Price of Anarchy in a Double-Sided Critical Goods Distribution System

David Sychrovský^1, Jakub Černý^2, Sylvain Lichau^3, and Martin Loebl^{1*}

^{1}Charles University, ^{2}Nanyang Technological University, ^{3}University of Bordeaux

Abstract

Measures of allocation optimality differ significantly when distributing standard tradable goods in peaceful times and scarce resources in crises. While realistic markets offer asymptotic efficiency, they may not necessarily guarantee fair allocation desirable when distributing the critical resources. To achieve fairness, mechanisms often rely on a central authority, which may act inefficiently in times of need when swiftness and good organization are crucial. In this work, we study a hybrid trading system introduced in (Jedličková, Loebl, and Sychrovský 2022) which combines fair allocation of buying rights with a market. An analogue of Price of Anarchy in this system, called frustration, is defined as a difference between the amount of goods the traders are entitled to according to their assigned buying rights and the amount of goods they are able to acquire in the market. Our contribution is the study of a realistic complex double-sided market mechanism for this system. The empirical analysis of this mechanism suggests that with the fairness mechanism present, the Price of Anarchy decreases.

1 Introduction

Most of the goods available to the general public are meant to increase the quality of life of individuals or count as luxuries, and are traded using standard market mechanisms. Other resources serve a more social purpose – when allocated well, they increase the well-being of the entire society like public housing, school seats, or healthcare products. Among those, some are desirable to be readily available to everyone, e.g., essential medicines, various equipment, or even vaccines that enable to reach herd immunity in the population only when enough people have developed protective antibodies against future infections. In times of need like disasters, local epidemics, or even conflicts and wars, these resources need to be distributed swiftly and in a highly organized manner to reach as many eligible people as possible in a limited timeframe.

Allocating such public resources is commonly reserved for governmental services and done at prices below market-clearing or even free of charge. However, leaving the competitive markets out of the allocation process often results in inefficiencies, both economic and temporal, caused by problems inherent to centralized planning (Moroney and Lovell 1997). On the other hand, real-world trading markets, frequently modeled as large double auctions with many sellers and buyers on each side, are capable of distributing the goods flexibly and reliably. The problem remains that even though, with increasing size, the participants are incentivized to be truthful (which leads to asymptotic efficiency (Cripps and Swinkels 2006)), the resulting goods reallocation is not necessarily socially optimal in terms of being available to everyone. Any discrepancy in wealth is then only exacerbated by crises similar to the coronavirus pandemic or the war in Ukraine we experienced in recent years. In such settings, scarce resources necessary for keeping the society up and running could be easily swayed by its more fortunate members, which has to be countered by carefully designed measures.

As an attempt to combine the best of both worlds, the following hybrid distribution system called Crisdis is suggested in (Jedličková, Loebl, and Sychrovský 2022): a trustworthy central authority provides a marketplace where buyers and sellers engage in two-sided repeated trading over a period of many days. At the beginning of each market day, the authority allocates buying rights to the participating members (e.g., individual hospitals), which are traded together with the goods. Everything else[1] is left up to the sellers and buyers themselves, with one requirement only: at the end of each trading day, each buyer needs to possess the number of rights greater or equal to the number of goods. The straightforward motivation for this arrangement is that the traders selling some of their assigned rights obtain extra funds, which they can use in future markets to satisfy their demand for the critical goods better. Another motivation is that the needs of individual participants used to allocate the rights can be evaluated independently by the central authority using real-time crisis data, thus sidestepping the bottleneck of many auction mechanisms – the proneness to strategic manipulation.

The utmost priority of Crisdis is to improve the accessibility of critical goods to all eligible buyers during crises in a trading system which is as realistic as possible. For this purpose, (Jedličková, Loebl, and Sychrovský 2022) introduced a measure of the social efficiency of the allocations realized by the semi-distributed system called *frustration*. Frustration

[*] contact author: loebl@kam.mff.cuni.cz

[1] In this context, we have in mind not just deciding on the price, but also storage, delivery, etc.

can be seen as a scaled negative difference between fairness and reality: for a participating buyer, it is the scaled difference between the (potential) allocation of rights to the buyer and the number of goods purchased by them if the value is at least zero, and zero otherwise. Assuming the market attains its equilibrium, the sum of frustrations of the traders describes the system's *Price of Anarchy*, i.e., the price the society pays for allocating the goods through the market and not directly as suggested by the fairness mechanism.

In (Jedličková, Loebl, and Sychrovský 2022), the authors study how frustration evolves during repeated interactions in the system under a single-sided auction mechanism based on activities of buyers. This work contributes by a thorough study of a more realistic double-sided mechanism.

1.1 Contributions

We study trading in a system consisting of a sequence of complex double-sided markets combined with a fairness mechanism designed to improve social good. Following (Jedličková, Loebl, and Sychrovský 2022), we focus on the well-known and thoroughly studied *contested garment distribution* for fairly allocating the rights (Aumann and Maschler 1985)[2]. As market mechanisms, in Section 3, we derive four different allocation strategies, ranging from random acceptable allocations to maximum clearing under average-price bids.

Our priority in this work is to study the behavior of a large complex system, which makes it difficult to analyze the traders' behavior theoretically. For this reason, in Section 4, we introduce a reinforced-learning algorithm in an attempt to approximate the system's equilibrium. In Section 5 we present the empirical results. First, we perform a thorough numerical analysis demonstrating how close to the equilibrium we are able to converge to. Then we carry out a series of ablation experiments, showing that the Price of Anarchy in the system without the fairness mechanism may be high. We confirm that together with intuitive governmental regulations akin to increased storing prices for the goods, the system with the fairness mechanism is able to decrease the Price of Anarchy. In the last part of the paper, we summarize the desired features of the trading system enlightened by the experiments.

1.2 Related work

Our work belongs to the literature on redistributive mechanisms, especially those mitigating inequalities. Perhaps the most related paper studies a two-sided market trading goods of homogeneous quality, optimizing the traders' total utilities (Dworczak, Kominers, and Akbarpour 2021). The difference lies in our explicit incorporation of the buyers' varying needs into the consideration and the fact that in our model, the valuations are common knowledge. This work was recently generalized into a setting with heterogeneous quality of tradable objects, more diverse measures of allocation optimality, and imperfect observations about the traders (Akbarpour, Dworczak, and Kominers 2020).

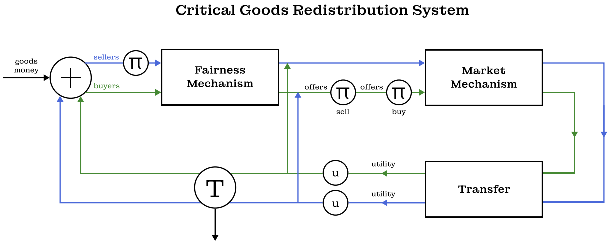

Figure 1: Fairness and market mechanisms positioned in a feedback loop of our redistribution system. One iteration of the loop corresponds to one Market. π refers to the strategies and T checks the termination condition.

Another related work presents multiple markets and non-market mechanisms for allocating a limited number of identical goods to several buyers (Condorelli 2013). The results are intuitive, as the author shows that when the buyers' willingness to pay coincides with the designer's allocation preferences, market mechanisms are optimal, and vice versa. In crises environments studied in our work, it is reasonable to assume that the critical resources are highly valuable to all participants, yet, some may lack the money to obtain them. Together with the fact that it is in society's interest to allocate the goods fairly, these results suggest that leaving the distribution solely to free markets is inadvisable.

2 Problem Definition

We assume the existence of a centralized marketplace where critical goods are traded periodically among the buyers and the sellers using an internal currency. We consider only one type of good and call it Good. To simplify the presentation we assume here that the Good is *divisible*[3]. We refer to each trading period as a *Market*. The structure of the entire system is depicted in Figure 1. In order to reduce volatility during trading, similarly as in (Jedličková, Loebl, and Sychrovský 2022), we introduce a new type of tradable resource called Right. In each Market, in order to buy the Good, the buyer also needs to possess an equivalent amount of the Right. The Rights are allocated to the buyers before the trading begins by a centralized *Fairness mechanism* using the sellers' offers. The traders then engage in a series of interactions resulting in their announcement of bids. A dedicated *Market mechanism* then allocates the Goods based on the bids.

The residual resources of Good are then transferred to the next Market, as we model the shortage of the critical Good for an extended amount of time. However, the residual of Right disappears after each Market. We call this finite sequence of Markets a *Crisis*.

2.1 One trading period: the Market

Formally, we model the trading of the goods during the Market as an imperfect-information double-auction represented as a sequential game $\mathcal{G} = (\mathcal{T}, \mathcal{M}, \mathcal{G}, \mathcal{D}, \phi, \mu, u, k)$.

[2] Our experimental results show that this fairness mechanism performs well in practice.

[3] However, the results can be translated to periodic trading of larger quantities of indivisible Good; see (Jedličková, Loebl, and Sychrovský 2022) for a detailed explanation.

The set of traders $\mathcal{T}$ consists of buyers B and sellers S; the sets of buyers and sellers are assumed to be disjoint. The set $\mathcal{M} = (M_1, M_2, \ldots, M_{|B|}) \in \mathbb{R}_0^{+,|\mathcal{T}|}$ determines the real and non-negative amount of money each buyer receives at the beginning of the Market. Similarly, the set $\mathcal{G} = (G_1, G_2, \ldots, G_{|S|}) \in \mathbb{R}_0^{+,|S|}$ specifies the real and non-negative amount of Good each seller is able to offer for trade. The demands $\mathcal{D} = (D_1, D_2, \ldots, D_{|B|}) \in \mathbb{R}_0^{+,|B|}$ describe the optimal real and non-negative amount of Good each buyer hopes to acquire during the trading. Function ϕ then implements the fairness mechanism, assigning real-valued, non-negative Rights $\mathcal{R} = (R_1, R_2, \ldots, R_{|B|}) \in \mathbb{R}_0^{+,|B|}$ to the individual buyers. To allocate the Rights, the mechanism needs to know the amount of Good put up for trade. This amount is given by the strategies of the sellers. For each seller $s \in S$, the set $\hat{\Pi}_s$ of their strategies contains pairs (v_s^G, p_s^G), where $v_s^G \leq G_s$ is the amount of Good offered at price p_s^G. The profile of one strategy per seller is denoted as $\hat{\pi}_S \in \hat{\Pi}_S$. The fairness mechanism is then formally defined as follows:

Definition 1. *For any sellers' strategy profile $\hat{\pi}_S$, the fairness mechanism is a function $\phi : \mathbb{R}_0^+ \times \mathbb{R}_0^{+,|B|} \to \mathbb{R}_0^{+,|B|}$ allocating Rights to each buyers, satisfying*

$$\begin{aligned}
&\sum_{b \in B} \phi_b(V, \mathcal{D}) = V && \forall\, \mathcal{D} \in \mathbb{R}_0^{+,|B|},\\
&D_b = 0 \Rightarrow \phi_b(V, \mathcal{D}) = 0 && \forall\, \mathcal{D} \in \mathbb{R}_0^{+,|B|}, \forall b \in B\\
&\phi_b(V, \alpha(\mathcal{D})) = \phi_{\alpha^{-1}(b)}(V, \mathcal{D}) && \forall\, \mathcal{D} \in \mathbb{R}_0^{+,|B|}, \forall b \in B, \forall \alpha,
\end{aligned}$$

where α is a permutation of buyers and $V = \sum_{v_s \in \hat{\pi}_S} v_s$.

In this work, we focus exclusively on a fairness mechanism implementing the *contested garment distribution* (CGD) (Aumann and Maschler 1985), as different fairness methods do not influence the strategizing of the traders significantly. Function μ is of more importance to us, allocating the resources after the bidding phase. The bidding is determined by the conditional strategies of the buyers. After observing the seller's offers, for a buyer $b \in B$, their set of strategies $\hat{\Pi}_b$ consists of tuples of three pairs $(v_b^R, p_b^R, \overline{v}_b^R, \overline{p}_b^R, \overline{v}_b^G, \overline{p}_b^G)$, where $v_b^R \leq \phi_b(V, \mathcal{D})$ and p_b^R are the amount and asking price of Right the buyer intends to sell, and $(\overline{v}_b^R, \overline{p}_b^R)$, $(\overline{v}_b^G, \overline{p}_b^G)$ are the amounts and bidding prices of Right and Good, respectively, the buyer wants to acquire after observing other buyers' offers for selling the Right. The profile of one conditional strategy per buyer is denoted as $\hat{\pi}_B \in \hat{\Pi}_B$. The market mechanism is then formally defined as:

Definition 2. *For any sellers' strategy profile $\hat{\pi}_S$ and any buyers' conditional strategy profile $\hat{\pi}_B$, the market mechanism is a function $\mu : \hat{\Pi}_S \times \hat{\Pi}_B \to \mathbb{R}_0^{+,|\mathcal{T}|} \times \mathbb{R}_0^{+,|B|} \times \mathbb{R}_0^{+,|\mathcal{T}|}$ which given bids of all traders returns a realization of trades, i.e., a reallocation of Good, Right and Money among the traders. We abuse the notation a little and write $\mu^G(\hat{\pi}_S, \hat{\pi}_B)$ and $\mu^M(\hat{\pi}_S, \hat{\pi}_B)$ to refer to the restrictions to reallocated Goods and Money, respectively.*

The choice of the market mechanism affects the strategizing of the traders to a great extent. We hence dedicate the entire next section to the study of multiple such mechanisms.

What remains is to define the utility function u. The sellers are motivated solely by profit. Thus, the utility they get from the Market is the amount of Money they receive. We refine this simple model by *adding negative utility* for the Good the seller has at the end of each Market. This penalty represents the societal desire for the sellers to sell most of the available critical Good, and include the state penalties which are usually in place during crises as well as damaged reputation[4]. Moreover, in case the Market terminates the Crisis, the sellers obtain also a small additional utility compensating for the Good they still keep stocked[5]. Formally,

$$u_s(\hat{\pi}_S, \hat{\pi}_B) = \begin{cases} \mu_s^M(\hat{\pi}_S, \hat{\pi}_B) + C_1 \mu_s^G(\hat{\pi}_S, \hat{\pi}_B) & \text{NT,} \\ u_s\text{-NT} + C_2 \mu_s^G(\hat{\pi}_S, \hat{\pi}_B) & \text{T,} \end{cases} \tag{1}$$

where NT/T denote non-terminal/terminal markets, and C_1 and C_2 are suitable constants. The utility of a buyer should incentivize them to keep a steady supply of Good throughout the Crisis. Therefore, after each trading period, they receive utility for the Good they have (up to their demand), which represents their regular consumption (e.g., per day). The buyers also receive some small utility C_3 per unit of Money they have at the end the Crisis. Formally,

$$u_b(\hat{\pi}_S, \hat{\pi}_B) = \begin{cases} \min\left\{D_b, \mu^G(\hat{\pi}_S, \hat{\pi}_B)\right\} & \text{NT,} \\ u_b\text{-NT} + C_3 \mu^M(\hat{\pi}_S, \hat{\pi}_B) & \text{T.} \end{cases} \tag{2}$$

Finally, we move to the description of the process of how the sellers and the buyers engage in trading in our iterated two-sided market. We assume that after the Rights are allocated, the bidding periods are repeated k-times, during which the buyers are allowed to alter their strategies. The full description follows:

1. Each seller declares the amount of Good for sale along with the selling price.
2. Each buyer is assigned Rights according to the fairness mechanism ϕ.
3. The trading is then repeated k times:
   - i Each buyer declares the amount of Right they are willing to sell along with the asking price.
   - ii Each buyer, given the available amounts and asking prices of Good and Right, declare their *bidding price and desired amount* of Good and Right separately.
   - iii The bids are cleared using the market mechanism μ.
   - iv The partial utilities are computed and shown to the traders.
4. The traders receive their final utilities as a sum of utilities from the individual trading periods.

[4] Without such penalty and if the Good is not perishable and the distribution crisis continues for a longer time, the strategic behavior of sellers would probably be to keep selling small amounts of the Good for very high prices.

[5] We assume here that the price at the end of the crisis does not immediately drop to zero.

An important aspect of our model is that the buyers can use the Money they obtained only in the next Market of the sequence. This gives the active buyers the advantage of buying the critical Good earlier than the passive buyers; the price of this advantage is the cost of buying additional rights.

2.2 Sequence of markets: the Crisis

We assume the trading takes place periodically, in a finite sequence of Markets, denoting, e.g., trading days. After each non-terminal trading, the sellers keep the unsold amount of Good and the buyers keep the Money and unconsumed amount of Good for the next Market. In contrast, the unused amount of Right is disposed of after each Market terminates. New Rights are then allocated in the next Market by the fairness mechanism according to the total amount of offered Good and estimated demands.

In this work, we model the full-blown crisis, leaving the boundary situations, i.e., the beginning and the end of a crisis, for future work. Hence we assume all traders are Markovian and base their strategies in the sequence of Markets on local observations exclusively, not conditioning their decision-making on past-Markets experiences. The conditional strategies (i.e., after making the observations) in the sequence and in individual Markets hence coincide. For any trader r, let us denote their set of unconditional strategies – functions taking the observations and outputting the conditional strategies – as Π_r. The set of strategy profiles $\pi = (\pi_1, \pi_2, \dots, \pi_{|\mathcal{T}|}), \quad \pi_1 \in \Pi_1, \pi_2 \in \Pi_2, \dots, \pi_{|\mathcal{T}|} \in \Pi_{|\mathcal{T}|}$ is then Π. As is usual, a situation in which no trader has an incentive to unilaterally change their strategies is called an *equilibrium*.

Definition 3. *Let $\pi^* \in \Pi$ be a profile of one unconditional strategy per each trader. We call π^* an equilibrium, if for any other profile $\pi \in \Pi$ and all traders r it holds that*

$$\sum_{i=1}^{T}\sum_{j=1}^{k} u_r(\pi^*|(i,j)) \geq \sum_{i=1}^{T}\sum_{j=1}^{k} u_r(\pi|(i,j)),$$

where $\overline{\pi}|(i,j), \overline{\pi} \in \Pi$ is the restriction of unconditional profile $\overline{\pi}$ to the corresponding conditional strategies in Market game $\mathcal{G}_i$ in the sequence, and trading period j.

During the entire crisis, we study how the amount of Good acquired by buyers evolves for different Market mechanisms and compare it to the amount of Rights assigned to them. The resulting discrepancy describes the inherent inequality in the system, formally defined as *frustration*.

Definition 4. *Let $\pi \in \Pi$ be a strategy profile and $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_t$ be the corresponding sequence of $t \leq T$ Markets with fairness mechanism ϕ and market mechanism μ. Then the frustration of buyer b after Market t is*

$$f_b^t(\pi) = \max\left\{\frac{\phi_b(\pi|(t,0)) - \sum_{j=1}^{k} \mu_b^G(\pi|(t,j))}{\phi_b(\pi|(t,0))}, 0\right\}.$$

The Price of Anarchy in the system is then the accumulated frustration the buyers experience in the sequence of $t \leq T$ Markets when the equilibrium π^* is reached, i.e.,

$$PoA^t = \frac{\sum_{i=1}^{t}\sum_{b \in B} f_b^i(\pi^*)}{t|B|}. \qquad \text{(PoA)}$$

.

3 Market Mechanisms

In this section, we study how to clear the bids in the Market, i.e., the mechanisms that can be used to schedule individual trades based on the inputs (bids) of the traders. We consider mechanisms with both *absolute* and *average* bidding prices, and we require that each market mechanism satisfies the following criteria for all inputs:

1. no trader sells more Good or Right than they offer;
2. no buyer buys more Good or Right than their declared desired amount;
3. no trader sells Good or Right for a lower price than their asking price;
4. no buyer buys Good or Right for a higher (or higher average) price than is their bidding price; and
5. no buyer can buy Rights from themselves.

Note that the last condition ensures that the desired amount of Right is actually what a buyer would expect. Without it, the buyer can trade virtually with themselves and thus get a lower amount of Right from the Market, even if they could buy more. There exist various market mechanisms which satisfy these properties.

Let us focus first on absolute bidding mechanisms which are those that prohibit *any* trades where the bidding price is larger than the asking price, i.e., $\overline{p}_b^G > p_s^G$ and/or $\overline{p}_b^R > p_{b'}^R$. The clearing constraints (compatibility of asking and bidding prices and possibly other constraints) will be represented by two bipartite graphs: $G_G = (B, S, E_G)$ which represents the compatibility for trading the Good and $G_R = (B_S, B_B, E_R)$ which represents the compatibility for trading the Right. Here, B_S and B_R are disjoint copies of B, B_S represents the sellers of Right and B_R represents the buyers of Right. A trader of B can be both a seller and a buyer of Right, but G_R does not connect their representing vertices by an edge. Both G_G, G_R are equipped with a positive real *weight* $w_G : V_G \to R$ and $w_R : V_R \to R$. The weights of the vertices naturally represent the individual amount (of Good or Right) offered for sale and the individual amount (of Good or Right) desired to buy.

Random allocation A simple random trading mechanism used for purchasing both Rights and Goods proceeds as follow. First, the buyers are randomly permuted. In this order, each buyer is given an randomly permuted lists of offers of the traders for Good and Right respectively. The buyer first trades Good with sellers, until he has no Right left. In the second stage, the buyer trades Good and Right in equal amount. This continues until they buy in total their acceptable volume, or there are no more offers. We also ensure at every step that the asking price is lower than their acceptable price, and the buyer purchases amount up to the amount offered by the other party.

This mechanism has a unsatisfactory property. Since the buyers are presented with offers in random order, they often do not buy the cheapest option. This can be realistic since no single buyer will be able to see all the offers and choose among them. However, if the trading proceeds sequentially, it is natural for the buyer to consider the cheapest offers first. This also gives incentive to the sellers and traders to make offers at a lower price.

Greedy allocation This algorithm is a modification of the random allocation which aims to address the issues mentioned in the last section. At the beginning, the buyers are sorted by the acceptable price of Good $\overline{p}_b^G$ in descending order. The mechanism again has two stages for each buyer. In the first stage, the buyer uses the Rights allocated to him to buy Goods, starting with the cheapest offer. When they have no Right left, they buy the same amount of Rights and Goods, again starting with the cheapest offers. We proceed until all offers are exhausted, or the buyer bought their acceptable volume and continue with the next buyer.

The Random and Greedy allocations are heuristics which can be implemented easily but do not necessarily lead to optimal allocation which clears maximum amount of bids.

Maximum clearing using absolute prices An allocation clearing maximum amount of bids where we also require that no Right is bought without buying equal amount of Good, can be obtained using network flows. We call a mechanism utilizing this approach *Maximum clearing*. Its advantage is that it works also for indivisible Good. Another advantage is that the result of the Maximum clearing allocation is the list of individual tradings with compatible asking and bidding prices. The final price of each individual trading may be chosen in various ways from this compatibility interval.

Theorem 1. *Maximum clearing allocation can be found efficiently using a reduction to the Max Flow problem. As a consequence, a Maximum clearing allocation is polynomial for both divisible and indivisible Good.*

The proof is deferred to Appendix A.

Maximum clearing using average prices In this variant of the Maximum clearing mechanism, we view the prices $\overline{p}_b^G$ and $\overline{p}_b^R$ as maximum *average* prices b is willing to pay.

Theorem 2. *Maximum clearing allocation with average bids can be found efficiently using a linear program.*

The proof is deferred to Appendix B.

4 Learning the System Equilibria

In this section, we describe a reinforced learning algorithm we use to obtain an approximation of the system's equilibrium. Because no analytical solution is known, we treat the entire interaction as a multi-agent reinforcement learning (MARL) problem as it is common in the literature (Fu et al. 2022; Liu et al. 2022; Perolat et al. 2022; Muller et al. 2019), with the assumption that the learning algorithm shall converge to a solution close to the equilibrium. We further verify the quality of the solution by computing its exploitability (Lanctot et al. 2017). The sellers and buyers are represented as agents who interact in the environment described in sections 2 and 3. Each agent is trained to maximize their own expected future utility in this environment.

4.1 Learning environment

The states of the environment relate to the information provided to the traders they may use to condition their strategies on, as described in subsection 2.1. A state of a seller in a given Market is determined by the amount of Good they have in stock. Note that the amount of Money the seller has is not relevant, as it does not impact their future strategy. In each moment, a buyer may be described by three values: the amount of Good, Right, and Money they possess.

During the learning process, the agents are not provided with the complete state of the Market. More specifically, we assume the sellers have access to the full state of the buyers, but not of the other sellers. This corresponds to sellers investing in some market research[6]. The buyers know the amount of Goods, Right and Money they have, and the amount and price of offered Goods and Rights. For simplicity, and to reduce the action space, we assume there exists a maximum price P the Good and Right can be offered at. Since the offered volume is bounded by the volume owned by a trader, the traders' actions fall in a closed interval.

Next, we focus on the traders' utilities. To clearly identify them, we need to specify constants C_i. Let us focus on the sellers first: their utility is given by two constants representing the price of storing the Good, and the expected future utility for the amount of Good in the terminal Market. We set the latter to be the market clearing price. This means the sellers expect to sell the Good for at least that price. The price of storing, C_1, may be chosen arbitrarily; however, it needs to be sufficiently high. If $|C_1|\frac{T}{2} < C_2$, it becomes beneficial for the sellers to keep the Good, and the selling price would thus be P. The buyers' utility is given in terms of the future expected utility for Money in the terminal Market. The relative penalty influences the mean utility a buyer obtains and again, it may be chosen arbitrarily. The future utility for Money is the utility for Good attainable with that Money, which is at least the utility for Good purchased at the maximum price P. C_3 should hence inversely depend on P.

4.2 Learning algorithm and network architecture

For training the agents' strategies we adopt an *actor-critic* algorithm called Twin-Delayed Deep Deterministic Policy Gradient (TD3) (Fujimoto, Hoof, and Meger 2018). We depict the pseudocode of the learning algorithm in Algorithm 1. The policy Π_t of each trader t is a random variable with a Gaussian distribution with mean and standard deviation represented by a neural network.

The architecture of neural networks we employ is shown in Figure 2. The buyers' actor needs to process the offers of the sellers and consecutively offer the Right for sale before processing the offers of other buyers. To accomplish that, the output of the first hidden layer is concatenated with the offers of the other buyers, and only the first hidden layer is used to predict the buyer's offer. In this way, the network can be used to obtain the buyer's offer without the offers of

[6] We are primarily interested in the case where buyers are hospitals. In such a scenario, it would not be difficult to obtain an accurate estimate of the funds and supply. The Rights assigned to each buyer are public information.

Algorithm 1: Equilibrium Learning Algorithm

```
$B \leftarrow$ set of buyers, $S \leftarrow$ set of sellers, $\mathcal{D} \leftarrow \{\}$
for $episode \in \{1, \dots N_{\text{sims}}\}$ do
    for $t \in \{1, \dots T\}$ do
        $G_S \leftarrow G_S + g$, $M_B \leftarrow M_B + m_B$
        $o_S, o_B \leftarrow$ observation of sellers/buyers
        $\pi_S \leftarrow \text{clip}(\Pi_S(o_S(s)), 0, 1)$
        $\overline{\pi}_B \leftarrow \text{clip}(\Pi_B(o_B(b), \pi_S), 0, 1)$
        $\pi_B \leftarrow \text{clip}(\Pi_B(o_B(b), \pi_S, \overline{\pi}_B), 0, 1)$
        Trade according to a market mechanism $\mu$
        Compute utilities $u_b, u_s$
        $\mathcal{D} \leftarrow \mathcal{D} \cup \{o_B, o_S, \pi_S, \pi_B, u_b, u_s\}$
        $G_B \leftarrow \max(G_B - d_B, 0)$
        if $t \bmod N_{\text{train}}$ is zero then
            Sample $batch \sim \mathcal{D}$
            Train on $batch$ using TD3
        end if
    end for
    Reset episode
end for
```

others influencing the result. Since the actions come from a bounded interval, the actors use a sigmoid activation function on the output layer on the means, which is then properly rescaled. The standard deviation uses the softplus activation.

Moreover, we enhance the vanilla TD3 algorithm with upgoing policy update (Vinyals et al. 2019) and reward clipping to $[-1, 1]$. To accelerate training, we also allow the sellers to share the same replay buffer $\mathcal{D}$. This makes the sellers' policies similar without using an identical actor.

5 Empirical Evaluation

Finally, we demonstrate the properties of our hybrid system with fairness and market mechanisms, and the effectiveness of our learning algorithm, on practical examples. First, we assess the quality of the learned solutions using NashConv, a measure of exploitability. In the second part, we analyze to which degree the incorporation of Rights affects how the Price of Anarchy of the approximated equilibrium evolves throughout the entire crisis. We evaluate systems combining three degrees of fairness with all four market mechanisms from Section 3. The variants of fairness we consider are systems with: (i) no distributed Rights (i.e., a free market), (ii) Rights and k=1 tradings; and (iii) Rights and k=2 tradings.

Experimental setting All experiments were conducted on a computational cluster with AMD EPYC 7532 CPUs running at 2.40GHz. We utilized only 5 of its 16 cores and 3GB of RAM. The code was implemented in Python using tensorflow 2.6, tensorflow-probability 0.15, mip 1.14, and numpy 1.21. The open-source CBC solver carried all LP computations. The complete list of all hyperparameters of Algorithm 1 can be found in Appendix C.

Experimental domain We consider a sequence of $T = 10$ Markets with four buyers and four sellers. We choose a prototypical setting where three of the four buyers receive significantly more funds then the last buyer. At the same time, this last buyer suffers a large demand, in most cases exceeding the demands of the others. We refer to the first three buyers as *rich* and to the last buyer as *poor*. We generate the instances of this setting by sampling the demands and the earnings of the buyers uniformly randomly from given intervals. For the rich buyers, the demand d_b in drawn from $\mathcal{U}(1, 2)$ and the earning $m_b \sim \mathcal{U}(4, 6)$. For a poor buyer, $d_b \sim \mathcal{U}(4, 6)$ and $m_b \sim \mathcal{U}(1, 2)$. The set of demands and earnings is then normalized such that $\mathbb{E}_{b \sim B}[d_b] = 1$ and $\mathbb{E}_{b \sim B}[m_b] = 1/8$. To fix a scale[7], we set the maximum price as $P = 1$. The constants in the utilities are then chosen as $C_1 = -1/8$, $C_2 = 1/2$ and $C_3 = 1/P = 1$.

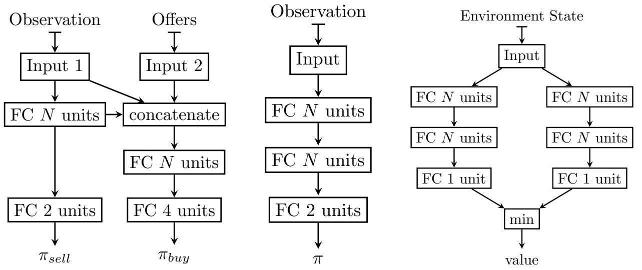

Figure 2: Architectures of the used neural networks: (Left) sellers' actor, (Middle) buyers' actor, and (Right) the critic.

5.1 Exploitability

We measure the quality of a candidate solution from episode t through its exploitability. For computing the exploitability we employ the notion of NashConv (Lanctot et al. 2017), given as $\sum_{i=1}^{T} \sum_{j=1}^{k} \sum_{\tau \in \mathcal{T}} u_\tau(\overline{\pi}|(i, j)) - u_\tau(\pi_t|(i, j))$. Here, $u_\tau(\pi_t|(i, j))$ denotes the utility of trader τ in market i and trade j under policy profile π_t. The policy profile $\overline{\pi}$ is then a profile of approximate best-responses to π_t. We train a best-response of each trader separately for 100 episodes, keeping the opponents' policies fixed, and starting from the policy of trader τ in π_t. Because obtaining the best-responses is immensely computationally demanding (the entire computation took about seven hours for each combination of the fairness mode and market mechanism), we chose one specific instance to assess the exploitability of, with earnings $m_b = (4/32, 5/32, 6/32, 1/32)$ and demands $d_b = (1/2, 1/2, 1/2, 5/2)$. In Figure 3 we present the results achieved with all four market mechanism in a system with Rights and $k = 1$. The results suggest the algorithm was able to reach a sufficiently close approximation of the equilibrium. Moreover, we verified the inclusion of Rights or the value of k do not have a significant effect on exploitability.

5.2 Price of Anarchy

In Figure 4 we depict the prices the society pays for distributing the critical Goods through a (regulated) market instead of centrally. All results are averaged over 10 instances and show also the standard errors. The top row compares

[7] This corresponds to choosing a currency such that the price of a unit of Good is at most one.

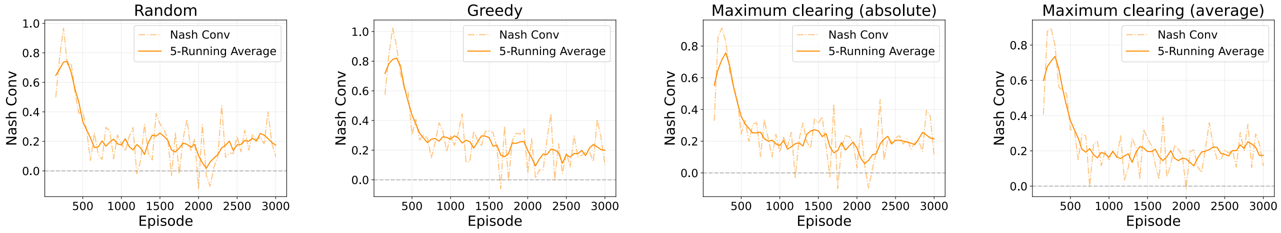

Figure 3: The exploitability of candidate solutions when learning the equilibrium in systems with Rights and k=1 trading period for four different market mechanisms. The dashed line shows the original values, the solid line highlights the 5-running mean.

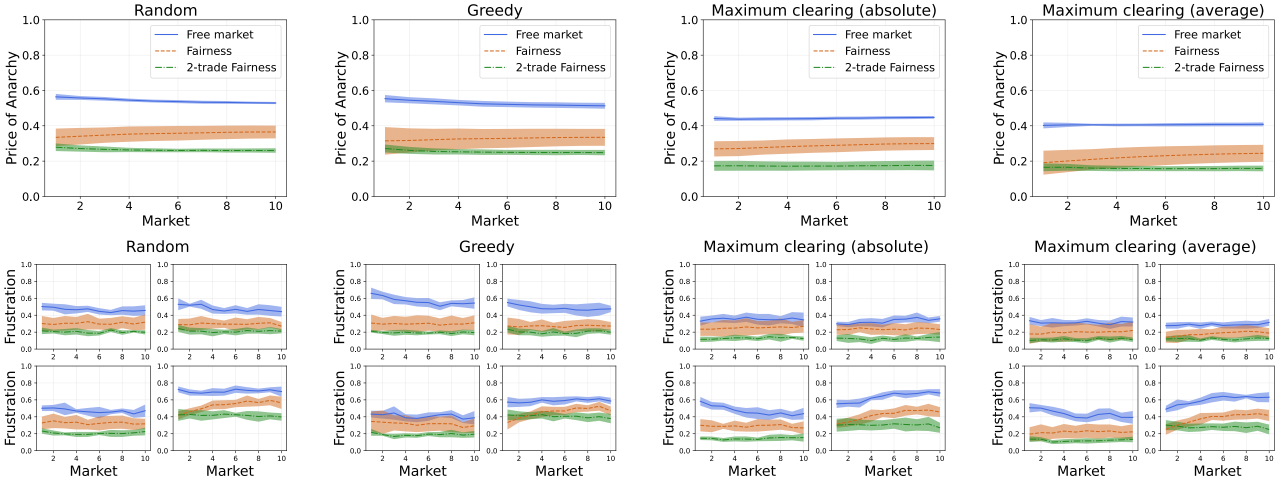

Figure 4: (Top) The Price of Anarchy and (Bottom) the individual frustrations of the four buyers in systems with three variants of fairness for four different market mechanisms. The poor buyer is always bottom right in the frustration graphs.

the Prices of Anarchy of systems with the three earlier described modes of fairness for the four introduced market mechanisms. Note that the PoA is always lower in the systems with Rights. Moreover, introducing a second trading period further decreases it. Another noteworthy observation is that maximum clearing allocations offer lower PoA than the other two, more basic mechanism.

The bottom row then shows the individual frustrations of the buyers. As expected, the poor buyer experiences the highest frustration. Otherwise the results observed with overall PoA clearly translate into the frustration of each buyer as well. Interestingly, the results suggest that introducing the fairness mechanism into the trading is beneficial not only for the poor buyer but for the rich buyers as well.

The computation of the approximated equilibrium over the period of 3000 episodes took about one hour for each instance, fairness mode, and market mechanism.

6 Conclusion

To the best of our knowledge, we are the first to introduce a system explicitly combining a double-sided market mechanism with a fairness mechanism allocating the buying rights for more socially just redistribution of critical goods during the times of need. We adopted the contested garment distribution as a baseline fair allocation and studied four separate market mechanism: random, greedy, absolute-prices maximum clearing, and average-prices maximum clearing. Our two main theoretical results show that the last two allocations can be computed in polynomial time. We then defined an analogue of Price of Anarchy (PoA) in our system as the sum of scaled differences between the amount of goods each trader was entitled to according to the fairness mechanism and the amount they were actually able to secure in the market, which we refer to as the individual frustrations. Furthermore, we developed a reinforcement-learning algorithm capable of approximating an equilibrium of the system in order to evaluate the PoA in practice. In the last part of our work, we show on a notorious example of a system with an underfunded and short-supplied buyer that introducing the buying rights may significantly decrease the frustrations, ergo, the PoA, especially for mechanisms prioritizing the amount of goods sold. Yet, it still remains an open question whether there exists a mechanism admitting zero PoA in the limit.

Future work We see two major limitations of our work. First, we focused on the full-blown crises and assumed a constant resupply of the goods over many trading periods. We would like to study more complex models akin to, e.g., the bullwhip effect. Second, we restricted our fairness model to the contested garment rule. Considering other models may change the system dynamics, and perhaps improve the PoA.

References

Akbarpour, M.; Dworczak, P.; and Kominers, S. D. 2020. Redistributive allocation mechanisms. *Available at SSRN 3609182*.

Aumann, R.; and Maschler, M. 1985. Game Theoretic Analysis of a Bankruptcy Problem from the Talmud. *Journal of Economic Theory 36, 195-213*.

Condorelli, D. 2013. Market and non-market mechanisms for the optimal allocation of scarce resources. *Games and Economic Behavior*, 82: 582–591.

Cripps, M. W.; and Swinkels, J. M. 2006. Efficiency of large double auctions. *Econometrica*, 74(1): 47–92.

Dworczak, P.; Kominers, S. D.; and Akbarpour, M. 2021. Redistribution through markets. *Econometrica*, 89(4): 1665–1698.

Fu, H.; Liu, W.; Wu, S.; Wang, Y.; Yang, T.; Li, K.; Xing, J.; Li, B.; Ma, B.; FU, Q.; and Wei, Y. 2022. Actor-Critic Policy Optimization in a Large-Scale Imperfect-Information Game. In *International Conference on Learning Representations*.

Fujimoto, S.; Hoof, H.; and Meger, D. 2018. Addressing function approximation error in actor-critic methods. 1587–1596.

Jedličková, A.; Loebl, M.; and Sychrovský, D. 2022. Critical Distribution System. *arXiv*, (2207.00898).

Lanctot, M.; Zambaldi, V.; Gruslys, A.; Lazaridou, A.; Tuyls, K.; Pérolat, J.; Silver, D.; and Graepel, T. 2017. A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning. *arXiv*, (1711.00832).

Liu, S.; Marris, L.; Hennes, D.; Merel, J.; Heess, N.; and Graepel, T. 2022. NeuPL: Neural Population Learning. *arXiv preprint arXiv:2202.07415*.

Moroney, J. R.; and Lovell, C. 1997. The relative efficiencies of market and planned economies. *Southern economic journal*, 1084–1093.

Muller, P.; Omidshafiei, S.; Rowland, M.; Tuyls, K.; Pérolat, J.; Liu, S.; Hennes, D.; Marris, L.; Lanctot, M.; Hughes, E.; Wang, Z.; Lever, G.; Heess, N.; Graepel, T.; and Munos, R. 2019. A Generalized Training Approach for Multiagent Learning. *CoRR*, abs/1909.12823.

Perolat, J.; de Vylder, B.; Hennes, D.; Tarassov, E.; Strub, F.; de Boer, V.; Muller, P.; Connor, J. T.; Burch, N.; Anthony, T.; et al. 2022. Mastering the Game of Stratego with Model-Free Multiagent Reinforcement Learning. *arXiv e-prints*, arXiv–2206.

Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; Oh, J.; Horgan, D.; Kroiss, M.; Danihelka, I.; Huang, A.; Sifre, L.; Cai, T.; Agapiou, J. P.; Jaderberg, M.; Vezhnevets, A. S.; Leblond, R.; Pohlen, T.; Dalibard, V.; Budden, D.; Sulsky, Y.; Molloy, J.; Paine, T. L.; Gulcehre, C.; Wang, Z.; Pfaff, T.; Wu, Y.; Ring, R.; Yogatama, D.; Wünsch, D.; McKinney, K.; Smith, O.; Schaul, T.; Lillicrap, T. P.; Kavukcuoglu, K.; Hassabis, D.; Apps, C.; and Silver, D. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 1–5.

A Proof of Theorem 1

Theorem 1. *Maximum clearing allocation can be found efficiently using a reduction to the Max Flow problem. As a consequence, a Maximum clearing allocation is polynomial for both divisible and indivisible Good.*

Proof. Given the disjoint copies of the graphs G_G, G_R, we construct an instance of the Max Flow problem as follows:

1. introduce two new vertices s, t;
2. join s by an arc (s, v) to each vertex v of B in G_G. Let the capacity $cap(s, v)$ of this arc be equal to the amount of the *remaining rights* of v, i.e., the assigned amount minus the amount intended to be sold. Clearly, each buyer b desires to buy at least $cap(s, v)$ of Good.
3. join s by an arc (s, v) to each vertex v of B_S in G_R. Let the capacity of this arc be equal to $w_R(v)$, i.e., the amount (possibly zero) of Right v intends to sell;
4. orient each edge of G_R towards B_B, the capacity of (x, y) being equal to $w_R(y)$, i.e., the amount of Right y intends to buy;
5. orient each edge of G_G towards S, the capacity of (x, y) being equal to $w_G(y)$, i.e., the amount of Good y intends to sell;
6. introduce a copy B' of B and join each vertex $v \in B_B$ of G_R by an arc (v, v') to its copy $v' \in B'$, its capacity being $w_R(v)$, i.e., the amount of Right y intends to buy;
7. join each $v' \in B'$ to S in the same way as its copy v is joined to S in G_G, orient these new edges towards S and let the capacity of each such arc terminating in $y \in S$ be $w_G(y)$, i.e., the amount of Good y intends to sell;
8. join each vertex y of S to t by the arc (y, t), its capacity being $w_G(y)$, i.e., the amount of Good y intends to sell.

This finishes the construction of the instance of the Max Flow problem. It is straightforward to see that max flow from s to t provides a clearing of bids with the maximum amount of the Good sold. Also, it is ensured that Right is bought along with the same amount of Goods. □

B Proof of Theorem 2

Theorem 2. *Maximum clearing allocation with average bids can be found efficiently using a linear program.*

Proof. We can find the maximum clearing allocation using the following linear program, where the variable $r_{b,b'}$ represents the amount of Right sold to b' by b, with $(b, b') \in E_R$ and the variable $g_{s,b}$ represents the amount of Good sold to b by s, with $(s, b) \in E_G$. We also introduce the variables m and M, representing the minimal, resp maximal, amount of goods bought by a buyer. Furthermore, we define c as $c = \epsilon * U$ where ϵ is the desired sensibility of the objective function and U an upper bound on $(M - m)$: $U = \max_{b \in B} \left(\min(d_b, \sum_{(s,b) \in E_G} w_G(s)) \right)$; c will be used to normalize $(M - m)$ in order to not interfere with the rest of the objective function. In our experiments, we used $c = \frac{1}{1000}$.

$$
\begin{array}{lll}
\max \sum_{(s,b)\in E_G} g_{s,b} - c(M-m) & & (1) \\
\text{s.t.} & & \\
\sum_{(s,b)\in E_G} g_{s,b} \leq r_b + \sum_{(b',b)\in E_R} r_{b',b} - v_b^R & \forall b \in B & (2) \\
\sum_{(s,b)\in E_G} g_{s,b} \leq \overline{v}_b^G & \forall b \in B & (3) \\
\sum_{(s,b)\in E_G} g_{s,b} \leq v_s^G & \forall s \in S & (4) \\
\sum_{(b,b')\in E_R} r_{b,b'} \leq v_b^R & \forall b \in B & (5) \\
m \leq \sum_{s\in S} g_{s,b} & \forall b \in B & (6) \\
M \geq \sum_{s\in S} g_{s,b} & \forall b \in B & (7) \\
\sum_{(s,b)\in E_G} g_{s,b} * p_s^G \leq \overline{p}_b^G \sum_{(s,b)\in E_G} g_{s,b} & \forall b \in B & (8) \\
\sum_{(b',b)\in E_R} p_{b'}^R * r_{b',b} \leq \overline{p}_b^R \sum_{(b',b)\in E_R} r_{b',b} & \forall b \in B & (9) \\
\sum_{(s,b)\in E_G} p_s^G g_{s,b} + \sum_{(b',b)\in E_R} p_{b'}^R r_{b',b} \leq M_b & \forall b \in B & (10) \\
g_{s,b} \geq 0 & \forall (s,b) \in E_G & (12) \\
r_{t,b} \geq 0 & \forall (t,b) \in E_R & (13)
\end{array}
$$

In this linear program, the objective function (1) maximizes the exchanges of goods, and spreads the distribution over the buyers. The constraints (2) and (3) then enforce that the buyers buy less good than they have rights, and the amount of good they buy does not exceed their demand v_b^G. The constraint (4) imposes a restriction on the amount of good the sellers may sell, ensuring it is at most v_s^G, i.e., the amount they committed themselves to be willing to sell. Similarly, the constraint (5) imposes that the buyers selling good sell at most the amount they intend to sell v_b^R. The constraint (6), resp (7), assures that m is lower, resp. higher, than the minimal, resp. maximal, amount of good bought by a buyer, and the sense of the objective function ensure that it will be exactly this quantity. The constraint (8) imposes that the buyers pay at most in average p_b^g for the goods. The constraint (9) forces that the buyers pay at most in average p_b^r for the rights. The constraint (10) is the budget constraint. □

C Hyperparameters

The experiments used the following values of parameters:

Actor learning rate	$3 \cdot 10^{-4}$
Critic learning rate	10^{-3}
Actor hidden layer size	32
Critic hidden layer size	256
Batch size	512
L2 penalty	10^{-2}
Discount factor	0.99
Target network update rate	0.002
Actor training frequency	3
Entropy penalty	$3 \cdot 10^{-3}$
Training episodes	3000
NashConv training episodes	100