Trust Is Risk:
A Decentralized Financial Trust Platform

Orfeas Stefanos Thyfronitis Litos[1] and Dionysis Zindros[2,⋆]

[1] National Technical University of Athens
[2] National and Kapodistrian University of Athens
olitos@corelab.ntua.gr, dionyziz@di.uoa.gr

Abstract. Reputation in centralized systems uses stars and review-based trust. Such systems require manual intervention and secrecy to avoid manipulation. In autonomous and open source decentralized systems this luxury is not available. Previous peer-to-peer reputation systems do not allow for financial arguments pertaining to reputation. We propose a concrete Sybil-resilient decentralized reputation system where direct trust is defined as lines-of-credit using bitcoin's 1-of-2 multisig. We introduce a new model for bitcoin wallets in which user coins are split among trusted associates. Indirect trust is subsequently defined transitively. This enables formal game theoretic arguments pertaining to risk analysis. We prove that risk and max flows are equivalent in our model. Our system allows for concrete financial decisions on the monetary amount a pseudonymous party can be trusted with. Through algorithmic trust redistribution, the risk incurred from making a purchase from a pseudonymous party in this manner remains invariant.

Keywords: decentralized · trust · web-of-trust · bitcoin · multisig · line-of-credit · trust-as-risk · flow · reputation

1 Introduction

Modern online marketplaces can be roughly categorized as centralized and decentralized. Two major examples of each category are ebay and OpenBazaar. The common denominator of established online marketplaces is that the reputation of each vendor and client is either expressed in the form of stars and user-generated reviews that are viewable by the whole network, or not expressed at all inside the marketplace and instead is entirely built on word-of-mouth or other out-of-band means.

Our goal is to create a decentralized marketplace where the trust each user gives to the rest of the users is quantifiable, measurable and expressable in monetary terms. The central concept used throughout this paper is that trust is equivalent to risk, or the proposition that *Alice*'s *trust* to

⋆ Research supported by ERC project CODAMODA, project #259152

another user *Bob* is defined to be the *maximum sum of money* that *Alice* can lose when *Bob* is free to choose any strategy he wants. To flesh out this concept, we will use *lines of credit* as proposed by Washington Sanchez [1], not to be confused with the synonymous financial product. Joining the network will be done by explicitly entrusting a certain amount of money to another user, say *Bob*. If *Bob* has already entrusted an amount of money to a third user, *Charlie*, then we indirectly trust *Charlie* since if the latter wished to play unfairly, he could have already stolen the money entrusted to him by *Bob*. Thus we can engage in economic interaction with *Charlie*. The currency used is Bitcoin [2].

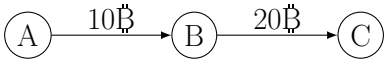

Fig.1: A trusts C 10฿

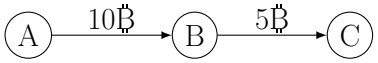

Fig.2: A trusts C 5฿

We thus propose a new kind of wallet where coins are not stored locally, but are placed in 1-of-2 multisigs, a bitcoin construction that permits any one of two pre-designated users to spend the coins contained therein [3]. We will use the notation $1/\{Alice, Bob\}$ to represent a 1-of-2 multisig that can be spent by either *Alice* or *Bob*.

Our approach changes the user experience in a subtle but drastic way. A user no more has to base her trust towards a store on stars, ratings or other dubious and non-quantifiable trust metrics. She can simply consult her wallet to decide whether the store is trustworthy and, if so, up to what value. This system works as follows: Initially *Alice* migrates her funds from P2PKH addresses in the UTXO [3] to 1-of-2 multisig addresses shared with friends she comfortably trusts. We call this direct trust. Our system is agnostic to the means players use to determine who is trustworthy for these direct 1-of-2 deposits.

Suppose that *Alice* is viewing the item listings of vendor *Charlie*. Instead of *Charlie*'s stars, *Alice* will see a positive value that is calculated by her wallet and represents the maximum monetary value that *Alice* can safely use to complete a purchase from *Charlie*. We examine exactly how this value is calculated in Trust Flow theorem (2). This monetary value reported by our system maintains the desired security property that, if *Alice* makes this purchase, then she is exposed to no more risk than she was willing to expose herself towards her friends. We prove this result in the Risk Invariance theorem (3). Obviously it will not be safe for *Alice* to buy anything from *Charlie* or any other vendor if she has entrusted no value to any other player.

We see that in TrustIsRisk the money is not invested at the time of the purchase and directly to the vendor, but at an earlier point in time and only to parties that are trustworthy for out-of-band reasons. The fact that this system can function in a completely decentralized fashion will become clear in the following sections. We prove this result in the Sybil Resilience theorem (5).

There are several incentives for a user to join this network. First of all, she can have access to a store that is otherwise inaccessible. Moreover, two friends can formalize their mutual trust by entrusting the same amount to each other. A large company that casually subcontracts other companies to complete various tasks can express its trust towards them using this method. A government can choose to entrust its citizens with money and confront them using a corresponding legal arsenal if they make irresponsible use of this trust. A bank can provide loans as outgoing and manage savings as incoming trust and thus has a unique opportunity of expressing in a formal and absolute way its credence by publishing its incoming and outgoing trust. Last but not least, the network can be viewed as a possible field for investment and speculation since it constitutes a completely new area for financial activity.

It is worth noting that the same physical person can maintain multiple pseudonymous identities in the same trust network and that multiple independent trust networks for different purposes can coexist. On the other hand, the same pseudonymous identity can be used to establish trust in different contexts.

2 The Trust Graph

We now engage in the formal description of the proposed system, accompanied by helpful examples.

Definition 1 (Graph). *TrustIsRisk is represented by a sequence of directed weighted graphs* $(\mathcal{G}_j)$ *where* $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j), j \in \mathbb{N}$. *Also, since the graphs are weighted, there exists a sequence of weight functions* (c_j) *with* $c_j : \mathcal{E}_j \to \mathbb{R}^+$.

The nodes represent the players, the edges represent the existing direct trusts and the weights represent the amount of value attached to the corresponding direct trust. As we will see, the game evolves in turns. The subscript of the graph represents the corresponding turn.

Definition 2 (Players). *The set* $\mathcal{V}_j = \mathcal{V}(\mathcal{G}_j)$ *is the set of all players in the network, otherwise understood as the set of all pseudonymous identities.*

Each node has a corresponding non-negative number that represents its capital. A node's capital is the total value that the node possesses exclusively and nobody else can spend.

Definition 3 (Capital). *The initial capital of A, $Cap_{A,0}$, is defined as the total value that belongs to A and exists in P2PKH in the UTXO at the beginning of the game.*

The capital of A is subsequently modified only during her turns, according to her actions. A rational player would like to maximize her capital in the long term. We also define a player's assets:

Definition 4 (Assets). *Sum of A's capital and outgoing trust.*

$$As_{A,j} = Cap_{A,j} + out_{A,j} \tag{1}$$

We consider the outgoing trust of a player as part of her assets. The formal definition of direct trust follows:

Definition 5 (Direct Trust). *Direct trust from A to B at the end of turn j, $DTr_{A\to B,j}$, is defined as the total amount of value that exists in $1/\{A,B\}$ multisigs in the UTXO in the end of turn j, where the money is deposited by A.*

$$DTr_{A\to B,j} = \begin{cases} c_j\left(A,B\right), & if\ (A,B) \in \mathcal{E}_j \\ 0, & else \end{cases} \tag{2}$$

This definition agrees with the title of this paper and coincides with the intuition and experimental results of [4] that the trust *Alice* shows to *Bob* in real-world social networks corresponds with the extent of danger in which *Alice* is ready to expose herself to in order to help *Bob*. An example graph with its corresponding transactions in the UTXO can be seen below.

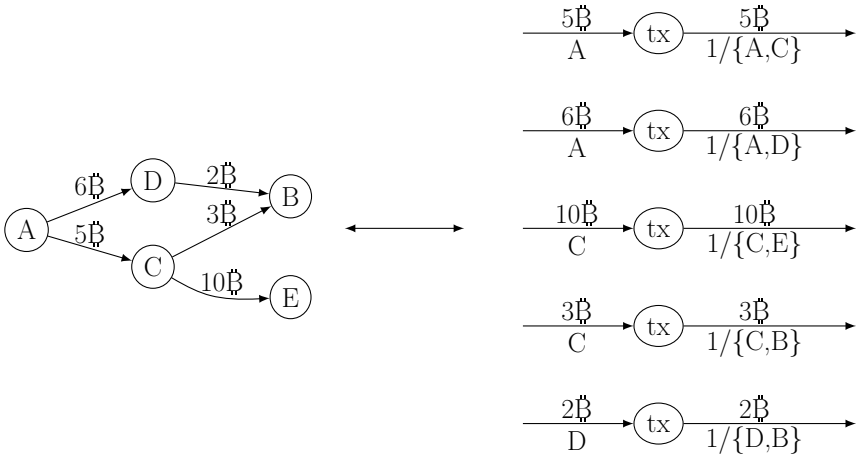

Fig.3: TrustIsRisk Game Graph and Equivalent Bitcoin UTXO

Any algorithm that has access to the graph $\mathcal{G}_j$ has implicitly access to all direct trusts of this graph. We use the notation $N^+(A)$ to refer to the nodes directly trusted by A and $N^-(A)$ for the nodes that directly trust A. We also use the notation $in_{A,j}, out_{A,j}$ to refer to the total incoming and outgoing direct trust respectively. For a reference of common definitions, see the Appendix.

3 Evolution of Trust

Definition 6 (Turns). *The game we are describing is turn-based. In each turn j exactly one player $A \in \mathcal{V}, A = Player(j)$, chooses one or more actions from the following two kinds:*
***Steal**(y_B, B): Steal value y_B from $B \in N^-(A)_{j-1}$, where $0 \leq y_B \leq DTr_{B\to A,j-1}$. Then:*

$$DTr_{B\to A,j} = DTr_{B\to A,j-1} - y_B$$

***Add**(y_B, B): Add value y_B to $B \in \mathcal{V}$, where $-DTr_{A\to B,j-1} \leq y_B$. Then:*

$$DTr_{A\to B,j} = DTr_{A\to B,j-1} + y_B$$

When $y_B < 0$, we say that A reduces her trust to B by $-y_B$. When $y_B > 0$, we say that A increases her trust to B by y_B. If $DTr_{A\to B,j-1} = 0$, then we say that A starts directly trusting B. If player A chooses no action in her turn, we say that she passes her turn. Also, let Y_{st}, Y_{add} be the total

value to be stolen and added respectively by A in her turn, j. For a turn to be feasible, it must hold that

$$Y_{add} - Y_{st} \leq Cap_{A,j-1} \ . \tag{3}$$

The capital is updated in every turn:

$$Cap_{A,j} = Cap_{A,j-1} + Y_{st} - Y_{add} \ .$$

Moreover, player A is not allowed to choose two actions of the same kind against the same player in one turn. The set of actions the player makes in turn j is denoted by $Turn_j$. The new graph that emerges by applying the actions on $\mathcal{G}_{j-1}$ is $\mathcal{G}_j$.

We use $prev\,(j)$ and $next\,(j)$ to denote the previous and next turn respectively played by $Player(j)$. A formal definition can be found in the Appendix.

Definition 7 (Damage). *Let j be a turn such that $Player\,(j) = A$.*

$$Damage_{A,j} = out_{A,prev(j)} - out_{A,j-1} \tag{4}$$

We say that A has been stolen value $Damage_{A,j}$ between $prev\,(j)$ and j. We omit turn subscripts if they are implied from the context.

Definition 8 (History). *We define History, $\mathcal{H} = (\mathcal{H}_j)$, as the sequence of all tuples containing the sets of actions and the corresponding player.*

$$\mathcal{H}_j = (Player\,(j)\,, Turn_j) \tag{5}$$

Knowledge of the initial graph $\mathcal{G}_0$ and the history amount to full comprehension of the evolution of the game. Building on the example of figure 3, we can see the resulting graph when D plays

$$Turn_1 = \{Steal\,(1, A)\,, Add\,(4, C)\} \ . \tag{6}$$

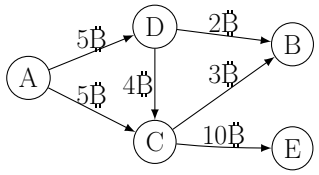

Fig.4: Game Graph after $Turn_1$ (6) on the Graph of figure 3

In the form presented here, TrustIsRisk is controlled by an algorithm that chooses a player, receives the turn that this player wishes to play and, if this turn is valid, executes it. These steps are repeated indefinitely. We assume players are chosen in a way that, after her turn, a player will eventually play again later.

```
TrustIsRisk Game
j = 0
while (True)
  j += 1
  v ←$ V_j
  Turn = vStrategy(G_0, v, (H)_{1...j-1})
  (G_j, Cap_{v,j}, H_j) = executeTurn(G_{j-1}, v, Cap_{v,j-1}, Turn)
```

`executeTurn()` checks the validity of `Turn` and substitutes it with an empty turn if invalid. Subsequently, it creates the new graph $\mathcal{G}_j$ and updates the history accordingly. For the routine code, see the Appendix.

4 Trust Transitivity

In this section we define some strategies and show the corresponding algorithms. Then we define the Transitive Game that represents the worst-case scenario for an honest player when another player decides to depart from the network with her money and all the money entrusted to her.

Definition 9 (Idle Strategy). *A player A is said to follow the idle strategy if she passes in her turn.*

```
Idle Strategy
Input : initial graph G_0, player A, history (H)_{1...j-1}
Output : Turn_j
idleStrategy(G_0, A, H) :
  return(∅)
```

The inputs and outputs are identical to those of `idleStrategy()` for the rest of the strategies, thus we avoid repeating them.

Definition 10 (Evil Strategy). *A player A is said to follow the evil strategy if she steals all incoming direct trust and nullifies her outgoing direct trust in her turn.*

```
evilStrategy(G_0, A, H) :
```

```
  Steals = ⋃_{v∈N⁻(A)_{j−1}} {Steal(DTr_{v→A,j−1}, v)}
  Adds = ⋃_{v∈N⁺(A)_{j−1}} {Add(−DTr_{A→v,j−1}, v)}
  Turn_j = Steals ∪ Adds
  return(Turn_j)
```

Definition 11 (Conservative Strategy). *Player A is said to follow the conservative strategy if she replenishes the value she lost since the previous turn, $Damage_A$, by stealing from others that trust her as much as she can up to $Damage_A$ and she takes no other action.*

```
consStrategy(G_0, A, H) :
  Damage = out_{A,prev(j)} - out_{A,j−1}
  if (Damage > 0)
    if (Damage >= in_{A,j−1})
      Turn_j = ⋃_{v∈N⁻(A)_{j−1}} {Steal(DTr_{v→A,j−1}, v)}
    else
      y = SelectSteal(G_j, A, Damage) #y = {y_v : v ∈ N⁻(A)_{j−1}}
      Turn_j = ⋃_{v∈N⁻(A)_{j−1}} {Steal(y_v, v)}
  else
    Turn_j = ∅
  return(Turn_j)
```

`SelectSteal()` returns y_v with $v \in N^-(A)_{j-1}$ such that

$$\sum_{v \in N^-(A)_{j-1}} y_v = Damage_{A,j} \wedge \forall v \in N^-(A)_{j-1}, y_v \leq DTr_{v \to A, j-1} \ . \quad (7)$$

Each conservative player can arbitrarily define how `SelectSteal()` distributes the $Steal()$ actions each time she calls the function, as long as (7) is respected.

As we can see, the definition covers a multitude of options for the conservative player, since in case $0 < Damage_{A,j} < in_{A,j-1}$ she can choose to distribute the $Steal()$ actions in any way she chooses.

The rationale behind this strategy arises from a real-world common situation. Suppose there are a client, an intermediary and a producer. The client entrusts some value to the intermediary so that the latter can buy the desired product from the producer and deliver it to the client. The

intermediary in turn entrusts an equal value to the producer, who needs the value upfront to be able to complete the production process. However the producer eventually does not give the product neither reimburses the value, due to bankruptcy or decision to exit the market with an unfair benefit. The intermediary can choose either to reimburse the client and suffer the loss, or refuse to return the money and lose the client's trust. The latter choice for the intermediary is exactly the conservative strategy. It is used throughout this work as a strategy for all the intermediary players because it models effectively the worst-case scenario that a client can face after an evil player decides to steal everything she can and the rest of the players do not engage in evil activity.

We continue with a very useful possible evolution of the game, the Transitive Game. In turn 0, there is already a network in place. All players apart from A and E follow the conservative strategy. Furthermore, the set of players is not modified throughout the Transitive Game, thus we can refer to $\mathcal{V}_j$ for any turn j as $\mathcal{V}$. Moreover, each conservative player can be in one of three states: Happy, Angry or Sad. Happy players have 0 loss, Angry players have positive loss and positive incoming trust, thus are able to replenish their loss at least in part and Sad players have positive loss, but 0 incoming trust, thus they cannot replenish the loss. These conventions will hold whenever we use the Transitive Game.

```
Transitive Game
Input : graph G_0, A ∈ V idle player, E ∈ V evil player
1  Angry = Sad = ∅ ; Happy = V \ {A, E}
2  for (v ∈ V \ {E})
3    Loss_v = 0
4  j = 0
5  while (True)
6    j += 1
7    v <-$ V \ {A}
8    Turn_j = vStrategy(G_0, v, (H)_{1...j-1})
9    executeTurn(G_{j-1}, Cap_{v,j-1}, Turn_j)
10   for (action ∈ Turn_j)
11     action match do
12       case Steal(y, w) do
13         exchange = y
14         Loss_w += exchange
15         if (v != E)
16           Loss_v -= exchange
```

```
17          if (w != A)
18            Happy = Happy \ {w}
19            if (in_{w,j} == 0)
20              Sad = Sad ∪ {w}
21            else
22              Angry = Angry ∪ {w}
23    if (v != E)
24      Angry = Angry \ {v}
25      if (Loss_v > 0)
26        Sad = Sad ∪ {v}        #in_{v,j} should be zero
27      if (Loss_v == 0)
28        Happy = Happy ∪ {v}
```

An example execution follows:

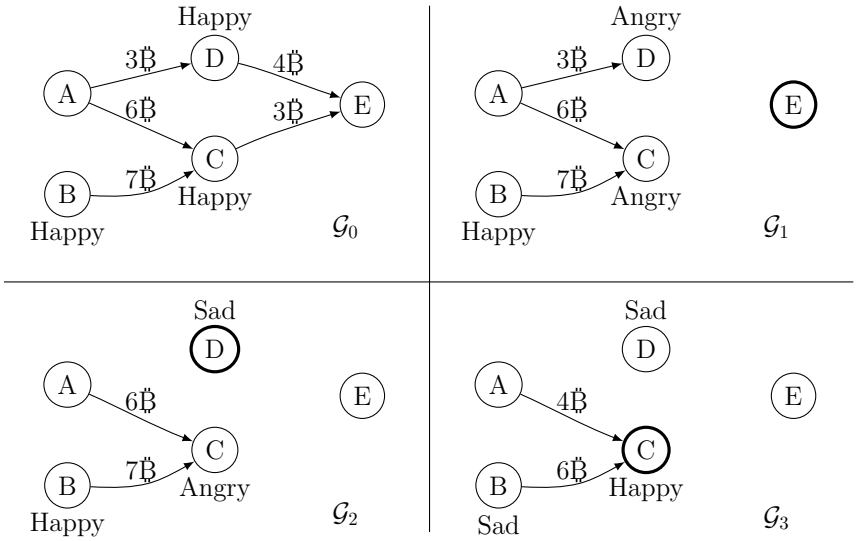

Fig.5: Turns of a `TransitiveGame`($\mathcal{G}_0, A, E$)

Let j_0 be the first turn on which E is chosen to play. Until then, all players will pass their turn since nothing has been stolen yet (see the Appendix (theorem 6) for a formal proof of this simple fact). Moreover, let $v = Player(j)$ and $j' = prev\,(j)$. The Transitive Game generates turns:

$$Turn_j = \bigcup_{w \in N^-(v)_{j-1}} \{Steal\,(y_w, w)\}\ , \tag{8}$$

where

$$\sum_{w \in N^-(v)_{j-1}} y_w = \min\left(in_{v,j-1}, Damage_{v,j}\right) \quad .$$

We see that if $Damage_{v,j} = 0$, then $Turn_j = \emptyset$.

From the definition of $Damage_{v,j}$ and knowing that no strategy in this case can increase any direct trust, we see that $Damage_{v,j} \geq 0$. Also, we can see that $Loss_{v,j} \geq 0$ because if $Loss_{v,j} < 0$, then v has stolen more value than she has been stolen, thus she would not be following the conservative strategy.

5 Trust Flow

Everything is in place to define the indirect trust, or simply trust, from one player to another.

Definition 12 (Indirect Trust). *The indirect trust from A to B after turn j is defined as the maximum possible value that can be stolen from A after turn j in the setting of* `TransitiveGame(`$\mathcal{G}_j$`,A,B)`*.*

It is $Tr_{A \to B} \geq DTr_{A \to B}$. The next theorem shows that $Tr_{A \to B}$ is finite.

Theorem 1 (Trust Convergence Theorem).
Consider a Transitive Game. There exists a turn such that all subsequent turns are empty:

$$\exists j' : \forall j \geq j', Turn_j = \emptyset \quad .$$

Proof Sketch. If the game didn't converge, the *Steal* () actions would continue forever without reduction of the amount stolen over time, thus they would reach infinity. However this is impossible, since there exists only finite total trust. □

Full proofs of all theorems and lemmas can be found in the Appendix.

In the setting of `TransitiveGame(`$\mathcal{G}$`,A,E)`, we make use of the notation $Loss_A = Loss_{A,j}$, where j is a turn that the game has converged. It is important to note that $Loss_A$ is not the same for repeated executions of this kind of game, since the order in which players are chosen may differ between executions and the conservative players are free to choose which incoming trusts they will steal and how much from each.

Let G be a weighted directed graph. We will investigate the maximum flow on this graph. For an introduction to the maximum flow problem see [5] p. 708. Considering each edge's capacity as its weight, a flow assignment $X = [x_{vw}]_{V \times V}$ with a source A and a sink B is valid when:

$$\forall (v, w) \in E, x_{vw} \leq c_{vw} \text{ and} \tag{9}$$

$$\forall v \in V \setminus \{A, B\}, \sum_{w \in N^+(v)} x_{wv} = \sum_{w \in N^-(v)} x_{vw} \ . \tag{10}$$

We do not suppose any skew symmetry in X. The flow value is $\sum_{v \in N^+(A)} x_{Av}$, which is proven to be equal to $\sum_{v \in N^-(B)} x_{vB}$. There exists an algorithm that returns the maximum possible flow from A to B, namely $MaxFlow\,(A, B)$. This algorithm evidently needs full knowledge of the graph. The fastest version of this algorithm runs in $O\,(|V||E|)$ time [6]. We refer to the flow value of $MaxFlow\,(A, B)$ as $maxFlow\,(A, B)$.

We will now introduce two lemmas that will be used to prove the one of the central results of this work, the Trust Flow theorem.

Lemma 1 (MaxFlows Are Transitive Games).
Let $\mathcal{G}$ be a game graph, let $A, E \in \mathcal{V}$ and $MaxFlow\,(A, E)$ the maximum flow from A to E executed on $\mathcal{G}$. There exists an execution of `TransitiveGame(`$\mathcal{G}, A, E$`)` *such that*

$$maxFlow\,(A, E) \leq Loss_A \ .$$

Proof Sketch. The desired execution of `TransitiveGame()` will contain all flows from the $MaxFlow\,(A, E)$ as equivalent $Steal\,()$ actions. The players will play in turns, moving from E back to A. Each player will steal from his predecessors as much as was stolen from her. The flows and the conservative strategy share the property that the total input is equal to the total output. □

Lemma 2 (Transitive Games Are Flows).
Let $\mathcal{H} =$ `TransitiveGame(`$\mathcal{G}, A, E$`)` *for some game graph $\mathcal{G}$ and $A, E \in \mathcal{V}$. There exists a valid flow $X = \{x_{wv}\}_{\mathcal{V} \times \mathcal{V}}$ on $\mathcal{G}_0$ such that*

$$\sum_{v \in \mathcal{V}} x_{Av} = Loss_A \ .$$

Proof Sketch. If we exclude the sad players from the game, the $Steal\,()$ actions that remain constitute a valid flow from A to E. □

Theorem 2 (Trust Flow Theorem).
Let $\mathcal{G}$ be a game graph and $A, E \in \mathcal{V}$. It holds that

$$Tr_{A \to E} = maxFlow\,(A, E) \ .$$

Proof. From lemma (1) we see that there exists an execution of the Transitive Game such that $Loss_A \geq maxFlow\,(A, E)$. Since $Tr_{A\to E}$ is the maximum loss that A can suffer after the convergence of the Transitive Game, we see that

$$Tr_{A\to E} \geq maxFlow\,(A, E) \quad . \tag{11}$$

Moreover, there exists an execution of the Transitive Game such that $Tr_{A\to E} = Loss_A$. From lemma (2), this execution corresponds to a flow. Thus

$$Tr_{A\to E} \leq maxFlow\,(A, E) \quad . \tag{12}$$

The theorem follows from (11) and (12). □

We note that the maxFlow is the same in the following two cases: When a player chooses the evil strategy and when the same player chooses a variation of the evil strategy where she does not nullify her outgoing direct trust.

Here we see another important theorem that gives the basis for risk-invariant transactions between different, possibly unknown, parties.

Theorem 3 (Risk Invariance Theorem). *Let $\mathcal{G}$ game graph, $A, B \in \mathcal{V}$ and l the desired value to be transferred from A to B, with $l \leq Tr_{A\to B}$. Let also $\mathcal{G}'$ with the same nodes as $\mathcal{G}$ such that*

$$\forall v \in \mathcal{V}' \setminus \{A\}, \forall w \in \mathcal{V}', DTr'_{v\to w} = DTr_{v\to w} \quad .$$

Furthermore, suppose that there exists an assignment for the outgoing trust of $A, DTr'_{A\to v}$, such that

$$Tr'_{A\to B} = Tr_{A\to B} - l \quad . \tag{13}$$

Let another game graph, $\mathcal{G}''$, be identical to $\mathcal{G}'$ except for the following change:

$$DTr''_{A\to B} = DTr'_{A\to B} + l \quad .$$

It then holds that

$$Tr''_{A\to B} = Tr_{A\to B} \quad .$$

Proof. The two graphs $\mathcal{G}'$ and $\mathcal{G}''$ differ only on the weight of the edge (A, B), which is larger by l in $\mathcal{G}''$. Thus the two $MaxFlow$s will choose the same flow, except for (A, B), where it will be $x''_{AB} = x'_{AB} + l$. □

It is intuitively obvious that it is possible for A to reduce her outgoing direct trust in a manner that achieves (13), since $maxFlow\,(A, B)$ is continuous with respect to A's outgoing direct trusts. We leave this calculation as part of further research.

6 Sybil Resilience

One of the primary aims of this system is to mitigate the danger for Sybil attacks [7] whilst maintaining fully decentralized autonomy.

Here we extend the definition of indirect trust to many players.

Definition 13 (Indirect Trust to Multiple Players). *The indirect trust from player A to a set of players, $S \subset \mathcal{V}$ is defined as the maximum possible value that can be stolen from A if all players in S follow the evil strategy, A follows the idle strategy and everyone else ($\mathcal{V} \setminus (S \cup \{A\})$) follows the conservative strategy. More formally, let choices be the different actions between which the conservative players can choose, then*

$$Tr_{A \to S,j} = \max_{j':j'>j,choices} \left[out_{A,j} - out_{A,j'} \right] \tag{14}$$

We now extend Trust Flow theorem (2) to many players.

Theorem 4 (Multi-Player Trust Flow).
Let $S \subset \mathcal{V}$ and T auxiliary player such that $\forall B \in S, DTr_{B \to T} = \infty$. It holds that

$$\forall A \in \mathcal{V} \setminus S, Tr_{A \to S} = maxFlow\left(A, T\right) \quad .$$

Proof. If T chooses the evil strategy and all players in S play according to the conservative strategy, they will have to steal all their incoming direct trust since they have suffered an infinite loss, thus they will act in a way identical to following the evil strategy as far as $MaxFlow$ is concerned. The theorem follows thus from the Trust Flow theorem. □

We now define several useful notions to tackle the problem of Sybil attacks. Let Eve be a possible attacker.

Definition 14 (Corrupted Set). *Let $\mathcal{G}$ be a game graph and let Eve have a set of players $\mathcal{B} \subset \mathcal{V}$ corrupted, so that she fully controls their outgoing direct trusts to any player in $\mathcal{V}$ and can also steal all incoming direct trust to players in $\mathcal{B}$. We call this the corrupted set. The players $\mathcal{B}$ are considered to be legitimate before the corruption, thus they may be directly trusted by any player in $\mathcal{V}$.*

Definition 15 (Sybil Set). *Let $\mathcal{G}$ be a game graph. Since participation in the network does not require any kind of registration, Eve can create any number of players. We will call the set of these players $\mathcal{C}$, or Sybil set. Moreover, Eve can arbitrarily set the direct trusts of any player in $\mathcal{C}$ to any player and can also steal all incoming direct trust to players in $\mathcal{C}$. However, players $\mathcal{C}$ can be directly trusted only by players $\mathcal{B} \cup \mathcal{C}$ but not by players $\mathcal{V} \setminus (\mathcal{B} \cup \mathcal{C})$, where $\mathcal{B}$ is a set of players corrupted by Eve.*

Definition 16 (Collusion). *Let $\mathcal{G}$ be a game graph. Let $\mathcal{B} \subset \mathcal{V}$ be a corrupted set and $\mathcal{C} \subset \mathcal{V}$ be a Sybil set, both controlled by Eve. The tuple $(\mathcal{B}, \mathcal{C})$ is called a collusion and is entirely controlled by a single entity in the physical world. From a game theoretic point of view, players $\mathcal{V} \setminus (\mathcal{B} \cup \mathcal{C})$ perceive the collusion as independent players with a distinct strategy each, whereas in reality they are all subject to a single strategy dictated by the controlling entity, Eve.*

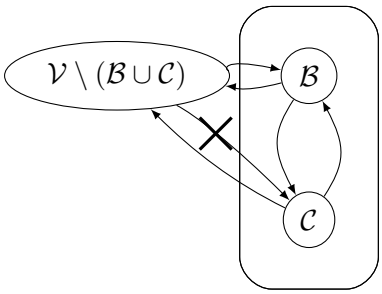

Fig.6: Collusion

Theorem 5 (Sybil Resilience).
Let $\mathcal{G}$ be a game graph and $(\mathcal{B}, \mathcal{C})$ be a collusion of players on $\mathcal{G}$. It holds that

$$Tr_{A \to \mathcal{B} \cup \mathcal{C}} = Tr_{A \to \mathcal{B}} \ .$$

Proof Sketch. The incoming trust to $\mathcal{B} \cup \mathcal{C}$ cannot be higher than the incoming trust to $\mathcal{B}$ since $\mathcal{C}$ has no incoming trust from players outside the collusion. □

We have proven that controlling $|\mathcal{C}|$ is irrelevant for Eve, thus Sybil attacks are meaningless.

With this we have successfully delivered our promise for a Sybil-resilient decentralized financial trust system with invariant risk when making a purchases.

7 Related Work

The topic of trust has been repeatedly attacked with several approaches: Purely cryptographic infrastructure where trust is rather binary and transitivity is not possible is explored in PGP [8]. A transitive web-of-trust for fighting spam is explored in Freenet [9]. Other systems require central trusted third parties, such as PKI [10] and Bazaar [11], or, in the case of BFT, authenticated membership [12]. While other trust systems attempt

to be decentralized, they do not prove any Sybil resilience properties and hence may be Sybil attackable. Such systems are FIRE [13], CORE [14] and others [15,16,17]. Other systems that define trust in a non-financial way are [18,19,20,21,22,23,24].

The Beaver marketplace [25] includes a trust model that relies on fees to discourage Sybil attacks. We chose to avoid fees in our system and mitigate Sybil attacks in a different manner. Our motivating application for exploring trust in a decentralized setting is the OpenBazaar marketplace. Transitive financial trust for OpenBazaar has previously been explored by [26]. That work however does not define trust as a monetary value. We are strongly inspired by [4] which gives a sociological justification for the central design choice of identifying trust with risk. We greatly appreciate the work in TrustDavis [27], which proposes a financial trust system that exhibits transitive properties and in which trust is defined as lines-of-credit, similar to our system. We were able to extend their work by using the blockchain for automated proofs-of-risk, a feature not available to them at the time. Our conservative strategy and the Transitive Game are very similar to the mechanism proposed by the economic paper [28] which also illustrates financial trust transitivity.

8 Further Research

While our trust network can form a basis for risk-invariant transactions in a pseudonymous and decentralized setting, more research is required to achieve other desirable properties. Some directions for future research are outlined below.

If *Alice* trusts *Bob* enough to make a purchase from him, she should not directly pay him the value of the good because then she will increase her trust towards *Bob*. She first has to reduce her outgoing direct trust in a manner such that the supposition (13) of Risk Invariance theorem is satisfied. The methods *Alice* can use to recalculate her outgoing trust will be discussed in a future paper.

The current description of TrustIsRisk refers to a static setting where the game evolves in turns. In each turn only one user changes the state of the network and the game is controlled by a central algorithm, the TrustIsRisk Game. In the dynamic setting, users should be able to play simultaneously, freely join, depart or disconnect temporarily from the network.

Our network evaluates indirect trust by computing the max flow in the graph of lines-of-credit. In order to do that, complete information about

the network is required. However, disclosing the network topology may be undesirable, as it subverts the identity of the participants even when participants are treated pseudonymously, as deanonymisation techniques can be used [29]. To avoid such issues, exploring the ability to calculate flows in a zero knowledge fashion may be desirable. However, performing network queries in zero knowledge may allow an adversary to extract topological information. More research is required to establish how flows can be calculated effectively in zero knowledge and what bounds exist in regards to information revealed in such fashion.

Our game theoretic analysis is simple. An interesting analysis would involve modelling repeated purchases with the respective edge updates on the trust graph and treating trust on the network as part of the utility function.

We are proposing a concrete financial game and not a theoretical concept. Thus its implementation as a wallet on any blockchain will be most welcome.

A simulation or actual implementation of TrustIsRisk, combined with analysis of the resulting dynamics can yield interesting experimental results. Subsequently, our trust network can be used in other applications, such as decentralized social networks [30].

1-of-2 multisigs correspond intuitively to simple directed weighted graphs. However it can be interesting to explore the trust relations that can arise by using other types of multisig, such as 1-of-3, as vessel for multi-party trust schemes. Our results do not necessarily hold for other multisigs and the simple relations now represented by directed weighted graphs have to be revised under a new kind of representation.

Appendix

1 Common Notation

Definition 17 (Neighbourhood).

1. *Let $N^+(A)_j$ be the set of players B that A directly trusts with any positive value at the end of turn j. More formally,*

$$N^+(A)_j = \{B \in \mathcal{V}_j : DTr_{A\to B,j} > 0\} \ . \tag{15}$$

$N^+(A)_j$ is called out neighbourhood of A on turn j. Let $S \subseteq \mathcal{V}_j$. Then

$$N^+(S)_j = \bigcup_{A\in S} N^+(A)_j \ . \tag{16}$$

2. *Let* $N^{-}(A)_j$ *be the set of players* B *that directly trust* A *with any positive value at the end of turn* j*. More formally,*

$$N^{-}(A)_j = \{B \in \mathcal{V}_j : DTr_{B \to A,j} > 0\} \ . \tag{17}$$

$N^{-}(A)_j$ is called in neighbourhood of A on turn j. Let $S \subseteq \mathcal{V}_j$. Then

$$N^{-}(S)_j = \bigcup_{A \in S} N^{-}(A)_j \ . \tag{18}$$

3. *Let* $N(A)_j$ *be the set of players* B *that either directly trust or are directly trusted by* A *with any positive value at the end of turn* j*. More formally,*

$$N(A)_j = N^{+}(A)_j \cup N^{-}(A)_j \ . \tag{19}$$

$N(A)_j$ is called neighbourhood of A on turn j. Let $S \subset \mathcal{V}_j$. Then

$$N(S)_j = N^{+}(S)_j \cup N^{-}(S)_j \ . \tag{20}$$

Definition 18 (Total Incoming/Outgoing Trust).

$$in_{A,j} = \sum_{v \in N^{-}(A)_j} DTr_{v \to A,j} \tag{21}$$

$$out_{A,j} = \sum_{v \in N^{+}(A)_j} DTr_{A \to v,j} \tag{22}$$

Definition 19 (Assets). *Sum of A's capital and outgoing trust.*

$$As_{A,j} = Cap_{A,j} + out_{A,j} \tag{23}$$

We consider the outgoing trust of a player as part of her assets. Here we add some concrete $Turn_j$ examples. Let $A = Player(j)$.

1.
$$Turn_j = \emptyset$$

2.
$$Turn_j = \{Steal\,(y, B)\,, Add\,(w, B)\}\ ,$$

given that

$$0 \leq y \leq DTr_{B\to A,j-1} \wedge -DTr_{A\to B,j-1} \leq w \wedge w - y \leq Cap_{A,j-1}\ .$$

3.
$$Turn_j = \{Steal\,(x, B)\,, Add\,(y, C)\,, Add\,(w, D)\}\ ,$$

given that

$$0 \leq x \leq DTr_{B\to A,j-1} \wedge -DTr_{A\to C,j-1} \leq y \wedge$$
$$\wedge - DTr_{A\to D,j-1} \leq w \wedge y + w - x \leq Cap_{A,j-1}\ .$$

4.
$$Turn_j = \{Steal\,(x, B)\,, Steal\,(y, B)\}$$

is not a valid turn because it contains two $Steal\,()$ actions against the same player. If

$$0 \leq x \wedge 0 \leq y \wedge x + y \leq DTr_{B\to A,j-1}\ ,$$

the correct alternative would be

$$Turn_j = \{Steal\,(x + y, B)\}\ .$$

Definition 20 (Previous/Next Turn). *Let $j \in \mathbb{N}$ a turn with $Player\,(j) = A$. We define $prev\,(j)\,, next\,(j)$ as the previous and next turn that A is chosen to play respectively. If j is the first turn that A plays, $prev\,(j) = 0$. More formally, if*

$$P = \{k \in \mathbb{N} : k < j \wedge Player\,(k) = A\}\ \textit{and}$$

$$N = \{k \in \mathbb{N} : k > j \wedge Player\,(k) = A\}\ ,$$

then we define $prev(j), next(j)$ *as follows:*

$$prev(j) = \begin{cases} \max P, & P \neq \emptyset \\ 0, & P = \emptyset \end{cases} \tag{24}$$

$$next(j) = \min N \tag{25}$$

$next(j)$ *is always well defined with the assumption that after each turn eventually everybody plays.*

2 Proofs, Lemmas and Theorems

Lemma 3 (*Loss* Equivalent to *Damage*).
Consider a Transitive Game. Let $j \in \mathbb{N}$ *and* $v = Player(j)$ *such that* v *is following the conservative strategy. It holds that*

$$\min(in_{v,j}, Loss_{v,j}) = \min(in_{v,j}, Damage_{v,j}) \quad .$$

Proof.
Case 1: Let $v \in Happy_{j-1}$. Then

1. $v \in Happy_j$ because $Turn_j = \emptyset$,
2. $Loss_{v,j} = 0$ because otherwise $v \notin Happy_j$,
3. $Damage_{v,j} = 0$, or else any reduction in direct trust to v would increase equally $Loss_{v,j}$ (line 14), which cannot be decreased again but during an Angry player's turn (line 16).
4. $in_{v,j} \geq 0$

Thus

$$\min(in_{v,j}, Loss_{v,j}) = \min(in_{v,j}, Damage_{v,j}) = 0 \quad .$$

Case 2: Let $v \in Sad_{j-1}$. Then

1. $v \in Sad_j$ because $Turn_j = \emptyset$,
2. $in_{v,j} = 0$ (line 25),
3. $Damage_{v,j} \geq 0 \wedge Loss_{v,j} \geq 0$.

Thus

$$\min(in_{v,j}, Loss_{v,j}) = \min(in_{v,j}, Damage_{v,j}) = 0 \quad .$$

If $v \in Angry_{j-1}$ then the same argument as in cases 1 and 2 hold when $v \in Happy_j$ and $v \in Sad_j$ respectively if we ignore the argument (1). Thus the theorem holds in every case. □

Proof of Theorem 1: Trust Convergence

First of all, after turn j_0 player E will always pass her turn because she has already nullified her incoming and outgoing direct trusts in $Turn_{j_0}$, the evil strategy does not contain any case where direct trust is increased or where the evil player starts directly trusting another player and the other players do not follow a strategy in which they can choose to $Add\,()$ trust to E. The same holds for player A because she follows the idle strategy. As far as the rest of the players are concerned, consider the Transitive Game. As we can see from lines 3 and 14 - 16, it is

$$\forall j, \sum_{v \in \mathcal{V}_j} Loss_v = in_{E,j_0-1} \ .$$

In other words, the total loss is constant and equal to the total value stolen by E. Also, as we can see in lines 1 and 26, which are the only lines where the Sad set is modified, once a player enters the Sad set, it is impossible to exit from this set. Also, we can see that players in $Sad \cup Happy$ always pass their turn. We will now show that eventually the $Angry$ set will be empty, or equivalently that eventually every player will pass their turn. Suppose that it is possible to have an infinite amount of turns in which players do not choose to pass. We know that the number of nodes is finite, thus this is possible only if

$$\exists j' : \forall j \geq j', |Angry_j \cup Happy_j| = c > 0 \wedge Angry_j \neq \emptyset \ .$$

This statement is valid because the total number of angry and happy players cannot increase because no player leaves the Sad set and if it were to be decreased, it would eventually reach 0. Since $Angry_j \neq \emptyset$, a player v that will not pass her turn will eventually be chosen to play. According to the Transitive Game, v will either deplete her incoming trust and enter the Sad set (line 26), which is contradicting $|Angry_j \cup Happy_j| = c$, or will steal enough value to enter the $Happy$ set, that is v will achieve $Loss_{v,j} = 0$. Suppose that she has stolen m players. They, in their turn, will steal total value at least equal to the value stolen by v (since they cannot go sad, as explained above). However, this means that, since the total value being stolen will never be reduced and the turns this will happen are infinite, the players must steal an infinite amount of value, which is impossible because the direct trusts are finite in number and in value. More precisely, let j_1 be a turn in which a conservative player is chosen and

$$\forall j \in \mathbb{N}, DTr_j = \sum_{w,w' \in \mathcal{V}} DTr_{w \to w',j} \ .$$

Also, without loss of generality, suppose that

$$\forall j \geq j_1, out_{A,j} = out_{A,j_1} \ .$$

In $Turn_{j_1}$, v steals

$$St = \sum_{i=1}^{m} y_i \ .$$

We will show using induction that

$$\forall n \in \mathbb{N}, \exists j_n \in \mathbb{N} : DTr_{j_n} \leq DTr_{j_1-1} - nSt \ .$$

Base case: It holds that

$$DTr_{j_1} = DTr_{j_1-1} - St \ .$$

Eventually there is a turn j_2 when every player in $N^-(v)_{j-1}$ will have played. Then it holds that

$$DTr_{j_2} \leq DTr_{j_1} - St = DTr_{j_1-1} - 2St \ ,$$

since all players in $N^-(v)_{j-1}$ follow the conservative strategy, except for A, who will not have been stolen anything due to the supposition.

Induction hypothesis: Suppose that

$$\exists k > 1 : j_k > j_{k-1} > j_1 \Rightarrow DTr_{j_k} \leq DTr_{j_{k-1}} - St \ .$$

Induction step: There exists a subset of the $Angry$ players, S, that have been stolen at least value St in total between the turns j_{k-1} and j_k, thus there exists a turn j_{k+1} such that all players in S will have played and thus

$$DTr_{j_{k+1}} \leq DTr_{j_k} - St \ .$$

We have proven by induction that

$$\forall n \in \mathbb{N}, \exists j_n \in \mathbb{N} : DTr_{j_n} \leq DTr_{j_1-1} - nSt \ .$$

However

$$DTr_{j_1-1} \geq 0 \wedge St > 0 \ ,$$

thus

$$\exists n' \in \mathbb{N} : n'St > DTr_{j_1-1} \Rightarrow DTr_{j_{n'}} < 0 \ .$$

We have a contradiction because

$$\forall w, w' \in \mathcal{V}, \forall j \in \mathbb{N}, DTr_{w \to w',j} \geq 0 \ ,$$

thus eventually $Angry = \emptyset$ and everybody passes. □

Proof of Lemma 1: MaxFlows Are Transitive Games

We suppose that the turn of $\mathcal{G}$ is 0. In other words, $\mathcal{G} = \mathcal{G}_0$. Let $X = \{x_{vw}\}_{\mathcal{V}\times\mathcal{V}}$ be the flows returned by $MaxFlow\,(A, E)$. For any graph G there exists a $MaxFlow$ that is a DAG. We can easily prove this using the Flow Decomposition theorem [31], which states that each flow can be seen as a finite set of paths from A to E and cycles, each having a certain flow. We execute $MaxFlow\,(A, E)$ and we apply the aforementioned theorem. The cycles do not influence the $maxFlow\,(A, E)$, thus we can remove these flows. The resulting flow is a $MaxFlow\,(A, E)$ without cycles, thus it is a DAG. Topologically sorting this DAG, we obtain a total order of its nodes such that $\forall$ nodes $v, w \in \mathcal{V} : v < w \Rightarrow x_{wv} = 0$ [5]. Put differently, there is no flow from larger to smaller nodes. E is maximum since it is the sink and thus has no outgoing flow to any node and A is minimum since it is the source and thus has no incoming flow from any node. The desired execution of Transitive Game will choose players following the total order inversely, starting from player E. We observe that $\forall v \in \mathcal{V} \setminus \{A, E\}$, $\sum\limits_{w\in\mathcal{V}} x_{wv} = \sum\limits_{w\in\mathcal{V}} x_{vw} \leq maxFlow\,(A, E) \leq in_{E,0}$. Player E will follow a modified evil strategy where she steals value equal to her total incoming flow, not her total incoming trust. Let j_2 be the first turn when A is chosen to play. We will show using strong induction that there exists a set of valid actions for each player according to their respective strategy such that at the end of each turn j the corresponding player $v = Player\,(j)$ will have stolen value x_{wv} from each in-neighbour w.

Base case: In turn 1, E steals value equal to $\sum\limits_{w\in\mathcal{V}} x_{wE}$, following the modified evil strategy.

$$Turn_1 = \bigcup_{v\in N^-(E)_0} \{Steal\,(x_{vE}, v)\}$$

Induction hypothesis: Let $k \in [j_2 - 2]$. We suppose that $\forall i \in [k]$, there exists a valid set of actions, $Turn_i$, performed by $v = Player\,(i)$ such that v steals from each player w value equal to x_{wv}.

$$\forall i \in [k], Turn_i = \bigcup_{w\in N^-(v)_{i-1}} \{Steal\,(x_{wv}, w)\}$$

Induction step: Let $j = k + 1, v = Player\,(j)$. Since all the players that are greater than v in the total order have already played and all of

them have stolen value equal to their incoming flow, we deduce that v has been stolen value equal to $\sum_{w \in N^+(v)_{j-1}} x_{vw}$. Since it is the first time v plays, $\forall w \in N^-(v)_{j-1}, DTr_{w \to v, j-1} = DTr_{w \to v, 0} \geq x_{wv}$, thus v is able to choose the following turn:

$$Turn_j = \bigcup_{w \in N^-(v)_{j-1}} \{Steal(x_{wv}, w)\}$$

Moreover, this turn satisfies the conservative strategy since

$$\sum_{w \in N^-(v)_{j-1}} x_{wv} = \sum_{w \in N^+(v)_{j-1}} x_{vw} \ .$$

Thus $Turn_j$ is a valid turn for the conservative player v.

We have proven that in the end of turn $j_2 - 1$, player E and all the conservative players will have stolen value exactly equal to their total incoming flow, thus A will have been stolen value equal to her outgoing flow, which is $maxFlow(A, E)$. Since there remains no Angry player, j_2 is a convergence turn, thus $Loss_{A,j_2} = Loss_A$. We can also see that if E had chosen the original evil strategy, the described actions would still be valid only by supplementing them with additional $Steal()$ actions, thus $Loss_A$ would further increase. This proves the theorem. □

Proof of Lemma 2: Transitive Games Are Flows

Let $Sad, Happy, Angry$ be as defined in the Transitive Game. Let $\mathcal{G}'$ be a directed weighted graph based on $\mathcal{G}$ with an auxiliary source. Let also j_1 be a turn when the Transitive Game has converged. More precisely, $\mathcal{G}'$ is defined as follows:

$$\mathcal{V}' = \mathcal{V} \cup \{T\}$$

$$\mathcal{E}' = \mathcal{E} \cup \{(T, A)\} \cup \{(T, v) : v \in Sad_{j_1}\}$$

$$\forall (v, w) \in \mathcal{E}, c'_{vw} = DTr_{v \to w, 0} - DTr_{v \to w, j_1}$$

$$\forall v \in Sad_{j_1}, c'_{Tv} = c'_{TA} = \infty$$

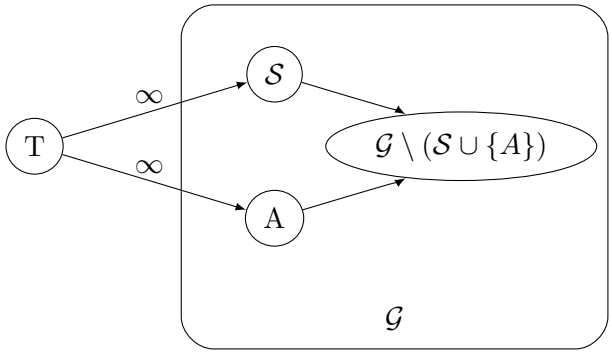

Fig.7: Graph $\mathcal{G}'$, derived from $\mathcal{G}$ with Auxiliary Source T.

In the figure above, $\mathcal{S}$ is the set of sad players. We observe that $\forall v \in \mathcal{V}$,

$$\begin{aligned} &\sum_{w \in N^-(v)' \setminus \{T\}} c'_{wv} = \\ &= \sum_{w \in N^-(v)' \setminus \{T\}} (DTr_{w \to v,0} - DTr_{w \to v,j_1}) = \\ &= \sum_{w \in N^-(v)' \setminus \{T\}} DTr_{w \to v,0} - \sum_{w \in N^-(v)' \setminus \{T\}} DTr_{w \to v,j-1} = \\ &= in_{v,0} - in_{v,j_1} \end{aligned} \tag{26}$$

and

$$\begin{aligned} &\sum_{w \in N^+(v)' \setminus \{T\}} c'_{vw} = \\ &= \sum_{w \in N^+(v)' \setminus \{T\}} (DTr_{v \to w,0} - DTr_{v \to w,j_1}) = \\ &= \sum_{w \in N^+(v)' \setminus \{T\}} DTr_{v \to w,0} - \sum_{w \in N^+(v)' \setminus \{T\}} DTr_{v \to w,j-1} = \\ &= out_{v,0} - out_{v,j_1} \ . \end{aligned} \tag{27}$$

We can suppose that

$$\forall j \in \mathbb{N}, in_{A,j} = 0 \ , \tag{28}$$

since if we find a valid flow under this assumption, the flow will still be valid for the original graph.
Next we try to calculate $MaxFlow\,(T, E) = X'$ on graph $\mathcal{G}'$. We observe that a flow in which it holds that $\forall v, w \in \mathcal{V}, x'_{vw} = c'_{vw}$ can be valid for the following reasons:

– $\forall v, w \in \mathcal{V}, x'_{vw} \leq c'_{vw}$ (Capacity flow requirement (9) $\forall e \in \mathcal{E}$)
– Since $\forall v \in Sad_{j_1} \cup \{A\}, c'_{Tv} = \infty$, requirement (9) holds for any flow $x'_{Tv} \geq 0$.
– Let $v \in \mathcal{V}' \setminus (Sad_{j_1} \cup \{T, A, E\})$. According to the conservative strategy and since $v \notin Sad_{j_1}$, it holds that

$$out_{v,0} - out_{v,j_1} = in_{v,0} - in_{v,j_1} \ .$$

Combining this observation with (26) and (27), we have that

$$\sum_{w \in \mathcal{V}'} c'_{vw} = \sum_{w \in \mathcal{V}'} c'_{wv} \ .$$

(Flow Conservation requirement (10) $\forall v \in \mathcal{V}' \setminus (Sad_{j_1} \cup \{T, A, E\})$)
– Let $v \in Sad_{j_1}$. Since v is sad, we know that

$$out_{v,0} - out_{v,j_1} > in_{v,0} - in_{v,j_1} \ .$$

Since $c'_{Tv} = \infty$, we can set

$$x'_{Tv} = (out_{v,0} - out_{v,j_1}) - (in_{v,0} - in_{v,j_1}) \ .$$

In this way, we have

$$\sum_{w \in \mathcal{V}'} x'_{vw} = out_{v,0} - out_{v,j_1} \text{ and}$$

$$\sum_{w \in \mathcal{V}'} x'_{wv} = \sum_{w \in \mathcal{V}' \setminus \{T\}} c'_{wv} + x'_{Tv} = in_{v,0} - in_{v,j_1} +$$

$$+ (out_{v,0} - out_{v,j_1}) - (in_{v,0} - in_{v,j_1}) = out_{v,0} - out_{v,j_1} \ .$$

thus

$$\sum_{w \in \mathcal{V}'} x'_{vw} = \sum_{w \in \mathcal{V}'} x'_{wv} \ .$$

(Requirement 10 $\forall v \in Sad_{j_1}$)
– Since $c'_{TA} = \infty$, we can set

$$x'_{TA} = \sum_{v \in \mathcal{V}'} x'_{Av} \ ,$$

thus from (28) we have

$$\sum_{v \in \mathcal{V}'} x'_{vA} = \sum_{v \in \mathcal{V}'} x'_{Av} \ .$$

(Requirement 10 for A)

We saw that for all nodes, the necessary properties for a flow to be valid hold and thus X' is a valid flow for $\mathcal{G}$. Moreover, this flow is equal to $maxFlow(T, E)$ because all incoming flows to E are saturated. Also we observe that

$$\sum_{v \in \mathcal{V}'} x'_{Av} = \sum_{v \in \mathcal{V}'} c'_{Av} = out_{A,0} - out_{A,j_1} = Loss_A \ . \tag{29}$$

We define another graph, $\mathcal{G}''$, based on $\mathcal{G}'$.

$$\mathcal{V}'' = \mathcal{V}'$$

$$E(\mathcal{G}'') = E(\mathcal{G}') \setminus \{(T, v) : v \in Sad_j\}$$
$$\forall e \in E(\mathcal{G}''), c''_e = c'_e$$

If we execute $MaxFlow(T, E)$ on the graph $\mathcal{G}''$, we will obtain a flow X'' in which

$$\sum_{v \in \mathcal{V}''} x''_{Tv} = x''_{TA} = \sum_{v \in \mathcal{V}''} x''_{Av} \ .$$

The outgoing flow from A in X'' will remain the same as in X' for two reasons: Firstly, using the Flow Decomposition theorem [31] and deleting the paths that contain edges $(T, v) : v \neq A$, we obtain a flow configuration where the total outgoing flow from A remains invariant, [3] thus

$$\sum_{v \in \mathcal{V}''} x''_{Av} \geq \sum_{v \in \mathcal{V}'} x'_{Av} \ .$$

Secondly, we have

$$\left.\begin{array}{l} \sum\limits_{v \in \mathcal{V}''} c''_{Av} = \sum\limits_{v \in \mathcal{V}'} c'_{Av} = \sum\limits_{v \in \mathcal{V}'} x'_{Av} \\ \sum\limits_{v \in \mathcal{V}''} c''_{Av} \geq \sum\limits_{v \in \mathcal{V}''} x''_{Av} \end{array}\right\} \Rightarrow \sum_{v \in \mathcal{V}''} x''_{Av} \leq \sum_{v \in \mathcal{V}'} x'_{Av} \ .$$

Thus we conclude that

$$\sum_{v \in \mathcal{V}''} x''_{Av} = \sum_{v \in \mathcal{V}'} x'_{Av} \ . \tag{30}$$

Let $X = X'' \setminus \{(T, A)\}$. Observe that

$$\sum_{v \in \mathcal{V}''} x''_{Av} = \sum_{v \in \mathcal{V}} x_{Av} \ .$$

[3] We thank Kyriakos Axiotis for his insights on the Flow Decomposition theorem.

This flow is valid on graph $\mathcal{G}$ because

$$\forall e \in \mathcal{E}, c_e \geq c''_e \ .$$

Thus there exists a valid flow for each execution of the Transitive Game such that

$$\sum_{v \in \mathcal{V}} x_{Av} = \sum_{v \in \mathcal{V}''} x''_{Av} \overset{(30)}{=} \sum_{v \in \mathcal{V}'} x'_{Av} \overset{(29)}{=} Loss_{A,j_1} \ ,$$

which is the flow X. □

Theorem 6 (Conservative World Theorem).
If everybody follows the conservative strategy, nobody steals any amount from anybody.

Proof. Let $\mathcal{H}$ be the game history where all players are conservative and suppose there are some *Steal* () actions taking place. Then let $\mathcal{H}'$ be the subsequence of turns each containing at least one *Steal* () action. This subsequence is evidently nonempty, thus it must have a first element. The player corresponding to that turn, A, has chosen a *Steal* () action and no previous player has chosen such an action. However, player A follows the conservative strategy, which is a contradiction. □

Proof of Theorem 5: Sybil Resilience
Let $\mathcal{G}_1$ be a game graph defined as follows:

$$\mathcal{V}_1 = \mathcal{V} \cup \{T_1\} \ ,$$

$$\mathcal{E}_1 = \mathcal{E} \cup \{(v, T_1) : v \in \mathcal{B} \cup \mathcal{C}\} \ ,$$

$$\forall v, w \in \mathcal{V}_1 \setminus \{T_1\}, DTr^1_{v \to w} = DTr_{v \to w} \ ,$$

$$\forall v \in \mathcal{B} \cup \mathcal{C}, DTr^1_{v \to T_1} = \infty \ ,$$

where $DTr_{v \to w}$ is the direct trust from v to w in $\mathcal{G}$ and $DTr^1_{v \to w}$ is the direct trust from v to w in $\mathcal{G}_1$.
Let also $\mathcal{G}_2$ be the induced graph that results from $\mathcal{G}_1$ if we remove the Sybil set, $\mathcal{C}$. We rename T_1 to T_2 and define $\mathcal{L} = \mathcal{V} \setminus (\mathcal{B} \cup \mathcal{C})$ as the set of legitimate players to facilitate comprehension.

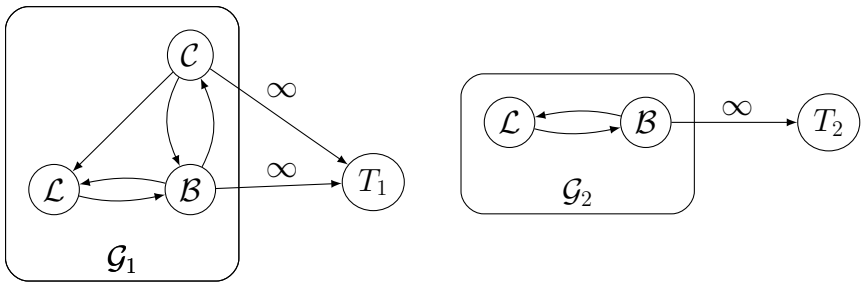

Fig.8: Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$

According to theorem (4),

$$Tr_{A\to\mathcal{B}\cup\mathcal{C}} = maxFlow_1\left(A, T_1\right) \wedge Tr_{A\to\mathcal{B}} = maxFlow_2\left(A, T_2\right) \quad . \tag{31}$$

We will show that the $MaxFlow$ of each of the two graphs can be used to construct a valid flow of equal value for the other graph. The flow $X_1 = MaxFlow\left(A, T_1\right)$ can be used to construct a valid flow of equal value for the second graph if we set

$$\begin{aligned} \forall v \in \mathcal{V}_2 \setminus \mathcal{B}, \forall w \in \mathcal{V}_2, x_{vw,2} &= x_{vw,1} \ , \\ \forall v \in \mathcal{B}, x_{vT_2,2} &= \sum_{w \in N_1^+(v)} x_{vw,1} \ , \\ \forall v, w \in \mathcal{B}, x_{vw,2} &= 0 \ . \end{aligned}$$

Therefore

$$maxFlow_1\left(A, T_1\right) \leq maxFlow_2\left(A, T_2\right)$$

Likewise, the flow $X_2 = MaxFlow(A, T_2)$ is a valid flow for $\mathcal{G}_1$ because $\mathcal{G}_2$ is an induced subgraph of $\mathcal{G}_1$. Therefore

$$maxFlow_1\left(A, T_1\right) \geq maxFlow_2\left(A, T_2\right)$$

We conclude that

$$maxFlow\left(A, T_1\right) = maxFlow\left(A, T_2\right) \ , \tag{32}$$

thus from (31) and (32) the theorem holds. □

3 Algorithms

This algorithm calls the necessary functions to prepare the new graph.

```
Execute Turn
Input : old graph G_{j-1}, player A ∈ V_{j-1}, old capital
    Cap_{A,j-1}, TentativeTurn
Output : new graph G_j, new capital Cap_{A,j}, new history H_j
executeTurn(G_{j-1}, A, Cap_{A,j-1}, TentativeTurn) :
  (Turn_j, NewCap) = validateTurn(G_{j-1}, A, Cap_{A,j-1},
      TentativeTurn)
  return(commitTurn(G_{j-1}, A, Turn_j, NewCap))
```

The following algorithm validates that the tentative turn produced by the strategy respects the rules imposed on turns. If the turn is invalid, an empty turn is returned.

```
Validate Turn
Input : old G_{j-1}, player A ∈ V_{j-1}, old Cap_{A,j-1}, Turn
Output : Turn_j, new Cap_{A,j}
validateTurn(G_{j-1}, A, Cap_{A,j-1}, Turn) :
  Y_st = Y_add = 0
  Stolen = Added = ∅
  for (action ∈ Turn)
    action match do
      case Steal(y, w) do
        if (y > DTr_{w→A,j-1} or y < 0 or w ∈ Stolen)
          return(∅, Cap_{A,j-1})
        else
          Y_st += y
          Stolen = Stolen ∪ {w}
      case Add(y, w) do
        if (y < -DTr_{A→w,j-1} or w ∈ Added)
          return(∅, Cap_{A,j-1})
        else
          Y_add += y
          Added = Added ∪ {w}
  if (Y_add - Y_st > Cap_{A,j-1})
    return(∅, Cap_{A,j-1})
  else
    return(Turn, Cap_{A,j-1} + Y_st - Y_add)
```

Finally, this algorithm applies the turn to the old graph and returns the new graph, along with the updated capital and history.

```
Commit Turn
Input : old G_{j-1}, player A ∈ V_{j-1}, NewCap, Turn_j
Output : new G_j, new Cap_{A,j}, new H_j
commitTurn(G_{j-1}, A, NewCap, Turn_j) :
  for ((v, w) ∈ E_j)
    DTr_{v→w,j} = DTr_{v→w,j-1}
  for (action ∈ Turn_j)
    action match do
      case Steal(y, w) do
        DTr_{w→A,j} = DTr_{w→A,j-1} − y
      case Add(y, w) do
        DTr_{A→w,j} = DTr_{A→w,j-1} + y
  Cap_{A,j} = NewCap; H_j = (A, Turn_j)
  return(G_j, Cap_{A,j}, H_j)
```

It is straightforward to verify the compatibility of the previous algorithms with the corresponding definitions.

References

1. Sanchez W.: Lines of Credit. https://gist.github.com/drwasho/2c40b91e169f55988618#part-3-web-of-credit (2016)
2. Nakamoto S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008)
3. Antonopoulos A. M.: Mastering Bitcoin: Unlocking Digital Cryptocurrencies. O'Reilly Media, Inc. (2014)
4. Karlan D., Mobius M., Rosenblat T., Szeidl A.: Trust and social collateral. The Quarterly Journal of Economics, pp. 1307-1361 (2009)
5. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.: Introduction to Algorithms (3rd ed.). MIT Press and McGraw-Hill (2009)
6. Orlin J. B.: Max Flows in O(nm) Time, or Better. STOC '13 Proceedings of the forty-fifth annual ACM symposium on Theory of computing, pp.765-774, ACM, New York, doi:10.1145/2488608.2488705 (2013)
7. Douceur J. R.: The Sybil Attack. International workshop on Peer-To-Peer Systems (2002)
8. Zimmermann P.: PGP Source Code and Internals. The MIT Press (1995)
9. Clarke I., Sandberg O., Wiley B., Hong T. W.: Freenet: A Distributed Anonymous Information Storage and Retrieval System. H. Federrath, Designing Privacy Enhancing Technologies pp. 46-66, Berkeley, USA: Springer-Verlag Berlin Heidelberg (2001)
10. What is a Public Key Infrastructure? http://www.net-security-training.co.uk/what-is-a-public-key-infrastructure/
11. Post A., Shah V., Mislove A.: Bazaar: Strengthening User Reputations in Online Marketplaces. Proceedings of NSDI'11: 8th USENIX Symposium on Networked Systems Design and Implementation, p. 183 (2011)
12. Lamport L., Shostak R., Pease M.: The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems (TOPLAS) 4.3, pp. 382-401 (1982)

13. Huynh T. D., Jennings N. R., Shadbolt N. R.: An Integrated Trust and Reputation Model for Open Multi-Agent Systems. Autonomous Agents and Multi-Agent Systems, 13(2), pp. 119-154 (2006)
14. Michiardi P., Molva R.: Core: a Collaborative Reputation Mechanism to Enforce Node Cooperation in Mobile Ad-hoc Networks. Advanced Communications and Multimedia Security, pp. 107-121, Springer US (2002)
15. Cannon L.: Open Reputation: the Decentralized Reputation Platform (2015) https://openreputation.net/open-reputation-high-level-whitepaper.pdf
16. Grünert A., Hudert S., König S., Kaffille S., Wirtz G.: Decentralized Reputation Management for Cooperating Software Agents in Open Multi-Agent Systems. ITSSA, 1(4), pp. 363-368 (2006)
17. Repantis T., Kalogeraki V.: Decentralized Trust Management for Ad-hoc Peer-to-Peer Networks. Proceedings of the 4th International Workshop on Middleware for Pervasive and Ad-hoc Computing, MPAC 2006, p. 6, ACM (2006)
18. Mui L., Mohtashemi M., Halberstadt A.: A Computational Model of Trust and Reputation. System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference, pp. 2431-2439 IEEE (2002)
19. Commerce B. E., Jøsang A., Ismail R.: The Beta Reputation System. Proceedings of the 15th Bled Electronic Commerce Conference (2002)
20. Suryanarayana G., Erenkrantz J. R., Taylor R. N.: An Architectural Approach for Decentralized Trust Management. IEEE Internet Computing, 9(6), pp. 16-23 (2005)
21. Visan A., Pop F., Cristea V.: Decentralized Trust Management in Peer-to-Peer Systems. 10th International Symposium on Parallel and Distributed Computing, pp. 232-239, IEEE (2011)
22. Suryanarayana G., Diallo M., Taylor R. N.: A Generic Framework for Modeling Decentralized Reputation-Based Trust Models. 14th ACM SigSoft Symposium on Foundations of Software Engineering (2006)
23. Caronni G.: Walking the web of trust. Enabling Technologies: Infrastructure for Collaborative Enterprises, WET ICE 2000, Proceedings, IEEE 9th International Workshops, pp. 153-158 (2000)
24. Penning H.P.: PGP pathfinder pgp.cs.uu.nl
25. Soska K., Kwon A., Christin N., Devadas S.: Beaver: A Decentralized Anonymous Marketplace with Secure Reputation (2016)
26. Zindros D. S.: Trust in Decentralized Anonymous Marketplaces (2015)
27. DeFigueiredo D. D. B., Barr E. T.: TrustDavis: A Non-Exploitable Online Reputation System. CEC, Vol. 5, pp. 274-283 (2005)
28. Fugger R.: Money as IOUs in Social Trust Networks & A Proposal for a Decentralized Currency Network Protocol. http://archive.ripple-project.org/decentralizedcurrency.pdf (2004)
29. Narayanan A., Shmatikov V.: De-anonymizing Social Networks. SP '09 Proceedings of the 2009 30th IEEE Symposium on Security and Privacy, pp. 173-187, 10.1109/SP.2009.22 (2009)
30. Konforty D., Adam Y., Estrada D., Meredith L. G.: Synereo: The Decentralized and Distributed Social Network (2015)
31. Ahuja R. K., Magnanti T. L., Orlin J. B.: Network Flows: Theory, Algorithms, and Applications. Prentice-Hall (1993) https://ocw.mit.edu. License: Creative Commons BY-NC-SA. (Fall 2010)
32. Jøsang A., Ismail R., Boyd C.: A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems, 43(2), pp. 618-644 (2007)