Proposition 6.1. *Let G_i be a subjective work graph with attacker j launching a tree Sybil attack σ_S. If R_i satisfies convergence of serial reports, parallel-report bound and path-responsiveness, then σ_S cannot be strongly beneficial in terms of reputation.*

Proof. We begin by only examining the first *layer* of the tree, given by the Sybils $\{s_{11}, \ldots, s_{1m_1}\}$. The given tree confined to this layer is a simple parallel Sybil attack and we know by the parallel-report bound that it must hold

$$\sum_{k=1}^{m_1} R_i(G_i', s_{1k}) \leq R_i(G_i, s_{11}).$$

The second layer of the tree attack can be interpreted as a number of Sybil attacks perpetrated by m_1 attackers. Again, we can apply the parallel-report bound and find that the profit of the second layer of each branch will be bounded by the profit of a serial attack with two Sybils . The serial-report bound ensures that the profit of the second layer will be bounded by the profit of the first layer, i.e.,

$$\sum_{k=1}^{m_2} R_i(G_i', s_{2k}) \leq \sum_{k=1}^{m_1} R_i(G_i', s_{1k}) \leq R_i(G_i, s_{11}).$$

We can continue this reasoning inductively and find that the profit of a tree Sybil attack with infinite layers, each containing finitely many Sybils, is bounded by the profit of an infinite serial Sybil attack. By convergence of serial reports, this profit will be finite. The only other way an attacker might attempt to obtain infinite reputation from a tree Sybil attack is by scaling one or more layers of the tree. Due to parallel-report responsiveness the profit of this attack is still finite and the attack cannot be strongly beneficial. □

If above weak representativeness is satisfied, the result holds in terms of work as well. Next, we introduce one additional property that bounds the profit of any passive Sybil attack by the profit of a tree attack multiplied by some constant. We call this property *multiple-path response bound.* We introduce the following operation to perform on a subjective work graph.

Let G_i be a subjective work graph with $k \in V_i$ such that there exist N directed paths $(P_n)_{n \leq N}$ connecting k to i. Now, define G_i' as an altered version of the subjective work graph of i, whereby the agent k is *split* into several agents $k_1, \ldots, k_N$, where every k_l $(l \leq N)$ is connected to i by exactly one path. G_i' is created by splitting k into as many nodes as there are paths connecting it to i. We begin with k_1 and remove all agents and edges that are part of any of the paths $P_2, \ldots, P_N$ while keeping all which are part of P_1. We now relabel k (as the end-point of P_1), k_1. Next, we add path P_2 to the graph. Any agent j (or edge e) in P_2 that is also part of P_1, is now duplicated into j_1 and j_2 such that $j_1 \in P_1$ and $j_2 \in P_2$, i.e., ($e_1 \in P_1$ and $e_2 \in P_2$). We continue this for all paths $P_1, \ldots, P_N$ and obtain G_i'.

Definition 6.6 (Multiple-Path Response Bound)**.** Let G_i be a subjective work graph with $k \in V_i$ such that there exist N paths $(P_n)_{n \leq N}$ connecting i and k and let G_i' be the subjective work graph obtained by performing the operation above on G_i. We say that the reputation mechanism R satisfies the multiple-path response bound if it holds

$$R_i(G_i, k) \leq \sum_{n=1}^{N} R_i(G_i', k_n).$$

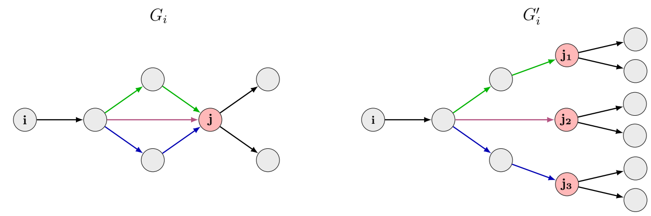

Figure 4: Example of multiple-path response bound applied to j from the view of i. j is connected to i via three paths and therefore split into three agents j_1, j_2, j_3.

Figure 4 and Figure 5 show examples of the multiple-path response bound applied to j in two different graph topologies. In Figure 4 there are three paths connecting i to j and applying the multiple-path response bound to j yields three agents, each connected to i via one path. In Figure 5 the graph contains a cycle. We interpret a cycle as infinite paths connecting i and j. Applying the multiple-path response bound graph transformation produces an infinite sequence of agents $(j_n)_{n \in \mathbb{N}}$ in G_i' where each j_n is connected to i via one path. This demonstrates that the given graph transformation can be applied to more sophisticated topologies as well.

As in our earlier restrictions on reputation mechanisms we claim that the multiple-path response bound property is not an arbitrary invention by us, but is in fact satisfied by all reputation mechanisms defined in [15] and plenty of the existing reputation mechanisms such as PageRank, Maxflow and Netflow [7]. Instead, we elaborate on the intuition behind this definition.

For a reputation mechanism to determine the cooperativeness of an agent, it needs to evaluate this agent's indirect contributions and consumption to/from i. i evaluates j by the incoming edges from i, whereby each path connecting j to i can be considered an indirect contribution and therefore, should influence the reputation score of j in i's subjective work graph. However, it is crucial for Sybil resistance that the effect of an additional path in the network should not exceed the effect that this additional path would have on $R_i(G_i, j)$ if it were the only path, as we do not want reputation to be gained disproportionately to the amount of work performed.

Next, we introduce transitive trust as a requirement to finally achieve Sybil-proofness.

Definition 6.7 (Transitive Trust)**.** Let G_i be a subjective work graph containing a directed path $P = (i, j_1, \ldots, j_n, j)$ of arbitrary length with strictly positive edge weights. We say R satisfies **transitive trust** if it holds

$$R_i(G_i, j_1), R_{j_1}(G_{j_1}, j_2), \ldots, R_{j_n}(G_{j_n}, j) > 0 \Rightarrow \; R_i(G_i, j).$$

We say that R satisfies **bounded transitive trust** if it also holds

$$R_i(G_i, j) \leq \min\left\{R_i(G_i, j_1), R_{j_1}(G_{j_1}, j_2), \ldots, R_{j_n}(G_{j_n}, j)\right\}.$$

If there are several (N) paths $(P_d)_{d \leq N}$ of lengths n_d given by $(i, j_1^d, \ldots, j_{n_d}^d, j)$ connecting i and j, then $R_i(G_i, j)$ must be bounded by the sum of the minimums given above (for each path).

$$R_i(G_i, j) \leq \sum_{d=1}^{N} \min\{R_{j_l^d}(G_{j_l^d}, j_{l+1}^d) \mid j_l^d \in P_d\}.$$

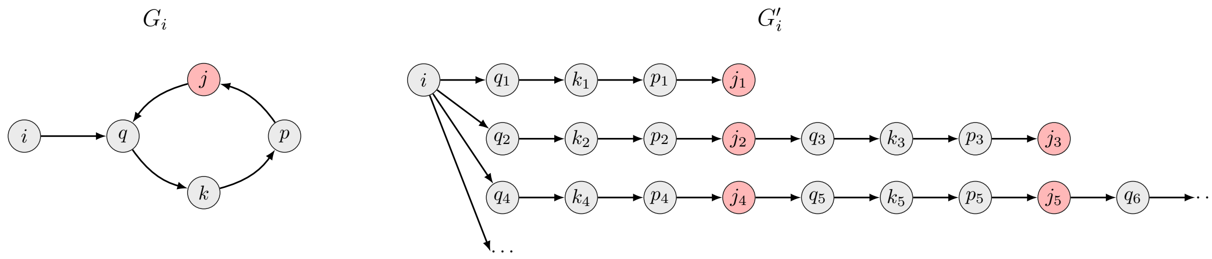

Figure 5: Example of multiple-path response bound applied to a graph with cycle. A loop implies infinite paths and therefore j is split into an infinite sequence $(j_n)_{n\in\mathbb{N}}$.

Definition 6.7 states that if i assigns j a reputation score greater than zero, and j assigns another agent k some reputation score greater than zero as well, then i assigns k some reputation score greater than zero as well. Bounded transitive trust implies that the reputation score k has with i must be bounded from above by the minimum of the reputation score i assigns j and the score j assigns k. It is a common property of reputation mechanisms, such as BarterCast and Netflow. Using these requirements we now prove Sybil-resistance of reputation mechanisms, as promised.

Lemma 6.2. *Let R_i satisfy the multiple-path response bound and bounded transitive trust. If j launches a passive Sybil attack σ_S. Then the profit $\omega^+_{rep}(\sigma_S)$ is bounded by the profit $\omega^+_{rep}(\sigma_{\tilde{S}})$ of a passive tree Sybil attack $\sigma_{\tilde{S}}$ multiplied by a constant $c < \infty$.*

Proof. First, assume there exists one directed path P connecting i to j. We can apply the multiple-path response bound to the Sybil region S yielding a new Sybil region $\tilde{S}$ in which every Sybil is connected to j via a single path. $\sigma_{\tilde{S}}$ is therefore a tree Sybil attack and the profit of the attack σ_S is bounded by the profit of $\sigma_{\tilde{S}}$. If there are finitely many directed paths $(P_n)_{n\leq N}$ connecting i and j then we can apply the multiple-path response bound to j, and obtain a subjective work graph G_i' with $j_1, \ldots, j_N$, each connected to i via a single path and committing the same Sybil attack. We obtain N equivalent Sybil attacks $\sigma_{S_1}, \ldots, \sigma_{S_m}$ and can apply the same procedure as we did in the case of a single path connecting i and j yielding N tree Sybil attacks $\sigma_{\tilde{S}_1}, \ldots, \sigma_{\tilde{S}_N}$. We can then infer the inequality $\omega^+_{rep}(\sigma_S) \leq N \cdot \omega^+_{rep}(\sigma_{\tilde{S}})$, where $\omega^+_{rep}(\sigma_{\tilde{S}})$ is the largest profit of the N tree Sybil attacks. Lastly, if there are infinite paths connecting i and j then the graph must contain a cycle and we can infer with the transitive trust property that $\sum_{n=1}^{\infty} R_i(G_i'', j_n) \leq \sum_{k\in N'(i)} R_i(G_i', k)$, where N_i' is the neighbourhood of i in V_i'. Hence, we conclude analogously to the case of finite paths $\omega^+_{rep}(\sigma_S) \leq \omega^+_{rep}(\sigma_{\tilde{S}}) \cdot \sum_{k\in N'(i)} R_i(G_i', k)$. □

Combining the results from Proposition 6.1 and Lemma 6.2, we argue that the profit of any arbitrary passive Sybil attack is finite.

Theorem 6.1. *Any reputation mechanism R satisfying path-responsiveness, multiple-path response bound, convergence of serial reports, the parallel-report bound, as well as bounded transitive trust is resistant to strongly beneficial passive Sybil attacks in terms of reputation.*

The proof follows directly from the proofs to Proposition 6.1 and Lemma 6.2. Using the properties of multiple-path response bound and bounded transitive trust, we obtain the following corollary.

Corollary 6.1. *Any reputation mechanism R satisfying path-responsiveness, multiple-path response bound, convergence of serial reports, the parallel-report bound, as well as bounded transitive trust is resistant to strongly beneficial active Sybil attacks in terms of reputation.*

Proof. Let σ_S be an active Sybil attack, then for any $c > 0$ we know there must be a bounded number of attack edges with edge weights larger than c. Therefore, we can apply the multiple-path response bound to each Sybil that is connected to an attack edge and obtain a finite number of passive Sybil attacks. The rest follows analogously to Theorem 6.1. □

If, in addition to the requirements stated above, R is weakly representative, then it is also resistant to strongly beneficial Sybil attacks in terms of work, as discussed in Section 4.

7 CONCLUSION

In this paper we have studied the Sybil-proofness of reputation mechanisms in multi-agent systems. We introduced rigorous metrics for the benefit of Sybil attacks, determined by the ratio of their cost and profit. While the goal was to bound the effect of Sybil attacks in terms of the work contributed and consumed by the attacker, these values were impractical to compute. We therefore introduced a pair of proxies, given in terms of the reputation obtained through the attack and through honest work. We introduced a requirement known as representativeness that ensures an equivalence between these two ratios. Using these metrics we revisited the impossibility result of Seuken and Parkes [11], pointing out an error which we attribute to ambiguity in their definitions of the attack benefit. We expanded on this result with two requirements we called parallel- and serial-report responsiveness and inverted the intuition behind these two requirements to obtain Sybil-resistance to parallel and serial attacks. We extrapolated our results to a combination of these two, known as tree attacks. Introducing a further requirement known as multiple-path response bound we achieve resistance to arbitrary attacks. Our bounds may seem loose, but a finite benefit ensures an attacker's contributions remain proportionate to its consumption, which is sufficient to protect the longevity of any multi-agent work system. In future work one may consider bounding the benefit of any attack by a fixed and finite value $c > 0$.

REFERENCES

[1] Ashwin R Bharambe, Cormac Herley, and Venkata N Padmanabhan. 2006. Analyzing and improving a bittorrent networks performance mechanisms. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*. IEEE, IEEE, 1–12.

[2] Alice Cheng and Eric Friedman. 2005. Sybilproof reputation mechanisms. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*. 128–132.

[3] Alice Cheng and Eric Friedman. 2006. Manipulability of PageRank under sybil strategies. In *First Workshop on the Economics of Networked Systems*. ACM.

[4] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. 2003. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*. 640–651.

[5] Brandon K Liu, David C Parkes, and Sven Seuken. 2016. Personalized hitting time for informative trust mechanisms despite sybils. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1124–1132.

[6] Michel Meulpolder, Johan A Pouwelse, Dick HJ Epema, and Henk J Sips. 2009. Bartercast: A practical approach to prevent lazy freeriding in p2p networks. In *2009 IEEE International Symposium on Parallel & Distributed Processing*. IEEE, 1–8.

[7] Pim Otte. 2016. *Sybil-resistant trust mechanisms in distributed systems*. Master's thesis. TU Delft.

[8] Pim Otte, Martijn de Vos, and Johan Pouwelse. 2020. TrustChain: A Sybil-resistant scalable blockchain. *Future Generation Computer Systems* 107 (2020), 770–780.

[9] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The pagerank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[10] Sven Seuken, Michel Meulpolder, DC Parkes, JA Pouwelse, J Tang, and DHJ Epema. 2014. Work accounting mechanisms: Theory and practice. In *Working Paper. Department of Informatics*. University of Zurich.

[11] Sven Seuken and David C Parkes. 2011. On the Sybil-proofness of accounting mechanisms. In *Workshop on the Economics of Networks, Systems, and Computation (NetEcon'11)*.

[12] Sven Seuken and David C Parkes. 2014. Sybil-proof accounting mechanisms with transitive trust. *Proceedings of the International Foundation for Autonomous Agents and Multiagent Systems* (2014).

[13] Sven Seuken, Jie Tang, and David C Parkes. 2010. Accounting mechanisms for distributed work systems. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

[14] Alexander Stannat and Johan Pouwelse. 2019. A Random Walk based Trust Ranking in Distributed Systems. *arXiv preprint arXiv:1903.05900* (2019).

[15] Jie Tang, Sven Seuken, and David C Parkes. 2010. Hybrid transitive trust mechanisms. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems.

[16] Atsushi Yamamoto, Daisuke Asahara, Tomoko Itao, Satoshi Tanaka, and Tatsuya Suda. 2004. Distributed pagerank: a distributed reputation model for open peer-to-peer network. In *2004 International Symposium on Applications and the Internet Workshops. 2004 Workshops*. IEEE, 389–394.