Mixing Coins of Different Quality: A Game-Theoretic Approach

Svetlana Abramova[1,2], Pascal Schöttle[1], and Rainer Böhme[1,2]

[1] Department of Computer Science, University of Innsbruck, Austria
[2] Department of Information Systems, University of Münster, Germany

Abstract. Cryptocoins based on public distributed ledgers can differ in their quality due to different subjective values users assign to coins depending on the unique transaction history of each coin. We apply game theory to study how qualitative differentiation between coins will affect the behavior of users interested in improving their anonymity through mixing services. We present two stylized models of mixing with perfect and imperfect information and analyze them for three distinct quality propagation policies: poison, haircut, and seniority. In the game of perfect information, mixing coins of high quality remains feasible under certain conditions, while imperfect information eventually leads to a mixing market where only coins of the lowest quality are mixed.

Keywords: Bitcoin, anonymity, blacklisting, policy, game theory.

1 Introduction

While public distributed ledgers serve as an essential backbone of many cryptocurrencies, they inherently act as a source of differentiation of coins by quality. Indeed, each individual unspent transaction output has its unique history recorded with cryptographic integrity protection in the public distributed ledger. Having the entire history publicly available makes cryptocoins non-fungible, that means distinguishable from each other in terms of the perceived quality. Coins generated in the genesis block or passed through credible exchanges might be more attractive to someone over coins whose transaction history contains patterns suggesting dubious activities in the past [17].

The traceability offered by public distributed ledgers has called the anonymity of financial transactions into question. For example, the most popular cryptocurrency, Bitcoin, was initially spoken of as a truly anonymous payment method. However, many studies [4,16,23,24] have shown that the blockchain infringes user privacy. In efforts to impede simple blockchain analyses, privacy-concerned users can cooperate with each other and mix their payments in a single transaction instead of sending multiple individual transactions. Many cryptocurrency protocols support such collective transactions, to which we refer here as *mixing transactions*. In general, mixing can be thought of as a privacy-enhancing overlay [15], which tangles the transaction's inner flows between incoming and outgoing funds. This makes money flows more difficult to trace [33].

If the history of transactions matters to users, mixing coins of good and bad quality in one single transaction bears the risk that good coins are exchanged for worse ones. A simple example of this would be a multi-input transaction, some inputs of which can be traced back to darknet markets or ransomware payments. Such a transaction may come under scrutiny of law enforcement as a payment possibly made by or to a criminal. Moreover, in the name of preventing financial crime, regulators may enforce transaction blacklisting. Although blacklisting is not explicitly implemented today, its ideas are already present in the Bitcoin system in various forms. For example, some wallet providers and exchanges allegedly denied or delayed transactions which tried to spend stolen funds or could be linked to darknet markets [5,12]. Blacklists are one possible source of qualitative differentiation.

Whenever coin quality matters, each downstream transaction must not lead to the loss of information about the quality of newly generated outputs. Rather, we must assume that some sort of a quality propagation policy is in place to allow for situations when coins of different quality are combined in one transaction. If qualitative differentiation between coins becomes common practice and a specific policy takes effect, users will always have to account for the risk of receiving coins of low quality. As this risk is especially amplified in the context of mixing, it is of particular interest to analyze how participants of mixing services will behave in these circumstances. Will they be willing to engage in mixing and, if yes, under what conditions? Will the market for mixing services persist after all?

Contributions. We apply game theory to study this scenario and formalize the game of mixing coins of different quality. Besides addressing the quality propagation effect, the model captures two main factors behind users' intentions to mix: privacy enhancement and financial compensation. While distinct propagation policies have been proposed in the literature, they are of limited value to both practice and research if their system-wide implications are not theoretically analyzed. To this end, we make several relevant contributions. Specifically, we devise a variant of the game for each of the policies and solve it under two regimes of perfect and imperfect information. This allows us to discuss the policy implications from the design perspective of distributed ledger protocols and simultaneously provide theoretical support for arguments brought into the debate around fungibility and privacy.

The remainder of the paper is organized as follows. Using Bitcoin as a prominent example, we begin with preliminaries on cryptocurrencies, mixing transactions and propagation policies in Section 2. Then, in Sections 3 and 4, we present and theoretically analyze the game for each introduced policy. The practical implications are discussed in Section 5. We briefly review related work in Section 6 and conclude with limitations and future research in Section 7.

2 Preliminaries

We use Bitcoin as running example, noting that the problem definition and solution approaches generalize to most cryptocurrencies known to date [7].

Bitcoin is a decentralized system that maintains a public, append-only ledger of confirmed transactions (known as the "*blockchain*") through collective efforts of a peer-to-peer network running a probabilistic consensus protocol [20]. Bitcoin addresses generated from public keys serve as account identifiers, whereas the knowledge of private keys indicates ownership of accounts and control over the coins in them. As no real-world identity is required to generate key pairs, each user may autonomously create an arbitrary number of Bitcoin addresses.

The blockchain stores a log of all valid transactions ever made in Bitcoin. A transaction is a digital record that consists of a list of inputs – references to existing addresses in the blockchain with a positive balance, and a list of outputs – addresses to which specified numbers of bitcoins are sent. Bitcoin is designed in a way that the total value of the inputs has to be spent in the outputs of a transaction. Otherwise, the difference is considered as a fee and paid to special nodes in the network (miners), who validate transactions and ensure a consistent and manipulation-resistant state of the blockchain.

2.1 Coin Mixing

A common thread of criticism of Bitcoin is the lack of full anonymity of payments [16]. With blockchain exploration tools at hand, one can browse through the complete transaction history and trace money flows back to their origins. Furthermore, experimental analyses of the limits of anonymity in Bitcoin [4,16,23,24] show that some users can be deanoynmized by applying appropriate heuristic techniques and consulting external information. Once a real name behind an address is found, the user's privacy might be jeopardized, as the blockchain allows a passive observer to look up other linked transactions [6]. Besides avoiding the re-use of public addresses, individuals seeking for greater anonymity may use available *mixing services*.

The concept of mixing coins of different users is fairly straightforward. Here, it refers to combining inputs and outputs of multiple parties in a single transaction. The current implementation of the protocol enables to build such collaborative transactions as it requires separate signatures for each public key specified in the transaction's inputs. With a sufficient number of participants engaged in a mixing transaction, it becomes harder to trace money flows by finding the connections between sending and receiving addresses. By extension, it gets even more difficult if mixing is done repeatedly. Nevertheless, individual values of inputs and outputs may still reveal enough information for a successful untangling of the transaction's inner flows [33].

The idea of and practical need for mixing has given rise to the emergence of special services and marketplaces designed to match supply and demand of anonymous transactions [17]. Here, we limit our focus on *CoinJoin*, as one specific example present in the Bitcoin system. In its simplest case, a CoinJoin transaction aggregates two or more inputs from two different users and contains at least two outputs of equal value. So, a blockchain observer cannot directly link these two outputs to the sending Bitcoin addresses. The larger the number of participants, the greater the anonymity of a CoinJoin transaction. However,

users interested in anonymity suffer from the necessity of finding other partners who are willing to participate in mixing at the same time. This limitation explains the presence of special mixing services and platforms [17], where individuals supply their bitcoins for use in mixing transactions in exchange for a small mixing fee.

2.2 Sources of Qualitative Differentiation

In addition to applications like colored coins [25] or possible cross-chain mixing in the future, black- and whitelisting are potential sources of qualitative differentiation between coins. For the sake of intuitive illustration, we nonjudgementally refer to blacklisting in this paper in order to model coins of different quality.

Although not fully implemented today, potential blacklisting of criminal transactions as well as the issue of fungibility are subjects of intense interest and ongoing debates. Since each output has an accessible and cryptographically verifiable history of ownership, Bitcoin is not fungible. Also the market participants' convention to treat bitcoins as if they were fungible has been repeatedly called into question. Statements published on Bitcointalk.org and relevant subreddit threads [5,30] illustrate this point:

> *"Looking to buy an old 50 BTC block. Where to buy? I'll pay in bitcoin. No FIAT/Alt coin. Willing to pay premium."*
>
> blockCollector, Nov 11, 2015

> *"BitPay is blacklisting certain bitcoins & rejecting customers. I'm certain others are doing it too. Fungibility is most pressing issue IMO."*
>
> TraderSteve, Sep 25, 2015

These examples support the conjecture underlying this work that coins differ in their quality. Transaction blacklisting followed by the devaluation of marked bitcoins has been suggested as a conceivable means of fighting financial crime [18,19]. In practice, this may be realized by enforcing the centralized actors (e. g., exchange services or wallet providers) or, alternatively, miners to consult blacklists and disregard those transactions that try to reclaim funds from criminal proceeds. The notorious story of a recently exploited vulnerability in the Decentralized Autonomous Organization (DAO), an Ethereum-based program, has clearly demonstrated the doubtfulness and disagreement in the community regarding how issues related to illicit use should be resolved and who would bear the burden of doing it [27].

Several obstacles impede the effectiveness of blacklisting as a policy tool. First, perpetrators can disguise the origins of money by resending their dirty coins through as many fake addresses as they need. Therefore, the application of blacklisting has to propagate through the entire transaction graph, rooted at the offending transaction. Second, as law enforcement takes time, ordinary users will inevitably face a risk of receiving allegedly clean coins that might be blacklisted by authorities later [19]. These facts call for a detailed elaboration of blacklisting propagation mechanisms and their effects on the ultimate quality of coins. This is especially crucial in case of multi-input transactions comprising of both high- and low-quality inputs.

2.3 Quality Propagation Policies

We consider three basic propagation techniques, termed as the **poison**, **haircut** and **seniority policies** [19], and assume that there is a consensus on one specific policy, implemented in the client software. To demonstrate the application of a propagation mechanism in each case, let us use an example of the transaction graph depicted in Figure 1. The transaction of interest Z references outputs of the two preceding transactions X and Y. Suppose, the transaction X is discovered to be a ransom payment and, consequently, all of its outputs are added to the blacklist. Under the **poison policy**, every output of the transaction that has at least one blacklisted predecessor is invalidated completely. Consequently, all outputs of Z will be blacklisted.

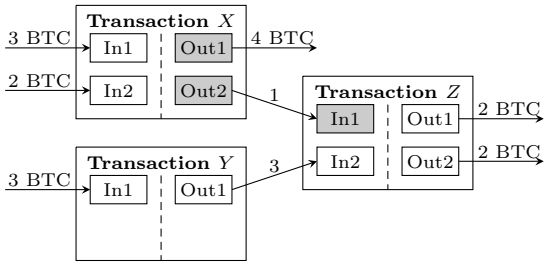


Fig. 1: Example of the transaction subgraph in Bitcoin. Gray areas indicate that both outputs of the transaction X are blacklisted.

The less drastic **haircut policy** dictates to devalue *all* outputs of a transaction proportionally to the total amount of blacklisted coins in its inputs. Thus, each output will contain an identical fraction of blacklisted coins, which is calculated as the fraction of the collective blacklisted value in the total transaction's input. Referring back to our example, both outputs of Z will thus have a partial devaluation of their nominal worth (25 percent, to be precise). As Bitcoin is divisible down to the smallest unit of one *satoshi* (worth 10^{-8} BTC), the haircut policy requires a special rule regulating blacklisting of minimum values and preventing money laundering through multiple tiny outputs. Such rule may dictate, for example, that the blacklisted value is rounded up to the full satoshi. (We ignore this quantization effect in the rest of this paper.)

Under the **seniority policy**, the output order and amounts determine how incoming blacklisted coins will be redistributed. Let us assume for simplicity that blacklisted coins are propagated in the order of the output list in a transaction (i. e., from top to bottom). Since the transaction X has one blacklisted input of the value 1 BTC, its first output of nominal value 2 BTC will be devalued by half. Similarly to the haircut policy and in contrast to the poison policy, the seniority regime does not change the total sum of blacklisted coins.

3 Model

Of particular interest for our study is to examine for each policy how users behave if coins of different qualities can hypothetically be mixed in one transaction. For that, we present two game-theoretic models of mixing, one with *perfect and complete* information and one with *imperfect and complete* information. A game of perfect information assumes that each player is aware of the prior actions chosen by other players, whereas imperfect information implies uncertainty regarding at least one move of another player. Complete information means that all players know all players' action sets and payoff functions [29, p. 136].

Two rational players A and B consider to transfer coins in a joint mixing transaction. Player A is a *privacy seeker* who initiates a mixing transaction, and B is a *privacy provider* who helps to establish a (minimum) anonymity set by participating in this transaction. Each player is endowed with an unlimited number of coins (i. e., transaction outputs) of different quality $q \in [0, 1]$. Coins with $q = 1$ are not on the blacklist and referred to as "good" or "clean" (e. g., coins passed through trusted exchanges), whereas coins with $q \neq 1$ are blacklisted or "bad" coins that can be linked to criminal activities. The term $(1 - q)$ can be alternatively interpreted as a fraction of the coin that has been devalued according to the applied policy.

The privacy seeker A desires more anonymity through mixing and pays player B the mixing fee $c \ll 1$ as a reward for joining a transaction. We assume that the fee for anonymity is payed out within the mixing transaction itself. Besides this financial compensation, player B also benefits from anonymity, as the mixing transaction anonymizes the identities of all participants. Player A selects $(1+c)$ coins of quality q_a, while player B chooses one coin of quality q_b. So, the move of player i is the choice of q_i. In addition, the players have the outside strategy not to engage in mixing at all, as both need to sign a mixing transaction before it can be broadcasted to the network. Note that we explicitly disregard transaction fees payed to miners and assume that each player transfers funds to (possibly multiple) destination addresses *under her control*. Thus, players A and B own afterwards funds of 1 and $(1 + c)$ nominal value, respectively. The quality of these funds may however change once a specific quality propagation policy takes effect. We use the notation q'_a and q'_b to denote the respective post-transaction quality factors.

In the presence of qualitative differentiation and everything else being constant, a rational player always tries to maximize her own utility, which corresponds in our setting to the maximization of the value of coins at disposal. The utility of each player is therefore measured in units of good coins and expressed by three relevant components: 1) the subjective value of anonymity the player attributes to a mixing transaction; 2) the post-transaction value of the funds held by the player; 3) the compensation fee paid by the privacy seeker A to the privacy provider B. We first define each component and later specify the payoff function of each player formally.

In reality, the perceived anonymity of an individual mixing transaction depends on multiple aspects (e. g., the number of participants, the number of inputs

and outputs and their exact quantities, repeated mixing etc.). Since transaction parties value anonymity differently and it is not trivial to quantify it, we express the benefit of (somewhat more) anonymity by a relative unit gain equal to one good coin. Suppose, without loss of generality,[3] that player A gains one unit of anonymity, whereas player B gains some level $\tau_b \in [0, 1]$. $\tau_b = 1$ indicates that both players value anonymity of a mixing transaction equally; $\tau_b = 0.5$ means that player B values it half as much as player A; $\tau_b = 0$ means that player B receives no benefit in terms of improved anonymity from mixing. Note that the gain in anonymity is discounted by the post-transaction qualities q'_a and q'_b. There is less value in having bad coins anonymized. Moreover, this avoids corner cases where players have incentives to mix at the risk of receiving bad coins.

The post-transaction qualities endogenously follow from the choice variables q_a and q_b and the applied quality propagation policy. Unlike the seniority policy, the poison and haircut policies allow us to formally define q'_a and q'_b as a function of the pre-transaction quality factors q_a and q_b. Under the poison policy, all coins are either good ($q_i = 1$) or bad ($q_i = 0$). Table 1 specifies the values of q'_a and q'_b for all possible combinations of q_a and q_b. Under the haircut policy, the levels of q'_a and q'_b are equal and, besides the choice variables q_a and q_b, depend on the parameter c. Since the fee (of quality q_a) is transferred in the mixing transaction, it influences the total transaction amount as well as the total level of blacklisted coins in the inputs.

Table 1: Post-transaction quality factors

Policy	$\boldsymbol{q_a}$	$\boldsymbol{q_b}$	$\boldsymbol{q'_a}$	$\boldsymbol{q'_b}$
Poison	1	1	1	1
	1	0	0	0
	0	1	0	0
	0	0	0	0
Haircut	$q_a \in [0, 1]$	$q_b \in [0, 1]$	$\frac{q_a \cdot (1+c)+q_b}{2+c}$	$\frac{q_a \cdot (1+c)+q_b}{2+c}$

With regard to the last component of the payoff function, the mixing fee has to be discounted by q_a in the payoff of player A and by q'_b in the payoff of player B in order to measure its value relatively to one clean coin. Thus, the players' payoffs π_i after successful mixing is given as follows:

$$\pi_A = 1 \cdot q'_a + 1 \cdot q'_a - c \cdot q_a = 2 \cdot q'_a - c \cdot q_a; \tag{1}$$

$$\pi_B = \tau_b \cdot q'_b + 1 \cdot q'_b + c \cdot q'_b = (\tau_b + 1 + c) \cdot q'_b. \tag{2}$$

[3] Otherwise, switch players A and B.

If one of the players disagrees to mix and chooses the outside option, the payoffs are as follows:

$$\pi_{A\perp} = 1 \cdot q_a + c \cdot q_a = (1 + c) \cdot q_a; \tag{3}$$

$$\pi_{B\perp} = 1 \cdot q_b. \tag{4}$$

4 Results

We first present the game of perfect and complete information for tractability and as a benchmark, before we consider the game of imperfect (and complete) information, in which the players choose coin qualities q_i simultaneously.

4.1 Perfect Information: Sequential Game

The model of perfect information assumes q_a and q_b to be common knowledge. This means that blacklists have to be public and always up-to-date, e. g., law enforcement agencies immediately discover and blacklist offending transactions. With public blacklists, each player can check the quality of the other player's coin before signing and broadcasting a mixing transaction to the network.

Figure 2 shows an extensive form of the sequential game by taking the poison policy as an example. The presented sequence of moves can be extended to the other two regimes, too, by considering a larger set of actions available to both players. Player A initiates the game by committing to the quality of her inputs q_a and the fee level c. Being informed about that choice, player B decides whether to mix with A or not. If B prefers to dismiss, the game is over. Otherwise, player B chooses the coin of a particular quality q_b and notifies A about it. Player A learns about the choice of B and makes the final move of the game. Reciprocally, if A rejects to mix with B, both players exit with the payoffs defined in Equations (3) and (4). Otherwise, players form a mixing transaction and get the payoffs as prescribed by (1) and (2). Under the seniority policy, players may additionally negotiate the order and amounts of transaction outputs until they reach a consensus or someone rejects to partner with.

Poison policy. We apply a backward induction procedure in order to analyze the game and find subgame perfect Nash equilibria. As the game is of perfect information, there are seven subgames in total, labeled Γ_1 through Γ_7 in Figure 2. Under backward induction, the subgames Γ_4–Γ_7 are solved first. In subgame Γ_4, player A agrees to mix the clean coin if $c \leq 0.5$ and exits otherwise. In subgame Γ_5, player A always exits due to the negative propagation of blacklisting. In subgames Γ_6 and Γ_7, player A is indifferent between the two available choices. Taking the respective equilibrium for each of the subgames Γ_4–Γ_7 and going backward in the game tree, we can see that the game has many subgame perfect Nash equilibria[4], all of which contain either the path:

$$(q_a = 1, q_b = 1 \text{ and mix})$$

[4] Note that the game of perfect information under the poison regime has even more Nash equilibria. However, these are not subgame perfect.

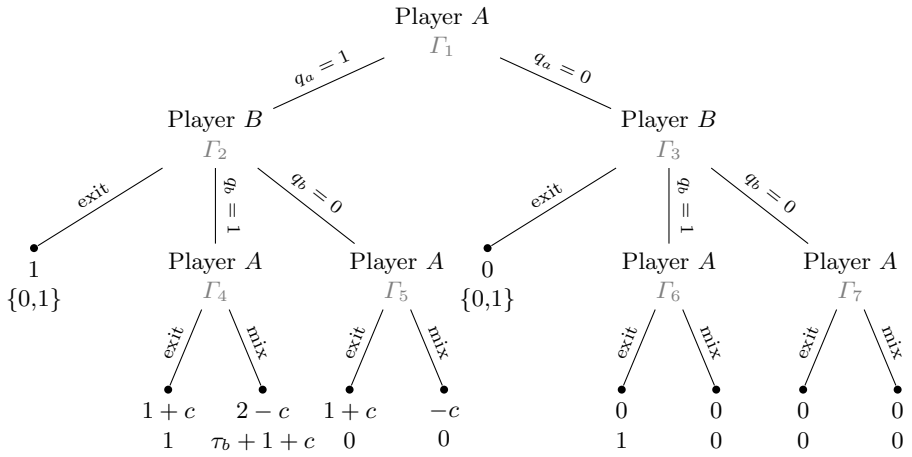


Fig. 2: **Poison policy:** game of perfect information in extensive form.

if $c \leq 0.5$; or otherwise the path:

$$(q_a = 1, q_b = 1 \text{ and exit}).$$

So, both players mix clean coins if $c \leq 0.5$, but player A refuses to pay a too high fee for anonymity.

Haircut policy. The haircut policy implies the presence of coins of any quality in the range of $[0, 1]$. Player A signs a mixing transaction if the payoff after mixing π_A is greater or equal than the payoff without mixing $\pi_{A\perp}$:

$$2 \cdot \frac{q_a \cdot (1+c) + q_b}{2+c} - c \cdot q_a \geq (1+c) \cdot q_a,$$
$$\frac{q_a}{q_b} \leq \frac{2}{2 \cdot c^2 + 3 \cdot c},\ q_b \neq 0,\ c \neq 0. \tag{5}$$

Analogously for player B:

$$(\tau_b + 1 + c) \cdot \frac{q_a \cdot (1+c) + q_b}{2+c} \geq 1 \cdot q_b,$$
$$\frac{q_a}{q_b} \geq \frac{(1-\tau_b)}{(1+c) \cdot (1+c+\tau_b)},\ q_b \neq 0. \tag{6}$$

Players A and B agree to mix with each other if both inequalities (5) and (6) hold. Figure 3 shows the corresponding game outcomes over the space defined by the quality ratio q_a/q_b and the fee c for three distinct values of $\tau_b \in \{0, 0.5, 1\}$. Region S_1 depicts all combinations of the model parameters which result in successful mixing for $\tau_b = 0$, regions $S1 \cup S2$ depict the same for $\tau_b = 0.5$; regions $S1 \cup S2 \cup S3$ apply to $\tau_b = 1$. In the corner case $q_b = 0$, mixing happens only if player A wishes to mix a bad coin, too.

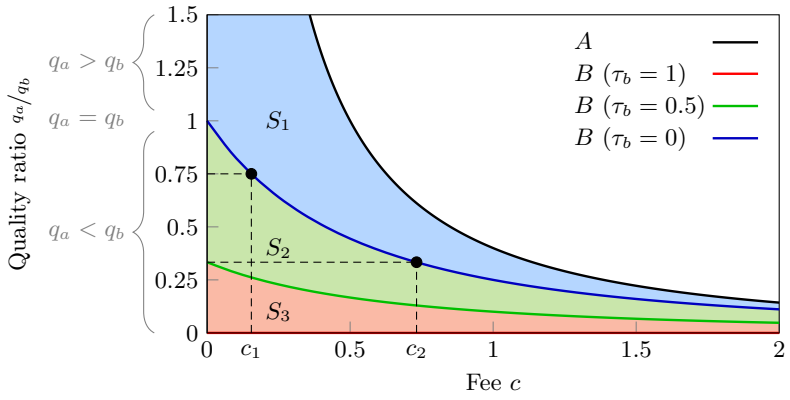


Fig. 3: **Haircut policy:** existence of equilibrium solutions in the game of perfect information as a function of the fee c and the quality ratio q_a/q_b for three different values of $\tau_b \in [0, 0.5, 1]$.

If player B values transaction anonymity as much as player A, i. e., $\tau_b = 1$, B is willing to partner with A regardless of the fee level c or player A's coin quality q_a. This is due to the fact that even in the worst possible case for player B ($q_a = 0,\ q_b = 1$, and $c = 0$), the degradation in the coin quality ($q'_b = 0.5$) is fully compensated by the gain in anonymity $\tau_b \cdot q'_b = 0.5$. Thus, the space of successful game outcomes in $S1 \cup S2 \cup S3$ is limited only by inequality (5), which corresponds to the uppermost line in Figure 3. Since the payoff of B is directly proportional to the quality factor of her coin, it is in her best interest to offer the good coin $q_b = 1$ for mixing. Therefore, the Nash equilibria in case $\tau_b = 1$ correspond to the set of action profiles $\{(q_a, q_b) \mid 0 < q_a \leq 1, q_b = 1\}$. If player B values anonymity half as much as player A, i. e., $\tau_b = 0.5$, her best response is defined by the *green line*. If A does not offer any fee, B chooses the coin of quality q_b, which is three times higher than q_a. If player A wishes to mix with the coin of higher quality, i. e., $q_b > 3q_a$, she has to compensate player B by offering a strictly positive fee. The exact level of c for a desirable quality ratio can be derived from inequality (6). If player B does not value the anonymity of a mixing transaction, i. e., $\tau_b = 0$, and there is no financial compensation c, she will supply the coin of the same quality ($q_a = q_b$).

In reality, however, the search cost associated with finding another transaction party with exactly the same quality level may be prohibitive. Similarly to conventional trading and payment markets [9], search frictions can be overcome by offering compensation to the enabling party with (slightly) better coins. If B does not have a coin of the required quality and can mix a coin of higher quality instead, she will agree to participate in the mixing transaction in exchange for a higher fee. The greater the difference in the quality of the coins of players A and B, the more A has to pay to B for joining a mixing transaction. This explains the monotonically decreasing shape of inequality (6) in Figure 3 when $\tau_b < 1$.

Seniority policy. This policy grants users more flexibility in controlling the effect of blacklisting propagation. Since players know the quality of all inputs, they can negotiate an internal structure of the mixing transaction until it is designed in such a way that low-quality fractions of inputs of both players appear at the beginning of the output list. Player A, as the privacy-seeking party, may also be willing to list some of her outputs first and sacrifice at the expense of gained anonymity up to half of the clean portion of her coins. Player B, who is interested in receiving the financial reward, will demand to list her address for the incoming fee c at the bottom of the output list. These order constraints may however leak sufficient information for successful matching of the relations between inputs and outputs of a mixing transaction. A passive observer of the blockchain may look up available blacklist data and, knowing the exact transacting amounts, may succeed in deanonymizing the mixing transaction.

The seniority policy can be reduced to the haircut policy if the players agree to split up blacklisted coins equally by randomly alternating the order of their outputs for the sake of anonymity. However, the players can be better off in terms of anonymity while maintaining the quality distribution if they adhere to one constant value for all (blacklisted and clean) transaction outputs. They can divide their input funds into multiple outputs of the same amount and randomize the order of their outputs within the upper subset of blacklisted outputs (*the blacklisted bin*) and the lower subset of clean outputs (*the clean bin*). This way, the attacker is left with a 50:50 % chance of correctly differentiating between the output of A and the output of B, whereas the players can preserve the original quality of their funds. The easiest, however impracticable solution would be to use one satoshi as the size of each output. In order to reduce the number of outputs by orders of magnitude, the players can express q_a and q_b as rational numbers in the standard form and use a reciprocal of the least common divisor of the denominators as constant value for all outputs.

Let us demonstrate one numerical example with $q_a = 3/4$ and $q_b = 1/2$ (the mixing fee is disregarded). Following the above logic, each player splits up her coin into four different outputs of the nominal value 0.25. The blacklisted bin will consist of one output of player A and two outputs of player B. The clean bin contains three outputs of player A and two outputs of player B. The order of the outputs within each bin must be random in order to get anonymity. As a result, the post-transaction quality factors q'_a and q'_b do not change and the players still enjoy transaction anonymity. Figure 4 in Appendix A illustrates this example (along with two other specific cases).

4.2 Imperfect Information: Simultaneous-Move Game

Since law enforcement agencies are unable to detect and mark illegal transactions in real time, there is always a risk that already confirmed transactions may get blacklisted later. Due to this time delay, users have to deal with the uncertainty about the quality of inputs when forming and signing a transaction. While they have more information about the origins and nature of their own coins (compared to passive observers of the public blockchain), they cannot know for sure qualities

of all other inputs. We model this more realistic case of mixing in a simultaneous-move game, in which players A and B choose q_a and q_b simultaneously.

Poison policy. Since all circulated coins are either good ($q_i = 1$) or bad ($q_i = 0$), players have only two possible strategies, which enables us to represent the model in normal form (see Table 2). The resulting payoffs are calculated by substituting the pre- and post-transaction quality factors (given in Table 1) in the payoff functions (1) and (2).

Table 2: **Poison policy:** game of imperfect information in normal form

		Player B	
		$q_b = 1$	$q_b = 0$
Player A	$q_a = 1$	$2 - c$, $\tau_b + 1 + c$	$-c$, 0
	$q_a = 0$	0, 0	0, 0

The presented model has two pure-strategy Nash equilibria ($q_a = 1, q_b = 1$) and ($q_a = 0, q_b = 0$). Note that the latter Nash equilibrium is weak, as player B gets the same payoff by changing her strategy to $q_b = 1$. Although the action profile ($q_a = 0, q_b = 0$) is a Nash equilibrium, it does not correspond to the social optimum of the game: the sum of the payoffs of both players reaches its maximum when ($q_a = 1, q_b = 1$). Similarly to Akerlof's classic market for lemons [3], the poison policy leads to a market failure because of adverse selection. Without knowing the true quality of coins, nobody is willing to mix good coins at the risk of encountering bad coins at least in one input of the mixing transaction.

Haircut policy. Over time, the haircut policy results in the circulation of coins of varying qualities. In the absence of an ability to perfectly differentiate coins by quality, users of mixing services will make decisions based on their expectations about the average quality of all coins observed in the mixing market. Substituting q_a and q_b with the expected average quality $\overline{q}$ in inequalities (5) and (6), respectively, the necessary conditions for players A and B to participate in the mixing transaction are as follows:

$$q_a \leq \frac{2 \cdot \overline{q}}{2 \cdot c^2 + 3 \cdot c}, \ c \neq 0; \tag{7}$$

$$q_b \leq \frac{(\tau_b + 1 + c)(1 + c)}{1 - \tau_b} \cdot \overline{q}, \ \tau_b \neq 1. \tag{8}$$

If player B values anonymity highly ($\tau_b = 1$), the mixing transaction happens regardless of the expected average quality and the fee level. The more interesting scenario is, however, when player B is solely motivated by the financial reward ($\tau_b = 0$). In this case, inequality (8) takes the form $q_b \leq (1 + c)^2 \cdot \overline{q}$. If player A does not pay a fee, player B has no incentive to supply a coin of quality better than the average quality $\overline{q}$. Otherwise, the fee incentivizes the privacy provider B to offer the coin of marginally higher quality.

It is reasonable to expect that criminals, who know with certainty which of their funds originate from illicit transactions, may engage in mixing for the purpose of money laundering. They have an incentive to supply coins of the worst quality in the hope of getting better ones. As other players can anticipate this behavior, the expected quality factor declines. This further drives owners of better coins out of the market and fuels the race to the bottom of $\overline{q}$, eventually leading to the presence of only bad coins in the mixing market. Consequently, there will be no equilibrium outcome with a strictly positive payoff for both players, and, given the absence of credible signaling mechanisms, the market for mixing coins of (marginally) good quality will not exist.

Seniority policy. Given the uncertainty regarding q_i, each player will prefer her addresses to be included at the bottom of the output list. Since the seniority policy can be reduced to the haircut regime, the above reasoning and solution can be applied here, too. Facing the risk of getting coins of worse quality, the players will prefer to mix rather bad coins than good ones.

It might seem at first glance that the seniority police allows for a modification of the model to a signaling game, because the output order can convey information and it is linked to payoff. In general, signaling games model strategic settings of incomplete information in which players can observe the actions of their opponents (*signals*) to make inferences about hidden information [28]. A fundamental principle is that signals must be costly to produce, or have costly consequences. This is what differentiates signals from "cheap talk" and guarantees their reliability. To enable mixing in more situations, players must be able to signal that they are committed to supply coins of high quality. A corresponding output order must be more costly to sign for players with bad coins than for players with good coins. However, as owners of bad coins have, in the strong sense, nothing to lose, the only possible signal is that of supplying low quality coins, which unfortunately does not lead to more mixing equilibria.

5 Discussion

This work is an attempt to conceptualize a formal model of the interplay of users in the presence of qualitative differentiation between cryptocoins. Although we motivate the game by taking the illustrative example of mixing services and blacklisting, the model (of imperfect information) can be generalized to a more common case where an individual user needs to decide whether to combine own coins of potentially different qualities in one multi-input transaction.

The regime of perfect information suggests a sequential game. It is applicable if blacklists are timely and public. Under these assumptions, mixing services persist. The poison policy dictates users to mix clean coins (if at all), while the haircut and seniority policies provides certain conditions under which users are also willing to mix coins of varying qualities. Moreover, the seniority policy can (approximately) be reduced to the haircut policy.

The regime of imperfect information suggests a simultaneous-move game. It leads to the failure of the market for mixing of (marginally) good coins under

the poison policy, and our preliminary results let us conjecture that the outcome applies to the haircut and seniority policies as well. (We plan to refine the analysis in a revised version of this work.) With uncertainty about a coin's quality and in the presence of criminals interested in using mixers for money laundering, owners of good coins have no incentive to seek anonymity at the risk of mixing with bad coins. In this regard, blacklisting can be viewed as an effective economic mechanism to make mixers less attractive or even to dry them out.

6 Related Work

Work on blacklisting coins is related to blacklisting content (or content providers), and therefore to Internet censorship. And it connects to anonymity online. Both are contentious topics; the former more than the latter.

Governments, Internet intermediaries, and organizational network administrators use many kinds of filtering techniques to intentionally limit or block access to online content, resources, or services [1]. Among them, blacklists of malware-infected or phishing sites are perhaps the best known and socially most accepted example. Although many empirical studies exist on the effectiveness, coverage, and sharing of phishing blacklists [26,31,32], there is a limited number of works examining them from a formal viewpoint. Edwards et al. [11] present a simple Markov model to study how malware infections might be contained through blacklisting, while Hofmeyr et al. [14] model potential policy interventions for controlling malware. They analyze the trade-off between prevented harm and collateral damage caused by blocking legitimate traffic.

Blacklisting has been put forward in the context of anonymous communication systems, such as Tor, JAP, or Mixminion, too. In [13], the authors formally define anonymous blacklisting systems and specify their security and privacy features. Such systems should allow users to authenticate anonymously with a service provider, while enabling the service provider to revoke access from abusive users without knowing their identities. Decentralized anonymity infrastructures (namely, mix-nets [10]) are also the focus of the paper by Acquisti et al. [2]. Since anonymity can be obtained only within an anonymity set [22], the authors explore with a game-theoretic approach the economic incentives of users to offer and use anonymity services.

In the growing literature on cryptocurrencies, the most closely related works can be classified into those that concern the implications of blacklisting and transaction risk scoring [19], and those that conduct various kinds of blockchain analyses in order to examine the (lack of) anonymity in the Bitcoin network [4,16,23,24]. Our paper draws on the ideas initially set out in [19], which discusses the potential use of blacklisting in Bitcoin and introduces the propagation policies. It is also inspired by works on the design [8] and use of centralized mixing services, as well as efforts to detect and break mixing schemes [33].

7 Concluding Remarks

This paper tackles the issue of non-fungibility of decentralized currencies and discusses its potential implications on the behavior of users. Specifically, it proposes a game-theoretic model of mixing coins of different quality under the regimes of perfect and imperfect information and analyzes three variants of it, one for each of three propagation policies. It finds the optimal strategies of players in the game of perfect information, and confirms that a Nash equilibrium in case of imperfect information is to mix bad coins only. Although the current operation of distributed ledgers is closer to the regime of imperfect information, we can still observe the existence of mixers. This is despite a surge of startups specializing in blockchain intelligence, allegedly to supply critical intermediaries, such as exchanges, with private blacklists. We conjecture that this discrepancy between theory and practice is due to several reasons, chiefly limited scope, lack of enforcement, or lack of reliability of existing blacklists. Alternative explanations include very high valuations of anonymity by some users, or simply nativity paired with luck of escaping negative experience.

There are several potential avenues for more rigorous and general models of mixing. First, the measurement of anonymity needs to be refined by taking other relevant transaction features into account. Second, the model needs to generalize to multiple players who choose inputs of arbitrary nominal value, but are constrained in terms of quality. Third, future research should elaborate more on market mechanisms for the survival of mixing services, e. g., by designing possible sanctions for the use of bad coins. The model can also be advanced by taking miner fees and the size of transactions into account. Finally, future work could examine whether it is possible to enforce side payments of the mixing fee without compromising the anonymity of any of the participants, and how this changes the game and its solutions.

Acknowledgments. The authors are grateful to Daniel G. Arce for his insightful comments on an earlier version of this paper. The authors are responsible for all remaining errors and omissions. This work was funded by the German Bundesministerium für Bildung und Forschung (BMBF) under grant agreement No. 13N13505 and by Archimedes Privatstiftung, Innsbruck, Austria.

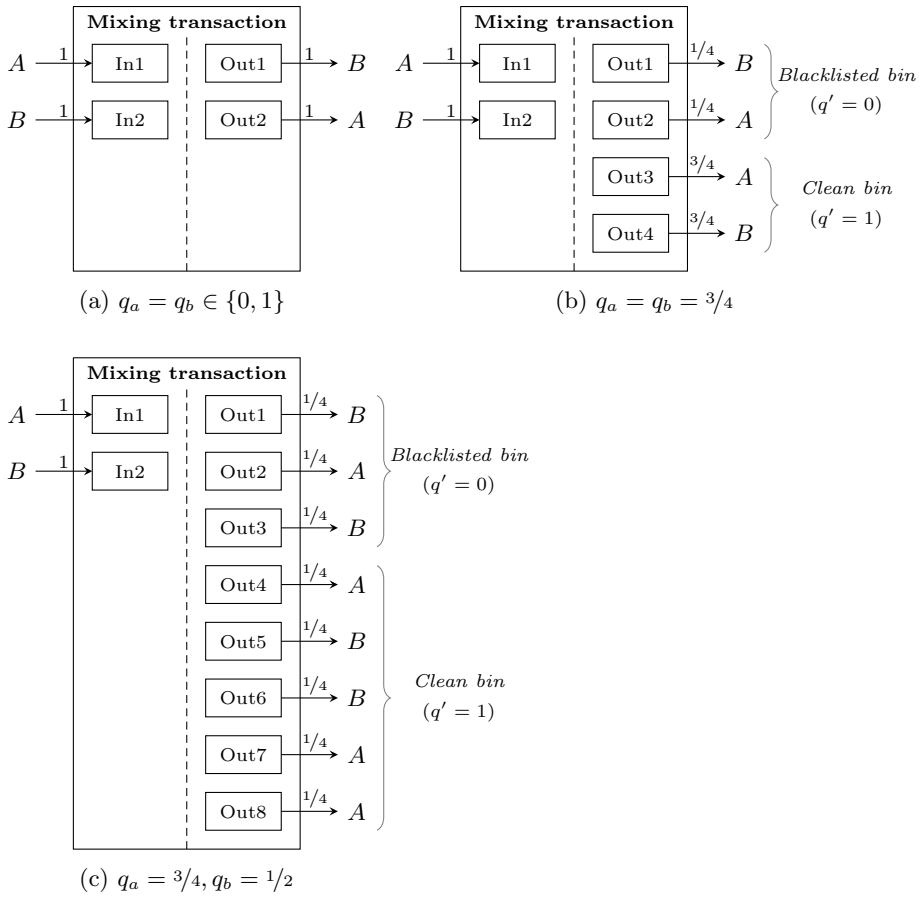
References

1. Aceto, G., Pescapé, A.: Internet Censorship Detection: A Survey. Computer Networks 83, 381–421 (2015)
2. Acquisti, A., Dingledine, R., Syverson, P.: On the Economics of Anonymity. In: Wright, R.N. (ed.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 2742, pp. 84–102. Springer, Berlin Heidelberg (2003)
3. Akerlof, G.A.: The Market for "Lemon": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics 84(3), 161–167 (1970)

4. Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T., Capkun, S.: Evaluating User Privacy in Bitcoin. In: Sadeghi, A.R. (ed.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 7859, pp. 34–51. Springer, Berlin Heidelberg (2013)
5. blockCollector: Looking to buy an old 50 BTC block. Where to buy? (2015), https://www.reddit.com/r/Bitcoin/comments/3sg8vm/looking_to_buy_an_old_50_btc_block_where_to_buy/, accessed on 14 November 2016
6. Böhme, R., Christin, N., Edelman, B., Moore, T.: Bitcoin: Economics, Technology, and Governance. Journal of Economic Perspectives 29(2), 213–238 (2015)
7. Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W.: SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In: Proceedings of the IEEE Symposium on Security and Privacy. pp. 104–121 (2015)
8. Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J.A., Felten, E.W.: Mixcoin: Anonymity for Bitcoin with Accountable Mixes. In: Christin, N., Safavi-Naini, R. (eds.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 8437, pp. 486–504. Springer, Berlin Heidelberg (2014)
9. Chacko, G., Jurek, J., Stafford, E.: The Price of Immediacy. The Journal of Finance 63(3), 1253–1290 (2008)
10. Chaum, D.L.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM 24(2), 84–90 (1981)
11. Edwards, B., Moore, T., Stelle, G., Hofmeyr, S., Forrest, S.: Beyond the Blacklist: Modeling Malware Spread and the Effect of Interventions. In: Proceedings of the 2012 Workshop on New Security Paradigms. pp. 53–66. ACM, New York, NY, USA (2012)
12. ExpertNeeded: Blockchain analysis help needed. Major money laundering case. (2016), https://bitcointalk.org/index.php?topic=1568048.0/, accessed on 14 November 2016
13. Henry, R., Goldberg, I.: Formalizing Anonymous Blacklisting Systems. In: Proceedings of the IEEE Symposium on Security and Privacy. pp. 81–95. IEEE Computer Society, Washington, DC, USA (2011)
14. Hofmeyr, S., Moore, T., Forrest, S., Edwards, B., Stelle, G.: Modeling Internet-Scale Policies for Cleaning up Malware. In: Schneier, B. (ed.) Economics of Information Security and Privacy III. pp. 149–170. Springer, New York (2013)
15. Meiklejohn, S., Orlandi, C.: Privacy-Enhancing Overlays in Bitcoin. In: Brenner, M., Christin, N., Johnson, B., Rohloff (eds.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 8976, pp. 127–141. Springer, Berlin Heidelberg (2015)
16. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A Fistful of Bitcoins: Characterizing Payments Among Men with No Names. In: Proceedings of the 2013 Conference on Internet Measurement Conference. pp. 127–140. ACM, New York, NY, USA (2013)
17. Möser, Böhme, R.: Join Me on a Market for Anonymity. In: Proceedings of the 15th Annual Workshop on the Economics of Information Security. Berkeley, CA, USA (2016)
18. Möser, M., Böhme, R., Breuker, D.: An Inquiry into Money Laundering Tools in the Bitcoin Ecosystem. In: APWG eCrime Researchers Summit (ECRIME). pp. 1–14. San Francisco, CA, USA (2013)
19. Möser, M., Böhme, R., Breuker, D.: Towards Risk Scoring of Bitcoin Transactions. In: Böhme, R., Brenner, M., Moore, T., Smith, M. (eds.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 8438, pp. 16–32. Springer, Berlin Heidelberg (2014)

20. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008), http://www.bitcoin.org/bitcoin.pdf, accessed on 14 November 2016
21. Noether, S.: Ring Confidential Transactions (2015), https://eprint.iacr.org/2015/1098.pdf, Cryptology ePrint Archive, Report 2015/1098, accessed on 14 November 2016
22. Pfitzmann, A., Hansen, M.: Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology (2005), Technical report
23. Reid, F., Harrigan, M.: An Analysis of Anonymity in the Bitcoin System. In: Altshuler, Y., Elovici, Y., Cremers, B.A., Aharony, N., Pentland, A. (eds.) Security and Privacy in Social Networks. pp. 197–223. Springer, New York (2013)
24. Ron, D., Shamir, A.: Quantitative Analysis of the Full Bitcoin Transaction Graph. In: Sadeghi, A.R. (ed.) Financial Cryptography and Data Security. Lecture Notes in Computer Science, vol. 7859, pp. 6–24. Springer, Berlin Heidelberg (2013)
25. Rosenfeld, M.: Overview of Colored Coins (2015), https://bitcoil.co.il/BitcoinX.pdf, accessed on 14 November 2016
26. Sheng, S., Wardman, B., Warner, G., Cranor, L.F., Hong, J., Zhang, C.: An Empirical Analysis of Phishing Blacklists. In: Proceedings of the 6th Conference on Email and Anti-Spam. CEAS'09 (2009)
27. Siegel, D.: Understanding The DAO Hack for Journalists (2016), https://medium.com/@pullnews/understanding-the-dao-hack-for-journalists-2312dd43e993, accessed on 14 November 2016
28. Spence, M.: Job Market Signaling. The Quarterly Journal of Economics 87(3), 355–374 (1973)
29. Tadelis, S.: Game Theory: An Introduction. Princeton University Press, Princeton, New Jersey, USA (2013)
30. TraderSteve: Bitpay is blacklisting certain bitcoins & rejecting customers. (2015), https://www.reddit.com/r/Bitcoin/comments/3mea6b/bitpay_is_blacklisting_certain_bitcoins_rejecting/, accessed on 14 November 2016
31. Tsalis, N., Virvilis, N., Mylonas, A., Apostolopoulos, T., Gritzalis, D.: Browser Blacklists: The Utopia of Phishing Protection. In: Obaidat, M.S., Holzinger, A., Filipe, J. (eds.) Proceedings of the 11th International Joint Conference on E-Business and Telecommunications. pp. 278–293. Springer International Publishing, Cham (2015)
32. Vasek, M., Weeden, M., Moore, T.: Measuring the Impact of Sharing Abuse Data with Web Hosting Providers. In: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security. pp. 71–80. WISCS '16, ACM, New York, NY, USA (2016)
33. Yanovich, Y., Mischenko, P., Ostrovskiy, A.: Shared Send Untangling in Bitcoin (2016), http://bitfury.com/content/5-white-papers-research/bitfury_whitepaper_shared_send_untangling_in_bitcoin_8_24_2016.pdf, working Paper, Bitfury Group Limited, accessed on 14 November 2016

A Seniority Policy



(a) $q_a = q_b \in \{0, 1\}$ (b) $q_a = q_b = 3/4$

(c) $q_a = 3/4, q_b = 1/2$

Fig. 4: **Seniority policy:** examples of mixing transactions in the perfect information regime: (a) shows the case when both coins are either good or bad (2 outputs); (b) – when both coins are of the same quality $q \in (0, 1)$ (4 outputs); (c) – when coins are of different quality (the number of outputs equals two times the least common divisor of the denominators of q_a and q_b expressed as rational numbers). The mixing fee is disregarded in these examples ($c = 0$).