

# Adaptive Neural Gossip Protocol (ANGP): Local Reputation via Self-Issued Certificates

Anton Toth

June 2026

## Abstract

ANGP is a fully decentralized protocol in which each peer maintains a local **Adaptive DAG** whose edge weights evolve according to a modified Hebbian rule based on delayed sensory feedback. Beyond collaborative learning, the protocol includes a reputation system built on **self-issued certificates**: each peer issues signed certificates for other peers based on cross-prediction error and then computes its local opinion of a peer as the **median** of the certificates it has itself issued for that peer. Because certificates are never forwarded and are stored only locally, an adversary cannot insert, delete, or modify an honest peer’s certificate set. This yields a provable **local non-manipulability** property that is independent of the number of Sybil or Byzantine identities the adversary controls.

The paper presents:

- the mathematical model of the Adaptive DAG and the Hebbian update rule;
- the secure gossip protocol (Ed25519, nonces, sequence numbers);
- the certificate-based reputation system and a formal proof of local non-manipulability;
- simulation results for majority attacks (10 honest vs 20 malicious) and for three mimicry strategies (Sybil mimic, Byzantine mimic, hybrid mimic) over 2000 steps.

Results show that under the evaluated scenarios, honest peers reach a local reputation of **1.0**, while all attacker categories are reduced to reputation **<0.02** after a finite number of steps, even when they mimic honest behavior for hundreds of steps before turning malicious.

## 1 Introduction

Distributed systems face two major challenges: *consensus* on a shared state and *reputation* to distinguish correct nodes from malicious ones. Classical solutions such as proof-of-work blockchains [1] or Byzantine fault-tolerant protocols [2] are either energy-intensive or require a known set of replicas. DAG-based ledgers [3] offer higher throughput but lack built-in learning. Holochain [4] provides an agent-centric model but does not adapt to changing behaviors.

Inspired by synaptic plasticity, ANGP merges a **plastic DAG** with a **reputation system based on direct observations**. Peers do not need global consensus. Each peer forms its own opinion about others exclusively from its own observations; that opinion is stored locally and cannot be forged by an adversary.

### Contributions:

- A formal definition of an Adaptive DAG with Hebbian update and attention propagation.
- A secure gossip protocol (Ed25519, nonces, sequence numbers).
- A reputation system where each peer issues certificates based on cross-prediction error and computes local reputation as the median of its own certificates.
- **Theorem 1 (Local Non-manipulability)**: An honest peer’s certificate set is write-only; the adversary cannot insert, delete, or modify it.
- Simulation results for majority attacks and three mimicry strategies, over 2000 steps.

## 2 Mathematical Model of the Adaptive DAG

### 2.1 Structure

An adaptive DAG is a tuple  $\mathcal{G} = (V, E, \mathbf{s}, \mathbf{w})$  where:

- $V$  – set of nodes (“neurons”), each with a timestamp.
- $E \subseteq V \times V$  – directed edges (parents  $\rightarrow$  child), acyclic.
- $\mathbf{s}_v \in \mathbb{R}^d$  – state vector of node  $v$  (continuous, changes over time).
- $w_{uv} \in [0, 1]$  – synaptic weight of edge ( $u \rightarrow v$ ).

### 2.2 Node Creation

At each time step, a peer receives a sensory input  $\mathbf{x}_t \in \mathbb{R}^d$ . It creates a new node  $v$  whose parents are the last  $k$  nodes (typically  $k = 5$ ). The initial state is:

$$\mathbf{s}_v^{(0)} = \tanh\left(0.7 \cdot \frac{\sum_{u \in \text{parents}(v)} w_{uv} \mathbf{s}_u}{\sum w_{uv}} + 0.3 \cdot \mathbf{x}_t\right)$$

The node’s identifier is  $\text{SHA512}/256(\text{len}(V) \parallel \mathbf{s}_v^{(0)} \parallel \mathbf{x}_t)$ .

### 2.3 Prediction and Error

The node predicts the next real output as the weighted average of its parents’ states:

$$\hat{\mathbf{y}}_v = \frac{\sum_u w_{uv} \mathbf{s}_u}{\sum w_{uv}}$$

After a short delay  $\delta t$ , the true output  $\mathbf{y}_t$  is observed. The local error is:

$$\epsilon_v = \|\hat{\mathbf{y}}_v - \mathbf{y}_t\|$$

**This error is the foundation of the reputation system.** It reflects how well a peer’s internal model can predict reality.

### 2.4 Modified Hebbian Update

For each edge ( $u \rightarrow v$ ):

$$\Delta w_{uv} = \alpha(\sigma_{uv}(1 - \epsilon_v)) - \beta w_{uv}(1 - \sigma_{uv})$$

where  $\sigma_{uv} = \frac{\mathbf{s}_u \cdot \mathbf{s}_v}{\|\mathbf{s}_u\| \|\mathbf{s}_v\|}$  is the cosine similarity. This is a variant of Oja’s rule [6].

- When  $\epsilon_v$  is small,  $(1 - \epsilon_v)$  is large  $\rightarrow$  weight increases (strengthening).
- When the error is large, the weight decreases (forgetting).
- The term  $\beta w_{uv}(1 - \sigma_{uv})$  provides normalization.

The node’s state is then updated:

$$\mathbf{s}_v \leftarrow \tanh(\mathbf{s}_v + \alpha(\mathbf{y}_t - \hat{\mathbf{y}}_v))$$

### 2.5 Distributed Attention

Each node has an attention score  $a_v$ . At each step:

- If  $\epsilon_v < 0.1$ ,  $a_v \leftarrow a_v + 0.05$ ;
- Otherwise,  $a_v \leftarrow 0.99 \cdot a_v$ .

If  $a_v > 0.5$ , attention propagates to its parents:  $a_u \leftarrow a_u + 0.1a_v$  (capped at 10.0). This mechanism makes nodes with small errors more “visible” in the network.

## 3 Secure Gossip Protocol

### 3.1 Cryptographic Identities

Each peer generates an Ed25519 key pair [5]. Its identifier is:

$$\text{ID} = \text{SHA512}/256(\text{public key})$$

To prevent massive Sybil attacks, an optional **Proof-of-Work** can be required upon registration:  $\text{SHA256}(\text{ID}||\text{nonce})$  must start with  $k$  zero bits (e.g.,  $k = 4$ ). This PoW is optional and does not affect the correctness of the reputation system.

### 3.2 Messages

All messages are signed and contain:

- `type: sync_weights, sync_request, sync_reply, pow_challenge, pow_response`
- `sender_id, seq_num, timestamp, nonce`
- `signature`

Messages larger than 1KB are compressed with zlib. Gossip occurs with probability 0.9, and the sender picks a random target (with reputation  $\geq 0.2$ ). It sends only the **last node** of its DAG (weights and a recent prediction). This reduces traffic compared to sending multiple nodes.

### 3.3 Cross-prediction and Certificates

When peer  $i$  receives a prediction from peer  $j$ , it computes the mean error  $\bar{\epsilon}_{ij}$  over the last 5 predictions received from  $j$ . Based on this error, it issues a **certificate**:

$$\text{Cert}_{ij} = \begin{cases} 1.0 & \text{if } \bar{\epsilon}_{ij} < 0.35 \quad (\text{positive}) \\ \max(0, 1 - 1.5\bar{\epsilon}_{ij}) & \text{if } \bar{\epsilon}_{ij} > 0.4 \quad (\text{negative}) \\ \text{no certificate} & \text{otherwise} \end{cases}$$

The certificate is a signed structure:  $\text{Cert}(i, j, \text{val}, \text{timestamp}, \text{nonce})$ .

**Crucial property:** certificates are **never forwarded**. Each peer stores them only locally in a set  $C_{ij}$  (all certificates that  $i$  has issued for  $j$ ).

### 3.4 Local Reputation

For a peer  $i$ , the local reputation of  $j$  is defined as:

$$R_i(j) = \text{median}(\{v \mid \text{Cert}(i, j, v) \text{ exists}\})$$

If no certificate exists,  $R_i(j) = 0.5$  (neutral). **Important:** The median is computed over *the entire history* of certificates issued by  $i$  for  $j$ . This choice gives higher weight to long-term consistent behavior; it also explains why a peer that was honest for a long time and then turns malicious retains a high reputation for a significant number of steps (the “history lock” effect). This effect is discussed in Section 7.2.

### 3.5 Robustness of the Median

The median is determined by the majority class of values. If the set contains many values close to 1.0 and few low values, the median will be close to 1.0, but not necessarily exactly 1.0 (e.g.,  $[1.0, 1.0, 0.85, 0.82, 0.80]$  has median 0.85). Hence we state: *The median remains dominated by the predominant certificate values; it does not flip unless the number of negative certificates exceeds half of the total.* This is sufficient for security: an honest peer will never be misclassified as malicious, and a malicious peer will eventually be recognized as such once it accumulates enough negative certificates.

### 3.6 Additional Security Mechanisms

- **Weight buffer:** each peer maintains a buffer of the last 5 weight updates from distinct senders and applies the median.
- **Consecutive negative penalty:** after 5 consecutive negative certificates, an additional penalty reduces reputation to 80% of its current value.
- **Inactivity:** if a peer has not sent any prediction for 60 seconds, its reputation is multiplied by 0.9.

## 4 Fundamental Security Theorem

[Local Non-manipulability of Reputation] For any honest peer  $i$  and any peer  $j$  (honest or malicious), the set of certificates  $C_{ij}$  stored locally by  $i$  is **write-only** by  $i$ . The adversary cannot:

- insert a fake certificate into  $C_{ij}$ ;
- delete or modify an existing certificate;
- influence the median calculation except indirectly through the observations that  $i$  makes about  $j$ .

Consequently, the reputation  $R_i(j)$  is determined exclusively by the *observations collected by  $i$  regarding  $j$*  (not by any external information). The adversary's power is limited to affecting the observed prediction errors  $\bar{e}_{ij}$ , but even then the median aggregation over a sliding window of 5 observations ensures that a single erroneous observation cannot flip the median if the majority are correct.

$C_{ij}$  is populated only when  $i$  calls the certificate issuance function. This function receives as argument the error  $\bar{e}_{ij}$  computed by  $i$  based on the predictions received. The adversary cannot force the function call (it does not control  $i$ 's execution thread) and cannot modify the stored value after issuance. Hence  $C_{ij}$  is a faithful record of  $i$ 's observation history. The median is a deterministic function of the values in  $C_{ij}$ .

[Independence of Compromised Fraction] The local reputation  $R_i(j)$  does **not** depend on the fraction of compromised peers in the network. Because certificates are never forwarded, the opinions of other peers (honest or malicious) never enter  $C_{ij}$ . Therefore the security of the reputation mechanism is *independent of whether the adversary controls 10%, 50%, or 90% of all peers*.

*Remark.* This corollary does **not** claim that ANGP solves the classical consensus problem; it only states that the local reputation mechanism itself is unaffected by the number of attackers. Global properties (e.g., convergence of all DAGs to the same model) may still depend on the connectivity of the honest subgraph.

## 5 Simulation Results

All simulations were implemented in Python with the following parameters:

- State dimension  $d = 4$
- $\alpha = 0.05$ ,  $\beta = 0.002$
- Positive threshold = 0.35, negative threshold = 0.4
- DAG limited to 200 nodes (bounded memory)
- Gossip sends only the last node
- Duration: 1000 or 2000 steps

### 5.1 Majority Attack (10 honest, 12 Sybil, 8 Byzantine)

*Observation:* Honest peers reach near-perfect local reputation; attackers are completely isolated in the honest peers' local views.

Step	Disagreement	Sybil rep	Byzantine rep	Honest rep
0	0.151	0.450	0.452	0.507
200	0.190	0.108	0.095	0.693
400	0.217	0.016	0.013	0.903
600	0.211	0.003	0.000	0.973
800	0.202	0.001	0.000	0.994
1000	0.208	0.000	0.000	<b>0.999</b>

Table 1: Majority attack results.

## 5.2 Mimicry Attacks (evaluated strategies)

We evaluated three mimicry strategies over 2000 steps:

- **Sybil mimic:** behaves honestly for 300 steps, then switches to a Sybil model (targeted cluster).
- **Byzantine mimic:** behaves honestly for 500 steps, then switches to Byzantine behaviour (random weights, random predictions).
- **Hybrid mimic:** behaves as a honest for 700 steps, then switches to Sybil-Byzantine behaviour.

Configuration: 10 honest, 6 static Sybil, 4 Sybil mimic, 2 Byzantine mimic, 2 hybrid mimic.

Results (selected steps):

Step	Honest rep	Static Sybil	Sybil mimic	Byzantine mimic	Hybrid mimic
0	0.507	0.452	0.451	0.453	0.453
200	0.796	0.055	0.929	0.770	0.853
400	0.982	0.000	0.923	1.000	0.948
600	1.000	0.000	0.324	1.000	1.000
800	1.000	0.000	0.051	0.688	1.000
1000	1.000	0.000	0.013	0.257	0.953
1200	1.000	0.000	0.003	0.039	0.826
1400	1.000	0.000	0.000	0.000	0.592
1600	1.000	0.000	0.000	0.000	0.180
1800	1.000	0.000	0.000	0.000	0.046
2000	1.000	0.000	0.000	0.000	0.011

Table 2: Mimicry attack results (2000 steps).

*Observations:*

- Sybil mimics have high reputation ( $\approx 0.93$ ) while they behave honestly, then gradually drop to 0 after 300–400 steps after the switch.
- Byzantine mimic remains at 1.0 for approximately 300 steps after switching (still being evaluated as honest), then drops to 0.339 by step 1000 and reaches 0.000 by step 1400.
- Hybrid mimic (honest until step 700) stays at high reputation (1.0) for about 300 steps after becoming Sybil-Byzantine, then slowly declines, reaching 0.011 by step 2000.

These results show that the **evaluated mimicry strategies** are eventually detected and isolated, although the median-over-entire-history introduces a delay (the history lock effect). This delay is a trade-off: it prevents an attacker from flipping a good peer’s reputation with a few bad observations, but it also allows a previously honest peer that turns malicious to retain high reputation for a number of steps proportional to its prior good behaviour.

System	Global consensus?	Local reputation integrity?
Bitcoin [1]	Yes (PoW chain)	No
IOTA [3]	Yes (tip selection)	No
Holochain [4]	Partial (agent-centric)	No
<b>ANGP</b>	<b>No</b>	<b>Yes</b> (median of local certificates)

Table 3: Comparison with existing systems.

## 6 Comparison with Other Systems

The following table compares ANGP with existing systems on two orthogonal dimensions: *global consensus* and *local reputation integrity*. This avoids comparing apples to oranges.

ANGP does not aim to replace global consensus protocols. Instead, it offers a new primitive: **non-manipulable local reputation** that can be used as a building block in decentralised learning, federated systems, and even blockchain ecosystems, or edge networks where global agreement is not required.

## 7 Discussion and Future Work

### 7.1 Strengths

- **Local non-manipulability** is formally proven and independent of the number of attackers.
- **No certificate forwarding** eliminates trust poisoning and gossip poisoning attacks.
- **Median aggregation** provides robustness to temporary observation errors.
- Simulations confirm that the protocol works under the evaluated adversarial scenarios.

### 7.2 Limitations

- **History lock effect:** Because the median is computed over the *entire certificate history*, a peer that was honest for a long period and then turns malicious will keep a high reputation for many steps. In our experiments, a Byzantine mimic remained at reputation 1.0 for about 300 steps after switching, and a hybrid mimic (Sybil then Byzantine) remained at 1.0 for about 300 steps after its second switch. This behaviour is intentional (it prevents an attacker from flipping a good peer’s reputation with a few bad observations), but it should be understood as a trade-off: the system reacts slowly to late-stage corruption.
- The protocol does **not** provide a global consensus on reputation; each peer has its own private view.
- The current simulation assumes a fully connected gossip overlay; large-scale deployment would require probabilistic gossip or a DHT.
- The convergence of the Adaptive DAG under adversarial conditions is not formally proven (only empirically shown).
- The mimicry experiments cover three specific strategies; other adaptive strategies may behave differently.

### 7.3 Relation Between Adaptive DAG and Reputation System

The two components (Adaptive DAG and reputation system) are largely independent. The reputation system does not influence the Hebbian updates; it only observes the predictions produced by the DAG. Thus the contribution of the paper is twofold: a plastic DAG for decentralised learning, and a novel local reputation mechanism. Their combination is natural (the DAG provides the predictions that feed the reputation system), but no emergent property requires them to be tightly coupled. This is acceptable as long as stated clearly.

## 7.4 Future Work

- Formal convergence analysis of the Hebbian DAG in the presence of Byzantine nodes.
- Scalability evaluation with probabilistic gossip (fanout constant) and DHT-based certificate storage.
- Real-network implementation (UDP, NAT traversal, churn) and benchmarking.
- Integration with federated learning frameworks to provide reputation for participant selection.

## 8 Conclusion

We have presented ANGP, a protocol that achieves **local reputation non-manipulability** through a novel combination of self-issued certificates and median aggregation. Each peer independently evaluates others based solely on its own observations; certificates are never forwarded, making it impossible for an adversary to alter an honest peer’s opinion. The protocol also includes an Adaptive DAG with Hebbian learning, Ed25519 signatures, and optional proof-of-work for Sybil admission control.

Simulations confirm that under the evaluated scenarios (majority attacks and three mimicry strategies), honest peers achieve local reputation 1.0 while all attackers drop below 0.02. The architecture appears suitable for deployment in sensor networks, edge computing environments, blockchain/DAG ecosystems and federated learning systems where global consensus is not required.

## Acknowledgments

The authors thank the open-source community for cryptographic libraries and simulation tools. The complete Python implementation is available as supplementary material.

## References

- [1] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” 2008.
- [2] M. Castro and B. Liskov, “Practical Byzantine Fault Tolerance,” in *OSDI*, 1999.
- [3] S. Popov, “The Tangle,” 2016.
- [4] Holochain Team, “Holochain: A Framework for Distributed Applications,” 2019.
- [5] D. Bernstein, “Curve25519: new Diffie-Hellman speed records,” 2006.
- [6] E. Oja, “Simplified neuron model as a principal component analyzer,” *J. Math. Biol.*, 1982.