global level for the sake of simplicity.

Next, there are major world cities. Or rather cities with the largest Upala communities. These cities are very well interconnected inside and a group of Upala developers has a lot of random connections in these cities too. The local and the global reputation here are almost equal.

Then there are big cities, small cities, villages, etc.

Every pair of cities (localities) has two connections in either way between them. Both connections have their own bandwidth. Bandwidths can be negative or positive.

Establishing a connection, connection weight

Connections can be established only by users. We call them ambassadors. Anybody can become an ambassador. A user just needs to go into another area and request a handshake.

A handshake establishes a connection. Ambassadors approve humans or reject handshakes when nobody shows up. A successful handshake adds weight

to the connection (this is not yet a bandwidth). Rejection subtracts from it. This way the connection from ambassador’s native location (A) to the remote one (B) is established and weighted.

Future ambassadors from the same location (A) also add or subtract weight to the connection (A to B). Previous ambassadors are punished or rewarded for their judgement. Majority punishes minority.

Ambassadors also receive scores (negative or positive) from their remote contacts through reputation propagation mechanism (as described above). But rewards and punishments are higher than within the local area.

Connection bandwidth

Every connection also has a bandwidth

which depends on city population, density of local connections, the distance between locations, ambassadors and locals reputation.

Two big cities need a lot of weight for their connections to get a good bandwidth. A small village can be visited by few ambassadors with a high score to get a good bandwidth to the big city.

Example

New York, Hartford, and Providence all have high internal interconnectedness and good bandwidth between them. New York has the best bandwidth to the global community.

Springfield is still in isolation. Moreover, Springfield had a massive bot attack recently so that it’s citizens decided to separate into their own cluster. Now there are two Springfields. One is inhabited with humans, the other with bots. But nobody from outside can tell, which is which.

Someone from Providence decides to visit Springfield. The ambassador performs 12 random handshakes and approves only those belonging to Springfield 1. A connection between Springfield 1 and Providence is established. But it’s weight is still too small to provide any sensible bandwidth.

More people start visiting Springfield. Some come from Providence others from Hartford. But they all now reject Springfield 1 citizens and approve Springfield 2 citizens during their random handshakes. The connection weight from Providence to Springfield 1 gradually turns into a negative number. Springfield 1 is considered a fraud. No special label needed — there is just no bandwidth.

Eventually, good bandwidth is established from Springfield 2 to both Providence and Hartford. Those who approved Springfield 1 get punished.

From local to global reputation score

In order to calculate the global score, the local score is router to the global level through connections between cities.

Reputation score points are lost on every connection. The lower the bandwidth the more points are lost. The system chooses the route with the best bandwidth. Connections with negative bandwidths are ignored.

Bot clusters cannot earn connectivity to anywhere.

Example

Alice has heard that her hometown Springfield has now a connection to the global community. And she would like to apply for a universal basic income experiment using her Upala ID.

In Springfield, she had 20 handshakes, 2 of which were rejected. Her local reputation score is 18.

Springfield had a lot of visits from Providence and Hartford. The bandwidths are 56 and 70% respectively. Good enough for the local score to be visible from the global level.

Providence to New York connection has 90% bandwidth, Hartford to New York — 89%, New York to global — 99%.

The best route is Springfield — Hartford — New York — Global. The bandwidth is 0,99 x 0,89

x 0,70 = 0,62.

Alice global score is 18 * 0,62 = 11,16. This is enough to be considered a human. Greetings, Alice!

Further research

- Incentives. Why would a user want to earn more reputation than it is needed to be considered a human?
- Incentives. How face-recognition nodes earn?
- Incentives. The protocol discriminates some disabled people (those who cannot walk, blind people).
- Incentives. Can we use uniqueness to create a blockchain?
- Attacks. Two real twins can create an army of clones.
- Attacks. A skillful makeup artist can try to create multiple personalities by cheating both humans (at handshakes) and face-recognition algorithm.
- Attacks. As a consequence of the previous, a skillful makeup artists can try to spam victims with their evil twins army.

Benefits

- Local communities can be initiated independently and included in the global community afterwards.
- An individual or an organization may choose to trust a representative of a local community before any connection to the global community is established. The community then starts to “exist” for this individual or organization.
- No complicated algorithms. We are on the path to a system where probability of person-hood can be calculated deterministically by anybody.
- A small number of interactions needed for a person to be confirmed as a human (random connections have more value than trusted connections).