SoK: Tools for Game Theoretic Models of Security for Cryptocurrencies

Sarah Azouvi
Protocol Labs
University College London

Alexander Hicks
University College London

Abstract

Cryptocurrencies have garnered much attention in recent years, both from the academic community and industry. One interesting aspect of cryptocurrencies is their explicit consideration of incentives at the protocol level, which has motivated a large body of work, yet many open problems still exist and current systems rarely deal with incentive related problems well. This issue arises due to the gap between Cryptography and Distributed Systems security, which deals with traditional security problems that ignore the explicit consideration of incentives, and Game Theory, which deals best with situations involving incentives. With this work, we offer a systematization of the work that relates to this problem, considering papers that blend Game Theory with Cryptography or Distributed systems. This gives an overview of the available tools, and we look at their (potential) use in practice, in the context of existing blockchain based systems that have been proposed or implemented.

1 Introduction

Since the deployment of Bitcoin in 2009, cryptocurrencies have garnered much attention from both academia and industry. Many challenges in this area have since been recognized, from privacy and scalability to governance and economics. In particular, the explicit consideration of incentives in the protocol design of cryptocurrencies (or "cryptoeconomics") has become an important topic.

The importance of economic considerations in security has been acknowledged since work by Anderson [8, 9], who recognized that many security failures could be explained by applying established ideas from Game Theory (GT) and Economics. However, the incentives at play tend to be external to the system design, and sometimes implicit, leading to failures when the the intended use of systems is misaligned with the incentives of users.

Cryptocurrencies, on the other hand, explicitly define some incentives in the design of their system, for example in the form of mining rewards, suggesting that they could be properly aligned and avoid traditional failures. Unfortunately, many attacks related to incentives have nonetheless been found for many cryptocurrencies [45, 46, 103], due to the use of lacking models. While many papers aim to consider both standard security and game theoretic guarantees, the vast majority end up considering them separately despite their relation in practice.

Here, we consider the ways in which models in Cryptography and Distributed Systems (DS) can explicitly consider game theoretic properties and incorporated into a system, looking at requirements based on existing cryptocurrencies.

Methodology

As we are covering a topic that incorporates many different fields coming up with an extensive list of papers would have been quite challenging, and would lead to an output of much greater length. In order to pick a representative subset of papers, we started by looking at existing surveys on the topic of Game Theory and Security [73,74,86,101,126,129], as well as specific book chapters on the topic e.g., Chapter 8 in the book by Nisan et al. [116]. We then looked at work published in popular Cryptography venues (e.g., IACR conferences), Security conferences (e.g., FC), as well as Distributed System venues (e.g., PODC) and interdisciplinary venues (e.g., WEIS, ACM Economics and Computation, P2PECON, GameSec), looking specifically for papers that cover both Game Theory and Security.

Different papers sometimes present different definitions for similar concepts, so we do not always include all these definitions. We also omit to present work on security and game theory that does not directly relate to what we discuss, e.g., the body of work by Tambe et al. [134] about physical security and allocation of limited security resources, or by Grossklags et al. [67] about security investments. Instead, we focus on some specific models of interest that we think are more appropriate, and match these with problems in the security of cryptocurrencies.

Our contributions

Our goal is to give an overview of the intersection of the three fields that are essential to the design of cryptocurrencies: Cryptography, Distributed Systems and Game Theory. Our contribution is an analysis of existing work that proposes solutions to this problem. Our analysis highlights new concepts introduced by these papers, as well as deficiencies. We do this in the context of security requirements that we formulate, arguing that they address deficiencies in existing security models that fail to cover all aspects of a decentralized cryptocurrency. We also discuss open challenges and how they could be addressed.

Section 3 introduces Game Theory and cryptocurrencies, and discusses security in the context of a decentralized system. In Section 4 we then look at the intersection of Cryptography and Game Theory, followed by the intersection of DS and Game Theory in Section 5. We then look at how these results are used in Section 6, where we look at proposed systems and their failures, tracing back to deficiencies identified in the two previous sections. In Section 7 we give a comparative table of some of the concepts presented in this paper and how they could be used in blockchain research, we discuss the open challenges posed by failures that are observed, and how they could be addressed. Finally, Appendix A collects formal definitions for the concepts presented in the paper.

2 Related work

The work that most closely resembles ours are the previous surveys bridging Computer Science and Game Theory [73, 74, 86, 101, 126, 129]. They were of great inspiration for this work, but they are quite outdated (dating back to 2002, 2005, 2007, 2008, 2010) given the recent output of research tied to cryptocurrencies and other blockchain based systems.

On the topic of blockchains, many SoK papers and surveys exist that cover consensus protocols and security [18, 22, 31, 59, 102, 133, 142]. These are very different from our work as we present general concepts and definitions related to designing decentralized systems with incentives. In particular, many concepts presented in this paper were not introduced in the context of consensus, but rather in the context of secure multiparty computation (MPC) or other problems tied to distributed systems. Most of the work presented in this SoK does not directly concern blockchains, which is the motivation behind this work.

3 Background

In this section we briefly introduce game theoretic tools that are mentioned throughout the paper such as solutions concepts and mechanism design (MD). For a complete introduction to Game Theory, the reader is invited to look at any of the books (or other resources) on the topic [82, 118]. For the sake of exposition we omit to cover concepts that are relevant e.g., correlated equilibria, Pareto efficiency and the single deviation test, as these do not appear in the papers we mention. We also discuss the interface between Game Theory and Cryptography and Security in practice, and briefly introduce Distributed Systems and cryptocurrencies.

3.0.1 Game Theory and Mechanism Design

A game is defined by a set of players and a set of actions for each player. A strategy for player is defined as a function from its local state to actions.

Game Theory uses solution concepts in order to predict the outcome of a game, the most well known is the *Nash equilibirum* (NE). A strategy is a Nash equilibrium if given that all the other players follow this strategy, player i is better off playing it as well.

It is often unrealistic to assume that players have complete information about the game they are in. A game where players do not always know what has taken place earlier in the game is said to have *imperfect information*. In the case where players do not know the *type* of the other players, which determines their utility function, the game is said to have *incomplete information*.

The players then have a probability distribution over the types of other players. Their beliefs are expressed as conditional probabilities based on the information they have, which they can update using Bayes' theorem when they gain new information, leading games of this form to be called *Bayesian games*.

Players now also think about expected payoffs, so the standard definition of a NE is no longer ideal but we can define a *Bayesian Nash equilibrium* (BNE) analogously by replacing utilities with expected utilities, although we still refer to them simply as utilities.

While Game Theory is typically about understanding the behavior of players in a given game, systems are usually designed and implemented with a goal in mind e.g., preventing double spending in cryptocurrencies. To achieve this, our goals can be expressed as a social choice function (SCF), a function that given the preference (or types) of all players outputs an outcome. For example, in a voting system, given all the ranked preferences of voters, a SCF will choose a candidate.

Once we have a target outcome in mind, the idea is to make sure that the incentives are designed in a way such that selfish players reach this outcome. In some ways, this can be thought of as reversing the basic idea of Game Theory, designing a game that leads to a specific outcome.

For example in a voting system we would like to design a system where given all the preferences of the players, the one chosen by the SCF is elected, one way to do this is to incentivize players to report their truthful preferences. A mechanism can also be viewed as a protocol, with the corresponding

game being thought of as having the protocol as the recommended strategy, and deviations from the protocol as other possible strategies.

It must be pointed out that doing this in practice is not always easy, as experimental Game theory reveals. Gneezy and Rustichini [63] looked at the effects of implementing fines at a nursery in order to reduce the rate at which parents collected their child late. Intuitively, this should motivate parents to arrive on time but parents instead interpreted the fine as a way of paying for extra childcare, and started coming even later. Furthermore, once the fine was removed as it was counter productive, the parents behavior did not revert back so the system had been irreversibly damaged.

A very important result in MD is the revelation principle that states that any social choice function that can be implemented by any mechanism can be implemented by a direct truthful mechanism. A mechanism is direct if players need only reveal their type/utility function to the designer of the game and truthful if players' best strategy is to reveal their true type.

3.0.2 Agents in Game Theory and Security

Both Game Theory and Security deal with the interaction of agents, but they differ noticeably in how they model these agents. Security deals with *adversaries*, agents that aim to circumvent or break a security property of the system. The value that an adversary attaches to their success is not usually given, as security should ideally be robust to any adversary, although they may have computational limitations. Game Theory, on the other hand, deals with *rational* (sometimes also called *selfish*) agents that assign a value to their goals, and would rather optimize their payoff than achieve an arbitrary goal, but they typically do not have restrictions (e.g., computational) like a security adversary would.

In practice, this translates to different assumptions being used when formally modelling a game or the security of a system, making it difficult to prove statements involving security and game theoretic properties. As both deal with different types of agents, a proof will involve complexity from both sides and quickly become hard to manage, leading to a disjoint treatment of both aspects in works that attempt to cover both. There are nonetheless inherent connections as security often uses game based proofs, although an adversary wins if they have a high enough probability of succeeding in their attack rather than if their utility is high enough. But if we assume that the adversary has a high payoff associated with the success of their attack, then we start to recover game theoretic intuition.

For proofs based on the simulation (ideal/real world) paradigm, some connections are also evident. The idea behind the simulation paradigm first originated in the context of secure computation. Goldreich et al. [64] introduced the idea of bypassing the need for a trusted third party (i.e., a mediator) in games of incomplete information, by replacing it with a protocol that effectively simulates it such that any information known by players at any step of game is the same as they would have known in an execution of the game involving the trusted third party. This was refined by Micali and Rogaway [109] in terms of ideal and secure function evaluation. The ideal function evaluation corresponds to the evaluation of the function with a trusted third party that receives the private inputs of the parties and evaluates the function before returning the result to each party, achieving the best possible result. The secure function evaluation involves the parties trying to replace the trusted third party with a protocol, which is considered secure if the parties cannot distinguish between the ideal and secure function evaluations, meaning that an adversary is not able to gain anything significant. Simulation has become a powerful tool for cryptographers [100] and Canetti's Universal Composability Model [32] expands on these ideas to provide a framework for secure composability of protocols.

3.0.3 Decentralization, incentives and security

Because the security of most decentralized systems, like cryptocurrencies, is linked not only to the security of the protocols, but also to having a majority of participants following the rules, decentralization and incentives have to be considered.

The ideal system does not depend in any way on any single party, which requires it to be decentralized. Troncoso et al. [136] give an overview of decentralized systems, defining a decentralized system as "a distributed system in which multiple authorities control different components and no single authority is fully trusted by all others". This highlights the fact that every component of the system should be decentralized, and in particular a single authority distributing its own system (or component) is not decentralized. This can be hard to achieve in practice, and the level of decentralization of a system should always be looked at critically. A decentralized system where all important parties are independent but under the jurisdiction of a single government may not truly be decentralized. All these independent parties may also depend on a very few hardware manufacturers (or other service providers).

Incentives are key to achieving an honest majority. Azouvi et al. [13] give an overview of the role incentives play in security protocols, including cryptocurrencies, highlighting the fact that achieving guarantees of equilibria on paper may not be meaningful in practice when the wrong assumptions and models are used. What does security mean in this context? Clearly, protocols that are cryptographically secure, and that achieve safety (i.e., the guarantee that nothing bad will happend) and liveness (i.e., the guarantee that somethind good will happen) are needed, otherwise nothing else would work. But if the security of the system also depends on achieving a high enough degree of decentralization, more is required. In particular, decentralization relates to the participants and their

behavior rather than solely the protocol. Any decentralized protocol can always be ran in a centralized manner, so it is not enough to design a system that can be used in a decentralized manner. Rather, the requirement is to design a system that is advantageous to use in decentralized manner.

Doing this naturally requires a better understanding of why users would want to be decentralized rather than try to gain more individual control of the system for themselves, so security is no longer just about the protocol itself, but also about how it can be used and how it is used.

Bitcoin represents an important innovation from classical consensus protocols as it is fully open and decentralized. In the Bitcoin consensus protocol (sometimes called Nakamoto consensus) participants can join and leave as they wish, and Sybils are handled through the use of *Proof-of-Work* (PoW).

The data structure that keeps track of the state of the system in Bitcoin is a chain of chronologically ordered blocks i.e., the blockchain, with each block containing a list of transactions. To win the right to append a block (and win the block reward), participants compete to solve a computational puzzle i.e., a PoW. They include in their block the solution to that puzzle, the PoW, such that other players can verify its correctness. This block then initiate a new puzzle to be solved. This process of creating new blocks is called *mining* and participants in this protocol are called *miners*.

The security of Bitcoin relies on a majority of the mining power (i.e., hashing power) in the network following the protocol, whether because they are honest or simply rational. Taking control of half of the computational power of Bitcoin for only an hour has a considerable cost (around 670k USD [2] as of May 2019), although it is with reach of potential adversaries. This cost depends on the hash rate of the network (i.e., the cost of mining) and the price of Bitcoin in USD as once mining is no longer profitable for some miners they are likely to stop mining, reducing the hashing power required to control a majority of the network.

This is one of the reason why Bitcoin's security is so tightly linked to incentives, as when mining is no longer worthwhile the security of the network decreases. Participation in the network is also rewarded by financial gain (through block rewards and transaction fees). The more participants there are, the harder it is to attack the network since the cost for mounting a 51% attack (where an adversary takes control of more than half of the computational power) increases. These financial motivations are thus also paramount.

Since Bitcoin's deployment, many alternative cryptocurrencies that similarly rely on a blockchain have emerged. The most popular of these is Ethereum [3], which differs from Bitcoin in that it provides a more complex scripting language meaning that rather than processing simple transactions, nodes in the system execute a script that allows users to perform multitude of functionalities (so-called *smart contracts*).

4 Cryptography and Game Theory

This section considers work at the intersection of Game Theory and Cryptography. Cryptography usually considers a worst-case adversary, but by relaxing this assumption, it is possible to design protocols that bypass impossibility results or achieve better efficiency than existing ones, while maintaining a realistic adversarial model.

4.1 Cryptography meets Game Theory: Rational Cryptography

Initiated by Dodis, Rabin and Halevi [39] rational cryptography is a subfield of cryptography that incorporates incentives in cryptographic protocols. In this context, new adversaries and their capabilities have to be defined, as well as how to account for incentives and how protocols can be proven secure for such adversaries.

First, we note that most of this work [11,40,56,65,71,89,90] focuses on multi-party secret sharing or secure function evaluation. Thus, no monetary incentive is usually considered. As pointed out by Dodis and Rabin [40], in a rational cryptographic context, the utilities of the players are usually dependent on cryptographic considerations such as: correctness (a player prefers to compute the function correctly), exclusivity (a player prefers that other players do not learn the value of the function correctly), privacy (a player does not want to leak information to other players), voyeurism (a player wants to learn as much as possible about the other parties).

In addition to the above, other interesting parameters can come to play in the adversary's utility function. For example, Aumann and Lindell [12] formalized the concept of *covert adversaries* that may deviate from the protocol but only if they are not caught doing so. As they argue, there are many obvious situations where parties cannot afford the effect of being caught cheating.

Covert adversaries are somehow similar to adding a punishment to the utility function. Rational players do not want to be caught cheating as the punishment decreases their utility. In Aumann and Lindell's setting the protocol detects the cheating, but in practice we need to incentivize participants to do so. Some work considers adding adversarial behavior together with rational adversaries [105], we consider this further in Section 5.

In terms of equilibria, the solution concepts proposed in these works are often extensions of a Nash Equilibria (NE), introduced in Section 3. For example, Halpern and Teague look for a NE that remains after other NE that are weakly dominated (i.e., at best only as good as others) are removed through iterated deletion, where all dominated strategies are removed at each step [71]. Asharov et al. [11] adapt the simulation-based definition to capture game-theoretic notions of (for example) fairness, meaning that one party learns the output of a computation if and only if the other does as well. As their

notions are weaker than standard cryptographic definitions, they can be achieved in some settings where impossibility results usually hold in traditional cryptography.

Following the work presented above, Garay et al. propose Rational Protocol Design (RPD) [57]. In this setting, they define a game between the designer of the protocol D and the attacker A. The game is parametrized by a multi-party functionality $\mathcal{F}$ and consists of two sequential moves. In the first step, D sends to A the description π of the protocol that honest parties are supposed to execute. In the second step, A chooses a polynomial-time interactive Turing machine (ITM) **Adv** to attack the protocol. The game is zero-sum in the original paper, but was later adapted to be a non-zero sum game in the context of Bitcoin [17].

In follow-up work [58], notions of fairness are also considered, and provide a mean of comparison between protocols i.e., which protocol is the fairer. Informally, a protocol π will be at least as fair as another protocol π_0 if the utility of the best adversary A attacking π (i.e., the adversary which maximizes $u_A(\pi, A)$) is no larger than the utility of the best adversary attacking π_0, except for some negligible quantity.

The solution concept introduced within the RPD framework is $\varepsilon-$subgame perfect equilibrium where the parties' utilities are ε close to their best response utilities. When it comes to security, the RPD framework defines the notion of attack-payoff security. Informally, attack payoff security states that an adversary has no incentive to deviate from the protocol.

Another concept, incentive compatibility, was introduced in a follow-up work of RPD [17]. Here, the definition is slightly different than the definition usually given within Mechanism Design (MD) where participants achieve the best outcome by revealing their true preferences. Informally, incentive compatibility states that agents gain some utility when participating in the protocol i.e., they choose to play instead of "staying at home".

Apart from the recent work of Badertscher, et al. [17], rational cryptography does not consider monetary payment.

One important drawback of RPD is that it does not consider the presence of irrational adversaries despite the fact that in security, we do not always know the motivation of an attacker. RPD uses a relaxed functionality to allow for some defined attacks but this may not cover all attacks, leaving the door open to potential attacks. The UC model does not automatically start accounting for all possible incentives - this is a clear flaw as we know that only arbitrarily considering incentives leads to failures (e.g., failure of considering outside incentives, or in general "soft" incentives like political and other external incentives [13]).

4.2 Game Theory meets Cryptography: computational games

Rather than starting from a cryptographic setting and incorporating game theoretic notions, as presented above, one can also start from a game theoretic setting and from there move towards cryptographic notions by considering the computational aspects of games. This approach is taken in a body of work by Halpern and Pass that considers *Bayesian machine games* first introduced in a preprint [69] that has later appeared in different forms [76, 78, 120], primarily venues focused on Economics rather than Security.

A Bayesian machine game (BMG) is defined very similarly to a standard Bayesian game (introduced in Section 3), it only differs in that it considers the complexity (in computation, storage cost, time or otherwise) of actions in the game. This is done by having players pick machines (e.g., a TM or ITM) that will execute their actions and defining a complexity function for that machine, which the utility takes into account.

A Nash equilibrium for a BMG is expressed in the usual way, but it now takes into account the machine profile rather than a strategy profile. There is, however, an important distinction to make between a standard Nash equilibrium and a Nash equilibrium in machine games, which is that the latter may not always exist. The necessary conditions for the existence of a Nash equilibrium in a machine game are given by Halpern and Pass [69] to be a finite type space, bounded machines and a computable game. A follow up paper by Halpern et al. [70] discusses the general question of the existence of a Nash equilibrium for resource bounded players.

So far, the discussion of computational games has not yet touched on security related issues, but Halpern and Pass prove an equivalence theorem that relates the idea of universal implementation in a BMG to the standard notion of secure computation in Cryptography [20, 64]. Intuitively, this goes back to the work of Goldreich, Micali and Widgerson [64] that first expressed (to the best of our knowledge) the idea of secure computation as the replacement of a mediator in a game that preserves an equilibrium.

A universal implementation corresponds to the idea that a BMG implements a mediator if whenever a set of players want to truthfully provide their input to the mediator, they also want to run their machine using the same input, preserving the equilibrium and action distribution. There are then multiple equivalence theorems of different strength (up to the information theoretic case), that relate flavors of secure computation to flavors of implementation. The relation is important, as it not only implies that secure computation leads to a form of game theoretic implementation, but also the reverse. This opens up the option that the guarantees of (some flavor of) secure computation could be achieved by considering the Game Theory of a problem, although it is not clear whether this process would be more efficient.

BMG have natural applications to known security problems. For example, dealing with covert adversaries as described by Aumann and Lindell [12] (introduced above) can be done by introducing a (two player, for example) mediated game where the honest strategy is to report your input to the mediator and output its reply (with utility $\frac{1}{2}$), and the string *punish* can be

output by a player to ensure the other receives payoff 0. Then any secure computation with respect to covert adversaries with deterrent (probability of getting caught cheating) $\frac{1}{2}$ is an implementation of the mediator as the expected utility of a cheating player will be $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}$, which is the same as that of the honest strategy.

5 Distributed Systems and Game Theory

In this section, we present the work that is at the intersection of Game Theory and Distributed Systems, looking at concrete problems that have been well studied. For each case we will illustrate important concepts and techniques used.

5.1 Algorithmic Mechanism Design

Algorithmic mechanism design (AMD) is concerned with designing games such that self-interested players achieve the game designer's goals, in the same way that distributed systems designers aim, for example, to achieve agreement in the presence of Byzantine players. AMD was first introduced by Nisan and Ronen [115] who proposed that an algorithm designer should ensure that the interests of participants in a distributed setting are best served by behaving correctly i.e., the algorithm designer should aim for incentive compatibility. The framework of Nisan and Ronen is defined for a centralized computation, but it has been extended to distributed algorithmic mechanism design (DAMD) following work by Feigenbaum et al. on cost sharing algorithms for multicast transactions [48]. This lead to further developments and applications of DAMD to interdomain routing, web caching, peer-to-peer file sharing, application layer overlay networks and distributed task allocation, which are summarized in a review by Feigenbaum and Shenker [50].

5.2 Public goods, free riding and hidden actions

Public goods, which are produced at a cost but available to use for free, naturally occur in distributed systems. In a public goods game, players choose to contribute a certain amount, with all contributions being combined and distributed among all players. Naturally, this can lead to players rationally deciding to contribute less to maximize their utility. They may even contribute nothing, which is generally referred to as *free riding*.

Varian first considered modeling the reliability of a system as a public good [138]. The reliability can either depend on the total effort (sum of the efforts exerted by the individuals), on the weakest link (minimum effort) or on the best shot (maximum effort). E.g., if there is a wall defending a city, its reliability can depend on the sum of all the work provided by the builders (total effort), on the lowest height (weakest link) or if we consider several walls, on the highest one (best shot). In the case of total effort the NE corresponds to all players free riding on the player with highest benefit-cost ratio. Moreover, the effort exerted in the NE is always lower than the social optimum i.e., the best outcome across all players.

Peer-to-peer file sharing also provides an interesting case of a networked system that has faced free riding [51]. As explained by Babaioff et al. [15], solutions to this problem could be based on a reputation system, barter or currency. Another approach, that would not need to keep any long term state information is to replace *indirect reciprocity* with *direct reciprocity*. For example, a file in BitTorrent is partitioned into smaller chunks, requiring repeat interactions among peers and enforcing more collaboration between them [37]. In practice, however, this has been shown to not be very effective as it is not robust to strategic agents [123] and induces free riding [84]. There is also the problem of dealing with newcomers, as an adversary can create new identities in order to abuse the system. Analyzing the incentives at play, Feldman et al. [52] suggest that penalizing all newcomers may be an effective way of dealing with the problem, as it is not possible to penalize only users abusing the system.

In addition to free-riding, there are many other parameters that a selfish player could abuse in a P2P file sharing system e.g., when to join or leave, who to connect to, untruthful sharing information, and so on. This is the problem of *hidden actions* i.e., how peers selfishly behave when their actions are hidden from the rest of the network. In order to analyze the degradation due to hidden actions, Babaioff et al. [15] apply the principal-agent framework, due to the similarity of the hidden action problem with that of *moral hazard*. This framework is used in economics when one entity, the *principal*, employs a set of n agents to take actions on its behalf. In order to capture the efficiency of a system in that framework, they define the *Price of Unaccountability* of a technology as the worst ratio between the principal's utility in the observable-actions case and the hidden actions case. Dealing with observable and hidden actions relates to the transparency of the system, which can be approached from a cryptographic point of view to ensure that agents all see the same set of actions [34]. Another solution, Karma [141], proposes a system for peer-to-peer resource sharing that avoids free riding, based on a combination of reputation system and consensus protocols.

Another important problem in Distributed Systems where rationality can cause problems is routing [127]. The problem is to find a path that minimizes the latency between a source and a target. One of the difficulties in doing so is that in a decentralized communication network it is not always possible to impose some routing strategy to nodes in order to, for example, regulate the load on a route. As highlighted in our background on Game Theory in Section 3, nodes usually act according to their own interests, which can be orthogonal to the overall optimal equilibrium. A game theoretic measure used by Roughgarden and Tardos in the context of routing

is the *Price of Anarchy* [127], which quantifies how much a system degrades due to selfish behavior. More formally, assuming we have a measure of the efficiency of each outcome, the Price of Anarchy is the ratio between the equilibrium and the optimal outcome. Inspired by this measure, Grossklags et al. [68] introduce the *Price of Uncertainty*, which measures the cost of incomplete information compared to that of complete information. An important observation is that assuming fixed possible losses, which is reasonable in the case of mining where one can at most lose the fixed cost hardware (and electricity) or stake, the more players are in the network, the less information matters. This also ties in to the value of information i.e., the possible change in utility from gaining information, which is defined for a computational setting by Halpern and Pass [75].

5.3 Consensus

5.3.1 Fault tolerance with rational players

We now look at the example of consensus. The approach here is to use incentives to bypass impossibility results on Byzantine Agreement or improve on existing constructions. In order to apply GT to DS, additional adjustments have to be made. For example, traditional GT considers deviation from only one agent (as in a NE) while in practice agents form coalition. In addition, in a DS it is important to consider multiple types of failures (e.g., processors may crash) that are not considered in GT.

In order to account for both of these requirements, the BAR model defined by Aiyer et al. [7] introduces three different types of players: Byzantine, altruistic (players that simply follow the rules) and rational players. In this case, the expected utility of a rational player is usually defined by considering the worst configuration of Byzantine players and the worst set of strategies that those Byzantine players could take, assuming all other non-Byzantine players obey the specified strategy profile. The goal of the BAR model is to provide guarantees similar to those of Byzantine fault tolerance to all rational and altruistic nodes, as opposed to all correct nodes. Two classes of protocols meet this goal, Incentive-Compatible Byzantine Fault Tolerant (IC-BFT) protocols and Byzantine Altruistic Rational Tolerant (BART) protocols. IC-BFT protocols, which are a subset of BART protocols, ensure that the protocol satisfies security and is the optimal one for rational nodes, while a BART protocol simply ensures security properties.

Groce et al. [66] introduce similar notions, perfect and statistical security, which state that in the presence of a rational adversary, the protocol still satisfies the security properties (e.g., consistency and correctness for consensus). They show feasibility results of information-theoretic (both perfect and statistical) Byzantine Agreement, assuming a rational adversary and complete or partial knowledge of the adversary preferences. Their protocols are also more efficient than traditional Byzantine Agreement protocols.

In the DAMD setting [49], participants are split into obedient, faulty, strategic and adversarial nodes of the network. The split follows the same lines as that of the BAR model, but separates the adversarial nodes from those that are faulty with no strategic goal. Computational restrictions here are expressed with regards to the solution concepts rather than the agents. This ties into topics in computational Game Theory, as a solution to a DAMD problem requires not only that incentive compatibility is achieved, but also that the solution be computationally tractable, which is not always the case. (The tractability of computing Nash equilibria, or approximations, is out of the scope of this paper.) The takeaway is that many solutions on paper are not straightforwardly obtained in an algorithmic setting, whether centralized or decentralized, and even approximations may not be enough.

5.3.2 Robustness

When it comes to adapting a NE to consider coalitions and irrational players Abraham et al. [5] extend the work of Halpern and Teague [72] to consider multiple players. They introduce the concept of robustness that encompasses two notions, resilience and immunity. Resilience captures the fact that a coalition of players has no incentive to deviate from the protocol, and is similar to the concept of collusion-proof NE [74]. Immunity captures the fact that even if some irrational players are present in the system, the utilities of the other players are not affected. An equilibrium that is both resilient to coalitions of up to k players, and immune to up to t irrational players is then said to be (k,t)-robust

Robustness is a very strong property, but it is hard to achieve in practice. Clement et al. [36] show that no protocol is (k,t)-robust if any node may crash and communication is necessary and costly. When designing cryptocurrencies, however, it is not unusual to consider that communication is free.

As discussed in Section 4.1 with covert adversaries, it can be helpful to add a form of punishment to enforce correct behavior by rational players. Halpern et al. [5] define a $(k,t)-$punishment strategy such that for any coalition of at most k players and up to t irrational players, as long as more than t players use the punishment strategy and the remaining players play the equilibrium strategy, then if up to k players collude, they will be worse off than they would have been if the rational players had played the equilibrium strategy. The idea is that by having more than t players use the punishment strategy is enough to stop k players colluding and deviating from the equilibrium strategy.

5.3.3 Price of Malice

As systems realistically involve rational and irrational players, it is important to consider how rational players react to the presence of irrational players. Moscibroda et al. [112] do this by considering a system with only rational and Byzantine players. They differentiate between an *oblivious* and *non oblivious model* i.e., whether selfish players know the existence of Byzantine players or not. They define a *Byzantine Nash equilibria* that extends NE in the case where irrational players are present. In a Byzantine Nash equilibria no selfish player can reduce their perceived expected cost, which depends on their information, by changing their strategy, given that the strategies of all other players are fixed.

In GT and MD, a concept very often discussed is the Price of Anarchy [126], which was introduced in the case of selfish routing. Moscibroda et al. [112] extend this to their setting, by defining the *Byzantine Price of Anarchy* that quantifies how much an optimal system degrades due to selfish behavior, when malicious players are introduced. More formally, it is the ratio between the worst social cost of a Byzantine Nash equilibrium and the minimal social cost, where the social cost of a strategy profile is the sum of all individual costs i.e., the optimality of each outcome.

The price of Malice is used to see how a system of purely selfish players degrades in the presence of malicious irrational players. More formally it is the ratio between the worst Byzantine Nash Equilibrium with malicious players and the Price of Anarchy in a purely selfish system.

Moscibroda et al. [112] also introduced the idea that Byzantine players can improve the overall system, which they called the *fear factor*. The intuition is that the rational players will adapt their strategies by fear of the actions of irrational players, rendering the overall system better. The example they introduce where this can be observed is virus inoculation. Based on the assumption that some players are irrational and will not get vaccinated, rational players will be incentivized to get vaccinated. In the case where everyone is rational, there is no equilibrium since as long as enough people get vaccinated, the rest of the population is safe. Thus irrational players here can make the overall system better.

6 Blockchains

We now consider blockchain based cryptocurrencies, which are an important example of systems involving aspects of both traditional security and game theoretic aspects. In this section we review the work that has been done by the security and distributed systems communities on blockchains that consider game-theoretic notions.

We start by reviewing the work that give some game-theoretic analysis of Bitcoin. We then illustrate how these analysis fell short with attacks that have been found on Bitcoin's incentives. We then gove an overview of the work that has been done on blockchain consensus protocols that considers the question of incentives and try to thwart these attacks. As we argued in Section **??**, having a system that is secure with some bounded number of Byzantine faults is not enough to have a decentralized system as decentralization cannot be assumed. Rather, incentives should be designed to ensure enough participation and prevent coalitions. We therefore also discuss work focusing on incentivizing decentralization. We then review the work on payment channels before finally presenting some notions of fairness with respect to blockchains' reward systems.

Along the way, we also highlight new concepts of interest that have been introduced in this field as well as how they relate to what we have previously discussed in this paper.

6.1 Game theoretic analysis of Bitcoin

Nakamoto's original Bitcoin paper [113] provided only informal security arguments but several papers have since formally argued the security of Bitcoin in different models [60,88,121], usually based in the simulation setting presented in Section 3, but without a consideration of incentives.

In early work in this area Kroll et al. [91] show that there is a NE in which all players behave consistently with Bitcoin's reference implementation, along with infinitely many equilibria in which they behave otherwise e.g., where they all agree to change a rule. Attacks like selfish mining [46,114,128] put this into question, showing that their model did not encompass behavior that could realistically occur. More recently, Fiat et al. [53] showed that the only possible pure equilibria in Bitcoin's mining are very chaotic (miners quitting and starting again periodically) or non-existent, depending on the configuration of players..

More recently Garay et al. [17] proved the security of Bitcoin in the RPD framework that was introduced in Section 4.1. Their approach is based on the observation that Bitcoin works despite its flaws, and they prove that Bitcoin is secure by relying on the rationality of players rather than an honest majority. This model inherits the flaws discussed in Section 4.1 e.g., they do not consider fully malicious players. Their model also does not encompass attacks on Bitcoin's incentive structure that we now describe.

6.2 Attacks on incentives

With time, many attacks related to incentives in cryptocurrencies have been found, typically involving either external incentives (e.g., in bribery attacks) or the unintended use of a cryptocurrency's technical mechanisms. The effect of these attacks results in lowering the power required for a 51% attack to less than 51%.

Selfish mining [46] involves a rational miner increasing their expected utility by withholding their blocks instead of broadcasting them to the rest of network, giving them an

advantage in solving the new proof-of-work and making the rest of the network waste computation by mining on a block that is not the top of the chain.

Inspired by techniques introduced by Gervais et al. [62], Sapirshtein et al. [128] use Markov Decision Processes (MDP) to find the optimal strategy when doing selfish mining. (MDP are used to help make decisions in a discrete state space where outcomes are partially random.) They show that with this strategy, an adversary could mount a 51% attack with less than 25% of the computational power. This problem is further studied by Hou et al. [81] using deep reinforcement learning. They suggest that selfish mining becomes less effective when performed by multiple adversaries. In addition to witholding their own block, miners are neither incentivize to propagate information (e.g. transactions or blocks) to the rest of network.

This is a problem that also exist in any P2P systems [43,98] that researchers have also looked into solving, using techniques similar to those proposed in Distributed Systems [6, 16, 44].

Another issue is the verifier dilemma [103], which shows that miners are not incentivized to verify the content of blocks, especially when this incurs an important computation on their end.

Mining gaps are another type of attack on incentives [33, 137] where the time between the creation of blocks increases because miners wait to include enough transactions (in order to get the transaction fees).

Bribery attacks are another family of attacks that are often thought of as an example of the *tragedy of the commons*, which describes a situation when individuals acting selfishly affect the common good [79]. In our context, it captures the fact that miners have to balance their aim to maximize their profit with the risk of affecting the long term health of the cryptocurrency they mine, potentially reducing its price and their profit.

Bonneau [23] first proposed that an adversary could mount a 51% attack at a much reduced cost by renting the necessary hardware for the length of the attack rather than purchasing it. More generally, a briber could pay existing miners to mine in a certain way, without ever needing to acquire any hardware. This lead to a series of papers [24, 99, 107, 135, 140] that show it is possible to introduce new incentives to an existing cryptocurrency, internally or externally, in ways that do not require trust between miners and briber (e.g., using smart contracts).

Ethereum's uncle reward mechanism (that allows blocks that were mined but not appended to the blockchain to later be referenced in another block for a reward) can be used to subsidize the cost of bribery attacks [107] and selfish mining [117, 125]. This is unfortunate, as they were originally introduced to aid decentralization [29] but have now been found to introduce incentives that work against this, by reducing the mining power required to perform certain attacks.

This puts into question the value of saying that a cryptocurrency is incentive compatible if new incentives can later be added. A cryptocurrency also does not exist in a vacuum, and external incentives can always manifest in adversarial ways. Goldfinger attacks, proposed by Kroll et al. [91], involve an adversary paying miners of a cryptocurrency to sabotage it by mining empty blocks. In some cases, even the threat of this type of attack can be enough to kill off a cryptocurrency, as users would not want their investments to disappear if the attack happens, and thus would not invest. As a Goldfinger attack can be implemented through a smart contract in another cryptocurrency [107], it is not inconceivable that this could be attempted in practice. This clearly shows that incentives from outside the cryptocurrency itself must be considered.

Budish [28] proposes an economic analysis of 51% attack and double spending and shows that the Nakamoto consensus has inherent economic limitations. In particular, he shows from a strictly economic point of view that the security of the blockchain relies on scarce, non-repurposable resources (i.e., ASICs) used by miners as opposed to Nakamoto's vision of "one-CPU-one-vote", and that the blockchain is vulnerable to sabotage at a cost linear in the amount of specialized computational equipment devoted to its maintenance.

6.3 Other blockain consensus protocols

As an alternative to existing systems, like Bitcoin and Ethereum, that have been shown to be vulnerable to the attacks we have just described, systems based on BlockDAGs rather than blockchains have been proposed [4, 14, 38, 130–132]. In this model, the data structure is a Directed Acyclic Graph (DAG) of blocks, meaning that each block can have more than one parent block. When creating a new block, a miner points to all the blocks that they are aware of, revealing their view of the blockchain. This exposes more of the decision making of the players and relates to the idea of hidden actions discussed in Section 5.

Due to some additional inherent flaws in Bitcoin e.g., scalability and energy consumption, new design papers are constantly proposed by both the academic community and industry, but many leave the treatment of incentives as future work. In particular, very few papers propose an incentive scheme associated with their consensus protocols [14, 87, 119, 122]. Moreover, the solution concepts considered in these papers are often overly-simplistic; e.g., some coalition proof NE that does not consider the impact of irrational players [21, 87, 119, 122]. Only Solidus [6] and Fantômette [14] consider robustness (introduced in Section 5).

In his draft work about incentives in Casper [30], Buterin introduces the *griefing factor* which is the ratio of the penalty incurred to the victim of an attack and the penalty incurred by the attacker. The idea of a griefing factor intuitively makes sense, as disputes in the real world can be resolved by fining a party according to the damages caused, and from a modelling point of view gives a quantifiable punishment that can be

explicitly taken into account when computing equilibria. He also proves that following the protocol in Casper is a NE as long as no player holds more than a third of the deposit at stake.

Due to the lack of formal model, it can be expected that more incentive related attacks will be proposed. For example, attacks on cryptocurrencies using PoS are now already appearing [26, 47, 85], further highlighting the need for better models.

Additionally to the consensus rules, another route to improving the incentivization of cryptocurrencies is through their transaction fees market. As pointed out by Lavi et al. [95] "competition in the fee market is what keeps the rational behavior of Bitcoin's users (partially) aligned with the goal of buying enough security for the entire system" and is thus crucial for its security. This problem is related to that of auction theory [94] and some of the literature of that field could be used here.

6.4 Incentivizing decentralization

Bitcoin has evolved to become different, in many ways, from the intended design and the idea of "one-CPU-one-vote" envisioned by Nakamoto. Because the price of mining has increased exponentially with the popularity of Bitcoin, miners have started forming mining pools, where they join their resources to mine, together, more blocks.

This is obviously a big threat to the security of cryptocurrencies as this could enable 51% attacks, which have already happened to other cryptocurrencies. As of November 2019, the most important 51% attack has targeted Ethereum Classic, which is the 16th largest cryptocurrency by market cap [1].

The centralization of cryptocurrencies' has been empirically analyzed by Gencer et al. [61] who measured how decentralized Bitcoin's and Ethereum's network are. They found that three or four mining pools control more than half of the hash power of the network. This highlights the need for further research studying this occurrence of centralization and how decentralization can be maintained in practice.

Several papers propose a game-theoretic analysis of the mining pools. Arnosti et al. [10] model hardware investments from miners as a game, Leonardos et al. [96] model mining as an *oceanic game*, used to analyze decision making in settings with small numbers of big players and large numbers of individually insignificant players. Lewenberg et al. [97] model the mining game as a *transferable utility coalitional game*, which allows players to form coalitions and to divide their payoffs amongst themselves. As a solution concept, they use the *core*, the set of feasible allocations that cannot be improved upon by a coalition, which describes stability in coalitional games. It captures the condition under which the agents would want to form coalitions rather than not i.e., whether there exist any sub-coalition where agents could have gained more on their own. This concept is often opposed to the *Shapley value* in Game Theory, which defines a fair way to divide the payment among the members of a coalition based on their respective contribution, but without any consideration for stability, unlike the core. Lewenberg et al. additionally define the *defection* function that captures the fact that not every agent subset can collaborate and form a new coalition. They show that mining pools are generally unstable, no matter how the revenue is shared, some miners would be incentivized to switch to a different pool.

Eyal [45] also studies the stability of mining pools and proposes an attack where pools infiltrate other pools to sabotage them by joining the pool and earning rewards, but without actually contributing i.e., not revealing when they find a PoW solution. There exists configurations in which this attack constitutes a NE and an example of a tragedy of the commons.

Mining pools can also attack each other through distributed denial of service (DDoS) attacks to lower the expected success of a competing pool (large ones in particular), rather than increasing their own computational power [83]. Over a two year period, Vasek et al. [139] found that 62.5% of mining pools accounting for more than 5% of the Bitcoin network power had been targeted, while only 17.1% of the smaller pools had been targeted. This has general implications for the mining ecosystem, as a peaceful equilibrium would require an increase to the cost of attacks and to the miner migration rate (miners switching pools), with no pool being significantly more attractive than others [93].

Brünjes et al. [27] introduce and study reward sharing schemes that promote the fair formation of stake pools in a PoS blockchain. They argue that a NE only considers *myopic players*, i.e., players who ignore the responses to their own actions. As a result, they consider the notion of *non-myopic Nash equilibrium* (based on previous work by Fiat et al. [54]), which captures the effects that a certain move will incur anticipating a strategic response from the other players.

Luu et al. [104] use smart contracts to decentralize mining by incuring mining fees lower than centralized mining pools. Miller et al. [111] present several definitions and constructions for "non-outsourceable" puzzles. Both papers use informal arguments to justify their construction as opposed to a formal model.

In a recent paper, Kwon et al. [92] propose a formal model for the decentralization of blockchains, and show that full decentralization is impossible unless there exists a Sybil cost.

6.5 Payment Channels

In order to overcome the scalability issues of Bitcoin, a new concept, referred to as *layer 2* or *payment channels* has been proposed [108]. The idea is that since the network cannot handle enough transactions, participants can take some transactions off-chain i.e., outside the main network, by opening a channel between themselves. This is done by locking a deposit on the blockchain, opening the channel and transact-

ing on the channel, then settling the overall balance of all transactions on-chain so that the blockchain will see only two transactions (locking the funds and settling the balance).

Several designs have been proposed to achieve this [41, 108, 110]. The high level idea is that participants will create evidence of each of their transactions (e.g., using signatures) so that whenever someone tries to cheat the other party can prove it and receive the cheating party's deposit as compensation.

In this setting, the security relies on the fact that cheating is easily detectable due to cryptographic evidence and on the financial punishment associated with it. So again, incentives are tightly linked to security. A few papers [42, 110] present formal models to analyze the security of these payment channels. They are based on the UC-model mentioned in Section 3 but do not consider utilities although it is an important part of the security of the system.

In order to facilitate payment channels, a routing solution has been proposed [124]. There are usually difficulties in this case, due to the need for collaterals to be locked by everyone on the routing path. This work is related to the one on selfish routing discussed in Section 5.

A problem with payment channels is the requirement for participants to be online to detect cheating i.e., the cheater broadcasting an old balance to the blockchain. McCorry et al. [106] propose delegating this task to a third party, a *watch tower*, but it is unclear how incentives should be designed in this context.

6.6 Fairness

Fairness in cryptocurrencies is implicitely captured by the notion of chain quality introduced by Garay et al. [60], which states that an adversary should not contribute more blocks to the blockchain than what they are supposed to i.e., proportionally to their computational power in the PoW setting.

Chen et al. [35] recently showed that a proportional reward system is the unique allocation rule that satisfies properties of symmetry, budget balance (weak or strong), sybil-proofness, and collusion-proofness, which are desirable.

In the PoS setting, Fanti et al. [47] define the notion of *equitability* that corresponds to how much a node's initial investment (i.e., stake) can grow or shrink, and address the problem of the "rich get richer" in PoS cryptocurrencies (which arguably also exists in PoW cryptocurrencies). They propose a geometric reward function that they prove is more equitable i.e., the distribution of stake stays more stable. In general, the problem of compounding of wealth is reinforced by the fact that early adopters of a cryptocurrency have a significant advantage, benefiting from the ease of mining (or staking) and greatly cheaper coin prices in the early days. Dealing with this is more of a macroeconomic problem that to the best of our knowledge has not yet received any attention.

7 Discussion

Within each section of this paper, we have reviewed models that draw on game theoretic tools and could thus be used to address problems in blockchain based cryptocurrencies. Table 1 gives a summary of some these models, and we will now discuss general points and lessons that can be drawn from the work so far.

The most immediate problem that most models try to address is the consideration of incentives at the level of security properties by encoding utility functions and some notion of equilibrium into a modified definition of the traditional security property. For example, RPD does this for UC security and the BAR model does this for distributed systems. Already, however, some differences emerge.

First, there are differences in the types of players. Different models assume players can be some subset of honest, rational, or Byzantine. How well the assumed types of players reflect the reality of the system will have a great impact on the usefulness of the model, regardless of its technical merits.

In particular, while honest and Byzantine players can be defined with respect to a protocol and whether or not they follow it, the meaning of a rational player is harder to pin down. This is because a player is in practice not rational only with respect to a protocol and the incentives in the system that the protocol is part of, but also with respect to a variety of possible external incentives. We have referred to this with respect to rational cryptography and bribery attacks, but it is a more general point that Ford and Böhme have explicitely highlighted [55].

Security models are built such that within the model, properties can remain true up to some amount of adversarial capabilities e.g., a third of Byzantine nodes and computational capabilities, but it is not clear how valid the models are when the assumed player types differ from reality. Moreover, comparing the models presented in this paper would require a concrete idea of what the correct assumptions that can be made about participants in cryptocurrencies are. Each model is constructed to work better than others within the constraints of their assumptions. As remarked by Box, "all models are wrong, but some are useful" [25]. It seems that the current limitation of the exisiting litterature is in understanding how useful these models are in practice rather than in introducing new models.

In Economics, Becker pointed out long ago that it was not always clear what rationality implied, because some observed behaviour was compatible with both rational and irrational behaviour [19], and concluded that perhaps irrationality deserved to be studied with more attention. Behavioural Economics and, more broadly, experimental data driven work has taken an important place in Economics to understand why certain models did not work in practice. Perhaps the same should be done here.

For example, mining rewards can be designed as part of

Table 1: Summary of the models surveyed in this work, along with the problem they seek to address and (if applicable) shortcomings.

Concept	Player types	Description	Shortcomings	Example, suggestions, and discussion
RPD	Rational, honest	Meta-game between the game designer and the adversary, which the adversary wins if it can exploit a vulnerabilities without decreasing its utility.	Does not consider irrational adversaries.	Could be used in Layer 2 since in that case a misbehaving participants will be penalized and the other participant will earn all of its money so Byzantine adversaries cannot harm honest players more than rational adversaries.
BMG	Rational	Considers games and the complexity of actions by modelling players as Turing machines	Results are primarily equivalence theorems with dense proofs rather than new results or simpler proof methods. No consideration of Byzantine adversaries.	This could be usefully applied to cryptocurrencies given that many situations are modeled as games involving computation e.g., mining.
Price of Unaccountability	Rational	Worst ratio between utilities in the observable-actions case and the hidden actions case.	Defined in the principal-agent model.	As differentiating between malicious behavior and genuine latency is hard, especially in PoS systems, the Price of Unaccountability may be a useful to evaluating this.
BAR	Byzantine, rational, honest	Introduces the idea of BART and IC-BFT protocols, which make it possible to bypass impossibility results in consensus protocols.		
(k,t)-robustness	Byzantine, rational, honest	Extension of NE that consider coalitions of both rational and Byzantine players.	The two concepts of immunity and robustness are treated separetely.	This model is better suited than NE to study the game theoretic aspects of consensus protocols that involve many players and coalitions.
Price of Malice and Byzantine Price of Anarchy	Byzantine, rational	Quantifies how much a system degrades with the presence of irrational players and relates to the concept of immunity introduced in (k,t)-robutsness.	Does not capture how different information sets impact the system.	Could inspire new measures that quantify the trade-off between blockchain and traditional consensus protocols. Blockchain-based systems are intended to be more scalable as they are meant to handle open participation, compared to classical consensus that requires the many messages to be exchanged, but in the case of Bitcoin this comes at the price of PoW so there is an incurred economic cost.
Fear factor	Byzantine, rational	Rational players are incentivized to follow the protocol by fear of Byzantine players. Similar to Price of Malice (if the system improves instead of deteriorates).	Same as above.	This could be used to argue against the verifier dilemma. Some users may be motivated to verify the content of blocks by fear that others will not.

a protocol to ensure some level of decentralization but the utility function of miners will also depend on their individual economic environment that the protocol cannot fix. In this case, relating the economics of miners to their relationship with the system could do more to ensure some level of decentralization than tweaking the rewards given out by the protocol.

Second, there are the way coalitions of players are treated. Pools are being observed and studied carefully in the community, but some important problems remain under-studied. It could be argued that in some cases pools somehow self-regulate as in, for example, the case of the Ghash.io Bitcoin pool that once got more than 51% of the hashing power but then decided to withdraw part of it [80]. This could be because of the fear of loss of confidence users of the system, which if it is justified, would mean that the incentives were to

some extent aligned to prevent a 51% attack.

Coalitions are also easily abstracted as an entity controlling a fixed share of power in the system but this ignores costs inherent to colluding such as, for example, communications costs as a coalition much reach some form of consensus on what actions it takes. In the same way that results are obtained for a whole system based on assumed proportions of player types in the system, perhaps useful results can be obtained based on the proportions of player types in a coalition with respect to that coallition.

8 Conclusion

Security researchers and cryptographers have been interested in incorporating game theoretic notions to their models for many years. In this work, we have highlighted existing concepts and explained how and where they could be used for specific applications.

The approach taken in most of the papers that we described here is to extend a field by for example incorporating utility functions (Rational Cryptography) or computation (Bayesian Machine Games). No completely new theory has appeared and it would be interesting to see a new theory built from the ground up to address considerations of incentives at all stages of the design process, rather than adapting existing models. We hope that this paper will give some inspiration towards new formal models.

References

[1] Coin market cap. https://www.coinmarketcap.com/coins.

[2] Crypto51. https://www.crypto51.app/.

[3] Ethereum. https://www.ethereum.org/.

[4] Snowflake to Avalanche: A Novel Metastable Consensus Protocol Family for Crypto. https://ipfs.io/ipfs/QmUy4jh5mGNZvLkjies1RWM4YuvJh5o2FYopNPVYwrRVGV.

[5] Ittai Abraham, Danny Dolev, Rica Gonen, and Joe Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Principles of Distributed Computing*, PODC '06, pages 53–62, New York, NY, USA, 2006. ACM.

[6] Ittai Abraham, Dahlia Malkhi, Kartik Nayak, Ling Ren, and Alexander Spiegelman. Solidus: An incentive-compatible cryptocurrency based on permissionless Byzantine consensus. *CoRR*, abs/1612.02916, 2016.

[7] Amitanand S. Aiyer, Lorenzo Alvisi, Allen Clement, Mike Dahlin, Jean-Philippe Martin, and Carl Porth. BAR Fault Tolerance for cooperative services. *SIGOPS Oper. Syst. Rev.*, 39(5):45–58, oct 2005.

[8] Ross Anderson. Why information security is hard-an economic perspective. In *Computer security applications conference, 2001. acsac 2001. proceedings 17th annual*, pages 358–365. IEEE, 2001.

[9] Ross Anderson and Tyler Moore. The economics of information security. *Science*, 314(5799):610–613, 2006.

[10] Nick Arnosti and S Matthew Weinberg. Bitcoin: A natural oligopoly. *arXiv preprint arXiv:1811.08572*, 2018.

[11] Gilad Asharov, Ran Canetti, and Carmit Hazay. Towards a game theoretic view of secure computation. In Kenneth G. Paterson, editor, *Advances in Cryptology – EUROCRYPT 2011*, pages 426–445, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[12] Yonatan Aumann and Yehuda Lindell. Security against covert adversaries: Efficient protocols for realistic adversaries. In *Theory of Cryptography Conference*, pages 137–156. Springer, 2007.

[13] Sarah Azouvi, Alexander Hicks, and Steven J. Murdoch. Incentives in security protocols. In Vashek Matyáš, Petr Švenda, Frank Stajano, Bruce Christianson, and Jonathan Anderson, editors, *Security Protocols XXVI*, pages 132–141, Cham, 2018. Springer International Publishing.

[14] Sarah Azouvi, Patrick McCorry, and Sarah Meiklejohn. Betting on blockchain consensus with fantomette. *CoRR*, abs/1805.06786, 2018.

[15] Moshe Babaioff, John Chuang, and Michal Feldman. Incentives in peer-to-peer systems. *Algorithmic Game Theory*, pages 593–611, 2007.

[16] Moshe Babaioff, Shahar Dobzinski, Sigal Oren, and Aviv Zohar. On bitcoin and red balloons. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 56–73. ACM, 2012.

[17] Christian Badertscher, Juan Garay, Ueli Maurer, Daniel Tschudi, and Vassilis Zikas. But why does it work? a rational protocol design treatment of bitcoin. Cryptology ePrint Archive, Report 2018/138, 2018.

[18] Shehar Bano, Alberto Sonnino, Mustafa Al-Bassam, Sarah Azouvi, Patrick McCorry, Sarah Meiklejohn, and George Danezis. Consensus in the age of blockchains. *arXiv preprint arXiv:1711.03936*, 2017.

[19] Gary S. Becker. Irrational behavior and economic theory. *Journal of Political Economy*, 70(1):1–13, 1962.

[20] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, STOC '88, pages 1–10, New York, NY, USA, 1988. ACM.

[21] Iddo Bentov, Rafael Pass, and Elaine Shi. Snow White: Provably secure proofs of stake. *IACR Cryptology ePrint Archive*, 2016:919, 2016.

[22] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten. SoK: Research perspectives and challenges for bitcoin and cryptocurrencies. In *2015 IEEE Symposium on Security and Privacy*, pages 104–121, May 2015.

[23] Joseph Bonneau. Why buy when you can rent? - Bribery attacks on bitcoin-style consensus. pages 19–26, 2016.

[24] Joseph Bonneau. Hostile blockchain takeovers (short paper). In *Bitcoin'18: Proceedings of the 5th Workshop on Bitcoin and Blockchain Research*, 2018.

[25] George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[26] Jonah Brown-Cohen, Arvind Narayanan, Christos-Alexandros Psomas, and S Matthew Weinberg. Formal barriers to longest-chain proof-of-stake protocols. *arXiv preprint arXiv:1809.06528*, 2018.

[27] Lars Brünjes, Aggelos Kiayias, Elias Koutsoupias, and Aikaterini-Panagiota Stouka. Reward sharing schemes for stake pools. *arXiv preprint arXiv:1807.11218*, 2018.

[28] Eric Budish. The economic limits of bitcoin and the blockchain. Technical report, National Bureau of Economic Research, 2018.

[29] Vitalik Buterin. Uncle rate and transaction fee analysis.

[30] Vitalik Buterin. Incentives in casper the friendly finality gadget, 2017. https://github.com/ethereum/research/blob/master/papers/casper-economics/casper_economics_basic.pdf.

[31] Christian Cachin and Marko Vukolić. Blockchains consensus protocols in the wild. *arXiv preprint arXiv:1707.01873*, 2017.

[32] Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. pages 136–145, 2001.

[33] Miles Carlsten, Harry Kalodner, S. Matthew Weinberg, and Arvind Narayanan. On the instability of bitcoin without the block reward. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 154–167, New York, NY, USA, 2016. ACM.

[34] Melissa Chase and Sarah Meiklejohn. Transparency overlays and applications. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 168–179, New York, NY, USA, 2016. ACM.

[35] Xi Chen, Christos Papadimitriou, and Tim Roughgarden. An axiomatic approach to block rewards. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies*, pages 124–131. ACM, 2019.

[36] A. Clement, J. Napper, H. Li, JP Martin, L. Alvisi, and M. Dahlin. Theory of BAR games. In *Brief Announcements: Proceedings of the Symposium on Principles of Distributed Computing (PODC 2007)*, Aug 2007.

[37] Bram Cohen. Incentives build robustness in bittorrent. In *Workshop on Economics of Peer-to-Peer systems*, volume 6, pages 68–72, 2003.

[38] George Danezis and Dave Hrycyszyn. Blockmania: from block dags to consensus. *CoRR*, abs/1809.01620, 2018.

[39] Yevgeniy Dodis, Shai Halevi, and Tal Rabin. A cryptographic solution to a game theoretic problem. In Mihir Bellare, editor, *Advances in Cryptology — CRYPTO 2000*, pages 112–130, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

[40] Yevgeniy Dodis, Tal Rabin, et al. Cryptography and game theory. *Algorithmic Game Theory*, pages 181–207, 2007.

[41] Stefan Dziembowski, Lisa Eckey, Sebastian Faust, and Daniel Malinowski. Perun: Virtual payment hubs over cryptocurrencies. In *Perun: Virtual Payment Hubs over Cryptocurrencies*, page 0. IEEE, 2017.

[42] Stefan Dziembowski, Sebastian Faust, and Kristina Hostáková. General state channel networks. In *Proceedings of the 25th ACM SIGSAC Conference on Computer and Communications Security*, CCS '18. ACM, 2018.

[43] Yuval Emek, Ron Karidi, Moshe Tennenholtz, and Aviv Zohar. Mechanisms for multi-level marketing. In

Proceedings of the 12th ACM conference on Electronic commerce, pages 209–218. ACM, 2011.

[44] Oğuzhan Ersoy, Zhijie Ren, Zekeriya Erkin, and Reginald L Lagendijk. Transaction propagation on permissionless blockchains: incentive and routing mechanisms. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pages 20–30. IEEE, 2018.

[45] Ittay Eyal. The miner's dilemma. In *Proceedings of the 2015 IEEE Symposium on Security and Privacy*, SP '15, pages 89–103, Washington, DC, USA, 2015. IEEE Computer Society.

[46] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. *Commun. ACM*, 61(7):95–102, jun 2018.

[47] Giulia Fanti, Leonid Kogan, Sewoong Oh, Kathleen Ruan, Pramod Viswanath, and Gerui Wang. Compounding of wealth in proof-of-stake cryptocurrencies. *arXiv preprint arXiv:1809.07468*, 2018.

[48] Joan Feigenbaum, Christos H Papadimitriou, and Scott Shenker. Sharing the cost of multicast transmissions. *Journal of Computer and System Sciences*, 63(1):21–41, 2001.

[49] Joan Feigenbaum, Michael Schapira, and Scott Shenker. Distributed algorithmic mechanism design. *Algorithmic Game Theory*, pages 363–384, 2007.

[50] Joan Feigenbaum and Scott Shenker. Distributed algorithmic mechanism design: Recent results and future directions. 2002.

[51] Michal Feldman and John Chuang. Overcoming free-riding behavior in peer-to-peer systems. *ACM sigecom exchanges*, 5(4):41–50, 2005.

[52] Michal Feldman, Christos Papadimitriou, John Chuang, and Ion Stoica. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on selected areas in communications*, 24(5):1010–1019, 2006.

[53] Amos Fiat, Anna Karlin, Elias Koutsoupias, and Christos Papadimitriou. Energy equilibria in proof-of-work mining. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pages 489–502, New York, NY, USA, 2019. ACM.

[54] Amos Fiat, Elias Koutsoupias, Katrina Ligett, Yishay Mansour, and Svetlana Olonetsky. Beyond myopic best response (in cournot competition). *Games and Economic Behavior*, 2013.

[55] Bryan Ford and Rainer Böhme. Rationality is self-defeating in permissionless systems. *arXiv preprint arXiv:1910.08820*, 2019.

[56] Georg Fuchsbauer, Jonathan Katz, and David Naccache. Efficient rational secret sharing in standard communication networks. In *Theory of Cryptography Conference*, pages 419–436. Springer, 2010.

[57] Juan Garay, Jonathan Katz, Ueli Maurer, Bjoern Tackmann, and Vassilis Zikas. Rational protocol design: Cryptography against incentive-driven adversaries. Cryptology ePrint Archive, Report 2013/496, 2013. http://eprint.iacr.org/2013/496.

[58] Juan Garay, Jonathan Katz, Bjoern Tackmann, and Vassilis Zikas. How fair is your protocol? a utility-based approach to protocol optimality. Cryptology ePrint Archive, Report 2015/187, 2015. https://eprint.iacr.org/2015/187.

[59] Juan Garay and Aggelos Kiayias. Sok: A consensus taxonomy in the blockchain era.

[60] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. The bitcoin backbone protocol: Analysis and applications. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015*, pages 281–310, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

[61] Adem Efe Gencer, Soumya Basu, Robbert van Renesse Ittay Eyal and, and Emin Gün Sirer. Decentralization in bitcoin and ethereum networks. In *Financial Cryptography and Data Security*. Springer Berlin Heidelberg, 2018.

[62] Arthur Gervais, Ghassan O. Karame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf, and Srdjan Capkun. On the security and performance of proof of work blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 3–16, New York, NY, USA, 2016. ACM.

[63] Uri Gneezy and Aldo Rustichini. A fine is a price. *The Journal of Legal Studies*, 29(1):1–17, 2000.

[64] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 218–229, New York, NY, USA, 1987. ACM.

[65] S Dov Gordon and Jonathan Katz. Rational secret sharing, revisited. In *International Conference on Security and Cryptography for Networks*, pages 229–241. Springer, 2006.

[66] Adam Groce, Jonathan Katz, Aishwarya Thiruvengadam, and Vassilis Zikas. Byzantine agreement with a rational adversary. In Artur Czumaj, Kurt Mehlhorn, Andrew Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming*, pages 561–572, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[67] Jens Grossklags, Nicolas Christin, and John Chuang. Secure or insure?: A game-theoretic analysis of information security games. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 209–218, New York, NY, USA, 2008. ACM.

[68] Jens Grossklags, Benjamin Johnson, and Nicolas Christin. The price of uncertainty in security games. In *Economics of Information Security and Privacy*, pages 9–32. Springer, 2010.

[69] J. Y. Halpern and R. Pass. Game Theory with Costly Computation. *ArXiv e-prints*, August 2008.

[70] J. Y. Halpern, R. Pass, and D. Reichman. On the Non-Existence of Nash Equilibrium in Games with Resource-Bounded Players. *ArXiv e-prints*, July 2015.

[71] Joseph Halpern and Vanessa Teague. Rational secret sharing and multiparty computation: Extended abstract. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 623–632, New York, NY, USA, 2004. ACM.

[72] Joseph Halpern and Vanessa Teague. Rational secret sharing and multiparty computation: Extended abstract. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 623–632, New York, NY, USA, 2004. ACM.

[73] Joseph Y. Halpern. Computer science and game theory: A brief survey. *CoRR*, abs/cs/0703148, 2007.

[74] Joseph Y. Halpern. Beyond Nash equilibrium: Solution concepts for the 21st century. *CoRR*, abs/0806.2139, 2008.

[75] Joseph Y Halpern. I don't want to think about it now: Decision theory with costly computation. In *Twelfth international conference on the principles of knowledge representation and reasoning*, 2010.

[76] Joseph Y. Halpern and Rafael Pass. Algorithmic rationality: Adding cost of computation to game theory. *SIGecom Exch.*, 10(2):9–15, June 2011.

[77] Joseph Y Halpern and Rafael Pass. Sequential equilibrium in computational games. In *IJCAI*, pages 171–176, 2013.

[78] Joseph Y Halpern and Rafael Pass. Algorithmic rationality: Game theory with costly computation. *Journal of Economic Theory*, 156:246–268, 2015.

[79] Garrett Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.

[80] Alex Hern. Bitcoin currency could have been destroyed by 51% attack. *The Guardian*.

[81] Charlie Hou, Mingxun Zhou, Yan Ji, Phil Daian, Florian Tramer, Giulia Fanti, and Ari Juels. Squirrl: Automating attack discovery on blockchain incentive mechanisms with deep reinforcement learning. *arXiv preprint arXiv:1912.01798*, 2019.

[82] Matthew O. Jackson. Mechanism theory. Survey, California Institute of Technology, 2003.

[83] Benjamin Johnson, Aron Laszka, Jens Grossklags, Marie Vasek, and Tyler Moore. Game-theoretic analysis of ddos attacks against bitcoin mining pools. In *International Conference on Financial Cryptography and Data Security*, pages 72–86. Springer, 2014.

[84] Seung Jun and Mustaque Ahamad. Incentives in bittorrent induce free riding. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 116–121. ACM, 2005.

[85] Sanket Kanjalkar, Joseph Kuo, Yunqi Li, and Andrew Miller. Short paper: I can't believe it's not stake! resource exhaustion attacks on pos. In *International Conference on Financial Cryptography and Data Security*. Springer, 2019.

[86] Jonathan Katz. Bridging game theory and cryptography: Recent results and future directions. In *Proceedings of the 5th Conference on Theory of Cryptography*, TCC'08, pages 251–272, Berlin, Heidelberg, 2008. Springer-Verlag.

[87] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. Ouroboros: A provably secure Proof of stake blockchain protocol. In *Annual International Cryptology Conference*, pages 357–388. Springer, 2017.

[88] Lucianna Kiffer, Rajmohan Rajaraman, and abhi shelat. A better method to analyze blockchain consistency. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 729–744, New York, NY, USA, 2018. ACM.

[89] Gillat Kol and Moni Naor. Cryptography and game theory: Designing protocols for exchanging information. In *Theory of Cryptography Conference*, pages 320–339. Springer, 2008.

[90] Gillat Kol and Moni Naor. Games for exchanging information. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 423–432. ACM, 2008.

[91] Joshua A. Kroll, Ian C. Davey, and Edward W. Felten. The economics of bitcoin mining , or bitcoin in the presence of adversaries. 2013.

[92] Yujin Kwon, Jian Liu, Minjeong Kim, Dawn Song, and Yongdae Kim. Impossibility of full decentralization in permissionless blockchains. *arXiv preprint arXiv:1905.05158*, 2019.

[93] Aron Laszka, Benjamin Johnson, and Jens Grossklags. When bitcoin mining pools run dry. In *International Conference on Financial Cryptography and Data Security*, pages 63–77. Springer, 2015.

[94] Ron Lavi. Computationally efficient approximation mechanisms. *Algorithmic Game Theory*, pages 301–329, 2007.

[95] Ron Lavi, Or Sattath, and Aviv Zohar. Redesigning bitcoin's fee market. *arXiv preprint arXiv:1709.08881*, 2017.

[96] Nikos Leonardos, Stefanos Leonardos, and Georgios Piliouras. Oceanic games: Centralization risks and incentives in blockchain mining. *arXiv preprint arXiv:1904.02368*, 2019.

[97] Yoad Lewenberg, Yoram Bachrach, Yonatan Sompolinsky, Aviv Zohar, and Jeffrey S. Rosenschein. Bitcoin mining pools: A cooperative game theoretic analysis. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '15, pages 919–927, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.

[98] Cuihong Li, Bin Yu, and Katia Sycara. An incentive mechanism for message relaying in unstructured peer-to-peer systems. *Electronic Commerce Research and Applications*, 8(6):315–326, 2009.

[99] Kevin Liao and Jonathan Katz. Incentivizing blockchain forks via whale transactions. In *International Conference on Financial Cryptography and Data Security*, pages 264–279. Springer, 2017.

[100] Yehuda Lindell. How to simulate it - A tutorial on the simulation proof technique. Cryptology ePrint Archive, Report 2016/046, 2016. http://eprint.iacr.org/2016/046.

[101] Nathan Linial. Games computers play: Game-theoretic aspects of computing. 01 2002.

[102] Ziyao Liu, Nguyen Cong Luong, Wenbo Wang, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. A survey on blockchain: A game theoretical perspective. *IEEE Access*, 7:47615–47643, 2019.

[103] Loi Luu, Jason Teutsch, Raghav Kulkarni, and Prateek Saxena. Demystifying incentives in the consensus computer. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 706–719, New York, NY, USA, 2015. ACM.

[104] Loi Luu, Yaron Velner, Jason Teutsch, and Prateek Saxena. Smartpool: Practical decentralized pooled mining. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 1409–1426, Vancouver, BC, 2017. USENIX Association.

[105] Anna Lysyanskaya and Nikos Triandopoulos. Rationality and adversarial behavior in multi-party computation. In Cynthia Dwork, editor, *Advances in Cryptology - CRYPTO 2006*, pages 180–197, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[106] Patrick McCorry, Surya Bakshi, Iddo Bentov, Andrew Miller, and Sarah Meiklejohn. Pisa: Arbitration outsourcing for state channels. Cryptology ePrint Archive, Report 2018/582, 2018. https://eprint.iacr.org/2018/582.

[107] Patrick McCorry, Alexander Hicks, and Sarah Meiklejohn. Smart contracts for bribing miners. *IACR Cryptology ePrint Archive*, 2018:581, 2018.

[108] Patrick Mccorry, Malte Möser, Siamak F. Shahandasti, and Feng Hao. Towards bitcoin payment networks. In *Proceedings, Part I, of the 21st Australasian Conference on Information Security and Privacy - Volume 9722*, pages 57–76, New York, NY, USA, 2016. Springer-Verlag New York, Inc.

[109] Silvio Micali and Phillip Rogaway. Secure computation. In *Annual International Cryptology Conference*, pages 392–404. Springer, 1991.

[110] Andrew Miller, Iddo Bentov, Ranjit Kumaresan, and Patrick McCorry. Sprites: Payment channels that go faster than lightning. *CoRR*, abs/1702.05812, 2017.

[111] Andrew Miller, Ahmed Kosba, Jonathan Katz, and Elaine Shi. Nonoutsourceable scratch-off puzzles to discourage bitcoin mining coalitions. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 680–691, New York, NY, USA, 2015. ACM.

[112] Thomas Moscibroda, Stefan Schmid, and Rogert Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Principles of Distributed Computing*, PODC '06, pages 35–44, New York, NY, USA, 2006. ACM.

[113] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. bitcoin.org/bitcoin.pdf.

[114] Kartik Nayak, Srijan Kumar, Andrew Miller, and Elaine Shi. Stubborn mining: Generalizing selfish mining and combining with an eclipse attack. In *Security and Privacy (EuroS&P), 2016 IEEE European Symposium on*, pages 305–320. IEEE, 2016.

[115] Noam Nisan and Amir Ronen. Algorithmic mechanism design. *Games and Economic behavior*, 35(1-2):166–196, 2001.

[116] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

[117] Jianyu Niu and Chen Feng. Selfish mining in ethereum. *arXiv preprint arXiv:1901.04620*, 2019.

[118] Guillermo Owen. *Game theory*. Saunders, 1968.

[119] Sunoo Park, Krzysztof Pietrzak, Albert Kwon, Joël Alwen, Georg Fuchsbauer, and Peter Gaži. SpaceMint: A cryptocurrency based on proofs of space. Cryptology ePrint Archive, Report 2015/528, 2015. https://eprint.iacr.org/2015/528.

[120] Rafael Pass and Joe Halpern. Game theory with costly computation: Formulation and application to protocol security. In *Proceedings of the Behavioral and Quantitative Game Theory: Conference on Future Directions*, BQGT '10, pages 89:1–89:1, New York, NY, USA, 2010. ACM.

[121] Rafael Pass, Lior Seeman, and Abhi Shelat. Analysis of the blockchain protocol in asynchronous networks. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 643–673. Springer, 2017.

[122] Rafael Pass and Elaine Shi. Fruitchains: A fair blockchain. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 315–324. ACM, 2017.

[123] Michael Piatek, Tomas Isdal, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani. Do incentives build robustness in bittorrent. In *Proc. of NSDI*, volume 7, 2007.

[124] Joseph Poon and Thaddeus Dryja. The bitcoin lightning network:scalable off-chain instant payments. https://lightning.network/lightning-network-paper.pdf, 2016.

[125] F. Ritz and A. Zugenmaier. The impact of uncle rewards on selfish mining in ethereum. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, pages 50–57, April 2018.

[126] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.

[127] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.

[128] Ayelet Sapirshtein, Yonatan Sompolinsky, and Aviv Zohar. Optimal selfish mining strategies in bitcoin. In Jens Grossklags and Bart Preneel, editors, *Financial Cryptography and Data Security*, pages 515–532, Berlin, Heidelberg, 2017. Springer Berlin Heidelberg.

[129] Yoav Shoham. Computer science and game theory. *Communications of the ACM*, 51(8):74–79, 2008.

[130] Yonatan Sompolinsky, Yoad Lewenberg, and Aviv Zohar. Spectre: A fast and scalable cryptocurrency protocol. Cryptology ePrint Archive, Report 2016/1159, 2016. https://eprint.iacr.org/2016/1159.

[131] Yonatan Sompolinsky and Aviv Zohar. Secure high-rate transaction processing in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 507–527. Springer, 2015.

[132] Yonatan Sompolinsky and Aviv Zohar. Phantom: A scalable blockdag protocol. Cryptology ePrint Archive, Report 2018/104, 2018. https://eprint.iacr.org/2018/104.

[133] Nicholas Stifter, Aljosha Judmayer, Philipp Schindler, Alexei Zamyatin, and Edgar Weippl. Agreement with satoshi–on the formalization of nakamoto consensus.

[134] Milind Tambe. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge University Press, 2011.

[135] Jason Teutsch, Sanjay Jain, and Prateek Saxena. When cryptocurrencies mine their own business. In *International Conference on Financial Cryptography and Data Security*, pages 499–514. Springer, 2016.

[136] Carmela Troncoso, Marios Isaakidis, George Danezis, and Harry Halpin. Systematizing decentralization and privacy: Lessons from 15 years of research and deployments. *Proceedings on Privacy Enhancing Technologies*, 2017(4):404 – 426, 2017.

[137] Itay Tsabary and Ittay Eyal. The gap game. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 713–728, New York, NY, USA, 2018. ACM.

[138] Hal Varian. System reliability and free riding. In *Economics of information security*, pages 1–15. Springer, 2004.

[139] Marie Vasek, Micah Thornton, and Tyler Moore. Empirical analysis of denial-of-service attacks in the bitcoin ecosystem. In *International conference on financial cryptography and data security*, pages 57–71. Springer, 2014.

[140] Yaron Velner, Jason Teutsch, and Loi Luu. Smart contracts make bitcoin mining pools vulnerable. In *International Conference on Financial Cryptography and Data Security*, pages 298–316. Springer, 2017.

[141] Vivek Vishnumurthy, Sangeeth Chandrakumar, and Emin Gun Sirer. Karma: A secure economic framework for peer-to-peer resource sharing. In *Workshop on Economics of Peer-to-peer Systems*, volume 35, 2003.

[142] Wenbo Wang, Dinh Thai Hoang, Zehui Xiong, Dusit Niyato, Ping Wang, Peizhao Hu, and Yonggang Wen. A survey on consensus mechanisms and mining management in blockchain networks. *arXiv preprint arXiv:1805.02707*, 2018.

A Glossary

In this appendix, we provide formal definitions for some of the concepts presented in the main body of the paper that are not formally defined.

A.1 Game Theory

To start off, we introduce the standard definitions for Bayesian games and mechanisms.

Bayesian game setting A Bayesian game setting is a tuple $(N, O, \Theta, \Pr, u)$, where:

- N is a finite set of n players;
- O is a set of outcomes;
- $\Theta = \Theta_1, \cdots, \Theta_n$ is a set of possible joint type vectors
- Pr is a (common prior) probability distribution on Θ; and
- $u = (u_1, \cdots, u_n)$, where $u_i : O \times \mathbb{R} \to \mathbb{R}$ is the utility function for each player i

Mechanism for a Bayesian game setting A mechanism for a Bayesian game setting (N, O, Θ, p, u) is a pair (A,M), where

- $A = A_1 \times \cdots \times A_n$, where A_i is the set of actions available to agent $i \in N$
- $M : A \to \mathcal{D}(O)$ maps each action profile to a distribution over outcomes

A.2 Game Theory and Cryptography

We now move on to concepts presented in Section 4.

$\varepsilon-$subgame perfect equilibrium [57] Let $\mathcal{G}_{\mathcal{M}}$ be an attack game. A staretgy profile (A, Π) is an $\varepsilon-$subgame perfect equilibrium in $\mathcal{G}_{\mathcal{M}}$ if: (1) for any $\Pi' \in \text{ITM}^n$, $u_D(\Pi', A(\Pi')) \leq u_D(\Pi, A(\Pi)) + \varepsilon$, and (2) for any $A' \in \text{ITM}$, $u_A(\Pi, A'(\Pi)) \leq u_A(\Pi, A(\Pi)) + \varepsilon$.

Attack-payoff security [57] Let $\mathcal{M} = (\mathcal{F}, \langle \mathcal{F} \rangle, v)$ be an attack model and let Π be a protocol that realizes functionality $\langle \mathcal{F} \rangle$. Π is attack-payoff secure in $\mathcal{M}$ if $\vec{U}^{\Pi, \langle \mathcal{F} \rangle} \overset{\text{negl}}{\leq} \vec{U}^{\Phi^{\mathcal{F}}, \langle \mathcal{F} \rangle}$ where $\Phi^{\mathcal{F}}$ is the "dummy" $\mathcal{F}$ hybrid protocol (i.e.,the protocol that forwards all inputs and outputs from the functionality $\mathcal{F}$, see Section 3) and $\bar{U}^{\Pi, \langle \mathcal{F} \rangle}$ is the maximized ideal expected payoff of an adversary.

Incentive compatibility [17] Let Π be a protocol and $\mathbb{P}$ be a set of PT protocols that have access to the same hybrids as Π. We say that Π is $\mathbb{P}-$ incentive compatible in the attack model $\mathcal{M}$ if and only if for some $\mathbf{Adv}$ $(\Pi, \mathbf{Adv})$ is a $(\mathbb{P}, \text{ITM})-$ subgame perfect equilibrium in the attack game defined by $\mathcal{M}$.

Bayesian Machine Game [69] A Bayesian machine game G is described by a tuple $(N, \mathcal{M}, \Theta, \Pr, \mathcal{C}_1, \ldots, \mathcal{C}_m, u_1, \ldots, u_2)$ where:

- N is the set of players, $\mathcal{M}$ is the set of possible machines
- $\Theta \subseteq (\{0,1\}^*)^{m+1}$ is the set of type profiles where the $(m+1)$st element in the profile corresponds to nature's type
- Pr is a distribution on Θ
- $\mathcal{C}_i$ is a complexity function
- $u_i : T \times (\{0,1\}^*)^m \times \mathbb{N} \to \mathbb{R}$ is player i's utility function.

Given a Bayesian machine game G, a machine profile $\vec{M}$, and $\varepsilon \geq 0$, M_i is an ε-best response to $\vec{M}_{-i}$ (the tuple consisting of all machines in $\vec{M}$ other than M_i) if, for every $M_i' \in \mathcal{M}$,

$$U_i^G[(M_i, \vec{M}_{-i})] \geq U_i^G[(M_i', \vec{M}_{-i})] - \varepsilon. \quad (1)$$

$\vec{M}$ is an ε-Nash equilibrium of G if, for all players i, M_i is an ε-best response to $\vec{M}_{-i}$. A Nash equilibrium is a 0-Nash equilibrium.

Universal implementation [69] Suppose that $\mathcal{G}$ is a set of n-player canonical games, $\mathcal{Z}$ is a subsets of N, $\mathcal{F}$ and $\mathcal{F}'$ are mediators, $M_1, \cdots, M_n$ are interactive machines, $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \to \mathbb{R}$. $(M, \mathcal{F}'$ is a $(\mathcal{G}, \mathcal{Z}, p)$-universal implementation of $\mathcal{F}$ with error ε if, for all n, all games $G \in \mathcal{G}$ with input length n and all $\mathcal{Z}' \subseteq \mathcal{Z}$ if $\vec{\Lambda}^{\mathcal{F}}$ is a $p(n,\cdot)$-robust $\mathcal{Z}'$-safe ε-NE in the mediated machine game $(G, \mathcal{F})$ then

1. (Preserving equilibrium) $\vec{M}$ is a $\mathcal{Z}'$-safe ε-NE in the mediated machine game $(G, \mathcal{F}')$
2. (Preserving Action Distributions) For each type profile $\vec{t}$, the action profile induced by $\vec{\Lambda}^{\mathcal{F}}$ in $(G, \mathcal{F})$ is identically distributed to then action profile induced by M in $(G, \mathcal{F}')$.

Sequential equilibrium in computational games [77] A pair $(\vec{M}, \mu)$ consisting of a machine profile $\vec{M}$ and a belief system μ is called a belief assessment. A belief assessment $(\vec{M}, \mu)$ is an interim (resp. ex ante) sequential equilibrium in a machine game G if μ is compatible with $\vec{M}$ and for all players i, states q of M_i, and machines M_i' compatible with M_i and q such that $(M_i, q, M_i') \in \mathcal{M}$ (the set of possible machines) (resp. (M_i, q, M_i') is a local variant of M_i), we have

$$U_i(\vec{M}|q,\mu) \geq U_i(((M_i, q, M_i'), \vec{M}_{-i})|q,\mu) \qquad (2)$$

A.3 Game Theory and Distributed Design

Finally, we give definitions for concepts presented in Section 5.

Incentive-Compatible Byzantine Fault Tolerant (IC-BFT) protocols [7] A protocol is IC-BFT if it guarantees the specified set of safety and liveness properties and if it is in the best interest of all rational nodes to follow the protocol exactly.

Byzantine Altruistic Rational Tolerant (BART) protocols [7] A protocol is BART if it guarantees the specified set of safety and liveness properties in the presence of all rational deviations from the protocol.

Perfect security [66] A protocol for broadcast or consensus is perfectly secure against rational adversaries controlling t players with utility U if for every t-adversary there is a strategy S such that for any choice of input for honest players 1. (S is tolerable): S induces a distribution of final outputs D in which no security condition is violated with nonzero probability, and 2. (S is Nash): For any strategy $S' \neq S$ with induced output distribution D' : $U(D) \geq U(D')$.

Statistical Security [66] A protocol for broadcast or consensus is statistically secure against rational adversaries controlling t players with utility U if for every t-adversary there is a strategy S such that for any choice of input for honest players S induces a distribution of final outputs D_k when the security parameter is k and the following properties hold: 1. (S is tolerable): no security condition is violated with nonzero probability in Dk for any k, and 2. (S is statistical Nash): for any strategy $S' \neq S$ with induced output distributions D_k' there is a negligible function $negl(.)$ such that $U(D_k) + negl(k) > U(D_k')$.

(k,t)-robustness [5] A strategy profile σ is a (k,t)-robust equilibrium if for all $C, T \subseteq N$, $C \cap T = \emptyset$, $|C| \leq k$, $|T| \leq t$ $\forall \tau_t \in \mathcal{S}_T$ $\forall \phi_C \in C$ we have: $u_i(\sigma_{-T}, \tau_T) \geq u_i(\sigma_{-C\cap T}, \phi_C, \tau_T)$

(k,t)-punishment [5] A joint strategy ρ is a (k, t)-punishment strategy with respect to σ if for all $C, T, P \subseteq N$ such that C, T, P are disjoint, $|C| \leq k$, $|T| \leq t$, and $|P| > t$, for all $\tau_T \in S_T$, for all $\phi_C \in S_C$, for all $i \in C$ we have $u_i(\sigma_T, \tau_T) > u_i(\sigma_{N-(C\cup T\cup P)}, \phi_C, \tau_T, \rho_P)$.