

# Trust Is Risk: A Decentralized Financial Trust Platform

Orfeas Stefanos Thyfronitis Litos<sup>1</sup> and Dionysis Zindros<sup>2,\*</sup>

<sup>1</sup> National Technical University of Athens  
`orfeas.litos@hotmail.com`

<sup>2</sup> University of Athens  
`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 in which 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 decision 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 marketplaces can be roughly categorized as centralized and decentralized. Two major examples are [ebay](#) and [OpenBazaar](#). [ebay](#) is centralized and as such it is vulnerable to ddos attacks [1] and can be considered as a single point of failure, it charges fees for the use of its services [2] and maintains a private database of personal data, but it can give money-back guarantees [3] since it is run by a single company that has a financial advantage in keeping its clients satisfied. On the other hand, [OpenBazaar](#) is a decentralized platform built on bitcoin [6], where individual stores or its [search engine](#) are vulnerable to ddos attacks [1], but not the platform as a whole. Additionally, it does not charge fees for its usage [4] and there

---

\* Research supported by ERC project CODAMODA, project #259152

is no central agent recording all the transactions alongside with private data [4] but it is possible for a buyer or a seller colluding with a moderator to scam the third party and there exists no central authority able to verify the truth of her claim and reimburse her [5]. Even though trust (or distrust) should be directed to each store individually, it is very likely that the whole platform will be discarded as untrusted by the scammed party.

Our goal is to create a distributed marketplace where the trust each user gives to the rest of the users is quantifiable, measurable and expressible in monetary terms. The central concept used throughout this paper is trust-as-risk, or the proposition that a user's *A* trust to another user *B* is defined to be the *maximum sum of money* *A* can lose when *B* is free to choose any strategy she wants. To flesh out trust-as-risk, we will use *lines of credit* as proposed by Washington Sanchez [11]. Joining the network will be done by explicitly entrusting a certain amount of money to another user. If the second user has already entrusted an amount of money to a store, then we trust indirectly the store, thus we can engage in economic interaction with said store.

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 [10]. We will use the notation  $1/\{A, B\}$  to represent a 1-of-2 multisig that can be spent by either of players *A*, *B*.

Our approach drastically changes the user experience. A user no more has to worry about stars and ratings to develop trust towards a store, she can simply check the trust *flowing* from her to the store (which will be a number expressed in bitcoins) and if this number exceeds the price of the product, she is safe to complete the transaction after modifying her direct trust towards her friends in an appropriate way, a process that can be handled by one of several algorithms that we propose, or in any custom way the user chooses to. On the other side, there is no guarantee that the store will complete its part of the transaction and no central authority can reimburse the money. However, it is possible for the defrauded user to make up for her loss by in turn stealing from other users that trust her, an action that will tautologically reduce their trust to her. The fact that this system can function in a completely distributed fashion will become clear in the following sections, along with the positive result that it is in principle Sybil resilient.

There are several incentives for a user to join this network. First of all, they can have access to a store that is otherwise inaccessible. More-

over, 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 its 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 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.

We should note that 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. Also the system we describe needs full network knowledge for all of the users, a situation that may be undesirable because of the privacy-invading implications it brings. Future work is needed to address both limitations.

## 2 A detailed example

The state of the game at any given moment can be represented by a directed weighted graph where each node represents a pseudonymous identity that is considered to correspond to exactly one user and each edge represents a direct trust, the weight of which is the value which the head entrusts the tail with. [Image]

Each node has a corresponding non-negative number that represents its capital. A node's capital is the total value that the node directly and exclusively controls. More technically, it is the number of bitcoins contained in P2PKH transactions in the UTXO of which the private keys are controlled by the node. A rational player would like to maximize her capital in the long term. Let's suppose that player *A* wants to buy a product that costs 10 from player *B*, but *A* has not given any direct trust to *B*. Does *A*'s friends trust *B* enough so that *A* can buy the product? In other words, should *B* decide to steal all its incoming direct trust and stop trusting directly other users and the rest of the players choose the conservative strategy of stealing from their respective incoming trusts just enough to replenish their loss [rationale needed], what would the eventual loss of *A* be? If it can be higher than the value of the product, then *A* is already exposed to a higher risk to *B*, thus paying for the product can be done without *A* being exposed to any additional risk towards *B*, as long as *A*

reduces the direct trust to her friends enough to reduce the risk by the value of the product. If the transaction completes in a satisfying manner,  $A$  can replenish the reduced trust to her friends, or can even choose to start directly trusting  $B$ .

Suppose that at the moment  $A$  is considering of performing a payment to him,  $B$  really decides to depart from the network, taking with him all the coins he can. A chain reaction starts throughout the network because each player that  $B$  exploited is expected to steal enough value from other users that directly trust her to replenish all the value she lost to  $B$ , the newly exploited users will act in the same way and so on, until a new equilibrium is reached (1). Going into detail, [Image]  
(Completion of example)

### 3 Definitions

#### Definition 1 (Graph).

*TrustIsRisk is represented by a sequence of weighted directed graphs  $(\mathcal{G}_j)$  where  $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j)$ ,  $j \in \mathbb{N}$ . Members of  $\mathcal{E}_j$  are tuples of two nodes from  $\mathcal{V}_j$ . More formally,  $e \in \mathcal{E}_j \Rightarrow \exists A, B \in \mathcal{V}_j : e = (A, B)$ . Also, since the graphs are weighted, there exists a sequence of functions  $(c_j)$  with  $c_j : \mathcal{E}_j \rightarrow \mathbb{R}^+$ .*

#### Definition 2 (Players).

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

#### Definition 3 (Capital of $A$ , $Cap_A$ ).

*Total amount of value that exists in P2PKH in the UTXO and can be spent by  $A$ . We also define  $Cap_{A,j}$  as the total amount of value that exists in P2PKH in the UTXO and can be spent by  $A$  during turn  $j$ .*

#### Definition 4 (Direct Trust from $A$ to $B$ After Turn $j$ , $DTr_{A \rightarrow B,j}$ ).

*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  $A$ .*

$$DTr_{A \rightarrow B,j} = \begin{cases} c_j(A, B), & \text{if } (A, B) \in \mathcal{E}_j \\ 0, & \text{if } (A, B) \notin \mathcal{E}_j \end{cases} \quad (1)$$

*A function or algorithm that has access to the graph  $\mathcal{G}_j$  has implicitly access to all direct trusts of this graph. The exception are the oracles, which in this case have access only to their incoming and outgoing direct trusts.*

**Definition 5 ((In/Out) Neighbourhood of  $A$  in Turn  $j$ ).**

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,

$$N^+(A)_j = \{B \in \mathcal{V}_j : DTr_{A \rightarrow B,j} > 0\} . \quad (2)$$

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

$$N^+(S)_j = \bigcup_{A \in S} N^+(A)_j \quad (3)$$

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,

$$N^-(A)_j = \{B \in \mathcal{V}_j : DTr_{B \rightarrow A,j} > 0\} . \quad (4)$$

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

$$N^-(S)_j = \bigcup_{A \in S} N^-(A)_j \quad (5)$$

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,

$$N(A)_j = N^+(A)_j \cup N^-(A)_j . \quad (6)$$

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

$$N(S)_j = N^+(S)_j \cup N^-(S)_j \quad (7)$$

4. Let  $S \subset \mathcal{V}_j$ . Let  $N(A)_{j,i}$  (respectively  $N^+(A)_{j,i}, N^-(A)_{j,i}, N(S)_{j,i}, N^+(S)_{j,i}, N^-(S)_{j,i}$ ) be the  $i$ -th element of set  $N(A)_j$  (respectively of  $N^+(A)_j, N^-(A)_j, N(S)_j, N^+(S)_j, N^-(S)_j$ ), according to an arbitrary but fixed enumeration of the set players.

**Definition 6 (Total Incoming/Outgoing Trust of  $A$  in Turn  $j$ ).**

$$in_{A,j} = \sum_{v \in N^-(A)_j} DTr_{v \rightarrow A,j} \quad (8)$$

$$out_{A,j} = \sum_{v \in N^+(A)_j} DTr_{A \rightarrow v,j} \quad (9)$$

**Definition 7 (Assets of  $A$  in turn  $j$ ,  $As_{A,j}$ ).**

*Sum of  $A$ 's capital and outgoing trust.*

$$As_{A,j} = Cap_{A,j} + out_{A,j} \quad (10)$$

*If the turn we refer to is obvious, it is possible to omit  $j$ .*

**Definition 8 (Turns).**

*The game we are describing is turn-based. In each turn  $j$  exactly one player  $A \in \mathcal{V}$ ,  $A = \text{Player}(j)$ , chooses an action (according to a certain strategy) that can be one of the following, or a finite combination thereof:*

1. *Steal value  $y_B$ ,  $0 \leq y_B \leq DTr_{B \rightarrow A,j-1}$  from  $B \in N^-(A)$ .*

$$DTr_{B \rightarrow A,j} = DTr_{B \rightarrow A,j-1} - y_B \quad (\text{Steal}(y_B, B)) \quad (11)$$

2. *Add value  $y_B$ ,  $-DTr_{A \rightarrow B,j-1} \leq y_B$  to  $B \in \mathcal{V}$ .*

$$DTr_{A \rightarrow B,j} = DTr_{A \rightarrow B,j-1} + y_B \quad (\text{Add}(y_B, B)) \quad (12)$$

*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  $DTr_{A \rightarrow B,j-1} = 0$ , 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*

$$Y_{add} - Y_{st} \leq Cap_{A,j-1} \quad (13)$$

*It holds that*

$$Cap_{A,j} = Cap_{A,j-1} + Y_{st} - Y_{add} \quad (14)$$

*Moreover, player  $A$  is not allowed to choose two actions of the same kind against the same player in the same turn.*

*The set of actions a player makes in turn  $j$  is represented with  $Turn_j$ .*

Let  $A = \text{Player}(j)$ .  $Turn_j$  Examples:

- 1.

$$Turn_j = \emptyset \quad (15)$$

- 2.

$$Turn_j = \{\text{Steal}(y, B), \text{Add}(w, B)\} \quad (16)$$

*given that*

$$DTr_{B \rightarrow A,j_2-1} \leq y \wedge -DTr_{A \rightarrow B,j_2-1} \leq w \wedge y - w \leq Cap_{A,j_2-1} \quad (17)$$

3.

$$Turn_j = \{Steal(x, B), Add(y, C), Add(w, D)\} , \quad (18)$$

given that

$$\begin{aligned} DTr_{B \rightarrow A, j_3-1} &\leq x \wedge -DTr_{A \rightarrow C, j_3-1} \leq y \wedge \\ \wedge -DTr_{A \rightarrow D, j_3-1} &\leq w \wedge x - y - w \leq Cap_{A, j_3-1} . \end{aligned} \quad (19)$$

4.

$$Turn_j = \{Steal(x, B), Steal(y, B)\} \quad (20)$$

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

$$x + y \leq DTr_{B \rightarrow A} , \quad (21)$$

the correct alternative would be

$$Turn_j = \{Steal(x + y, B)\} . \quad (22)$$

**Definition 9 (Previous/Next Turn of a Player).**

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

$$P = \{k \in \mathbb{N} : k < j \wedge Player(k) = A\} \text{ and} \quad (23)$$

$$N = \{k \in \mathbb{N} : k > j \wedge Player(k) = A\} , \quad (24)$$

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

$$prev(j) = \begin{cases} \max P, & P \neq \emptyset \\ 0, & P = \emptyset \end{cases} \quad (25)$$

$$next(j) = \min N \quad (26)$$

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

**Definition 10 ( $A$  is Stolen  $x$ ).**

Let  $j, j'$  be two consecutive turns of  $A$  ( $next(j) = j'$ ). We say that  $A$  has been stolen a value  $x$  between  $j$  and  $j'$  if

$$out_{A,j} - out_{A,j'} > 0 . \quad (27)$$

If turns are not specified, we implicitly refer to the current and the previous turns.

**Definition 11 (History).**

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

$$\mathcal{H}_j = (\text{Player}(j), \text{Turn}_j) \quad (28)$$

**Definition 12 (Conservative Strategy).**

A player  $A$  is said to follow the conservative strategy in turn  $j$  if for any value  $x$  that has been stolen from her since the previous turn she played, she substitutes it in her turn by stealing from others that trust her value equal to  $\min(x, \text{in}_{A,j})$  and she takes no other action. More formally, let  $j' = \text{prev}(j)$ ,  $\text{Damage}_j = \text{out}_{A,j'} - \text{out}_{A,j-1}$  and  $N^-(A)_j = \{v_1, \dots, v_k\}$ . It holds that

$$\begin{aligned} \text{Strategy}(A) = \text{Conservative} &\Rightarrow \forall j \in \mathbb{N} : \text{Player}(j) = A, \\ \text{Turn}_j &= \begin{cases} \emptyset, & \text{Damage}_j \leq 0 \\ \bigcup_{i=1}^k \{\text{Steal}(y_i, v_i)\}, & \text{Damage}_j > 0 \end{cases} \end{aligned} \quad (29)$$

In the second case, it holds that

$$\sum_{i=1}^k y_i = \min(\text{in}_{A,j-1}, \text{Damage}_j) \quad (30)$$

As we can see, the definition covers a multitude of options for the conservative player, since in case  $0 < \text{Damage}_j < \text{in}_{A,j-1}$  she can choose to distribute the  $\text{Steal}(s)()$  in any way she chooses, as long as

$$\forall i, y_i \leq \text{DTr}_{N^-(A)_{j,i} \rightarrow A,j-1} \wedge \sum_{i=1}^{|N^-(A)_j|} y_i = \text{Damage}_j \quad (31)$$

The oracle remembers  $\text{PrevOutTrust} = \text{out}_{A,j'}$  for  $j' = \text{prev}(j)$  and can observe all incoming and outgoing direct trusts of player  $A$ . We note that  $N(A)_{j-1} = N(A)_j$ .

**Conservative Oracle**

Input : previous graph  $\mathcal{G}_{j-1}$

Output :  $\text{Turn}_j$

$\text{consOracle}(\mathcal{G}_{j-1}) :$

$$1 \quad \text{NewOutTrust} = \sum_{v \in N^+(A)_{j-1}} \text{DTr}_{A \rightarrow v,j-1}$$



```

2   NewInTrust =  $\sum_{v \in N^-(A)_{j-1}} DTr_{v \rightarrow A, j-1}$ 
3   Damage = PrevOutTrust - NewOutTrust
4   if (Damage > 0)
5       if (Damage >= NewInTrust)
6           Turnj =  $\emptyset$ 
7           for ( $v \in N^-(A)_{j-1}$ )
8               Turnj = Turnj  $\cup$  {Steal( $DTr_{v \rightarrow A, j-1}, v$ )}
9       else
10          ( $y_1, \dots, y_{|N^-(A)_{j-1}|}$ ) = SelectSteal( $DTr_{N^-(A)_{j-1,1} \rightarrow A, j-1},$ 
11           $\dots, DTr_{N^-(A)_{j-1, |N^-(A)_{j-1}|} \rightarrow A, j-1},$  Damage)
12          Turnj =  $\emptyset$ 
13          for ( $i \in [|N^-(A)_{j-1}|]$ )
14              Turnj = Turnj  $\cup$  {Steal( $y_i, N^-(A)_{j-1, i}$ )}
15      else
16          Turnj =  $\emptyset$ 
17      return(Turnj)

```

SelectSteal() returns  $y_i$  with  $i \in [|N^-(A)_j|]$  such that

$$\sum_{i=1}^{|N^-(A)_j|} y_i = \text{Damage} \wedge \forall i, y_i \leq DTr_{N^-(A)_{j,i} \rightarrow A} . \quad (32)$$

**Definition 13 (Idle strategy).**

A player  $A$  is said to follow the idle strategy if she passes in her turn. More formally, it holds that

$$\text{Strategy}(A) = \text{Idle} \Rightarrow \forall j \in \mathbb{N} : \text{Player}(j) = A, \text{Turn}_j = \emptyset . \quad (33)$$

Idle Oracle

Input : previous graph  $\mathcal{G}_{j-1}$

Output : Turn<sub>j</sub>

idleOracle( $\mathcal{G}_{j-1}$ ) :

1 return( $\emptyset$ )

**Definition 14 (Evil strategy).**

A player  $A$  is said to follow the evil strategy if she steals value

$$y_v = DTr_{v \rightarrow A, j-1} \forall v \in N^-(A)_j \quad (34)$$

(steals all incoming direct trust) and reduces her trust to  $w$  by

$$DTr_{A \rightarrow w, j-1} \forall w \in N^+(A)_j \quad (35)$$

(nullifies her outgoing direct trust) in her turn. More formally, it holds that

$$\begin{aligned} \text{Strategy}(A) = \text{Evil} &\Rightarrow \forall j \in \mathbb{N} : \text{Player}(j) = A, \\ \text{Turn}_j &= \left( \bigcup_{v \in N^-(A)_{j-1}} \{ \text{Steal}(DTr_{v \rightarrow A, j-1}, v) \} \right) \cup \\ &\quad \left( \bigcup_{w \in N^+(A)_{j-1}} \{ \text{Add}(-DTr_{A \rightarrow w, j-1}, w) \} \right). \end{aligned} \quad (36)$$

We again note that  $N(A)_{j-1} = N(A)_j$ .

**Evil Oracle**

**Input** : previous graph  $\mathcal{G}_{j-1}$

**Output** :  $\text{Turn}_j$

**evilOracle**( $\mathcal{G}_{j-1}$ ) :

```

1   $\text{Turn}_j = \emptyset$ 
2  for ( $v \in N^-(A)_{j-1}$ )
3     $\text{Turn}_j = \text{Turn}_j \cup \{ \text{Steal}(DTr_{v \rightarrow A, j-1}, v) \}$ 
4  for ( $w \in N^+(A)_{j-1}$ )
5     $\text{Turn}_j = \text{Turn}_j \cup \{ \text{Add}(-DTr_{A \rightarrow w, j-1}, w) \}$ 
6  return( $\text{Turn}_j$ )
```

**Definition 15 (Indirect Trust from  $A \in \mathcal{V}_j$  to  $B \in \mathcal{V}_j$ ,  $Tr_{A \rightarrow B, j}$ ).**

Maximum possible value that can be stolen from  $A$  if  $B$  follows the evil strategy,  $A$  follows the idle strategy and everyone else ( $\mathcal{V} \setminus \{A, B\}$ ) follows the conservative strategy. More formally, if

$$\begin{aligned} \text{Strategy}(A) &= \text{Idle} \wedge \text{Strategy}(B) = \text{Evil} \wedge \\ &\wedge \forall v \in \mathcal{V} \setminus \{A, B\}, \text{Strategy}(v) = \text{Conservative} \end{aligned} \quad (37)$$

and choices are the different actions between which the conservative players can choose, then

$$Tr_{A \rightarrow B, j} = \max_{j' : j' > j, \text{choices}} [\text{out}_{A, j} - \text{out}_{A, j'}] \quad (38)$$

**Definition 16 (Indirect Trust from  $A \in \mathcal{V}_j$  to  $S \subset \mathcal{V}_j$ ,  $Tr_{A \rightarrow S, j}$ ).**  
*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, if  $S \subset \mathcal{V}$ ,*

$$\begin{aligned} Strategy(A) &= Idle \wedge \forall E \in S, Strategy(E) = Evil \wedge \\ &\wedge \forall v \in \mathcal{V} \setminus (S \cup \{A\}), Strategy(v) = Conservative \end{aligned} \quad (39)$$

*and choices are the different actions between which the conservative players can choose, then*

$$Tr_{A \rightarrow S, j} = \max_{j': j' > j, configurations} [out_{A, j} - out_{A, j'}] \quad (40)$$

**Definition 17 (Collusion).**

*A collusion of players  $S \subset \mathcal{V}$  is a set of players that is entirely controlled by a single entity in the physical world. From a game theoretic point of view, other players ( $v \in \mathcal{V} \setminus S$ ) 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.*

## 4 Theorems-Algorithms

The following algorithm has read access to direct trusts in  $\mathcal{G}_{j-1}$  and write access to direct trusts in  $\mathcal{G}_j$ .

**Execute Turn**

Input : old graph  $\mathcal{G}_{j-1}$ , player  $A \in \mathcal{V}_{j-1}$ , old capital  $Cap_{A, j-1}$ , ProvisionalTurn

Output : new graph  $\mathcal{G}_j$ , new capital  $Cap_{A, j}$ , new history  $\mathcal{H}_j$

```

1 executeTurn( $\mathcal{G}_{j-1}$ ,  $A$ ,  $Cap_{A, j-1}$ , ProvTurn) :
2   ( $Turn_j$ , NewCap) = validateTurn( $\mathcal{G}_{j-1}$ ,  $A$ ,  $Cap_{A, j-1}$ ,
   ProvisionalTurn)
3   return(commitTurn( $\mathcal{G}_{j-1}$ ,  $A$ ,  $Turn_j$ , NewCap))

```

**Validate Turn**

Input : old graph  $\mathcal{G}_{j-1}$ , player  $A \in \mathcal{V}_{j-1}$ , old capital  $Cap_{A, j-1}$ , ProvisionalTurn

Output :  $Turn_j$ , new capital  $Cap_{A, j}$

```

1 validateTurn( $\mathcal{G}_{j-1}$ ,  $A$ ,  $Cap_{A, j-1}$ , ProvisionalTurn) :
2    $Y_{st} = 0$ 
3    $Y_{add} = 0$ 

```

```

4   for (action ∈ ProvisionalTurn)
5       action match do
6           case Steal(y,w) do
7               if (y > DTrw→A,j-1 || y < 0)
8                   return(∅, CapA,j-1)
9               else
10                  Yst = Yst + y
11            case Add(y,w) do
12                if (y < -DTrA→w,j-1
13                    return(∅, CapA,j-1)
14                else
15                    Yadd = Yadd + y
16            if (Yadd - Yst > CapA,j-1)
17                return(∅, CapA,j-1)
18            else
19                return(ProvisionalTurn, CapA,j-1 + Yst - Yadd)

```

#### Commit Turn

Input : player old graph  $\mathcal{G}_{j-1}$ ,  $A \in \mathcal{V}_{j-1}$ ,  $Turn_j$ , NewCap

Output : new graph  $\mathcal{G}_j$ , new capital  $Cap_{A,j}$ , new history  $\mathcal{H}_j$

```

1   commitTurn( $\mathcal{G}_{j-1}$ , A, Turnj, NewCap) :
2       for ((v, w) ∈  $\mathcal{E}_j$ )
3           DTrv→w,j = DTrv→w,j-1
4       for (action ∈ Turnj)
5           action match do
6               case Steal(y,w) do
7                   DTrw→A,j = DTrw→A,j-1 - y
8               case Add(y,w) do
9                   DTrA→w,j = DTrA→w,j + y
10          CapA,j = NewCap
11           $\mathcal{H}_j$  = (Player(j), Turnj)
12          return( $\mathcal{G}_j$ , CapA,j,  $\mathcal{H}_j$ )

```

#### TrustIsRisk Game

```

1   j = 0
2   while (True)
3       j = j + 1
4       v =  $\mathcal{V}_j$ 
5       ProvisionalTurn = vOracle( $\mathcal{G}_{j-1}$ )
6       ( $\mathcal{G}_j$ , Capv,j,  $\mathcal{H}_j$ ) = executeTurn( $\mathcal{G}_{j-1}$ , v, Capv,j-1,

```

We continue with an 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. This convention will hold whenever we use the Transitive Game.

#### Transitive Game

Input : graph  $\mathcal{G}_0$ ,  $A \in \mathcal{V}_0$  idle player,  $E \in \mathcal{V}_0$  evil player

Output : history  $\mathcal{H}$

```

1 Angry =  $\emptyset$ 
2 Happy =  $\emptyset$ 
3 Sad =  $\emptyset$ 
4 for ( $v \in \mathcal{V}_0 \setminus \{E\}$ )
5    $Loss_v = 0$ 
6   if ( $v \neq A$ )
7     Happy = Happy  $\cup \{v\}$ 
8 j = 0
9 while (True)
10  j = j + 1
11   $v = \mathcal{V}_j \setminus \{A\}$ 
12   $Turn_j = vOracle(\mathcal{G}_{j-1})$  #vOracle() = idleOracle()
13  executeTurn( $\mathcal{G}_{j-1}$ ,  $Cap_{v,j-1}$ ,  $Turn_j$ )
14  for (action  $\in Turn_j$ )
15    action match do
16      case Steal(y, w) do
17        exchange = y
18         $Loss_w = Loss_w + exchange$ 
19        if ( $v \neq E$ )
20           $Loss_v = Loss_v - exchange$ 
21        if ( $w \neq A$ )
22          Happy = Happy  $\setminus \{w\}$ 
23          if ( $in_{w,j} == 0$ )
24            Sad = Sad  $\cup \{w\}$ 
25          else
26            Angry = Angry  $\cup \{w\}$ 
27  Angry = Angry  $\setminus \{v\}$ 
28  if ( $in_{v,j} = 0 \ \&\& \ Loss_v > 0$ )
29    Sad = Sad  $\cup \{v\}$ 
30  if ( $Loss_v = 0$ )
31    Happy = Happy  $\cup \{v\}$ 

```

We observe that the set of players is not modified throughout the Transitive Game, thus we can refer to  $\mathcal{V}_j$  for any turn  $j$  as  $\mathcal{V}$  in any setting where we use the Transitive Game.

Let  $j_0$  be the first turn on which  $E$  is chosen to play. Until then, according to theorem 3, all players will pass their turn. Moreover, let  $j' = \text{prev}(j)$  and  $\{v_1, \dots, v_k\} = N^-(A)_j$ . Given that

$$\text{Damage}_{v,j} = \text{out}_{v,j'} - \text{out}_{v,j} \quad , \quad (41)$$

the algorithm generates turns:

$$\text{Turn}_j = \begin{cases} \emptyset, & \text{Damage}_{v,j-1} = 0 \\ \bigcup_{i=1}^k \{\text{Steal}(y_i, v_i)\}, & \text{Damage}_j > 0 \end{cases} . \quad (42)$$

In the second case, it is

$$\sum_{i=1}^k y_i = \min(\text{in}_{v,j-1}, \text{Damage}_{v,j-1}) \quad . \quad (43)$$

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

**Lemma 1 (Loss Equivalent to Damage).**

Let  $j \in \mathbb{N}, v \in \mathcal{V}_j \setminus \{A, E\}, v = \text{Player}(j)$ . It holds that

$$\min(\text{in}_{v,j}, \text{Loss}_{v,j}) = \min(\text{in}_{v,j}, \text{Damage}_{v,j}) \quad . \quad (44)$$

*Proof.* – Let  $v \in \text{Happy}_{j-1}$ . Then

1.  $v \in \text{Happy}_j$  because  $\text{Turn}_j = \emptyset$ ,
2.  $\text{Loss}_{v,j} = 0$  because otherwise  $v \notin \text{Happy}_j$ ,
3.  $\text{Damage}_{v,j} = 0$ , or else any reduction in direct trust to  $v$  would increase equally  $\text{Loss}_{v,j}$  (line 18), which cannot be decreased again but during an Angry player's turn (line 20).
4.  $\text{in}_{v,j} \geq 0$

Thus

$$v \in \text{Happy}_{j-1} \Rightarrow \min(\text{in}_{v,j}, \text{Damage}_{v,j}) = \min(\text{in}_{v,j}, \text{Loss}_{v,j}) = 0 \quad . \quad (45)$$

– Let  $v \in \text{Sad}_{j-1}$ . Then

1.  $v \in Sad_j$  because  $Turn_j = \emptyset$ ,
2.  $in_{v,j} = 0$  (lines 28 - 29),
3.  $Damage_{v,j} \geq 0 \wedge Loss_{v,j} \geq 0$ .

Thus

$$v \in Sad_{j-1} \Rightarrow \min(in_{v,j}, Damage_{v,j}) = \min(in_{v,j}, Loss_{v,j}) = 0 . \quad (46)$$

- Let  $v \in Angry_{j-1} \wedge v \in Happy_j$ . Then the same argument as in the first case holds, if we ignore the argument (1).
- Let  $v \in Angry_{j-1} \wedge v \in Sad_j$ . Then the same argument as in the second case holds, if we ignore the argument (1).

□

**Theorem 1 (Trust Convergence Theorem).**

*Consider a Transitive Game and let  $j_0$  be the first turn that  $E$  is chosen to play. Given that all players eventually play, there exists a turn  $j' > j_0$  such that*

$$\forall j \geq j', Turn_j = \emptyset . \quad (47)$$

*Proof Sketch.* If the game didn't converge, the *Steal* () actions would continue forever, thus they would reach infinity. However this is impossible, since there exists only finite total trust.

*Proof.* First of all,

$$\forall j > j_0 : Player(j) = E \Rightarrow Turn_j = \emptyset \quad (48)$$

because  $E$  has already nullified his 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  $E$ , thus player  $E$  can do nothing  $\forall j > j_0$ . We also see that

$$\forall j > j_0 : Player(j) = A \Rightarrow Turn_j = \emptyset \quad (49)$$

because of the idle strategy that  $A$  follows. As far as the rest of the players are concerned, consider the Transitive Game. As we can see from lines 5 and 18 - 20, it is

$$\forall j, \sum_{v \in \mathcal{V}_j} Loss_v = in_{E,j_0-1} . \quad (50)$$

In other words, the total loss is constant and equal to the total value stolen by  $E$ . Also, as we can see in lines 3 and 29, 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

$$\exists j' : \forall j \geq j', |Angry_j \cup Happy_j| = c > 0 \wedge Angry_j \neq \emptyset . \quad (51)$$

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 29), 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

$$\forall j \in \mathbb{N}, DTr_j = \sum_{w, w' \in \mathcal{V}} DTr_{w \rightarrow w', j} . \quad (52)$$

Also, without loss of generality, suppose that

$$\forall j \geq j_1, out_{A,j} = out_{A,j_1} . \quad (53)$$

In  $Turn_{j_1}$ ,  $v$  steals

$$St = \sum_{i=1}^m y_i . \quad (54)$$

We will show using induction that

$$\forall n \in \mathbb{N}, \exists j_n \in \mathbb{N} : DTr_{j_n} \leq DTr_{j_1-1} - nSt . \quad (55)$$

Base case: It holds that

$$DTr_{j_1} = DTr_{j_1-1} - St . \quad (56)$$



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

$$DTr_{j_2} \leq DTr_{j_1} - St = DTr_{j_1-1} - 2St , \quad (57)$$

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

Induction hypothesis: Suppose that

$$\exists k > 1 : j_k > j_{k-1} > j_1 \Rightarrow DTr_{j_k} \leq DTr_{j_{k-1}} - St . \quad (58)$$

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

$$DTr_{j_{k+1}} \leq DTr_{j_k} - St . \quad (59)$$

We have proven by induction that

$$\forall n \in \mathbb{N}, \exists j_n \in \mathbb{N} : DTr_{j_n} \leq DTr_{j_1-1} - nSt . \quad (60)$$

However

$$DTr_{j_1-1} \geq 0 \wedge St > 0 , \quad (61)$$

thus

$$\exists n' \in \mathbb{N} : n'St > DTr_{j_1-1} \Rightarrow DTr_{j_{n'}} < 0 . \quad (62)$$

We have a contradiction because

$$\forall w, w' \in \mathcal{V}, \forall j \in \mathbb{N}, DTr_{w \rightarrow w', j} \geq 0 , \quad (63)$$

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

In games where there exists one evil  $E$ , one idle player  $A$  and the rest of the players are conservative, we define  $Loss_A = Loss_{A,j}$ , 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 will steal and how much from each.

Let  $G$  be a weighted directed graph. According to [8] p. 709, if we consider each edge's capacity as its weight ( $\forall e \in E(G), c_e = c(e)$ ), we say that a flow assignment  $X = [x_{vw}]_{V(FG) \times V(FG)}$  with a source  $A$  and a sink  $B$  is valid when:

$$\forall (v, w) \in E(FG), x_{vw} \leq c_{vw} \quad (64)$$

and

$$\forall v \in V(FG) \setminus \{A, B\}, \quad \sum_{w \in N^+(v)} x_{wv} = \sum_{w \in N^-(v)} x_{vw} . \quad (65)$$

**Lemma 2 (MaxFlows Are Transitive Games).**

Let  $\mathcal{G}_{j_0}$  be a game graph at a specific turn  $j_0$ , let  $A, E \in \mathcal{V}_{j_0}$  and  $\text{maxFlow}(A, E)$  the maximum flow from  $A$  to  $E$  executed on  $\mathcal{G}_{j_0}$ . There exists an execution of  $\text{TransitiveGame}(\mathcal{G}_{j_0}, A, E)$  such that

$$\text{maxFlow}(A, E) \leq \text{Loss}_A . \quad (66)$$

*Proof.*

Without loss of generality, we suppose that the interesting turn is 0. In other words,  $j_0 = 0$ . Let  $X = \{x_{vw}\}_{\mathcal{V} \times \mathcal{V}}$  be the flows returned by the execution of the  $\text{MaxFlow}(A, B)$  algorithm on  $\mathcal{G}_0$ . It is known that for any directed weighted graph  $G$  there exists a  $\text{MaxFlow}$  over  $G$  that is a DAG [citation needed]. We also know that we can apply the topological sort algorithm to any DAG and obtain a total ordering of its nodes with the following property:  $\forall$  nodes  $v, w$ , it holds that  $v < w \Rightarrow x_{wv} = 0$  [citation needed]. Put differently, there is no flow from larger to smaller nodes. We execute the topological sort on  $X$  and obtain a total order of the nodes, such that  $B$  is the maximum and  $A$  is the minimum node.  $B$  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 algorithm 4 will choose players following the total order, starting from player  $B$ . We observe that  $\forall v \in \mathcal{V} \setminus \{A, B\}, \sum_{w \in \mathcal{V}} x_{wv} = \sum_{w \in \mathcal{V}} x_{vw} \leq \text{maxFlow}(A, B) \leq \text{in}_{B,0}$ . Player  $B$  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  $v = \text{Player}(j)$  will have stolen value  $x_{wv}$  from each in neighbour  $w$ . Base case: In turn 1,  $B$  steals value equal to  $\sum_{w \in \mathcal{V}} x_{wB}$ , following the modified evil strategy.

$$\text{Turn}_1 = \bigcup_{v \in N^-(B)_0} \{\text{Steal}(x_{vB}, v)\} \quad (67)$$

Induction hypothesis: Let  $k \in [j_2 - 2]$ . We suppose that  $\forall j \in [k]$ , there exists a valid set of actions,  $\text{Turn}_j$ , performed by  $v = \text{Player}(j)$  such

that  $v$  steals from each player  $w$  value equal to  $x_{wv}$ .

$$\forall j \in [k], \text{Turn}_j = \bigcup_{w \in N^-(v)_{j-1}} \{\text{Steal}(x_{wv}, w)\} \quad (68)$$

Induction step: Let  $j = k + 1, v = \text{Player}(j)$ . 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,

$\forall w \in N^-(v), \text{DTr}_{w \rightarrow v, j-1} = \text{DTr}_{w \rightarrow v, 0} \geq x_{wv}$ , thus  $v$  is able to choose the following turn:

$$\text{Turn}_j = \bigcup_{w \in N^-(v)_{j-1}} \{\text{Steal}(x_{wv}, w)\} \quad (69)$$

Moreover, this turn satisfies the conservative strategy since

$$\sum_{w \in N^-(v)_{j-1}} x_{wv} = \sum_{w \in N^+(v)_{j-1}} x_{vw} . \quad (70)$$

Thus  $\text{Turn}_j$  is a valid turn for the conservative player  $v$ .

We have proven that in the end of turn  $j_2 - 1$ , player  $B$  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  $\text{maxFlow}(A, B)$ . Since there remains no Angry player, it is obvious that  $j_2$  is a turn that Transitive Game has converged thus  $\text{Loss}_{A, j_2} = \text{Loss}_A$ . It is also obvious that if  $B$  had chosen the original evil strategy, the described actions would still be valid only by supplementing them with additional  $\text{Steal}()$  actions, thus  $\text{Loss}_A$  would further increase. We have proven that there exists a valid execution of the Transitive Game where

$$\text{Loss}_A \geq \text{maxFlow}(A, B) . \quad (71)$$

□

### Lemma 3 (Transitive Games Are Flows).

Let  $\mathcal{H} = \text{TransitiveGame}(\mathcal{G}, A, B)$  for some game graph  $\mathcal{G}$  and  $A, B \in \mathcal{V}$ . Let *Sad*, *Happy*, *Angry* be as defined in the Transitive Game. There exists a valid flow  $X = \{x_{wv}\}_{\mathcal{V} \times \mathcal{V}}$  on  $\mathcal{G}_0$  such that

$$\sum_{v \in \mathcal{V}} x_{Av} = \text{Loss}_A . \quad (72)$$

*Proof.* 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:

$$\mathcal{V}' = \mathcal{V} \cup \{T\} \quad (73)$$

$$\mathcal{E}' = \mathcal{E} \cup \{(T, A)\} \cup \{(T, v) : v \in \text{Sad}_{j_1}\} \quad (74)$$

$$\forall (v, w) \in \mathcal{E}, c'_{vw} = DTr_{v \rightarrow w, 0} - DTr_{v \rightarrow w, j_1} \quad (75)$$

$$\forall v \in \text{Sad}_{j_1}, c'_{Tv} = c'_{TA} = \infty \quad (76)$$

(Image) We observe that  $\forall v \in \mathcal{V}$ ,

$$\sum_{w \in N^-(v) \setminus \{T\}} c'_{vw} = \sum_{w \in N^-(v) \setminus \{T\}} (DTr_{w \rightarrow v, 0} - DTr_{w \rightarrow v, j_1}) = in_{v, 0} - in_{v, j_1} \quad (77)$$

and

$$\sum_{w \in N^+(v) \setminus \{T\}} c'_{vw} = \sum_{w \in N^+(v) \setminus \{T\}} (DTr_{v \rightarrow w, 0} - DTr_{v \rightarrow w, j_1}) = out_{v, 0} - out_{v, j_1} . \quad (78)$$

We can suppose that

$$\forall j \in \mathbb{N}, in_{A, j} = 0 , \quad (79)$$

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  $MaxFlow(T, B) = X'$  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 (64)  $\forall e \in \mathcal{E}$ )
- Since  $\forall v \in \text{Sad}_{j_1} \cup \{A\}, c'_{Tv} = \infty$ , requirement (64) holds for any flow  $x'_{Tv} \geq 0$ .
- Let  $v \in \mathcal{V}' \setminus (\text{Sad}_{j_1} \cup \{T, A, B\})$ . According to the conservative strategy and since  $v \notin \text{Sad}_{j_1}$ , it holds that

$$out_{v, 0} - out_{v, j_1} = in_{v, 0} - in_{v, j_1} . \quad (80)$$

Combining this observation with (77) and (78), we have that

$$\sum_{w \in \mathcal{V}'} c'_{vw} = \sum_{w \in \mathcal{V}'} c'_{wv} . \quad (81)$$

(Flow Conservation requirement (65)  $\forall v \in \mathcal{V}' \setminus (\text{Sad}_{j_1} \cup \{T, A, B\})$ )

– Let  $v \in Sad_{j_1}$ . Since  $v$  is sad, we know that

$$out_{v,0} - out_{v,j_1} > in_{v,0} - in_{v,j_1} . \quad (82)$$

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

$$x'_{Tv} = (out_{v,0} - out_{v,j_1}) - (in_{v,0} - in_{v,j_1}) . \quad (83)$$

In this way, we have

$$\sum_{w \in \mathcal{V}'} x'_{vw} = out_{v,0} - out_{v,j_1} \text{ and} \quad (84)$$

$$\begin{aligned} \sum_{w \in \mathcal{V}'} x'_{vw} &= \sum_{w \in \mathcal{V}' \setminus \{T\}} c'_{vw} + x'_{Tv} = in_{v,0} - in_{v,j_1} + \\ &+ (out_{v,0} - out_{v,j_1}) - (in_{v,0} - in_{v,j_1}) = out_{v,0} - out_{v,j_1} . \end{aligned} \quad (85)$$

thus

$$\sum_{w \in \mathcal{V}'} x'_{vw} = \sum_{w \in \mathcal{V}'} x'_{vw} . \quad (86)$$

(Requirement 65  $\forall v \in Sad_{j_1}$ )

– We set

$$x'_{TA} = \sum_{v \in \mathcal{V}'} x'_{Av} , \quad (87)$$

thus from (79) we have

$$\sum_{v \in \mathcal{V}'} x'_{vA} = \sum_{v \in \mathcal{V}'} x'_{Av} . \quad (88)$$

(Requirement 65 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  $maxFlow(T, B)$  because all incoming flows to  $B$  are saturated. Also we observe that

$$\sum_{v \in \mathcal{V}'} x'_{Av} = \sum_{v \in \mathcal{V}'} c'_{Av} = out_{A,0} - out_{A,j_1} = Loss_A . \quad (89)$$

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

$$\mathcal{V}'' = \mathcal{V}' \quad (90)$$

$$E(\mathcal{G}'') = E(\mathcal{G}') \setminus \{(T, v) : v \in Sad_j\} \quad (91)$$

$$\forall e \in E(\mathcal{G}''), c_e'' = c_e' \quad (92)$$

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

$$\sum_{v \in \mathcal{V}''} x_{Tv}'' = x_{TA}'' = \sum_{v \in \mathcal{V}''} x_{Av}'' . \quad (93)$$

The outgoing flow from  $A$  in  $X''$  will remain the same as in  $X'$  for two reasons:

No capacity reachable by  $A$  is modified and  $T$  has no incoming flow, thus

$$\sum_{v \in \mathcal{V}''} x_{Av}'' \geq \sum_{v \in \mathcal{V}'} x_{Av}' [\text{citation needed}] \text{ and} \quad (94)$$

$$\left. \begin{array}{l} \sum_{v \in \mathcal{V}''} c_{Av}'' = \sum_{v \in \mathcal{V}'} c_{Av}' = \sum_{v \in \mathcal{V}'} x_{Av}' \\ \sum_{v \in \mathcal{V}''} c_{Av}'' \geq \sum_{v \in \mathcal{V}''} x_{Av}'' \end{array} \right\} \Rightarrow \sum_{v \in \mathcal{V}''} x_{Av}'' \leq \sum_{v \in \mathcal{V}'} x_{Av}' \quad (95)$$

Thus we conclude that

$$\sum_{v \in \mathcal{V}''} x_{Av}'' = \sum_{v \in \mathcal{V}'} x_{Av}' . \quad (96)$$

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

$$\sum_{v \in \mathcal{V}''} x_{Av}'' = \sum_{v \in \mathcal{V}} x_{Av} . \quad (97)$$

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

$$\forall e \in \mathcal{E}, c_e \geq c_e'' . \quad (98)$$

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

$$\sum_{v \in \mathcal{V}} x_{Av} = \sum_{v \in \mathcal{V}''} x_{Av}'' \stackrel{(96)}{=} \sum_{v \in \mathcal{V}'} x_{Av}' \stackrel{(89)}{=} Loss_{A,j_1} , \quad (99)$$

which is the flow  $X$ . □

**Theorem 2 (Trust flow theorem).**

*Let  $\mathcal{G}$  be a game graph and  $A, B \in \mathcal{V}$ . It holds that*

$$Tr_{A \rightarrow B} = maxFlow(A, B) . \quad (100)$$

*Proof.*

From lemma 2 we see that there exists an execution of the Transitive Game such that

$$Loss_A = maxFlow(A, B) . \quad (101)$$

Since  $Tr_{A \rightarrow B}$  is the maximum loss that  $A$  can suffer after the convergence of the Transitive Game, we see that

$$Tr_{A \rightarrow B} \geq maxFlow(A, B) . \quad (102)$$

Moreover, there exists an execution of the Transitive Game such that

$$Tr_{A \rightarrow B} = Loss_A . \quad (103)$$

From lemma 3, this execution corresponds to a flow. Thus

$$Tr_{A \rightarrow B} \leq maxFlow(A, B) . \quad (104)$$

We conclude that

$$Tr_{A \rightarrow B} = maxFlow(A, B) . \quad (105)$$

□

**Theorem 3 (Conservative World Theorem).**

*If everybody follows the conservative strategy, nobody steals any amount from anybody.*

*Proof.*

Suppose that we are interested in graphs  $\mathcal{G}_j$ .

Let  $(j_k)$  an increasing sequence of positive integers,

$$\text{let } S_{j_k} \subseteq N^-(Player(j_k))_{j_k} \text{ and} \quad (106)$$

$$\text{let } \forall v \in S_{j_k}, y_{v, j_k} > 0 .$$

Suppose that there exists a subseries of History,  $(Turn_{j_k})$ , where

$$Turn_{j_k} = \bigcup_{v \in S_{j_k}} \{Steal(y_{v, j_k}, v)\} , \quad (107)$$

This subseries must have an initial element,  $Turn_{j_1}$ . However,  $Player(j_1)$  follows the conservative strategy, thus somebody must have stolen from her as well, so  $Player(j_1)$  cannot be the initial element. We have a contradiction, thus there cannot exist a series of stealing actions when everybody is conservative. □

Note: The  $\text{maxFlow}$  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.

**Theorem 4 (Trust to Multiple Players).**

Let  $S \subset \mathcal{V}$  and  $T$  auxiliary player such that

$$\forall B \in S, DTr_{B \rightarrow T} = \infty . \quad (108)$$

It holds that

$$\forall A \in \mathcal{V} \setminus S, Tr_{A \rightarrow S} = \text{maxFlow}(A, T) . \quad (109)$$

*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, thus they will act in a way identical to following the evil strategy as far as  $\text{MaxFlow}$  is concerned, thus, according to the Trust Flow Theorem, it holds that

$$\forall A \in \mathcal{V} \setminus S, Tr_{A \rightarrow T} = \text{maxFlow}(A, T) = Tr_{A \rightarrow S} . \quad (110)$$

□

One of the primary aims of this system is to mitigate the danger for sybil attacks whilst maintaining fully decentralized autonomy. Let Eve be a possible attacker. 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 invest any amount she chooses, thus she can arbitrarily set the direct trusts of any player  $C \in \mathcal{C}$  to any player  $P \in \mathcal{V}$  and can also steal all incoming direct trust to these players. Additionally, we give Eve a set of players  $B \in \mathcal{B}$  that she has corrupted (the corrupted set), so she fully controls their direct trusts to any player  $P \in \mathcal{V}$  and can also steal all incoming direct trust to these players. The players  $B \in \mathcal{B}$  are considered to be legitimate before the corruption, thus they may be directly trusted by any player  $P \in \mathcal{V}$ . However, players  $C \in \mathcal{C}$  can be trusted only by players  $D \in \mathcal{B} \cup \mathcal{C}$  and not by players  $A \in \mathcal{V} \setminus (\mathcal{B} \cup \mathcal{C})$ . In short,

$$\forall D \in \mathcal{B} \cup \mathcal{C}, P \in \mathcal{V}, \text{ Eve controls } DTr_{D \rightarrow P}, \text{ can steal } DTr_{P \rightarrow D} \quad (111)$$

and can change  $|\mathcal{C}|$ . Furthermore, the following restriction true:

$$\forall C \in \mathcal{C}, P \in \mathcal{V}, DTr_{P \rightarrow C} = 0 . \quad (112)$$



**Lemma 4 (No Evil Edges in the *MinCut*).**

Let  $S \subset \mathcal{V}$ ,  $A \notin S$ . When calculating  $\text{MaxFlow}(A, S)$ , it is impossible to have an edge  $(v, w) \in \text{MinCut} : v \in S$ .

*Proof.* Let  $T$  be the auxiliary node. It is

$$\forall v \in S, c_{vT} = \infty . \quad (113)$$

We can see that  $\text{out}_A < \infty$  and thus

$$\text{maxFlow}(A, S) < \infty . \quad (114)$$

Since all edges in the *MinCut* are saturated and due to (114), we have

$$\nexists v \in S : (v, T) \in \text{MinCut} . \quad (115)$$

Suppose that

$$\exists v \in S, w \in \mathcal{V} : (v, w) \in \text{MinCut} . \quad (116)$$

Then this edge must be saturated, that is  $x_{vw} = c_{vw} > 0$ . However, there exists an alternative flow configuration  $X'$  where

$$\begin{aligned} \forall (u, u') \in \mathcal{E} \setminus \{(v, w), (v, T)\}, x'_{u,u'} &= x_{u,u'} , \\ x'_{vw} &= 0 \text{ and} \\ x'_{vT} &= x_{vT} + x_{vw} , \end{aligned} \quad (117)$$

which is valid because

$$\left. \begin{aligned} \sum_{w \in N^+(v)} x_{vw} &= \sum_{w \in N^+(v)} x'_{vw} \\ c_{vT} &= \infty \end{aligned} \right\} \Rightarrow x'_{vT} \leq c_{vT} \quad (118)$$

and  $X'$  is maximum as well because it carries exactly the same flow as  $X$ . Thus

$$(v, w) \notin \text{MinCut} . \quad (119)$$

□

**Theorem 5 (Sybil Resilience).**

Let  $\mathcal{B} \cup \mathcal{C} \subset \mathcal{V}$  with  $\mathcal{B} \cap \mathcal{C} = \emptyset$  be a collusion of players who are controlled by an adversary, Eve. Eve also controls the number of players in the Sybil set  $\mathcal{C}$ ,  $|\mathcal{C}|$ . Players  $C \in \mathcal{C}$  are not directly trusted by players outside the collusion, contrary to players  $B \in \mathcal{B}$ , the corrupted set, who may be directly trusted by any player in  $\mathcal{V}$ . It then holds that

$$\text{Tr}_{A \rightarrow \mathcal{B} \cup \mathcal{C}} = \text{Tr}_{A \rightarrow \mathcal{B}} . \quad (120)$$

*Proof.* Suppose two separate games  $\mathcal{G}_1, \mathcal{G}_2$  with

$$(\mathcal{B} \subset \mathcal{V}(\mathcal{G})_1 \wedge \mathcal{B} \subset \mathcal{V}(\mathcal{G})_2) \wedge (\mathcal{C} \subset \mathcal{V}_1 \wedge \mathcal{C} \cap \mathcal{V}_2 = \emptyset) . \quad (121)$$

Let  $T_1 \in \mathcal{V}_1, T_2 \in \mathcal{V}_2$  be two auxiliary players such that

$$\forall D \in \mathcal{B} \cup \mathcal{C}, DTr_{D \rightarrow T_1} = \infty , \quad (122)$$

$$\forall B \in \mathcal{B}, DTr_{B \rightarrow T_2} = \infty . \quad (123)$$

Suppose also that there exist  $|\mathcal{B} \cup \mathcal{C}|$  consecutive turns for the first game and  $|\mathcal{B}|$  consecutive turns for the second game during which all the colluding players choose actions according to the evil strategy. More formally, suppose that

$$\begin{aligned} & \exists j \in \mathbb{N} : \forall d_1 \in [|\mathcal{B} \cup \mathcal{C}|], Player(j + d) \in \mathcal{B} \cup \mathcal{C} \wedge \\ & \wedge \forall d_1, d_2 \in [|\mathcal{B} \cup \mathcal{C}|], d_1 \neq d_2, Player(j + d_1) \neq Player(j + d_2) \wedge \\ & \wedge \forall d \in [|\mathcal{B} \cup \mathcal{C}|], Strategy(Player(j + d)) = Evil \wedge \\ & \wedge \forall j' \in [j] Player(j') \notin \mathcal{B} \cup \mathcal{C} \end{aligned} \quad (124)$$

for the first and likewise for the second game.

According to theorem (4),

$$Tr_{A \rightarrow \mathcal{B} \cup \mathcal{C}} = maxFlow(A, T_1) \wedge Tr_{A \rightarrow \mathcal{B}} = maxFlow(A, T_2) . \quad (125)$$

From lemma (4), we know that

$$\forall (v, w) \in MinCut, v \notin \mathcal{B} \cup \mathcal{C} \quad (126)$$

and thus

$$\begin{aligned} & e \in MinCut_1 \Rightarrow e \in \mathcal{E}_2 \wedge \\ & \wedge e \in MinCut_2 \Rightarrow e \in \mathcal{E}_1 \wedge \\ & \wedge \forall e \in MinCut_1 \cup MinCut_2, c_1(e) = c_2(e) . \end{aligned} \quad (127)$$

Thus the flow  $X_1$  resulting from  $maxFlow(A, T_1)$  can be used to construct a valid flow of equal value for the second case if we set

$$\forall v \in \mathcal{V}_2 \setminus \mathcal{B}, \forall w \in \mathcal{V}_2, x_{vw,2} = x_{vw,1} , \quad (128)$$

$$\forall v \in \mathcal{B}, x_{vT_2,2} = \sum_{w \in N^+(v)} x_{vw,1} , \quad (129)$$

$$\forall v, w \in \mathcal{B} \cup \mathcal{C}, x_{vw,2} = 0 . \quad (130)$$

Likewise, the flow  $X_2$  resulting from  $\text{maxFlow}(A, T_2)$  can be used to construct a valid flow of equal value for the first case if we set

$$\forall v \in \mathcal{V}_1 \setminus (\mathcal{B} \cup \mathcal{C}), \forall w \in \mathcal{V}_1, x_{vw,1} = x_{vw,2} , \quad (131)$$

$$\forall v \in \mathcal{B}, x_{vT_1,1} = \sum_{w \in N^+(v)} x_{vw,2} , \quad (132)$$

$$\forall v \in \mathcal{C}, \forall w \in \mathcal{V}_1, x_{vw,1} = 0 . \quad (133)$$

From these two observations, we deduce that there exists a function, say  $F_2(X_1)$ , that transforms the  $\text{MaxFlow}_1$  of the first graph into a valid flow for the second graph that has the same amount of flow as  $\text{MaxFlow}_1$  and there also exists a similar function  $F_1(X_2)$  that transforms the  $\text{MaxFlow}_2$  of the second graph into a valid flow for the first graph that has the same amount of flow as  $\text{MaxFlow}_2$ . Suppose that

$$\text{maxFlow}_1 < \text{maxFlow}_2 . \quad (134)$$

Then

$$F_1(\text{MaxFlow}_2) > \text{maxFlow}_1 , \quad (135)$$

which is a contradiction. Likewise, suppose that

$$\text{maxFlow}_1 > \text{maxFlow}_2 . \quad (136)$$

Then

$$F_2(\text{MaxFlow}_1) > \text{maxFlow}_2 , \quad (137)$$

which is also a contradiction. We conclude that

$$\text{maxFlow}(A, T_1) = \text{maxFlow}(A, T_2) , \quad (138)$$

thus

$$\text{Tr}_{A \rightarrow \mathcal{B}} = \text{Tr}_{A \rightarrow \mathcal{B} \cup \mathcal{C}} . \quad (139)$$

□

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

## 5 Related Work

## 6 Further Research

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

## 6.1 Zero knowledge

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 [12]. 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.

## Appendix

### References

1. Patrikakis C., Masikos M., Zouraraki O.: Distributed Denial of Service Attacks. The Internet Protocol Journal, Vol. 7, N. 4, <http://www.cisco.com/c/en/us/about/press/internet-protocol-journal/back-issues/table-contents-30/dos-attacks.html> (2004)
2. Standard ebay selling fees, <http://pages.ebay.com/help/sell/fees.html>
3. ebay money back guarantees, <http://pages.ebay.com/ebay-money-back-guarantee/questions.html>
4. What is OpenBazaar, <https://blog.openbazaar.org/what-is-openbazaar/>
5. Can Bitcoin and Multisig Reduce Identity Theft and Fraud?, <https://blog.openbazaar.org/can-bitcoin-and-multisig-reduce-identity-theft-and-fraud/>
6. Nakamoto S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008)
7. Kahn Arthur B.: Topological sorting of large networks. Communications of the ACM Vol. 5, Issue 11, pp. 558-562, ACM, New York (1962)
8. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.: Introduction to Algorithms (3rd ed.). MIT Press and McGraw-Hill (2009) [1990]
9. Zindros D. S.: Trust in decentralized anonymous marketplaces (2015)
10. Buterin V.: Bitcoin Multisig Wallet: The Future of Bitcoin. Bitcoin Magazine (2014), <https://bitcoinmagazine.com/articles/multisig-future-bitcoin-1394686504>
11. Sanchez W.: Lines of Credit (2016) <https://gist.github.com/drwashio/2c40b91e169f55988618#part-3-web-of-credit>
12. Biryukov A., Khovratovich D., Pustogarov I.: Deanonymisation of clients in Bitcoin P2P network. arXiv:1405.7418 [cs.CR] (2014) <https://arxiv.org/pdf/1405.7418v3.pdf>