It's on Me! The Benefit of Altruism in BAR Environments

Edmund L. Wong, Joshua B. Leners, and Lorenzo Alvisi

Department of Computer Science, The University of Texas at Austin
1616 Guadalupe, Suite 2.408
Austin, TX 78701 USA
{elwong,leners,lorenzo}@cs.utexas.edu

Abstract. Cooperation, a necessity for any peer-to-peer (P2P) cooperative service, is often achieved by rewarding good behavior now with the promise of future benefits. However, in most cases, interactions with a particular peer or the service itself eventually end, resulting in some last exchange in which departing participants have no incentive to contribute. Without cooperation in the last round, cooperation in any prior round may be unachievable. In this paper, we propose leveraging altruistic participants that simply follow the protocol as given. We show that altruism is a simple, necessary, and sufficient way to incentivize cooperation in a realistic model of a cooperative service's last exchange, in which participants may be Byzantine, altruistic, or rational and network loss is explicitly considered. By focusing on network-level incentives in the last exchange, we believe our approach can be used as the cornerstone for incentivizing cooperation in any cooperative service.

Keywords: P2P, cooperative services, game theory, BAR

1 Introduction

Establishing and maintaining cooperation between peers in decentralized services spanning multiple administrative domain (MAD) is hard [17, 21]. Because participants may be selfish and withhold resources unless contributing is in their best interest, these services must provide sufficient incentives for participants to contribute. These incentive structures must of course be resilient against buggy or malicious peers; however, they must also be robust against a more subtle threat: an overabundance of good will from the unselfish peers who simply follow the protocol run by the service. It is, after all, the unselfishness of correct peers—as codified in the protocol they obediently follow—that allows selfish peers to continue receiving service without contributing their fair share. Yet, the efforts of well-meaning peers alone may be insufficient to sustain the service. Further, asking these peers to increase their contribution to make up for free-riders may backfire: even well-meaning peers, if blatantly taken advantage of, may give in to the temptation of joining the ranks of the selfish, leading in turn to more defections and to the service's collapse.

Although real MAD systems include a sizable fraction of correct and unselfish peers [2], their impact on the incentive structure of MAD services is not well understood. The BAR model [3] does explicitly account for these peers—they are the *altruistic* peers, who, together with the selfish *rational* peers and the potentially disruptive *Byzantine* peers, give the model its acronym—but existing BAR-tolerant systems have side-stepped the challenge of altruism by designing protocols that neither depend on nor leverage the presence of altruistic peers.[1] This paper asks the following question: *Can we leverage the good will of altruistic nodes and still motivate rational participants to cooperate?* We find that not only is altruism not antithetical to rational cooperation, but that, in a fundamental way, rational cooperation can only be achieved in the presence of altruism. To do so, we distill the issue to a rational peer's *last* opportunity to cooperate.

The last exchange. Rational peers are induced to cooperate with another peer (or, more generally, with a service) by the expectation that, if they cooperate, they will receive future benefit. However, in most cases, interaction with a particular peer or with the service itself eventually comes to an end. In this last exchange, rational peers do not have incentive to contribute, as doing so incurs cost without any future benefit. Unfortunately, rational cooperation throughout the protocol often hinges on this critical last exchange: the lack of incentive to cooperate at the end may, in a sort of reverse domino effect, demotivate rational peers from cooperating in *any* prior exchange.

Most current systems address this problem in one of three ways (or some combination of them). Some systems [3, 16] assume that rational peers interact with the service forever, and thus future incentives always exist; others [13–15] assume rational peers deviate only if their increase in utility is above a certain threshold; others, finally, try to threaten rational peers with the possibility of losing utility if they deviate. For instance, in BAR Gossip [16], peers that do not receive the data they expect *pester* the guilty peer by repeatedly requesting the missing contribution.

Unfortunately, each of these approaches relies on somewhat unrealistic assumptions. Few relationships in life are infinite in length; worse, as we will show later in this paper, with a lossy network and the possibility of Byzantine peers it may be impossible to incentivize cooperation even in an infinite-length protocol. The real possibility of penny-pinching peers can undermine any system that assumes no deviation unless their expected gain is "large enough." Finally, threats such as pestering are effective only when they are credible: to feel threatened, a peer must believe that it will be rational for the other peer to pester. Since pestering incurs cost for the initiator as well as for the receiver, it is surprisingly hard to motivate rational peers to pester in the first place. For example, pestering in BAR Gossip is credible only under the rather implausible assumption that a peer, even when faced with enduring silence, will never give up on an

[1] Gossip-based BAR-tolerant streaming protocols [15, 16] do rely on an altruistic source for seeding the stream but otherwise model the gossiping peers as either rational or Byzantine.

unresponsive peer and forever continue to attribute a peer's lack of contribution to the unreliability of the network [16].

Our contributions. We model the last-exchange problem as a finite-round game between two peers, $\mathcal{P}_1$ and $\mathcal{P}_2$; neither peer expects to interact with the other beyond this exchange. We assume $\mathcal{P}_1$ holds a contribution (*e.g.*, some information) that is of value to $\mathcal{P}_2$; however, contributing yields no expectation of further benefit for $\mathcal{P}_1$. We are interested in studying whether $\mathcal{P}_2$ can nonetheless induce a selfish $\mathcal{P}_1$ to contribute by threatening to pester it if $\mathcal{P}_1$ fails to do so. Pestering is an attractive threat because it is simple and does not require the involvement of a third party. We want to determine whether it can be made a credible threat under realistic system assumptions, unlike in BAR Gossip [16]. In each round, $\mathcal{P}_1$ is given a choice whether to contribute or not; in response, $\mathcal{P}_2$ may pester $\mathcal{P}_1$. Peers communicate through a lossy channel and therefore do not necessarily share the same view of the ongoing game. For instance, $\mathcal{P}_1$ may have contributed, but $\mathcal{P}_2$ may not have received the contribution.

We show that, without requiring implausible network assumptions or the specter of never-ending pestering, the presence of altruistic peers is both necessary and sufficient to make pestering a credible threat and motivate rational peers to contribute. In particular:

- We prove that there exists no equilibrium strategy where rational peers contribute if all peers are either rational or Byzantine—even if we allow for an infinite number of pestering rounds.
- We show that the presence of altruistic peers is sufficient to transform pestering into a credible threat. Intuitively, if rational peers have sufficiently high beliefs that they may be interacting with an altruistic peer, they are motivated to pester, making it in turn preferable for rational peers to contribute.

The fraction of altruistic peers sufficient to sustain rational contribution depends on several system parameters, including the probability of network loss, the fraction of Byzantine peers in the system, and the behavior that rational peers expect from altruistic and Byzantine peers. Exploring this space through a simulator we find that:

- Altruistic peers make rational cooperation easy to achieve under realistic conditions. In particular, we find that even if less than 10% of the population is altruistic, rational peers are incentivized to cooperate in a system where the network drops 5% of all packets and Byzantine peers make up over 50% of the remainder of the population.
- Prodigal altruistic peers do harm rational cooperation: if altruistic peers contribute every time they are pestered, then we cannot always achieve rational cooperation; when we do, it requires an implausibly high fraction of altruistic peers. This is good news: the less foolishly generous is the altruistic behavior sufficient to incentivize rational contribution, the more feasible it is to design systems with a sustainable population of altruistic peers.

– The uncertainty introduced by network loss is both a bane and a boon. On the one hand, it significantly complicates the analysis of a peer's optimal strategy because each peer does not know what the other has observed. On the other, it lowers the threshold for rational cooperation by leaving open some possibility that the other peer may be altruistic, even when the observed behavior suggests otherwise.

Organization of paper. After presenting in Section 2 the game theoretic framework used to analyze the last exchange problem, we show in Section 3 that rational cooperation is impossible in the absence of altruistic peers. We proceed to derive, in Section 4, conditions under which altruism is sufficient to elicit rational cooperation in the last exchange and, in Section 5, use simulations to study the implications of these conditions on the design of cooperative services. We discuss related work in Section 6 before concluding in Section 7. Because of space limitations, we omit proofs or provide proof sketches of most of our results; detailed proofs can be found in a companion technical report [23].

2 Formalizing the last exchange problem

We consider cooperative services that can be modeled as a collection of peer-to-peer pairwise exchanges, in which two players $\mathcal{P}_1$ and $\mathcal{P}_2$ communicate over unreliable channels. In particular, we focus on the last exchange between $\mathcal{P}_1$ and $\mathcal{P}_2$; we are interested in studying under which conditions $\mathcal{P}_2$ can induce a selfish $\mathcal{P}_1$ to contribute with the threat of pestering.

We model this last exchange as a $(T+1)$-round stochastic sequential game, which is similar to a repeated game except that it allows players' payoffs to change. This flexibility is critical to model the intuition that $\mathcal{P}_2$ benefits from $\mathcal{P}_1$'s contribution only the first time $\mathcal{P}_2$ receives it. In each round, $\mathcal{P}_1$ moves first by choosing between two actions: contribute (denoted by c) or do nothing (n). $\mathcal{P}_2$ follows by choosing between two actions: pester (p) or do nothing (n). Since our analysis of the game often relies on the number of rounds remaining rather than on the round number, we think of the game as starting with round T and ending with round 0.

While doing nothing has neither cost nor benefit, $\mathcal{P}_1$ incurs a cost s_c in every round in which it contributes and a cost r_p in every round in which it is pestered; $\mathcal{P}_2$ incurs a cost r_c in every round it receives a contribution and a cost s_p in every round it pesters $\mathcal{P}_1$. A non-Byzantine $\mathcal{P}_2$ starts off being *destitute*, *i.e.*, $\mathcal{P}_2$ does not have $\mathcal{P}_1$'s contribution. A destitute $\mathcal{P}_2$ receives a one-time benefit $b_c \gg r_c + s_p$ the first time it receives $\mathcal{P}_1$'s contribution; now no longer destitute, $\mathcal{P}_2$ gains no further benefit from receiving further copies of the contribution.

Network loss, signals, and utilities. To model the unreliable channel through which $\mathcal{P}_1$ and $\mathcal{P}_2$ communicate, we adopt from game theory the concept of *private signals*: for every action a played by some player, both players privately observe some (possibly different) resulting *signal*. Specifically, let ρ, $0 < \rho < 1$, be the rate of network loss, which we assume to be common knowledge. When $\mathcal{P}_i$ plays

a, $\mathcal{P}_i$ observes a, and its peer $\mathcal{P}_{-i}$ observes a with probability $1 - \rho$ and n otherwise. Thus, players do not always observe their peer's actions accurately and cannot rely on their peer accurately observing their own actions.

The sequence of signals observed by $\mathcal{P}_i$ until round t defines $\mathcal{P}_i$'s *history* h_i^t. At the beginning of the game, h_1^T is the empty sequence; h_2^T consists instead of the signal s_1^T corresponding to $\mathcal{P}_1$'s action in round T. In round $k < T$, h_1^k is obtained by appending to $\mathcal{P}_1$'s prior history h_1^{k+1} the pair of signals observed by $\mathcal{P}_1$ in round $k{+}1$: $h_1^k = (h_1^{k+1}, s_1^{k+1}, s_2^{k+1})$. Similarly for $\mathcal{P}_2$, $h_2^k = (h_2^{k+1}, s_2^{k+1}, s_1^k)$.

$\mathcal{P}_1$ and $\mathcal{P}_2$'s utilities are defined with respect to the signals they observe:

$$u_1(C, \hat{P}) = -\left(|\hat{P}|r_p + |C|s_c\right) \qquad u_2(P, \hat{C}) = H[|\hat{C}| - 1]b_c - \left(|P|s_p + |\hat{C}|r_c\right)$$

where C and $\hat{P}$ are the sets of rounds in which $\mathcal{P}_1$ respectively contributed and observed $\mathcal{P}_2$ pester; P and $\hat{C}$ are the sets of rounds in which $\mathcal{P}_2$ respectively pestered and observed $\mathcal{P}_1$ contribute; and $H[n]$ is the unit step function.[2]

Strategies, types, beliefs, and equilibrium. All players are assigned an initial *strategy*, *i.e.*, the protocol. The strategy that a player ultimately follows, however, depends on the player's *type*:

- Byzantine (**B**): These players play an arbitrary strategy.
- Altruistic (**A**): These players follow the strategy assigned to them initially.
- Rational (**R**): These players follow a strategy only if deviating unilaterally from it does not increase their utility.

As noted earlier, our game is stochastic: the payoffs of a non-Byzantine $\mathcal{P}_2$ change depending on whether $\mathcal{P}_2$ is destitute (and wants $\mathcal{P}_1$ to contribute) or not (and wants $\mathcal{P}_1$ to do nothing). For convenience, we abuse notation and introduce two additional types, **D** and ¬**D**, to characterize the state of a non-Byzantine player $\mathcal{P}_2$, depending on whether or not it is destitute. Note that if $\mathcal{P}_1$ contributes, a $\mathcal{P}_2$ of type **D** will observe c, and hence change to type ¬**D**, with probability $1 - \rho$.

We are interested in finding strategies in which a rational players is not able to deviate and increase its utility *ex-ante*, *i.e.*, in expectation over its peers' types and strategies. Determining whether a given strategy is a rational player's best response, however, is complicated in our game because, while each player knows its own type, it does not know for certain the type of its peer. Instead, $\mathcal{P}_i$ starts with initial *beliefs* $\mu_i(\theta)$ representing the probabilities that $\mathcal{P}_i$ assigns to the statement that its peer $\mathcal{P}_{-i}$ is of type θ. We assume that, for all $i \in \{1, 2\}$, $\mu_i(\theta)$ equals an initial value $\mu(\theta)$, which is common knowledge, and that the beliefs of a rational $\mathcal{P}_i$'s evolve based on the history h_i^t it has observed; we use $\mu_i(\theta|h_i^t)$ to denote $\mathcal{P}_i$'s conditional beliefs.

For a given set of beliefs, a rational player's strategy $\sigma_{\mathbf{R},i}$ depends on the specific strategy that it expects its peer to adopt—which, in turn, depends on the peer's type. A rational player expects an altruistic $\mathcal{P}_i$'s strategy $\sigma_{\mathbf{A},i}$ to be

[2] $H[n] = 0$ if $n < 0$; else $H[n] = 1$.

identical to the initially assigned protocol and a rational player to follow $\sigma_{\mathbf{R},i}$ (assuming that it is a best response). If $\mathcal{P}_i$ is Byzantine, however, its strategy $\sigma_{\mathbf{B},i}$ can in principle be arbitrary, significantly complicating the task of identifying a rational player's best response ex-ante. We address this difficulty by restricting the beliefs that a rational player can hold vis-à-vis Byzantine behaviors. In particular, we assume that a rational player does not expect to be able to influence $\sigma_{\mathbf{B},i}$ through its actions, *i.e.*, a rational player expects to observe a Byzantine peer $\mathcal{P}_i$ do nothing in round t with some probability $\beta_i^t \geq \rho$ that does not depend on $\mathcal{P}_i$'s current history h_i^t. Thus, rational players, instead of considering an arbitrary $\sigma_{\mathbf{B},i}$, best-respond to this expected Byzantine strategy $\bar{\sigma}_{\mathbf{B},i}$. While this restriction sacrifices the generality of Byzantine behavior, it models the reasonable distrust that a rational player is likely to harbor towards a Byzantine peer's threats and promises. If $\mathcal{P}_i$ is not Byzantine, however, the expectation is that its strategy will depend on its observed history h_i^t; we use $\sigma_{\theta,i}(a|h_i^t)$ to denote the conditional probability that a is played by $\mathcal{P}_i$ of type θ given h_i^t.

Formally, let $\sigma_i = (\bar{\sigma}_{\mathbf{B},i}, \sigma_{\mathbf{A},i}, \sigma_{\mathbf{R},i})$ denote the strategies that a rational player expects $\mathcal{P}_i$ to adopt, depending on $\mathcal{P}_i$'s type; $\sigma = (\sigma_1, \sigma_2)$ denote the strategy profile that describes the (expected) strategies for $\mathcal{P}_1$ and $\mathcal{P}_2$; and $\mu = (\mu_1, \mu_2)$ denote the belief profile that describes the beliefs μ_1 and μ_2 held by a rational $\mathcal{P}_1$ and $\mathcal{P}_2$. We are interested in perfect Bayes equilibrium: a strategy profile and set of beliefs (σ^*, μ^*) such that for all $i \in \{1,2\}$, $\mu_i^*(\theta|h_i^t)$ is computed using Bayes rule whenever h_i^t is reached via a signal that may be observed with positive probability; and for all histories h_i^t and strategies $\sigma'_{\mathbf{R},i}$:

$$E^{(\sigma^*_{\mathbf{R},i}, \sigma^*_{-i})}[u_i|h_i^t] \geq E^{(\sigma'_{\mathbf{R},i}, \sigma^*_{-i})}[u_i|h_i^t]$$

where $E^{(\sigma_{\mathbf{R},i}, \sigma^*_{-i})}[u_i|h_i^t]$ is a rational $\mathcal{P}_i$'s expected utility from playing $\sigma_{\mathbf{R},i}$ with beliefs μ_i^*, with both strategy and beliefs conditional on h_i^t, while its peer $\mathcal{P}_{-i}$ plays σ^*_{-i}. To lighten the already substantial notation, we drop $\bar{\sigma}_{\mathbf{B},i}$ and $\sigma_{\mathbf{A},i}$ from our expressions whenever it is obvious what these strategies are.

We assume that all players are limited to actions in the strategy space. This can be accomplished in practice if actions outside of the strategy space generate a proof of misbehavior [3, 9] and if the associated punishments (*e.g.*, financial penalties) are sufficient to deter rational players. Finally, we assume that a rational $\mathcal{P}_1$ does not try to avoid pestering by severing its network connection: if losing a fraction of bandwidth from pestering is undesirable, disconnecting and losing all of it is even less desirable.

3 The need for altruism

Altruism is not only sufficient to incentivize cooperation, it is necessary. In this section, we assume that there is no altruism, and we show that, as a result, rational players never pester or contribute.

Theorem 1 *There exists no equilibrium in which rational $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively contribute and pester.*

Proof. (Sketch) Suppose such an equilibrium (σ^*, μ^*) exists. Then there exists some rounds t_c and t_p such that $\mathcal{P}_1$ and $\mathcal{P}_2$ contribute and pester, respectively, with some positive probability for the last time. However, a rational $\mathcal{P}_1$ never contributes after round t_p since $\mathcal{P}_1$ incurs cost by contributing, yet there is no further threat of pestering; thus, $t_c \geq t_p$.[3] On the other hand, a rational $\mathcal{P}_2$ only pesters until round $t_c + 1$ since $\mathcal{P}_2$ incurs cost by pestering with no chance of further contribution from $\mathcal{P}_1$; thus $t_p > t_c$. Contradiction.

Theorem 1 only holds when the game lasts for a finite number of rounds. When there exists no bound on the number of rounds, a weaker, yet in practice still crippling, result holds. We summarize the main result here; the model of the infinitely-repeated game, which generalizes the finitely-repeated model, and details of the results can be found in the companion technical report [23].

Theorem 2 *In the infinitely-repeated game, suppose a non-destitute $\mathcal{P}_2$ always prefers to do nothing and rational players expect that there exists some positive fraction of Byzantine peers that either (a) when playing as $\mathcal{P}_1$, never contributes; or (a) when playing as $\mathcal{P}_2$, plays the same strategy played by a destitute rational $\mathcal{P}_2$. Then there exists no pure equilibrium in which rational $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively contribute and pester.*[4]

Proof. (Sketch) If some Byzantine $\mathcal{P}_2$ pesters as if playing the destitute rational strategy, despite $\mathcal{P}_1$'s contributions, then $\mathcal{P}_1$'s belief that $\mathcal{P}_2$ is Byzantine eventually grows arbitrarily close to 1. Similarly, if a Byzantine $\mathcal{P}_1$ never contributes despite $\mathcal{P}_2$'s incessant pestering, then $\mathcal{P}_2$ becomes increasingly certain $\mathcal{P}_1$ is Byzantine. It can be shown that a player's belief in its peer being Byzantine eventually grows sufficiently high such that the expected utility of contributing (in the first case) or pestering (in the second) is lower than that of doing nothing. By showing a bound of the number of rounds in which a rational $\mathcal{P}_1$ contributes or $\mathcal{P}_2$ pesters, it follows, using an argument similar to the finitely-repeated game, that this bound must be 0.

4 Altruism to the rescue

We now show that altruism is sufficient to incentivize rational peers to, respectively, pester and contribute by constructing a cooperative strategy profile and proving that it is an equilibrium. We start by specifying the *altruistic strategy*:

- $\sigma_{\mathbf{A},1}$: $\mathcal{P}_1$ contributes during round T. During round $t < T$, $\mathcal{P}_1$ contributes, only if pestered, with probability $(1-\alpha)/(1-\rho)^2$, where α is a known parameter such that $0 < (1-\alpha)/(1-\rho)^2 \leq 1$.[5]

[3] Recall that we count rounds in reverse.

[4] We can also show there exist no mixed equilibria that put positive probability on a finite number of histories. We leave other mixed strategies to future work.

[5] Hence, if $\mathcal{P}_2$ pesters an altruistic $\mathcal{P}_1$ during round t, $\mathcal{P}_2$ expects to observe a contribution in round $t-1$ with probability $1-\alpha$.

– $\sigma_{\mathbf{A},2}$: For any round $t > 0$, $\mathcal{P}_2$ pesters if and only if $\mathcal{P}_2$ is destitute.

In practice, all players are initially given the altruistic strategy. Although we cannot guarantee that a rational $\mathcal{P}_1$ will follow $\sigma_{\mathbf{A},1}$, we can (and will) prove that, under the expectation that its peer $\mathcal{P}_{-i}$ of type θ plays $\sigma_{\theta,-i}$,[6] a rational $\mathcal{P}_i$ will play the following *rational strategy*:

– $\sigma_{\mathbf{R},1}$: During round t, $\mathcal{P}_1$ contributes if and only if $t > \lceil s_c/((1-\rho)^2 r_p) \rceil - 1$ (*i.e.*, if being pestered is sufficiently expensive to overcome the cost of contributing), $\mathcal{P}_1$ observes pestering (for round $t < T$), and $\mathcal{P}_1$'s belief that $\mathcal{P}_2$ is destitute exceeds some threshold $\bar{\mu}_1^t$.
– $\sigma_{\mathbf{R},2}$: Same as $\sigma_{\mathbf{A},2}$.

In a perfect Bayes equilibrium, whether a rational player deviates or not depends on its beliefs for all histories, both those on and off the equilibrium path. In our desired equilibrium, almost every history has some positive probability of being observed: a rational $\mathcal{P}_2$ expects that an altruistic $\mathcal{P}_1$ contributes with positive probability (if $\mathcal{P}_2$ pestered); destitute $\mathcal{P}_2$ always pester; and, as a result of network loss, doing nothing is always observable with positive probability from either player. Thus, for most histories, Bayes rule can be applied to calculate a rational player's beliefs. For those histories that are not observed with positive probability, beliefs can be assigned which support the rational strategy as an equilibrium strategy. The details, omitted here for lack of space, can be found in the companion technical report [23].

Generally, there may exist multiple strategies that result in a cooperative equilibrium. We believe that our rational strategy represents a sensible design point: incentivizing a rational $\mathcal{P}_1$ to contribute in every round would require $\mathcal{P}_1$ to start with an unrealistically low belief in $\mathcal{P}_2$ being Byzantine. Fortunately, this is unnecessary: we show in Section 5 that the rational strategy results in $\mathcal{P}_1$ often contributing multiple times.

4.1 When does a rational $\mathcal{P}_2$ pester?

Intuitively, $\mathcal{P}_2$ pesters only if it is destitute *and* it believes that $\mathcal{P}_1$ is sufficiently altruistic (and thus willing to contribute, even in the final rounds).

Lemma 3 *If a rational $\mathcal{P}_2$ has received a contribution, $\mathcal{P}_2$ does nothing.*

Proof. (Sketch) If $\mathcal{P}_2$ already has the contribution, $\mathcal{P}_2$ receives no further benefit from receiving another contribution. In fact, pestering and receiving another contribution only incurs cost.

[6] Thus, all our lemmas and theorems should be prefaced by "In our cooperative equilibrium…".

Theorem 4 *A rational, destitute $\mathcal{P}_2$ pesters in round $t > 0$ if its belief $\mu_2(\boldsymbol{A}|h_2^t)$ that $\mathcal{P}_1$ is altruistic is such that:*

$$\mu_2(\boldsymbol{A}|h_2^t) > \frac{s_p}{\alpha^{t-1}(1-\alpha)(b_c - r_c) + (1-\alpha^{t-1})s_p} \tag{1}$$

Proof. (Extended sketch) By contradiction. Consider some equilibrium (σ^*, μ^*) in which a rational, destitute $\mathcal{P}_2$ does nothing in round t despite holding beliefs $\mu_2(\mathbf{A}|h_2^t)$ that satisfies condition (1). Construct an alternate rational strategy $\sigma'_{\mathbf{R},2}$ in which $\mathcal{P}_2$ instead pesters in round t; if $\mathcal{P}_2$ receives a contribution in round $t-1$, then $\mathcal{P}_2$ does nothing for the remainder of the game; if it does not, $\sigma'_{\mathbf{R},2}$ and $\sigma^*_{\mathbf{R},2}$ are identical for the remaining rounds.

Consider $\mathcal{P}_2$'s difference in expected utility between playing $\sigma^*_{\mathbf{R},2}$ and $\sigma'_{\mathbf{R},2}$. There are three cases. If $\mathcal{P}_1$ is Byzantine, the expected difference in utility between $\sigma^*_{\mathbf{R},2}$ and $\sigma'_{\mathbf{R},2}$ is s_p. If $\mathcal{P}_1$ is rational, it can be shown that $\mathcal{P}_2$ has a better chance of receiving a contribution in the future if $\mathcal{P}_2$ pesters now (versus doing nothing); hence, if $\mathcal{P}_2$ is destitute starting from h_2^t, then the expected difference in utility between $\sigma^*_{\mathbf{R},2}$ and $\sigma'_{\mathbf{R},2}$ is at most s_p. Finally, if $\mathcal{P}_1$ is altruistic, then the expected utility, starting from $\mathcal{P}_2$'s turn in round $t-1$, of playing $\sigma^*_{\mathbf{R},2}$ or $\sigma'_{\mathbf{R},2}$ is the same; let $V(\mathbf{A}, t-1)$ represent this utility. The expected difference in utility between $\sigma^*_{\mathbf{R},2}$ and $\sigma'_{\mathbf{R},2}$ when facing an altruistic $\mathcal{P}_1$ is then

$$s_p - (1-\alpha)(b_c - r_c - V(\mathbf{A}, t-1))$$

It can be shown that pestering an altruistic $\mathcal{P}_1$ until $\mathcal{P}_2$ gets the contribution or $t = 0$ is in $\mathcal{P}_2$'s best interest; thus, for $i < t$, $V(\mathbf{A}, i) \leq -s_p + (1-\alpha)(b_c - r_c) + \alpha V(\mathbf{A}, i-1)$, where $V(\mathbf{A}, 0) = 0$. Solving the recursion and using condition (1) we find that the expected difference in utility between $\sigma^*_{\mathbf{R},2}$ and $\sigma'_{\mathbf{R},2}$ is at most

$$s_p - \mu_2(\mathbf{A}|h_2^t)(\alpha^{t-1}(1-\alpha)(b_c - r_c) + (1-\alpha^{t-1})s_p) < 0$$

This implies that $\mathcal{P}_2$ prefers to play $\sigma'_{\mathbf{R},2}$ over $\sigma^*_{\mathbf{R},2}$. Contradiction.

4.2 When does a rational $\mathcal{P}_1$ contribute?

It is obvious that $\mathcal{P}_1$ never contributes when the threat of pestering does not offset the cost of contributing.

Lemma 5 $\mathcal{P}_1$ *does nothing for rounds $t \leq \tau$, where*

$$\tau = \left\lceil \frac{1}{(1-\rho)^2} \frac{s_c}{r_p} \right\rceil - 1 \tag{2}$$

The next two results limit when $\mathcal{P}_1$ contributes. Even after it contributes, if $\mathcal{P}_1$ is unsure whether or not $\mathcal{P}_2$ is still destitute, a combination of several factors (a highly lossy network that may have dropped the contribution, a high cost for being pestered, enough rounds to make pestering a sufficiently costly threat) may induce $\mathcal{P}_1$ to contribute again, without being pestered.

Lemma 6 gives a condition under which $\mathcal{P}_1$ contributes only if a non-Byzantine $\mathcal{P}_2$ is known to be destitute. It follows in Theorem 7 that, after the first round, $\mathcal{P}_1$ contributes only if pestered in the prior round.

Lemma 6 *Let $t < T$ be the current round, where*

$$T < \frac{1-\rho+\rho^2}{\rho^2(1-\rho)^2}\frac{s_c}{r_p} \tag{3}$$

If $\mathcal{P}_1$ contributed in the past and has not been pestered since, then $\mathcal{P}_1$ does nothing in round t.

Theorem 7 *Let $t < T$ be the current round, and suppose that $\mathcal{P}_1$ observed nothing from $\mathcal{P}_2$ in round $t+1$. Then $\mathcal{P}_1$ does nothing in round t if (3) holds.*

Proof. (Sketch) By contradiction. Consider some round t in which $\mathcal{P}_1$ contributes despite having done nothing and observed $\mathcal{P}_2$ do nothing in round $t+1$ (if $\mathcal{P}_1$ contributed, Lemma 6 holds). It can be shown that since (1) $\mathcal{P}_1$'s belief in $\mathcal{P}_2$ being destitute is strictly non-decreasing when observing $\mathcal{P}_2$ do nothing and (2) the number of expected pesters decreases with the number of remaining rounds, $\mathcal{P}_1$ increases its utility by contributing in round $t+1$ instead of round t. Contradiction.

We now consider the conditions under which $\mathcal{P}_1$ actually contributes. In every round, $\mathcal{P}_1$ must make a choice:

- Pay the cost of contributing now (s_c), hoping to stop a non-Byzantine $\mathcal{P}_2$ from pestering in the future. The savings are a function of the remaining rounds and the beliefs about $\mathcal{P}_2$.
- Delay contributing, at the risk of being pestered (with cost at most $(1-\rho)r_p$), hoping to glean more about $\mathcal{P}_2$'s type.

Procrastination has its lure. Since we are considering strategies where a non-Byzantine $\mathcal{P}_2$ always pesters (minus the last round) whereas a Byzantine $\mathcal{P}_2$ may not, every additional signal can drastically affect $\mathcal{P}_1$'s expected utility and possibly save $\mathcal{P}_1$ the cost of contributing. Moreover, doing nothing now does not preclude $\mathcal{P}_1$ from contributing in the future. Yet, we find that if $\mathcal{P}_1$ has sufficiently strong belief that $\mathcal{P}_2$ is destitute, procrastination is something best put off until tomorrow: for every round sufficiently removed from the end of the game, there exists a belief threshold above which contributing yields a higher expected utility for $\mathcal{P}_1$.

Theorem 8 *Let h_1^t be the history of $\mathcal{P}_1$ in round t, where $\tau < t \leq T$ and τ is defined by condition (2). A rational $\mathcal{P}_1$ contributes if its belief that $\mathcal{P}_2$ is destitute exceeds some threshold $\bar{\mu}_1^t \leq 1$. In particular, if $\mu_1(\boldsymbol{D}|h_1^t) \geq \bar{\mu}_1^t$, $\mathcal{P}_1$ contributes; otherwise, $\mathcal{P}_1$ does nothing.*

Proof. (Extended sketch) By induction on t. The base case, $t = \tau + 1$, is simple: since $\mathcal{P}_1$ never contributes after round $\tau + 1$, we can calculate the threshold at which $\mathcal{P}_1$ prefers to contribute. For the inductive step, we assume the theorem holds for all t, $\tau < t \leq t_0$; we prove $t = t_0+1$ by contradiction. If a threshold does not exist, there must be some belief $\mu^*(\mathbf{D}|h_1^t)$ in which $\mathcal{P}_1$ prefers to contribute and higher beliefs in which $\mathcal{P}_1$ prefers to do nothing. It can be shown that we can always find a belief $\mu_1'(\mathbf{D}|(h_1')^t)$, arbitrarily close to $\mu^*(\mathbf{D}|h_1^t)$ from above, such that:

- $\mathcal{P}_1$ does nothing in round t; and
- $\mathcal{P}_1$ plays the same actions in subsequent rounds after observing the same non-empty sequence of signals as if $\mathcal{P}_1$ had started with belief $\mu^*(\mathbf{D}|h_1^t)$.

It follows that $\mathcal{P}_1$'s expected utility of playing action a_1^t followed by the optimal strategy σ^* with either belief $\mu_1^*(\theta|h_1^t)$ or $\mu_1'(\theta|(h_1')^t)$ must be equal; let $V(a_1^t, \theta)$ be this expected continuation utility. By Lemmas 3 and 6, we know that $V(a_1^t, \neg\mathbf{D}) = 0$. Given belief $\mu_1^*(\theta|h_1^t)$, $\mathcal{P}_1$ prefers to contribute; given belief $\mu_1'(\theta|h_1^t)$, $\mathcal{P}_1$ prefers to do nothing. This implies that

$$-s_c \geq \mu_1^*(\mathbf{D}|h_1^t)(V(n,\mathbf{D}) - \rho V(c,\mathbf{D})) + \mu_1^*(\mathbf{B}|h_1^t)(V(n,\mathbf{B}) - V(c,\mathbf{B}))$$
$$-s_c < \mu_1'(\mathbf{D}|(h_1')^t)(V(n,\mathbf{D}) - \rho V(c,\mathbf{D})) + \mu_1'(\mathbf{B}|(h_1')^t)(V(n,\mathbf{B}) - V(c,\mathbf{B}))$$

Using these conditions, the fact that $\mathcal{P}_1$ is never better off contributing to a Byzantine $\mathcal{P}_2$ (*i.e.*, $V(n,\mathbf{B}) - V(c,\mathbf{B}) + s_c \geq 0$), and Lemma 6, we can derive a contradiction.

4.3 The rational strategy is an equilibrium strategy.

Theorem 9 *Assume an altruistic $\mathcal{P}_i$ plays the altruistic strategy $\sigma_{\mathbf{A},i}$. Let T be constrained by condition (3) and condition (1) hold in all rounds but the last one. Then the rational strategy is an equilibrium strategy.*

Proof. By Lemmas 3 and 5 and Theorems 4, 7, and 8.

5 Characterizing the equilibrium

To understand the implications of Section 4 on the design of cooperative services, we explore, through simulation, the parameter space for which our cooperative equilibrium holds. We ask the following questions:

1. What fraction of altruistic peers suffices to motivate $\mathcal{P}_2$ to pester? The shaded areas in Figure 1 show (for different rates of network loss, different initial beliefs about the likelihood of $\mathcal{P}_1$ being Byzantine, and different worth of receiving a contribution) the fraction of altruistic peers that suffices to trigger $\mathcal{P}_2$'s pestering, as a function of the probability $((1-\alpha)/(1-\rho)^2)$ that an altruistic $\mathcal{P}_1$ will contribute if pestered ($\mathcal{P}_1$'s *generosity*). We assume that $\mathcal{P}_2$ believes an altruistic $\mathcal{P}_1$ follows the altruistic strategy; a Byzantine $\mathcal{P}_1$ never contributes;

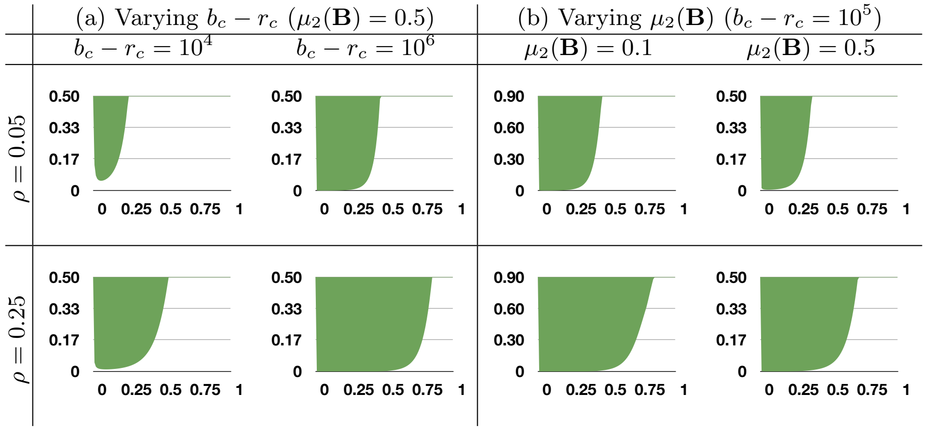
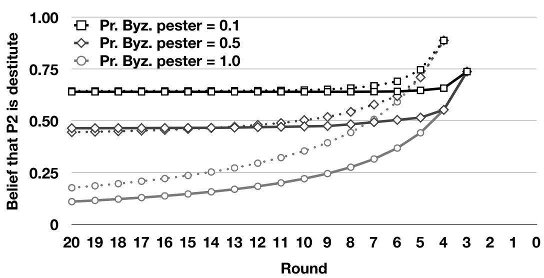


Fig. 1: Sufficient initial beliefs for a rational $\mathcal{P}_2$ in its peer $\mathcal{P}_1$ being altruistic to incentivize $\mathcal{P}_2$ to pester (y-axis, shaded area) for varying amounts of altruistic generosity (x-axis). Simulation run with $s_p = 1$, $T = 20$.



Pr. Byz. pester	$\rho = 0.05$	$\rho = 0.25$
0.1	14	17
0.5	5	9
1.0	2	3

Fig. 2: Left: the belief thresholds of $\mathcal{P}_1$'s beliefs; solid lines represent $\rho = 0.05$; dotted lines represent $\rho = 0.25$. Right: the maximum number of times $\mathcal{P}_1$ contributes for each of these thresholds. Simulation run with $s_c/r_p = 2$, $T = 20$.

and a rational $\mathcal{P}_1$ only contributes in round T. This is a conservative estimate on the fraction of altruistic peers sufficient to motivate $\mathcal{P}_2$; in practice, the actual amount is likely to be lower than we report. As expected, for a given level of generosity, it is easier to incentivize $\mathcal{P}_2$ if the value of the contribution increases and the likelihood of $\mathcal{P}_1$ being Byzantine decreases. Given a highly lossy network, $\mathcal{P}_2$ is also more willing to continue pestering, as it is more willing to attribute to network loss its failure to receive a contribution.

2. How are $\mathcal{P}_1$'s beliefs and the rate of network loss affecting $\mathcal{P}_1$'s willingness to contribute? $\mathcal{P}_1$ may contribute only if its belief that $\mathcal{P}_2$ is destitute is above a certain threshold. We show in Figure 2 how that threshold changes over the course of a game in which $T = 20$. For six configurations, obtained by taking the cross product of two rates of network loss and three different probabilities that a Byzantine $\mathcal{P}_2$ will pester, we plot the belief threshold and report the number of times that $\mathcal{P}_1$ contributes. For a given round, we assign $\mathcal{P}_1$ some initial belief that $\mathcal{P}_2$ is destitute and construct the game tree to determine whether that initial belief is sufficient to motivate $\mathcal{P}_1$ to contribute in that round; we use binary search to approximate the threshold value. As expected, when the game has only few rounds left and the cost from being pestered is not enough to overcome the cost of contributing, there is no threshold above which $\mathcal{P}_1$ contributes. Note also that increasing ρ increases the belief threshold required to convince $\mathcal{P}_1$ to contribute (as it reduces the expected threat from pestering) but also makes $\mathcal{P}_1$ more likely to contribute when pestered, since past contributions are more likely to have been dropped. Also, the belief threshold increases as the likelihood of a Byzantine $\mathcal{P}_2$ pestering decreases, since it becomes more in $\mathcal{P}_1$'s interest to delay contribution, waiting to see whether $\mathcal{P}_2$ will pester. However, when $\mathcal{P}_1$ observes pestering, its belief that $\mathcal{P}_2$ is destitute increases, and $\mathcal{P}_1$ becomes more more willing to contribute. Finally, decreasing the relative cost of contributing (s_c/r_p) has an obvious effect on $\mathcal{P}_1$'s likelihood to contribute (not shown).

3. Too much generosity? An intriguing conclusion from Figure 1 is that altruistic generosity can make it much harder to motivate $\mathcal{P}_2$ to pester. The reason is that the more generous altruistic peers are, the easier it is for a rational $\mathcal{P}_2$ to determine, from observed signals, whether $\mathcal{P}_1$ is altruistic or not, which in turn affects whether $\mathcal{P}_2$ continues to pester. Figure 1 shows the effects that an altruistic peer's generosity has on cooperation. For higher levels of altruistic generosity, we can only guarantee cooperation if such generosity is offset by a high ρ or $b_c - r_c$. Altruistic generosity becomes a more obvious discriminant if a Byzantine $\mathcal{P}_1$ never contributes, but it becomes less conspicuous with higher rates of network loss, which affects the observed generosity from $\mathcal{P}_2$'s perspective. As expected, $\mathcal{P}_2$ is more willing to pester given a more valuable contribution.

6 Related work

Incentive-compatible systems and protocols. There has been much work in incentive-compatible systems (*e.g.*, [3, 4, 13, 14, 16, 20]). None of these systems assume the existence of altruistic players, and only a few [3, 16, 20] consider Byzantine peers. Our techniques can be applied to many of these systems. For example, BAR Gossip [16], FOX [14], and PropShare [13] can use altruism to incentivize key exchange. Our technique can be used towards implementing BAR Gossip's "fair-enough" exchange and may provide insight into the larger fair exchange problem [11, 19]. Finally, rational secret sharing [10] faces a similar problem to the last exchange. However, without a pestering mechanism, our work is not directly applicable.

Irrationality in incentive-compatible protocols. Eliaz [7] proposed the generalization of Nash equilibrium to scenarios where some number of peers may be Byzantine. Aiyer et al. [3] generalized this to the BAR model, which introduced the possibility of altruistic peers and on which our model is based. Abraham et al. [1] describe (k, t)-robust equilibrium, a solution concept in which a rational player does not deviate despite the possibility of collusion by groups up to size k and up to t "irrational" agents that may play any strategy. Similarly, Martin [18] introduces an equilibrium concept in which rational players do not deviate regardless of Byzantine or altruistic players' actions. Our work differs from previous work by showing the need for altruism to address a key problem in cooperative services and considering real-world issues such as network costs and lossy links. Vassilakis et al. [22] study how altruism affects content sharing in P2P services at the application level. Their approach complements our own; we focus on network-level incentives and issues (such as lossy links) that motivate participants to actually send the content they share at the application level. Their work does not address Byzantine participants.

Game-theory. There has been extensive work that has covered imperfect knowledge, private signaling, and the use of altruism in game theory. The use of altruism to achieve cooperation in the finitely-repeated prisoner's dilemma game was first proposed by Kreps et al. [12]. It was shown that reputations could be maintained even when there was imperfect observation of actions [8]. Cripps et al. later showed that, under certain conditions, reputations cannot be maintained forever unless the action played by the irrational player was part of a rational player's equilibrium strategy [5, 6]. None of the previous work consider both the possibility of Byzantine and altruistic players. Many of them also assume that actions or their corresponding signals can either be observed at least publicly [5, 8], if not perfectly [12]. More importantly, the focus of this work is the existence (or nonexistence) of equilibrium under general conditions. We focus on the application of theory to a specific problem and a realistic model that we believe to be applicable to many distributed protocols.

7 Conclusion

Despite the presence of altruistic peers in real-world MAD systems, little attention has been given to their role in establishing rational cooperation. In this paper, we take the first step in understanding their function by showing that altruism is necessary and sufficient to motivate rational cooperation in the crucial last exchange between MAD peers. Our results suggest that, while a small fraction of altruistic peers is sufficient to spur rational peers into action even in systems with a large fraction of Byzantine peers, overly generous altruistic peers can irreparably harm rational cooperation.

Acknowledgments. We are grateful to Tom Wiseman and Joe Halpern for many illuminating discussions, and to the anonymous referees. This work is supported by NSF Grant No. 0905625.

References

1. Abraham, I., Dolev, D., Gonen, R., Halpern, J.: Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation.

In: PODC '06. pp. 53–62 (Jul 2006)
2. Adar, E., Huberman, B.A.: Free riding on Gnutella. First Monday 5(10), 2–13 (Oct 2000), http://www.firstmonday.org/issues/issue5_10/adar/index.html
3. Aiyer, A.S., Alvisi, L., Clement, A., Dahlin, M., Martin, J.P., Porth, C.: BAR fault tolerance for cooperative services. In: SOSP '05. pp. 45–58 (Oct 2005)
4. Cohen, B.: Incentives build robustness in BitTorrent. In: First Workshop on the Economics of Peer-to-Peer Systems (Jun 2003)
5. Cripps, M.W., Mailath, G.J., Samuelson, L.: Imperfect monitoring and impermanent reputations. Econometrica 72(2), 407–432 (2004)
6. Cripps, M.W., Mailath, G.J., Samuelson, L.: Disappearing private reputations in long-run relationships. J. of Economic Theory 127(1), 287–316 (May 2007)
7. Eliaz, K.: Fault tolerant implementation. Rev. of Econ. Studies 69, 589–610 (Aug 2002)
8. Fudenberg, D., Levine, D.K.: Maintaining a reputation when strategies are imperfectly observed. Review of Economic Studies 59(3), 561–79 (July 1992)
9. Haeberlen, A., Kouznetsov, P., Druschel, P.: PeerReview: Practical accountability for distributed systems. In: SOSP '07. pp. 175–188 (Oct 2007)
10. Halpern, J., Teague, V.: Rational secret sharing and multiparty computation. In: Proc. 36th STOC. pp. 623–632 (2004)
11. Kremer, S., Markowitch, O., Zhou, J.: An intensive survey of non-repudiation protocols. Computer Communications 25(17), 1606–1621 (Nov 2002)
12. Kreps, D., Milgrom, P., Roberts, J., Wilson, R.: Rational cooperation in the finitely repeated prisoners' dilemma. J. of Economic Theory 27(2), 245–52 (Aug 1982)
13. Levin, D., LaCurts, K., Spring, N., Bhattacharjee, B.: BitTorrent is an auction: analyzing and improving BitTorrent's incentives. SIGCOMM Comput. Commun. Rev. 38(4), 243–254 (2008)
14. Levin, D., Sherwood, R., Bhattacharjee, B.: Fair file swarming with FOX. In: IPTPS '06 (Feb 2006)
15. Li, H., Clement, A., Marchetti, M., Kapritsos, M., Robinson, L., Alvisi, L., Dahlin, M.: FlightPath: Obedience vs choice in cooperative services. In: OSDI '08. pp. 355–368 (Dec 2008)
16. Li, H.C., Clement, A., Wong, E., Napper, J., Roy, I., Alvisi, L., Dahlin, M.: BAR Gossip. In: OSDI '06. pp. 191–204 (Nov 2006)
17. Locher, T., Moor, P., Schmid, S., Wattenhofer, R.: Free riding in bittorrent is cheap. In: HotNets '06 (Nov 2006)
18. Martin, J.P.: Leveraging altruism in cooperative services. Tech. Rep. MSR-TR-2007-76, Microsoft Research (Jun 2007)
19. Pagnia, H., Gärtner, F.C.: On the impossibility of fair exchange without a trusted third party. Tech. Rep. TUD-BS-1999-02, Darmstadt University of Technology, Department of Computer Science, Darmstadt, Germany (Mar 1999)
20. Peterson, R.S., Sirer, E.G.: Antfarm: efficient content distribution with managed swarms. In: NSDI '09. pp. 107–122 (2009)
21. Piatek, M., Isdal, T., Anderson, T., Krishnamurthy, A., Venkataramani, A.: Do incentives build robustness in BitTorrent? In: NSDI '07. pp. 1–14 (Apr 2007)
22. Vassilakis, D.K., Vassalos, V.: An analysis of peer-to-peer networks with altruistic peers. Peer-to-Peer Networking and Applications 2(2), 109–127 (2009)
23. Wong, E.L., Leners, J.B., Alvisi, L.: It's on me! The benefit of altruism in BAR environments. Technical Report TR-10-08, UT Austin (Feb 2010)