SN Computer Science (2021) 2:272
https://doi.org/10.1007/s42979-021-00650-4

ORIGINAL RESEARCH

A Faithful Mechanism for Incremental Multi-Agent Agreement Problems with Self-Interested and Privacy-Preserving Agents

Farzaneh Farhadi[1] · Nicholas R. Jennings[1]

Received: 1 September 2020 / Accepted: 16 April 2021 / Published online: 11 May 2021

Abstract
Distributed multi-agent agreement problems (MAPs) are central to many multi-agent systems. However, to date, the issues associated with encounters between self-interested and privacy-preserving agents have received limited attention. Given this, we develop the first distributed negotiation mechanism that enables self-interested agents to reach a socially desirable agreement with limited information leakage. The agents' optimal negotiation strategies in this mechanism are investigated. Specifically, we propose a reinforcement learning-based approach to train agents to learn their optimal strategies in the proposed mechanism. Also, a heuristic algorithm is designed to find close-to-optimal negotiation strategies with reduced computational costs. We demonstrate the effectiveness and strength of our proposed mechanism through both game theoretical and numerical analysis. We prove theoretically that the proposed mechanism is budget balanced and motivates the agents to participate and follow the rules faithfully. The experimental results confirm that the proposed mechanism significantly outperforms the current state of the art, by increasing the social-welfare and decreasing the privacy leakage.

Keywords Multi-agent agreement problem · Incentive mechanism design · Faithful distributed mechanism · Privacy awareness · Reinforcement learning

Introduction

In a multi-agent agreement problem (MAP) multiple agents have to make an agreement, yet they may have different preferences for different possible outcomes. Thus, it is of central importance to be able to aggregate the preferences so as to make a socially desirable agreement. In this paper, we focus on a class of MAP where the following two conditions are present: (1) Agents have private information about their preferences and may be reluctant to share them; (2) Decisions are incremental in that the needs for new agreements arise over time and the new decisions must be consistent with the old ones. This class is known as Private Incremental multi-agent agreement problems (piMAPs) [62] and is focused on many real-world applications including resource allocation [9], distributed scheduling [71], electronic commerce [86] and logistics [79].

To solve piMAPs, agents need to negotiate with each other until an agreement is reached or until they find out that there is no feasible solution to the problem. However, the assumption of private information places limits on the information that the agents may be willing to exchange in the negotiation. Specifically, privacy-preserving agents face a tradeoff between the amount of information they reveal and their desire to reach an acceptable agreement [31, 48]. For example, exchanging no information minimizes the amount of information revealed but is unlikely to lead to an agreement, whereas all agents revealing all their constraints maximizes the chance of finding an optimal agreement but at the cost of all privacy.

This paper is a significantly extended version of our recent conference paper [26]. The work is supported and funded by Samsung Electronics R&D Institute UK (SRUK).

"This article is part of the topical collection "Advances in Multi-Agent Systems Research: EUMAS 2020 Extended Selected Papers" guest edited by Nick Bassiliades and Georgios Chalkiadakis".

✉ Farzaneh Farhadi
f.farhadi@imperial.ac.uk

Nicholas R. Jennings
n.jennings@imperial.ac.uk

[1] Department of Computing, Imperial College London, London, UK

The problem of designing a negotiation mechanism that strikes a balance between outcome efficiency and privacy becomes particularly challenging when the agents are self-interested and are willing to better their outcome by manipulating the protocol. Self-interested agents, when trusted to perform an action in a mechanism, may betray that trust by not performing the action as required. For example, they may provide false information during the negotiation or may not pass an agent's message on to another if these deviations increase their individual welfare. When dealing with self-interested agents, the mechanism has to provide appropriate incentives to the agents to make sure that they cannot profit from manipulating the mechanism.

In the standard incentive mechanism design setting [7], there exists a trusted central authority that first solicits all agents' private information and then constructs a solution for the problem that maximizes the global objective function. In this setting, the goal is to design an incentive-compatible mechanism that induces agents to reveal their information truthfully. Incentive compatibility is a required feature for mechanisms in which information is distributed but the algorithm is executed centrally. However, in a distributed setting like piMAP, where both information and algorithm execution are distributed, a stronger notion of obedience is required. This leads to the concept of faithfulness for distributed mechanisms [83]. A mechanism is faithful if agents have incentives to obey all the mechanism rules, including truthful information revelation, honest computation, and faithful algorithm execution.

A common approach for designing either incentive-compatible or faithful mechanisms is to impose economic incentives on self-interested agents by using monetary payments [15, 76]. The use of money, however, is prohibited or unnatural in many real-life agreement problems, including political decision making or allocating "public" goods like school admissions [35, 36]. The lack of monetary payments as a means to incentivize agents makes mechanism design in such settings more challenging. However, recently, a class of simple non-monetary mechanisms, namely mechanisms based on artificial currencies [22], has gained considerable attention due to their successful use in practice [10, 77, 78]. Such mechanisms involve endowing agents with a budget of an artificial currency and then organizing a monetary mechanism with payments in terms of artificial currencies. The main challenge of designing such mechanisms is that an artificial currency has no independent valuation outside the setting of the mechanism; therefore, the mechanism must define a usage for the currency to make it valuable to the agents. The main goal of this paper was to design a faithful distributed incentive mechanism for piMAP based on an artificial currency.

In addition to faithfulness, we would like our mechanism to satisfy voluntary participation [7] and individual budget balance. The former is a standard requirement for an incentive mechanism which requires that the participation in the mechanism results in at least the same expected utility as not participating overall. The latter is an extended version of budget balance that requires each agent's expected payment to be zero at all on- and off-equilibrium paths. The standard notion of budget balance is not appropriate for distributed mechanisms without a central authority (or bank), as in these mechanisms, the payments are always made within the network from one agent to another; hence, the budget balance is always satisfied at the network level. In this paper, we would like to design a mechanism that keeps all the individuals' budgets balanced. We term this extended version of budget balance *individual budget balance (IBB)*. Imposing individual budget balance motivates agents to put all their effort on achieving a good agreement and not on collecting reward. (For more details, see "Individual rationality and individual budget balance".)

Against this background, we propose the first faithful distributed incentive mechanism for piMAPs that satisfies voluntary participation and individual budget balance. Our mechanism is based on the *score-voting* idea which has been used in the literature for designing centralized incentive mechanisms [56] (see "Related literature" for more details). Specially, we design a distributed score-based multi-round (DSM) negotiation mechanism in which at each round, one agent, called the initiator, offers a set of possible agreements to its neighbors, called responders, and asks them to score the offers. The initiator evaluates the offers based on the received scores and makes a decision. To motivate responders to score offers truthfully, the initiator promises some future rewards in terms of an artificial currency. Rewards are determined based on the scores the responders give to the offers and designed so as to incentivize them to give true scores. The second role of the rewards is to drive the initiator away from its selfish behaviour and motivate it to make socially optimal decisions based on the received scores. We guarantee faithfulness of the mechanism by setting non-manipulable rules. To test the effectiveness of the proposed mechanism, several benchmarks are adopted to demonstrate the performance in terms of social-welfare, privacy loss and convergence speed, where the privacy loss is measured by a novel privacy metric introduced in the paper. The results show that the minimum and maximum number of agreements that the agents are allowed to discuss at each round are control parameters that balance the tradeoff between these performance metrics.

For the proposed mechanism, we determine the optimal strategies of both the initiator and the responders at each round. To find the initiator's optimal offering strategy, we formulate the problem as a Markov Decision Process (MDP). To solve the MDP without requiring complete information about the expected future utilities, we devise a reinforcement learning algorithm that allows the initiator to learn the expected utilities and eventually its optimal offering strategy, in a trial-and-error fashion. We also provide a heuristic offering strategy that eliminates the need for learning and hence significantly reduces the time-complexity. We show by numerical simulations that this heuristic strategy performs very closely to the optimal policy.

The rest of the paper is organized as follows: After a review of the main relevant literature ("Related literature"), a description of the general model of the multi-agent agreement problem is given in "Strategic multi-agent negotiation over a communication graph". In "Privacy leakage", we present a metric to quantify the privacy leakage in a negotiation mechanism. In "The distributed score-based multi-round negotiation mechanism", we formulate piMAP as a mechanism design problem and introduce a novel distributed incentive mechanism to solve it. We establish the properties of the proposed mechanism in "Properties of the mechanism". In "Initiator's optimal offering strategy", we study the initiator's behavior in the mechanism and present both an optimal and a heuristic, but less complex, strategy to maximize its utility. In "Numerical results", we evaluate our proposed mechanism by simulations compared to several benchmarks. We conclude our paper in "Conclusions and future work". Short proofs are in the main text, while more technical proofs are deferred to the Appendix.

Related Literature

Multi-agent agreement problems are a special form of the well known distributed constraint optimization problem (DCOP) [61] for modeling multi-agent coordination tasks. DCOPs assume that a set of decision variables are distributed among a set of agents and constraints among the variables require agents to coordinate their decisions. As a strict subset of DCOP intended to model "agreement", MAPs require that constraints between variables belonging to different agents are limited to equality constraints.

Several distributed algorithms designed originally for general DCOPs (and hence applicable to MAPs) currently exist. Most of these algorithms have been proposed to solve classical DCOPs where all decision variables and constraints are known a priori and agents are *fully cooperative*. These algorithms can be classified as being either exact or non-exact, based on whether they can guarantee to return the optimal solution or they trade optimality for shorter running times, producing near-optimal solutions. Some representative examples of the first group are SyncBB [41], ADOPT [61], DPOP [74], AFB [34], BnB-ADOPT [96], and PT-FB [52]. Some recent examples of the non-exact algorithms are GDBA [67], BMS [80], BnB-FMS [53], Max-Sum-ADVP [102], D-Gibbs [64], ACO-DCOP [13], and AED [55].

All the above-mentioned algorithms are offline meaning that full information about the variables and constraints must be available in advance and then the decisions about all variables are made simultaneously. However, piMAP is incremental with some new decision variables and/or constraints being introduced over time. There are a number of algorithms that handle such dynamic DCOP problems [42, 75, 89, 97]. Most of this work responds to the changes of the problem by resolving the DCOP every time such changes occur [75, 89, 97]. Such algorithms are not suitable for incremental MAPs in which the previous agreements should not change when the need for a new agreement arises.

There is also a strand of work focusing specifically on MAPs and designing more specialized (and potentially more efficient) algorithms for agreement problems [2, 8, 16, 18–20, 47, 100, 104]. These MAPs are often studied in the domain of meeting scheduling where the attendees of each meeting must agree on a time and/or a place for the meeting [15, 104]. Some research efforts tackle meeting scheduling MAPs as offline problem [8, 18, 47], but the most relevant to our work are those approaches that study incremental MAPs [2, 16, 19].

The main and most well-known technique that is used in the literature for reaching an agreement in incremental settings is negotiation [11, 14, 44, 62, 72, 82]. For instance, in [14] the authors present a negotiation mechanism that helps agents to make incremental agreements efficiently. However, most of the available negotiation mechanisms are designed for cooperative environments where the agents obey the rules without questioning or challenging them [11, 14, 44, 62, 82]. In competitive environments where agents are self-interested, each negotiation mechanism induces a game among agents [50, 81]. A number of papers use game theoretic techniques to study the interactions of self-interested agents in a negotiation game and to determine the outcome [5, 72]. Another class of works use mechanism design theory to design negotiation mechanisms so as to achieve desired outcomes [15, 98].

There is a long tradition of using mechanism design techniques to manage distributed systems [15, 25, 27, 69]. Most of this work concentrates on problems where the information is distributed but the mechanism is executed centrally by a trusted entity. In such settings, voting mechanisms are widely used to facilitate agreement [33]. In [3, 56], the authors present two sufficient conditions for a non-monetary voting mechanism, namely *neutrality* and *elementary monotonicity* that guarantee its incentive compatibility. Neutrality means that every voting alternative is treated in the same way, i.e., the "names" of the alternatives do not matter [56]. Elementary monotonicity requires that if an alternative a is the outcome at a particular voting mechanism, then it is also the outcome of the vote where a single voter increases its vote to a. Scoring correspondences are a class of voting mechanisms that satisfy neutrality and elementary monotonicity and hence are incentive compatible [56]. In a scoring correspondence, the central authority asks each self-interested agent to give a score to each possible option. Then, it selects the option with the maximum aggregate score. The set of feasible scores that the agents are allowed to give to the options is determined by the central authority and distinguishes the different scoring correspondences.

The voting mechanisms, as well as the other incentive mechanisms presented in [15, 25, 27, 69], presume the existence of a central trusted entity. The first steps in providing a decentralized incentive mechanism were presented in [28, 29]. In these mechanisms, both information and algorithm execution are distributed; however, the agents are assumed to have strategic behaviors only for information revelation, and not for mechanism execution. Starting from [70], a few research works have attempted to design faithful mechanisms that are completely robust to manipulation [59, 60, 73, 76, 83, 90, 91]. In [83], the authors introduce a general decomposition technique that splits a distributed algorithm into disjoint phases, each of which are provably robust against rational manipulation. This decomposition technique is powerful because it can allow an exponential reduction in the number of joint manipulation actions that must be checked in a faithfulness proof. In [76], the authors integrate the DPOP algorithm with the Vickrey–Clarke–Groves (VCG) mechanism and introduce the first DCOP algorithm, named as M-DPOP, that provides a faithful distributed implementation for efficient social choice. The works of [90] and [91] provide two other VCG-based mechanisms for faithful implementation of dual decomposition and average consensus algorithms. The idea of faithful distributed implementation has been applied to real-world problems such as smart grid [60], electricity pricing [59], and wireless spectrum auctions [73]. However, in all available works in this area, agents are assumed to be privacy-neutral and hence have no objection to sharing their private information with others. Therefore, these mechanisms are not appropriate for piMAP where agents are privacy-preserving and seek to reach an agreement with minimum privacy leakage.

Privacy is recognized as a key motivating factor in the design of several multi-agent algorithms, and researchers have introduced several types of privacy concerns that agents may have during a multi-agent encounter [24, 37, 39, 40, 46, 48, 49, 88, 99]. In [46, 88], privacy is considered as the negotiation subject, however, the privacy loss caused by the exchange of messages during the algorithm is neglected. In [49], four notions of privacy are introduced to capture the agents' privacy attitudes in the message exchange part of a DCOP algorithm: agent privacy where agents hide their identities, topology privacy where agents hide the topological structure of the constraint graph, constraint privacy where agents hide the constraints from the ones who are not involved, and decision privacy where agents hide their final decisions (see [24, 49] for more details). Out of these four privacy notions, constraint privacy, which has drawn the most attention in past research [21, 54, 85], is most relevant to our work. This is because in piMAPs, the information that the agents want to keep confidential is about their preferences and hence the constraints they must respect to satisfy their goals.

One approach to guarantee constraint privacy has been to use cryptographic techniques [99], but the required use of multiple external servers may not always be desirable, available or justifiable for its benefit. Instead, a second approach has attracted significant attention, where researchers provide metrics for measuring the extent of constraint privacy loss in multi-agent algorithms [30, 57, 84]. The most general work in this regard is [54] which proposed a unifying quantitative framework, known as the Valuation of Possible States (VPS), for quantifying the privacy loss in a variety of multi-agent algorithms. This framework led to entropy-based [8], proportional [54], and state-guessing metrics [38] that have been widely used in the literature. However, these metrics are unable to model the agents' different and even contradictory attitudes towards information sharing with different agents. One of our contributions in this paper is to introduce a privacy metric that resolves this drawback (see "Privacy leakage").

As discussed above, both privacy issues and the selfish behavior of agents have been studied in a wide range of works, separately. However, there are only a few works that deal with agents that are both privacy-aware and self-interested [1, 43, 51, 65, 66, 101]. In [43, 51, 65, 66, 101] the authors focus on offline problems, while the problem studied in [1] is online. These works often consider differential privacy as their privacy notion and try to design incentive mechanisms such that the impact of a single agent's revealed information on the final outcome is small. This notion is suitable for centralized settings when the agents have complete trust in a central entity, but not in other agents. In such settings, the privacy-sensitive agents would like to make sure that announcing the final decision by the central entity does not disclose their private information. This notion is not applicable for settings without a central authority. Therefore, the tools developed for designing differentially-private incentive mechanisms cannot be used to design our distributed incentive mechanism.

The mechanism designed in our paper can be viewed as a negotiation mechanism with an artificial or virtual currency. Virtual currencies were first introduced in online games and social networks as a means to buy and sell virtual goods without making use of real money, thus avoiding security issues, taxation, and mistrust [6]. Then, the idea has been adopted in non-cooperative networks to provide non-monetary incentive mechanisms. The representative examples of virtual currency systems have been used in non-cooperative systems are Bitcoin [63], Nuglets [12], and WhoPay [95]. The idea of incentivizing agents to make desirable actions by granting them some rewards in terms of an artificial currency has been applied to a few negotiation mechanisms [4]. However, all this work is for centralized settings. Our paper is distinct from this literature in that our proposed mechanism is designed for distributed settings with no central authority.

Strategic Multi-Agent Negotiation Over a Communication Graph

We formalize the private incremental multi-agent agreement problem (piMAP) with self-interested agents as a network in which at each instant of time, a selfish agent may need to make an agreement with some of its neighbors. Such agents negotiate with their neighbors to reach socially-acceptable agreements. The agents' preferences over different agreements, which are their own private information, determine their strategies in the negotiation.

We model a multi-agent system by an undirected graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where the nodes $\mathcal{V}=\{1,2,\dots,V\}$ represent the agents, and the edges $\mathcal{E}\subseteq\{\{i,j\}|i,j\in\mathcal{V}, i\neq j\}$ represent reciprocal relationships among agents. An edge $\epsilon_{ij}=\{i,j\}\in\mathcal{E}$ implies that agents i and j are socially-connected to each other and may need to enter into a joint agreement. Social connections could include family, friends, co-workers and neighbors. We define the set of agent i's connections as $N_i=\{j\in\mathcal{V}|\{i,j\}\in\mathcal{E}\}$.

At some point t in time, an agent $I(t)\in\mathcal{V}$ may need to make an agreement with a group $R_I(t)\subseteq N_{I(t)}$ of its connections. The subject of the agreement could be anything such as the time and/or the place of a meeting, a schedule for using a shared resource, or the allocation of a task. When the need for an agreement arises, agent $I(t)$ initiates a negotiation process $\mathcal{N}(t)$ and contacts other agents $R_I(t)$. In this negotiation, agent $I(t)$ is called the initiator and other agents $R_I(t)$ are called the responders. The set of all participants of the negotiation $\mathcal{N}(t)$ is denoted by $A(t)=R_I(t)\cup\{I(t)\}$. We denote the set of all possible outcomes of the negotiation $\mathcal{N}(t)$ by $\mathcal{O}(t)=\{O_{t,0},O_{t,1},\dots,O_{t,o(t)}\}$, where $O_{t,0}$ represents the disagreement outcome and $o(t)$ represents the number of possible agreements at $\mathcal{N}(t)$.

The agents are assumed to be self-interested, meaning they have some preferences over the outcomes and attend to their desires without any regard to the preferences of others. Each agent $i\in A(t)$ has a valuation function

$$V_i(t):\mathcal{O}(t)\to[\underline{V},\bar{V}]\cup\{-\infty\},\tag{1}$$

that assigns a value to each possible outcome of the negotiation $\mathcal{N}(t)$. A valuation $V_i(O_{t,k})$ represents the value of agreeing on outcome $O_{t,k}$ for agent i, where $V_i(O_{t,k})=-\infty$ means that the outcome $O_{t,k}$ does not satisfy agent i's constraints and hence is infeasible for it. Any outcome $O_{t,k}$ with $V_i(O_{t,k})\neq-\infty$ is defined to be feasible for agent i. We denote the minimum and maximum values of feasible outcomes by $\underline{V}$ and $\bar{V}$, respectively.

Reaching an agreement is the agents' main and uncompromisable goal in an agreement problem. Therefore, directing the negotiation according to their own preferences is valuable for the agents as long as it does not prevent the agreement from being reached. This phenomenon can be clearly seen in the meeting scheduling application. In the meeting scheduling, all agents' first priority is to set a meeting. Therefore, the agents are reluctant to prevent a meeting from being scheduled under the pretext of their individual preferences. We capture this fact by the following assumption.

Assumption 1 *Disagreement has an infinite cost for the agents* [1]

$$V_i(O_{t,0})=-\infty,\forall i,t.\tag{2}$$

Agents are not aware of the other agents' valuation functions $V_i(.)$. However, they have a common belief about the strictness of each agent's feasibility constraints. We define each agent i's strictness coefficient $d_i\in[0,1]$ as the probability by which an arbitrary outcome is infeasible for agent i, i.e. $d_i=\mathbb{P}\big(V_i(O_{t,k})=-\infty\big)$. Agents with higher strictness coefficients are more strict in their preferences and hence fewer agreements can satisfy their requirements. In a meeting scheduling application, for example, the strictness coefficient is equivalent to the agent's calendar density which is the proportion of busy hours in its calendar. During a negotiation, agents can build an estimation of their neighbors' strictness coefficients based on the messages they send. As time goes on, the estimations are likely to become more accurate and approach the actual d_is. In this paper, we assume that each agent has a rich history of interactions with its neighbors and hence has an accurate estimation of their strictness coefficients.[2]

In addition to the final outcome, the agents are also concerned about the following two matters:

[1] There are other approaches in the literature that map the disagreement outcome to a finite negative value [62]. Designing a faithful and privacy-preserving incentive mechanism under such setting is a potential area for further investigation.

[2] In "Initiator's optimal offering strategy", we will discuss what will happen if this information is not available.

(a) the speed of convergence; and
(b) the amount of information they share during the negotiation process.

Speed of convergence refers to the number of rounds Z needed until the negotiation process converges towards an outcome. Agents prefer negotiations that take fewer rounds to finish, as they require agents to use less communication resources. We denote the value agent i assigns to the communication resources it utilizes in each round of the negotiation by a bargaining cost $\beta_i \geq 0$. Therefore, the cost incurred by agent i when it participates in Z rounds of negotiation is $\beta_i Z$.

The agents are privacy-aware and prefer not to share their private information with others. We denote agent i's privacy sensitivity when it shares its information with agent j by $\theta_{ij} \geq 0$, where a higher θ_{ij} means that agent i is less keen to reveal its information to agent j. The amount of agent i's information that is leaked by sending a set of messages $M_{i\to j}$ to agent j in negotiation $\mathcal{N}(t)$ is captured by a leakage function $L_{i,j}(M_{i\to j})$. We will detail how to design the leakage function in "Privacy leakage".

Based on the discussion above, we model agent i's utility from the negotiation $\mathcal{N}(t)$ as follows:

$$U_i^t(O_{t,k}, Z, \{M_{i\to j}\}_{j\in A(t)}) = V_i(O_{t,k}) - \beta_i Z - \sum_{\substack{j \in A(t)\\ j \neq i}} \theta_{ij} L_{i,j}(M_{i\to j}). \tag{3}$$

When agent i participates in negotiation $\mathcal{N}(t)$, it cares about the utility U_i^t it receives in $\mathcal{N}(t)$ as well as the expected utilities it can collect in future negotiations. At each time t, agent i has a belief μ_i^t about the number of future negotiations it might need to participate. Therefore, it chooses its strategy at time t so as to maximize its expected total utility from time t onward. Before modeling the agents' expected total utilities mathematically, let's first introduce the following assumption and discuss how to handle the issues caused by it.

Assumption 2 *Each agent i's personal parameters, including its value function $V_i(.)$, its bargaining cost β_i, its privacy sensitivities θ_{ij}, for $j \neq i$, and its belief function μ_i^t, for all t, are its own private information and cannot be observed by any other agent. These private valuations will be addressed as agents' types.*

According to (3) and Assumption 2, agents' preferences over different negotiation strategies are determined based on their types which are unknown to others. Therefore, to assure that all selfish agents follow their required actions in the negotiation, including information-revelation, computation and message passing, a faithful incentive mechanism must be designed so as to provide appropriate incentives to agents of any type to follow their prescribed strategies.

Due to practical concerns discussed in "Introduction", we restrict attention to non-monetary incentive mechanisms with an artificial currency [22]. For ease of reference, in this paper, we call this type of currency "the convenience point". In this class of mechanisms, the agents receive rewards in terms of convenience points, based on their level of cooperation. For example, the rewards granted to an agent i could be an increasing function of the number of other agents' proposals that are accepted by agent i during the negotiation. The mechanism must define a usage for convenience points to make them valuable for the agents. This goal is often achieved by making some desirable actions that agents can do in the negotiation, such as making or rejecting a proposal, costly.

The convenience point is the link between the agents' current decisions and the utility they can gain in the future. The more cooperatively an agent i behaves in the current negotiation, the more convenience points and hence more negotiation power it will have in the future. To capture this phenomenon, we model the expected total utility from time t onward for an agent i with budget of $b_i(t)$ convenience points at time t as

$$\mathbb{E}[U_i^{t:\infty}|b_i(t)] = \mathbb{E}[U_i^t] + \sum_k \mu_i^t(k)\mathbb{E}[U_i^{t+1:\infty}|k \text{ negotiations}, b_i(t+1)], \tag{4}$$

where U_i^t is agent i's instant profit at negotiation $\mathcal{N}(t)$ derived by (3), $\mu_i^t(k)$ is agent i's belief for participating in k negotiations after time t, and $\mathbb{E}[U_i^{t+1:\infty}|k \text{ negotiations}, b_i(t+1)]$ is agent i's expected utility from time $t+1$ onwards if it participates in k negotiations with the initial budget of $b_i(t+1)$. The function $\mathbb{E}[U_i^{t+1:\infty}|k \text{ negotiations}, b_i(t+1)]$ is increasing in terms of the budget $b_i(t+1)$, as an agent with a higher budget can influence the negotiations more significantly and hence better manipulate the process to its own advantage.

Each agent i is a long-run optimizer and hence chooses its negotiation strategy so as to maximize (4). Based on (4), there is a tradeoff between the agent's current-negotiation utility and the utility it can derive from future negotiations. To increase the total profit, agent i can spend more points and influence the current negotiation to its own advantage, however, doing this reduces its remaining points for the future negotiations and hence reduces its future utilities. The tipping point of this tradeoff at any time t depends on (i) the budget $b_i(t)$ of convenience points that is available to agent i at time t, and (ii) the belief function μ_i^t agent i has about the number of future negotiations. An agent with a low budget that believes that it should participate in many negotiations in the future might prioritize collecting points over reaching a desirable agreement in the current negotiation. However, an agent with a large number of points in its pocket or one that believes that it is participating in its last negotiation, might be not worried about spending its points to drive the negotiation to its most preferable agreement.

Our goal in this paper is to design an incentive mechanism based on convenience points to motivate agents with any budget and any belief to act faithfully in the negotiations. Designing the reward function, that determines how the points are distributed among the agents, and the cost function, that determines how the points can be used by the agents to influence the outcomes, are our main tools to achieve this goal. We discuss our design method thoroughly in "The distributed score-based multi-round negotiation mechanism". Before going to "The distributed score-based multi-round negotiation mechanism", however, in the next section, we complete the definition of agents' utility function by introducing a privacy leakage function $L(.)$ that captures the agents' willingness to hide their private information.

Privacy Leakage

As discussed in "Strategic multi-agent negotiation over a communication graph", the agents are privacy-aware and incur some cost from the leakage of their private information to others. For instance, in a meeting scheduling application, the agents might have some private events in their calendars that they do not want to share with others. For example, an employee may need to go out of the office on Thursday morning for a personal matter and does not want their employer to know this. Therefore, in a negotiation to set a work meeting, they prefer not to announce their unavailability on Thursday morning. Similarly, in negotiations for solving task allocation problems an agent may want to hide its inability to do a particular task from the others.

The above examples show that in many real-life negotiations, the agents consider the feasibility or infeasibility of different possible agreements for them as private and would like to hide this information from others. In the subsection below, we develop a privacy metric that measures the agents' privacy loss in a negotiation and then discuss the main features of this metric in "Properties of the privacy leakage function."

The Privacy Metric

We make use of the Valuation of Possible States (VPS) framework [54] to quantify the privacy loss from each agent $i \in A(t)$ to each agent $j \neq i$ during a negotiation. In VPS, agent i's private information is modeled as a state $s \in S$, where S is a set of all possible states that i may occupy. At each instant of time, agent i puts a value on each possible belief that agent j could have about i's state. Agent i's privacy loss during a negotiation with respect to j is the difference in the valuations of agent j's belief before and after the negotiation.

Now, to apply VPS to the MAP studied in this paper, we define a belief function $Bel_{j,i} : \mathcal{O}(t) \times \mathcal{M} \rightarrow [0,1]$ that returns the belief of agent j in the feasibility of different outcomes for agent i as a function of the messages it has received so far. We denote the set of messages agent j received from agent i before the start of the negotiation $\mathcal{N}(t)$ by $M^0_{i\rightarrow j}(t)$. This set includes all the messages agent j receives from i in previous negotiations. Keeping track of the entire history of messages can become cumbersome, but fortunately each agent's belief about other agents is a sufficient statistic for the complete history. Therefore, in practice, the agents only need to store their beliefs and update them by applying Bayes rule when a new message arrives.[3]

To employ the idea of VPS, we also need to define a value function $\mathbb{V}_{ij}$ that captures the value agent i assigns to each belief of agent j. Then, the amount of agent i's privacy that is leaked to agent j during negotiation $\mathcal{N}(t)$ can be defined as follows:

$$L_{i,j}(M_{i\rightarrow j}) = \mathbb{V}_{ij}(Bel_{j,i}(., M^0_{i\rightarrow j}(t))) - \mathbb{V}_{ij}(Bel_{j,i}(., M^0_{i\rightarrow j}(t) \cup M_{i\rightarrow j})), \tag{5}$$

where $Bel_{j,i}(., M) = (Bel_{j,i}(O_{t,1}, M), \ldots, Bel_{j,i}(O_{t,o(t)}, M))$ is a belief vector that represents agent j's belief about the feasibility of all possible agreements for agent i.

As discussed in "Related literature", the value functions that have been previously used in the literature are unable to model the agents' different and even contradictory attitudes towards information sharing with different agents. In practice, each agent i may have different preferences related to the beliefs of different agents. For example, in a meeting scheduling problem, an agent may want to hide their unavailability to meet at a certain time from their boss, but prefer their other colleagues to know their unavailability so that they can help to push the meeting to another date.

To model this behavior, we define an ideal belief function $D_{i,j} : \mathcal{O}(t) \rightarrow [0,1]$ that represents how agent i likes agent j to think about it. The ideal beliefs are generally audience-dependent as the agents' privacy concerns may vary across different audiences. Agent i uses ideal function $D_{i,j}$ as a touchstone to determine the value of belief $Bel_{j,i}(., M)$ of agent j. The comparison of the beliefs can be done based on any vector norm $\|.\|$. In this work, we choose weighted L_1 norm to measure the distance between $Bel_{j,i}(., M)$ and $D_{i,j}$. The reasons of this selection are: (1) the weighted L_1 norm

[3] In any particular negotiation with selfish agents, each agent j also has a belief about whether the messages it has received from i are correct or misleading. However, our goal in this paper is to design a mechanism in which agents always tell the truth. Therefore, the agents can always trust in the information they receive and use that to build their beliefs.

simplifies the calculations by treating the leak of information about different outcomes independently, and (2) the use of a weighted norm allows us to capture the fact that some bits of information may be much more sensitive than others (e.g. an agent may wish to keep several secrets, but some are much more important than others).

The weighted L_1 norm of a column vector $\boldsymbol{q}$ is

$$\|\boldsymbol{q}\|_{1,w} = \|\boldsymbol{w}^T\boldsymbol{q}\|_1 = \sum_k w_k|q_k|, \tag{6}$$

where $\boldsymbol{w} > 0$ is a fixed column weight vector. Using this norm, we define the value function $\mathbb{V}_{ij} : [0,1]^{o(t)} \to \mathbb{R}$ as

$$\begin{aligned}\mathbb{V}_{ij}(Bel_{j,i}(.,M)) &= -\|Bel_{j,i}(.,M) - D_{i,j}(.)\|_{1,\boldsymbol{w}_i} \\ &= -\sum_{k=1}^{o(t)} w_{i,k}\left|Bel_{j,i}(O_{t,k},M) - D_{i,j}(O_{t,k})\right|.\end{aligned} \tag{7}$$

This function takes value 0 when agent j has the ideal belief $D_{i,j}(.)$ and takes a negative value with magnitude of the distance between $Bel_{j,i}(.,M)$ and $D_{i,j}(.)$, otherwise. The weight vector $\boldsymbol{w}_i > 0$, can be arbitrarily selected by each agent i.

Substituting (7) in (5), the privacy loss of agent i to agent j during negotiation $\mathcal{N}(t)$ can be derived as

$$\begin{aligned}L_{i,j}(M_{i\to j}) = \|&Bel_{j,i}(., M^0_{i\to j}(t) \cup M_{i\to j}) - D_{i,j}\|_{1,\boldsymbol{w}_i} \\ &- \|Bel_{j,i}(., M^0_{i\to j}(t)) - D_{i,j}\|_{1,\boldsymbol{w}_i}.\end{aligned} \tag{8}$$

Properties of the Privacy Leakage Function

The privacy metric of "The privacy metric" has two main features.

Feature 1. The ideal belief function $D_{i,j}$ is a powerful tool for modeling the agents' different attitudes towards their privacy information. For example,

- $D_{i,j}(O_{t,k}) = 1$ (or $D_{i,j}(O_{t,k}) = 0$) implies that agent i wishes to persuade agent j that outcome $O_{t,k}$ is feasible (infeasible) for it; e.g. during a negotiation to set a meeting, an employee aims to persuade their boss that they are available to meet at any time the boss wants.
- $D_{i,j}(O_{t,k}) = 0.5$ means that agent i wants to maximize agent j's uncertainty (entropy) about its personal information; e.g. in a negotiation before a common-value auction, the auctioneer is very careful to reveal no information about its reserve price [92]. The reserve price is a threshold indicating the lowest price the auctioneer is willing to accept for selling the good. Disclosure of this information discourages some bidders with low budgets from participating. As a result, their information about the value of the good plays no role in the auction even though it may be relevant for the valuation of other bidders. The consequence is to prevent some sales from being made even though the aggregate information would imply that a transaction should occur. This decreases the auctioneer's revenue and hence is not in its interest.
- $D_{i,j}(O_{t,k}) = \mathbb{1}(V_i(O_{t,k}) = -\infty)$, where $\mathbb{1}(.)$ is the indicator function, implies that agent i wants to twist the truth and distort reality; e.g. to set a trap for a repairman thief, a detective may request a repair and then, during the negotiation to set a time, tell the repairman that no one is at home at times they actually are.

The ability to model the agents' different attitudes towards their privacy information is an important feature which is missing in the existing privacy metrics. Considering a fixed and identical attitude towards sharing information with all other agents is simply unrealistic.

Feature 2. The privacy metric $L_{i,j}$ defined in (8) is able to model the possible correlations among the feasibility of different agreements for an agent. In many negotiations, the possible outcomes are correlated. In this contect, outcomes $O_{t,k}$ and $O_{t,k'}$ are said to be correlated if the feasibility of one of them gives some information about the feasibility of the other. For example, in meeting scheduling, an agent may have some side information about the pattern of agent i's calendar. This side information could be the length or repeat frequency of agent i's meetings, or the length of breaks it normally has between meetings. In this situation, knowing that agent i is free or busy at a time slot may reveal some information about its availability at other time slots.

Task allocation is another example in which correlation is significant. Consider an outcome in which both tasks 1 and 2 are assigned to agent i. If this outcome is infeasible for agent i, we can conclude that agent i is unable to perform either task 1 or 2 or it cannot do both of the tasks simultaneously. In this case, any outcome in which both tasks 1 and 2 are assigned to agent i is definitely infeasible for it. Moreover, this information decreases other agents' beliefs on the feasibility of outcomes that assign either task 1 or 2 to agent i.

As seen from the above examples, the indirect leakage of information is present in many real life problems. Therefore, it is essential for a privacy metric to be able to model this feature. The privacy leakage function defined in (8) achieves this goal through the posterior beliefs $Bel_{j,i}(., M^0_{i\to j}(t) \cup M_{i\to j})$ that agents have about the feasibility of undiscussed outcomes for other agents. In other VPS-based metrics available in the literature [54], the value function $\mathbb{V}$ is defined based on the number of states that others believe to be possible for each agent i. Therefore, these metrics cannot model the correlation unless the indirect leakage of information is strong enough to completely remove the possibility of an agent i's state in others' points of view. In contrast to this literature, our privacy metric can model even weak correlations.

In connection with this feature, it is of interest to take a look at a special case where the agreements are independent. That means, discussing each outcome $O_{t,k}$ provides no information about the feasibility of other outcomes. In this case, suppose that each agent j has a prior belief $1 - d_i$ that an outcome $O_{t,k}$ is feasible for agent i, i.e. $Bel_{j,i}(O_{t,k}, M^0_{i\to j}(t)) = 1 - d_i$. Then, the costs for privacy loss when agent i reveals feasibility (i.e. $Bel_{j,i}(O_{t,k}, M^0_{i\to j}(t) \cup M_{i\to j}) = 1$) or infeasibility (i.e. $Bel_{j,i}(O_{t,k}, M^0_{i\to j}(t) \cup M_{i\to j}) = 0$) of this outcome are

$$\begin{aligned} L_{i,j}(O_{t,k} \text{ is feasible for } i) &= w_{i,k}\Big(1 - D_{i,j}(O_{t,k}) - \Big|1 - d_i - D_{i,j}(O_{t,k})\Big|\Big) \\ &= w_{i,k}(\min(d_i, 2(1 - D_{i,j}(O_{t,k})) - d_i)), \end{aligned} \tag{9}$$

and

$$\begin{aligned} L_{i,j}(O_{t,k} \text{ is infeasible for } i) &= w_{i,k}\Big(D_{i,j}(O_{t,k}) - \Big|1 - d_i - D_{i,j}(O_{t,k})\Big|\Big) \\ &= w_{i,k}(\min(1 - d_i, 2D_{i,j}(O_{t,k}) - 1 + d_i)), \end{aligned} \tag{10}$$

respectively. These results will be used in "Initiator's optimal offering strategy7" to design the initiator's optimal negotiation strategy.

Feature 3. In Feature 2, we discussed that the agents may have some side information about the correlation among the feasibility of different agreements for each other. This information is assumed to be common knowledge and fixed. However, in some applications, the agents may receive private signals over time that are informative about other agents' preferences. These signals may impact the agents' beliefs about others and hence their privacy losses.

When such private information is available, we denote the posterior belief of an agent j about the feasibility of agreement $O_{t,k}$ for agent i by $Bel_{j,i}(O_{t,k}, M^0_{i\to j}(t) \cup M_{i\to j}, S^n_j)$, where S^n_j represents the signals received by agent j up to round n of negotiation $\mathcal{N}(t)$. Using this notation, we can derive the amount of agent i's privacy that is leaked to agent j during negotiation $\mathcal{N}(t)$ as

$$L_{i,j}(M_{i\to j}, S^n_j) = \mathbb{V}_{ij}(Bel_{j,i}(., M^0_{i\to j}(t), S^0_j)) - \mathbb{V}_{ij}(Bel_{j,i}(., M^0_{i\to j}(t) \cup M_{i\to j}, S^n_j)). \tag{11}$$

The main difference between this case and the former one without private signals is that in the latter case each agent's utility depends not only on its own private information, but also on other agents' private information. This property is known in the literature as interdependent valuations [27, 58]. Agents with interdependent valuations, choose their actions so as to maximize the expected utility they can collect, where the expectation is taken with respect to the signals other agents may receive.

The privacy leakage function presented in this section completes the description of the agents' preferences in our model. In the next section, we present an incentive mechanism to solve the problem of designing a faithful negotiation protocol for privacy-aware agents.

The Distributed Score-Based Multi-Round Negotiation Mechanism

In this section, we present a distributed negotiation mechanism that privacy-aware agents can employ to reach an agreement. This mechanism is faithful, and gives sufficient incentives to the selfish participants to follow its rules. The mechanism is individually rational (IR) as well, meaning that the utilities the agents get in this mechanism are at least as much as the utilities they get when they do not participate in the mechanism [7]. This feature which is also known as voluntary participation, is important as the selfish agents are not forced to join the mechanism. We first describe informally the idea behind our mechanism in "Key intuitions" and then go on to the formal description in "Formal description".

Key Intuitions

As we discussed in "Related literature", scoring correspondences have been widely used in the literature for solving agreement problems. However, these mechanisms are not directly applicable for distributed piMAPs with self-interested agents due to the following two drawbacks: (1) Scoring correspondences are designed for centralized settings where a trusted central entity exists and is responsible for running the mechanism. This class of mechanisms provides appropriate incentives for the agents to score the candidates truthfully, by promising that the central authority selects the option with the maximum aggregate score. However, in a distributed setting, the agents act strategically and if any of them asks others to score some outcomes, it is clear that it will select its own most preferable outcome among those that are feasible for others. (2) In scoring correspondences, the agents' preferences over all possible options are collected in one round to calculate the socially-optimal outcome. In these mechanisms, similar to the direct mechanisms, the privacy leakage is too high to be meaningful in an agreement problem with privacy-aware agents.

To solve these drawbacks, we propose a novel Distributed Score-based Multi-round (DSM) negotiation mechanism based on an artificial currency. Making payments in terms of an artificial currency, which we call "convenience points" ("Strategic multi-agent negotiation over a communication graph"), is our main tool to provide a faithful implementation of scoring correspondences in distributed settings. The main role of the convenience points is to make the selfish initiator sensitive to the responders' opinions. In this mechanism, the number of points (i.e. subsidy) the initiator must give to an agent to persuade it to accept an agreement is inversely proportional to the score s it gave to that agreement. Therefore, the total cost incurred by the initiator to get everyone's consent for an agreement is inversely proportional to the responders' total satisfaction with that agreement. This cost aligns the initiator's total utility with the social-welfare; that is, the higher satisfaction an agreement provides to the participants, the more attractive it is to the initiator.

In general, adding a virtual currency to a scoring correspondence can ruin its incentive compatibility. This is because the responders are not concerned only about the final agreement anymore, but also about the points they can collect. For example, a reward function that is decreasing in terms of the scores might motivate the responders to give lower scores to the offers. To avoid this problem, we design a reward function that is (1) decreasing in terms of the score the agent gives to the selected agreement, and (2) increasing in terms of the degree of flexibility it shows in the negotiation. This means that if a responder announces to be generally more satisfied with the offers, it will get more points if one of its undesirable agreements is selected. Using this idea, we design the reward function such that the tradeoff between the benefits of giving low and high scores to the offers will make truth-telling the best strategy for the responders.

As discussed in "Strategic multi-agent negotiation over a communication graph", a mechanism with an artificial currency must define a usage for the currency to make it valuable for the agents. In our mechanism, the usage of convenience points is twofold. If an agent acts as an initiator in a negotiation, it needs convenience points to get responders' consent for the agreement it would like to select. The second usage is for responders who want to score the offers made by the initiator. Scoring is considered to be costly with the costs inversely proportional to the given scores. Therefore, the number of convenience points that an agent has is an indicator of the impact it can make in the future negotiations.

Formal Description

The DSM negotiation mechanism is multi-round, where each round consists of three stages: (1) Proposal, (2) Scoring, and (3) Assessment (see Fig. 1). Each round of the mechanism starts with the proposal stage in which the initiator proposes some agreements to the responders. The responders score the proposals at stage 2. Then, the initiator assesses the responses at stage 3. If the initiator finds any of the agreements appealing at this stage, it announces the agreement publicly and goes to the Reward stage where it should give responders' rewards. Otherwise, it goes to the next round and repeats stages 1–3 until a suitable agreement is found or no agreements remain undiscussed. In the following subsections, we discuss the main stages of our mechanism in more detail.

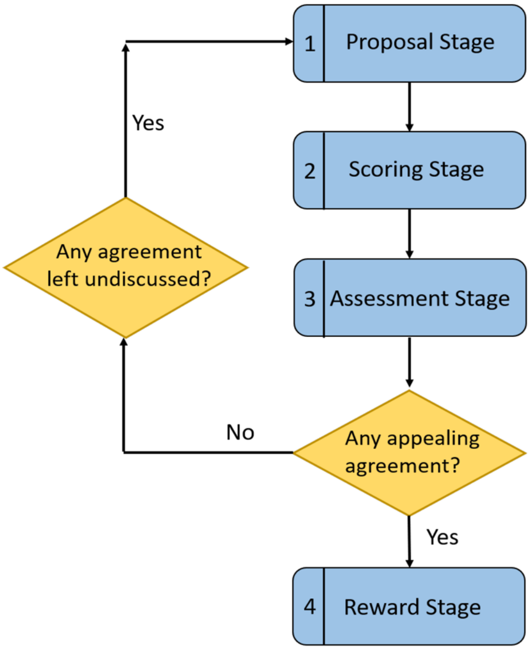


Fig. 1 An overview (not detailed) flowchart of the DSM mechanism

Proposal Stage

The initiator offers between $L_{\min}$ and $L_{\max}$ agreements to the responders. We denote the number of proposals it makes at round n of negotiation $\mathcal{N}(t)$ by $L_n(t)$. These proposals are denoted by $\{O_n^1, \dots, O_n^{L_n(t)}\} \subseteq \mathcal{O}(t)$. The initiator extracts the responders' opinions about these agreements at the scoring stage. Then, it decides whether any of the proposed agreements are suitable. If the initiator finds none of the agreements appealing, it is allowed to go to the next round and make some new offers, provided that it followed the rules at round n and offered at least $L_{\min}$ agreements. Otherwise, the negotiation terminates at the end of round n, independent of whether or not agreement is achieved. This rule is set to encourage the initiator to make at least $L_{\min}$ offers at each round of the negotiation.

Scoring Stage

Responders are asked to score each of the offers $O_n^1, \dots, O_n^{L_n(t)}$ on a scale of 0 to $D-1$, where 0 means “infeasible”, 1 means “Feasible but very unsatisfactory” and $D-1$ means “Feasible and completely satisfactory”. The number of satisfaction levels $D \geq 2$ represents the number of rating levels that are used by the agents to evaluate the suitability of different agreements. The responders who give score $s \in \{1, 2, \dots, D-2\}$ to an agreement $O_{t,k}$ accept it only if the initiator compensates them for the hardship they endure by giving them some convenience points. Two examples of hardship could be attending a meeting after work hours and doing a job for which the agent has no passion or interest. The agents use these convenience points to score future offers.

We denote the score vector given by responder i at round n by $\mathbf{s}_{i,n} = (s_{i,n}^1, \dots, s_{i,n}^{L_n(t)})$, where $s_{i,n}^j \in \{0, 1, \dots, D-1\}$ indicates how satisfied responder i is with the j-th agreement offered at round n.

We define some parameters related to the responder i’s scores at round n, as follows:

1. Frequency $H_{i,n}^s$ of each score s: This parameter is defined as the number of agreements to which responder i gives score s; that is

$$H_{i,n}^s = \sum_j \mathbb{1}(s_{i,n}^j = s). \tag{12}$$

2. Availability level $A_{i,n}$: This is the number of agreements responder i has announced to be feasible for it; that is

$$A_{i,n} = \sum_{s=1}^{D-1} H_{i,n}^s. \tag{13}$$

3. Degree of Flexibility $F_{i,n}$: The flexibility responder i shows at round n of negotiation is defined as

$$F_{i,n} = \sum_{s=1}^{D-1} (A_{i,n}+1)^{s-1} H_{i,n}^s. \tag{14}$$

This function gives the decimal value of number $(H_{i,n}^{D-1}, \dots, H_{i,n}^1)$ in base $A_{i,n}+1$. Therefore, a greater value of $F_{i,n}$ means responder i has announced to be more satisfied with the agreements offered at round n. Function F is invertible; meaning that for each i and n, given $A_{i,n}+1$, the vector $(H_{i,n}^{D-1}, \dots, H_{i,n}^1)$ can be reconstructed from the flexibility $F_{i,n}$.

Scoring is costly. The agents need to spend $C(s)$ points to give a score s to an offer. The cost function $C(s)$ is given by

$$C(s) = (D-s-1)\,\mathrm{sign}\,(s), \tag{15}$$

where sign (s) is equal to + 1, − 1, or 0, as s is positive, negative, or zero, respectively. Based on (15), the agents can report infeasibility of an agreement and their complete satisfaction, free of charge, by giving scores 0 and $D-1$, respectively. However, reporting “feasibility but lack of complete satisfaction” requires spending points. The number of points an agent needs to spend to give score s to an agreement is a decreasing linear function of s. Using (15) and (12), we can derive the total cost to responder i, for giving scores $\mathbf{s}_{i,n}$ to the offers of round n as

$$C_{i,n} = \sum_{s=0}^{D-1} C(s) H_{i,n}^s = C(A_{i,n}, F_{i,n}). \tag{16}$$

As discussed earlier, the frequency variables $H_{i,n}^s$, $s \in \{1, \dots, D-1\}$ can be uniquely determined by the availability level $A_{i,n}$ and the degree of flexibility $F_{i,n}$. Therefore, the cost $C_{i,n}$ of agent i’s scoring at round n can be considered as a function of $A_{i,n}$ and $F_{i,n}$, and hence represented as $C(A_{i,n}, F_{i,n})$.

The initiator collects the points used by responders for scoring at round n, if it finds a suitable agreement at round n. Otherwise, the points are refunded to the responders.

Assessment Stage

After receiving the responders’ scores $\mathbf{s}_{i,n}$, $i = 1, \dots, M$, the initiator evaluates all offers $\{O_n^1, \dots, O_n^{L_n(t)}\}$ and decides which, if any, of them are suitable to be selected. Since the initiator is selfish, it does this evaluation based on its own utility. According to (4), the initiator’s goal is to maximize its forward utility

$$\mathbb{E}\big[U_I^{t:\infty} | b_I(t)] = U_I^t + \sum_k \mu_I^t(k) \mathbb{E}\big[U_I^{t+1:\infty} | k \text{ negotiations}, b_I(t+1)\big], \tag{17}$$

where the first term on the right represents the initiator’s instant profit in negotiation $\mathcal{N}(t)$ and the second term represents its expected utility in the future negotiations. Based on (3), the initiator’s instant profit from selecting any agreement O_n^j at round n of negotiation $\mathcal{N}(t)$ is

$$U_I^t(O_n^j) = \begin{cases} V_I(O_n^j) - \beta_I n - \sum_{i \in R_I(t)} \theta_{Ii} L_{I,i}(M_{I \to i}), & \text{If } s_{i,n}^j > 0 \quad \text{for all } i, \\ -\infty, & \text{Otherwise.} \end{cases} \tag{18}$$

The final agreement must be feasible for all participants. Therefore, the initiator’s utility would be $-\infty$, if the agreement is not feasible for at least one responder. For the agreements that are feasible for all agents, the initiator's instant profit is computed as the difference between the value it gains and the costs it incurs. The initiator gains value $V_I(O^j_n)$ from making agreement O^j_n, and incurs costs of $\beta_I n$ and $\sum_{i \in R_I(t)} \theta_{Ii} L_{I,i}(M_{I \rightarrow i})$, from running the mechanism for n rounds and losing its privacy in communications with different responders, respectively.

The cost terms in the initiator's instant utility function (18) are independent of the selected outcome O^j_n; they are just affected by the number of execution rounds and the offers the initiator makes at each round. Therefore, the initiator considers the cost terms when it is deciding whether to choose an agreement at this round or go to the next round; but it disregards them when it is comparing the offers within one round. When the initiator decides round n is the final round of negotiation, it chooses an agreement O^j_n that is feasible for everyone and maximizes $V_I(O^j_n) + \sum_k \mu^t_I(k) \mathbb{E}\left[U_I^{t+1:\infty} | k \text{ negotiations}, b_I(t) + B_I\right]$, where the second term is an increasing function of the initiator's point income B_I.

The initiator's point income B_I is the difference between the points it collects during the negotiation and the points it spends. The initiator neither spends or collects any points at the pre-final rounds. However, at the final round, it collects the points that responders spend for scoring the offers and awards some points to the responders that are not completely satisfied with the final choice. The number of points that must be awarded to a responder i when agreement O^j_n is selected is determined by a reward function $r(s^j_{i,n}, A_{i,n}, F_{i,n}, L_n(t))$. This function determines each responder i's reward based on its score $s^j_{i,n}$ to the selected outcome O^j_n, the availability $A_{i,n}$ and the flexibility $F_{i,n}$ it announces for the whole offer set, and the number $L_n(t)$ of offers made by the initiator at round n. In "Reward stage", we will fully explain the design of the reward function $r(.)$. This function is designed so as to be decreasing in terms of the score $s^j_{i,n}$ and increasing in terms of the flexibility $F_{i,n}$.

From the discussions above, we can derive the initiator's points income in the negotiation $\mathcal{N}(t)$ as

$$B_I = \sum_{i \in R_I(t)} C(A_{i,n}, F_{i,n}) - \sum_{i \in R_I(t)} r(s^j_{i,n}, A_{i,n}, F_{i,n}, L_n(t)). \tag{19}$$

The reward function is a decreasing function of the score the responder gives to the selected outcome. Therefore, the initiator's point income B_I is directly proportional to the responders' total satisfaction with the final agreement. For a fixed set of offers, the initiator can collect more points by choosing an agreement that is more favorable to the responders. However, the initiator has a personal preference $V_I(.)$ over the agreements, as well. Therefore, the selfish initiator that maximizes the sum of $V_I(.)$ and an increasing function of B_I faces a tradeoff between its own preferences and the social-welfare. It can determine the best point of this tradeoff by solving the following optimization problem:

$$\begin{aligned} \max_{j \in \{1,\dots,L_n(t)\}} \; & V_I(O^j_n) + \sum_k \mu^t_I(k) \mathbb{E}\Big[U_I^{t+1:\infty} | k \text{ negotiations}, \\ & b_I(t) + \textstyle\sum_{i \in R_I(t)} C(A_{i,n}, F_{i,n}) - \sum_{i \in R_I(t)} r(s^j_{i,n}, A_{i,n}, F_{i,n}, L_n(t))\Big], \\ \text{s.t.} \quad & s^j_{i,n} \geq 0, \quad \forall i \in R_I(t). \end{aligned} \tag{20}$$

Let j^* denote the solution of the optimization problem (20). $O^{j^*}_n$ is the best agreement for the initiator among those offered at round n. However, the initiator has an option to not select any agreement at round n and move forward to the next round. In this case, the mechanism does not let the initiator go back to the agreements it previously offered. This rule is designed to encourage the initiator to choose the final agreement as soon as possible. If the initiator rejects a feasible agreement at round n and moves to the next round, there is a risk that no other feasible agreement can be found and hence disagreement happens. To avoid this risk, the initiator prefers to make an agreement as soon as it can.

Reward Stage

When the initiator finds an agreement $O^{j^*}_n$ appealing at round n, the negotiation moves to the final stage, which is called the reward stage. In this stage, the initiator awards the promised rewards to the responders. Rewards are determined based on the scores the responders gave to the offers and must be designed so as to incentivize responders to give true scores.

To incentivize agents to score the offered agreements truthfully, we design reward function $r(.)$ such that it satisfies the following conditions:

$$\sum_{s=1}^{D-1} P(s, A, F, L) r(s, A, F, L) - C(A, F) = 0, \tag{21}$$
$$\forall L \leq L_{\max}, \forall A \leq L, \forall F \geq (A+1)^{D-2},$$

(a) where $P(s, A, F, L)$ is the probability that a responder with flexibility F that announced feasibility of A out of L agreements assigns to the fact that one of the agreements it scored s will be selected by the initiator.

(b) $r(s, A, F, L)$ is a decreasing function of s for $s > 0$.

(c) $r(D-1, A, F, L) = 0$, for all $A, F, L \leq L_{\max}$.

(d) $r(s, A, F, L) = \infty$, for all $s, A, F, L > L_{\max}$.

(e) $r(.)$ is invariant to shifting of the scores. That is, $\mathbf{s}'_{i,n} = (\mathbf{s}_{i,n} + c) \operatorname{sign}(\mathbf{s}_{i,n})$, where $c \in \{1, \dots, D-2\}$, implies that $r(s', A', F', L) = r(s, A, F, L)$.[4]

The intuitions behind the above conditions are as follows. Condition (a) guarantees that provided the agent gives

[4] It is clear that by this transformation, we have $A' = A$.

score $D-1$ to at least one offer (i.e. $F \geq (A_{i,n}+1)^{D-2}$), the expected reward it gets minus the number of points it spends is zero. This expectation is computed based on the agent's belief about the likely effectiveness of its scores on the selection of the final agreement. Condition (b) means that the agents that are less satisfied with the final agreement receive higher rewards. Condition (c) ensures that a responder that is announced to be completely satisfied with the chosen agreement receives no reward. Condition (d) means that the initiator that wants to offer more than L_{max} agreements in one round, needs to pay an infinite number of points to the responders. This condition is set to prevent the initiator from violating the upper bound L_{max}. Condition (e) determines the reward for scores with $F < (A_{i,n}+1)^{D-2}$ and guarantees that the reward function is only sensitive to the relative scores the agent gives to the offers and not on the absolute values. Based on the definition provided in condition (e), we call scores $\mathbf{s}$ and $\mathbf{s}'$ shifted versions of each other, if (1) they mark the same agreements as infeasible, and (2) they differ only by a constant factor in the feasible agreements.

Theorem 1 *For any fixed belief profile* $\{P(s,A,F,L)\}_{s,A,F,L}$ *with*

$$\sum_{s=1}^{D-2} P(s,A,F,L) \neq 0, \tag{22}$$

for all $A \leq L \leq L_{\max}$ *and* F *such that* $\bmod\ (F,(A_{i,n}+1)^{D-2}) > 0$, *the system of equations defined in* (21) *has a solution that satisfies (b)–(e).*

Theorem 1 proves that for a fixed belief profile $\{P(s,A,F,L)\}_{s,A,F,L}$, we can find a reward function $r(.)$ that satisfies all the desirable conditions (a)–(e). However, probabilities $\{P(s,A,F,L)\}_{s,A,F,L}$ are not independent of the reward function, but a function of it. The reason is as follows. The probabilities $\{P(s,A,F,L)\}_{s,A,F,L}$ depend on the initiator's strategy in selecting the final agreement. The optimization problem (20) shows that the initiator's strategy is determined based on the reward function $r(.)$. Therefore, to derive a belief profile and reward function that are consistent with each other, we have to run Algorithm 1. This algorithm works by first considering an arbitrary reward function $r(.)$ that satisfies conditions (b)–(e). These conditions are weak and easily satisfied. Then it calculates probabilities $\{P(s,A,F,L)\}_{s,A,F,L}$ that match with the selected reward function and updates function $r(.)$ based on equation (21) and conditions (b)–(e). This procedure repeats until convergence is reached. Theorem 1 ensures that the algorithm will never stick in Line 5 because of not finding a solution to the set of equations (21). In practice, Algorithm 1 converges to a solution in a few iterations.[5]

Algorithm 1: Reward Design

```
1 Initialize reward function $r(.)$ such that it satisfies conditions (b)-(e);
2 $err \leftarrow \infty$;
3 while $err > th$ do
4     Calculate probabilities $\{P(s,A,F,L)\}_{s,A,F,L}$ based on $r(.)$;
5     $r_{new} \leftarrow$ Solution of the set of equations (21) that satisfies conditions (b)-(e);
6     $err \leftarrow Norm(r - r_{new})$;
7     $r \leftarrow r_{new}$;
8 end
```

We represent the DSM negotiation mechanism designed in this section by $\Gamma = (L_{\min}, L_{\max}, D, r(.))$. The corresponding pseudo-code of this mechanism is given by Algorithm 2. Briefly, when the need for an agreement arises, the initiator $I(t)$ starts a negotiation process by offering some agreements to the responders $R_I(t)$ (Lines 2–7). The number of offers at each round is one of the initiator's decision variables. Receiving the offers, responders use their convenience points to express their preferences over the offers (Line 8). Then, the initiator evaluates each offer based on the utility it provides to it (Line 9). If the initiator finds any of the agreements acceptable, it will announce it and terminate the negotiation by paying the rewards (Lines 10–13). Otherwise, it will go to the next round if $L_n(t) \geq L_{min}$ (Lines 14–21). At each instant, the agreement indicator $a_g \in \{0, \pm 1\}$ shows whether the negotiation is in process ($a_g = 0$), ends by agreement ($a_g = 1$), or ends by disagreement ($a_g = -1$). The MATLAB code of negotiation based on the DSM mechanism is publicly available at https://github.com/ffarhadi20/Distributed-Score-based-Multiround-Mechanism.

[5] The formal proof of the convergence of Algorithm 1 is a challenging open question and is left for our future work.

Algorithm 2: DSM mechanism $\mathit{\Gamma} = (L_{min}, L_{max}, D, r(.))$

```
for time $t = 1, 2, \ldots$ do
    if The need for an agreement arises then
        The negotiation process $\mathcal{N}(t)$ starts;
        The round number $n \leftarrow 1$;
        The agreement indicator $a_g \leftarrow 0$;
        while $a_g = 0$ do
            Proposal stage: Initiator offers $L_n(t) \leq L_{max}$ agreements to the
              responders. it offers at least $L_{min}$ agreements if possible;
            Scoring stage: Each responder $i$ gives an integer score between 0 to
              $D - 1$ to each offered agreement; i.e. $\mathbf{s}_{i,n} = (s^1_{i,n}, \ldots, s^{L_n(t)}_{i,n})$;
            Assessment stage: Initiator solves the optimization problem (20)
              and finds the agreement $O^{j^*}_n$ that maximizes its utility;
            if Initiator prefers $O^{j^*}_n$ to moving to the next round then
                $O^{j^*}_n$ is selected as the final agreement;
                $a_g \leftarrow 1$;
                Reward stage: Initiator awards $r(s^j_{i,n}, A_{i,n}, F_{i,n}, L_n(t))$ points
                  to each responder $i$;
            else
                if $L_n(t) \geq L_{min}$ then
                    $n \leftarrow n + 1$;
                else
                    Disagreement arises;
                    $a_g \leftarrow -1$;
                end
            end
        end
    end
end
```

Properties of the Mechanism

In "The distributed score-based multi-round negotiation mechanism", we designed a mechanism to be followed by the agents to make an agreement. However, since selfish agents always try to maximize their utilities, they may cheat and deviate from the specified rules if it is advantageous. For example, at the proposal stage, the initiator may offer less than $L_{\min}$ agreements, when it can do otherwise, or at the reward stage the initiator may refuse to pay the responders' rewards.

Definition 1 The initiator is faithful to the mechanism $\Gamma = (L_{\min}, L_{\max}, D, r(.))$, if it follows all the following rules:

(I1) At a proposal stage, it offers no less than $L_{\min}$ and no more than L_{max} agreements, when it can do so.

(I2) At a proposal stage, it chooses an offer size such that the maximum budget needed for selecting a feasible agreement is below its available budget.

(I3) At an assessment stage, it selects the solution of the optimization problem (20) as the final agreement, if the feasible set is not empty.

(I4) At the reward stage, it awards the promised rewards to the responders.

Satisfaction of Rule (I2) is an important property for a mechanism which is known in the literature as budget feasibility [68]. In a budget feasible mechanism, no one commits to make a payment that it cannot afford.

Responders take actions only at the scoring stage. At this stage, each responder should give scores to the offers according to its preference orderings. We define the faithful behavior of the responders in Definition 2. A responder can deviate from its faithful behavior by violating the conditions stated in this definition.

Definition 2 Responder i is faithful to the mechanism $\Gamma = (L_{\min}, L_{\max}, D, r(.))$, if its score vector $\mathbf{s}_{i,n}$ at each round n satisfies the following conditions:

(R1) For each $j = 1, \dots, L_n$, $s^j_{i,n} = 0$ if and only if the agreement O^j_n is infeasible for responder i.

(R2) The scores are non-decreasing in the agreements' values, i.e. $V_i(O^j_n) > V_i(O^k_n)$ implies that $s^j_{i,n} \geq s^k_{i,n}$.

(R3) The scores are as discriminatory as possible. That is, agreements with different values get different scores, as long as both the number of satisfaction levels D and the responder's budget b_i allow.

In "Initiator's faithfulness" and "Responders' faithfulness", we prove that the mechanism Γ is faithful for the initiator and the responders, respectively, meaning that the mechanism provides sufficient incentives to them to not deviate from the rules. Then, in "Individual rationality and individual budget balance", we discuss the additional interesting properties of the mechanism Γ. All of the proofs are given in the appendix.

Initiator's Faithfulness

We prove that the initiator can gain no benefit by deviating from the rules (I1)–(I4). The initiator plays an active role in three out of four stages of the mechanism (i.e. proposal, assessment, and reward). Thus, in the first instance, it may seem difficult to prove the initiator's loyalty to the algorithm's rules. However, in the design process, having in mind that the initiator is selfish, we established some control policies into the mechanism so as to guarantee the initiator's faithfulness. The most prominent control policies of the mechanism are as follows:

(C1) The initiator is not allowed to move forward to round $n + 1$, if $L_n < L_{\min}$. Therefore, the initiator proposes at least L_{min} offers at each round, if possible, to preserve the chance of continuing the negotiation;

(C2) $r(s, A, F, L) = \infty$, for $L > L_{\max}$. With a finite budget, the initiator cannot get responders' consent for any agreement, if it offered more than $L_{\max}$ outcomes. This feature forces the initiator to observe the upper limit $L_{\max}$;

(C3) There is no returning to the past options. The mechanism does not let the initiator go back to the agreements it previously rejected. Therefore, to prevent disagreement, the initiator prefers to select a feasible agreement as soon as it finds one.

The control policies discussed above help us to state the following theorem.

Theorem 2 *The mechanism* $\Gamma = (L_{\min}, L_{max}, D, r(.))$ *is faithful for the initiator.*

We can see from proof of Theorem 2 that control policies (C1)–(C3) have important roles in preventing the initiator from deviating from desirable rules (I1)–(I4). We provide two examples below to illustrate how the results would change if any of these control policies were lifted.

Example 1 If (C1) was not in place, an initiator with a low bargaining cost and high privacy sensitivity would prefer to deviate from (I1) and offer the options one by one. In this case, since the initiator offers the options in order of its own satisfaction, the final outcome would be the best option for the initiator that is feasible for all responders. This outcome is not socially desirable as it is similar in spirit to the outcome of dictator games, where one agent (i.e. the initiator) has all the power and the other agents (i.e. responders) have no opportunity to express their preferences. Control policy (C1) has been developed to prevent this undesirable behavior.

Example 2 If (C3) was not in place, an initiator with a low bargaining cost and low privacy sensitivity would prefer to deviate from (I3) and negotiate all the available options with the responders to find the agreement that exactly maximizes its own utility. This perfectionism lowers the convergence speed and led the responders to lose all their privacy.

It is important to note that in addition to control policies (C1)–(C3), the reward function designed in "Reward stage" has an important role in providing the incentive for the initiator to follow the rules (I1)–(I4). If the initiator did not have to pay rewards to the responders based on their satisfaction from the final outcome, it would not care about the responders' preferences and only take the feasibility or infeasibility of different outcomes for responders into consideration. Then, the initiator would choose the best agreement for itself that is feasible for all responders. By designing a suitable reward function which is inversely proportional to the responders' satisfaction from the final outcome, we drive the initiator away from its selfish behaviour and motivate it to make socially-optimal decisions.

Responders' Faithfulness

We now prove that the responders have appropriate incentives to follow rules (R1)–(R3).

Lemma 1 *The responders do not have any incentive to lie about the feasibility of agreements for them, i.e. giving a rating of* 0 *to an agreement* O^j_n *is efficient for a responder i if and only if agreement* O^j_n *is infeasible for it.*

Lemma 1 proves that condition (R1) of Definition 2 is satisfied. To prove satisfaction of conditions (R2) and (R3), we need the two following lemmas. Lemma 3 states an important property of the proposed mechanism that is key to proving faithfulness.

Lemma 2 *It is optimal for each responder i to announce its complete satisfaction for at least one feasible offer, if it exists.*

Lemma 3 *A responder's expected net point income at each round is* 0*, provided that it gives score* $D-1$ *to at least one feasible offer.*

As a result of Lemmas 2 and 3, when a responder is deciding about the scores it should give to the offers, it can neglect the points and only take into account the effect of its scores on the selection of the final agreement. This property helps us to prove the next two lemmas.

Lemma 4 *It is never optimal for a responder to give a higher rating to an agreement it likes less.*

Lemma 5 *The scores are as discriminatory as possible. That is, as long as the number of satisfaction levels and the responder's budget allow, it is optimal for the responder to give unequal scores to agreements with unequal values.*

Based on Lemmas 1–5, we can state the following main theorem.

Theorem 3 *For any* $L_{\min}$*,* $L_{\max}$*, and* D*, the mechanism* $\Gamma = (L_{\min}, L_{\max}, D, r(.))$ *where reward function* $r(.)$ *is derived by Algorithm 1 is faithful for the responders.*

Proof This is directly derived from Lemmas 1–5. □

Individual Rationality and Individual Budget Balance

In "Initiator's faithfulness" and "Responders' faithfulness", we proved that DSM is faithful and can be executed by selfish agents with no need for a trusted controller. In this section, we show that this mechanism also satisfies two other desirable properties, namely individual rationality and individual budget balance. The individual rationality is mainly due to Assumption 1, and the individual budget balance is mainly due to the appropriate design of the reward function.

Individual Rationality: The DSM mechanism is individually rational for both the initiator and the responders. That is, each participant prefers the outcome of the mechanism to the utility it gets when it does not participate. The reason is simple: The participants cannot reach an agreement unless all of them participate in the negotiation process. Therefore, to avoid disagreement, which is assumed to be the worst outcome for all agents (see Assumption 1), the agents prefer to participate in the mechanism.

Individual Budget Balance: The DSM mechanism is neither profitable nor loss-making in terms of the points. The convenience points are just a tool for aligning the initiator's objective with the responders'. Therefore, the agents' long-term objectives are not to collect points, but rather making socially-acceptable agreements.

Theorem 4 *The DSM mechanism is individually budget balanced (IBB). That is, at each round of the mechanism, each participant's expected point income is zero.*

Proof In the DSM mechanism, all responders trade their points with the initiator. Therefore, the initiator's point income is the negative of the sum of the responders' point income. Based on Lemmas 2 and 3, the responders' expected point incomes are zero at the optimal strategy. Therefore, the initiator's expected point income is zero as well. This proves the individual budget balance of the DSM mechanism. □

This feature is particularly useful as it ensures that consecutive negotiations do not lead to one agent losing all its points or one agent gaining a large number of extra points. This property keeps agents' negotiation power balanced and hence can be interpreted as the fairness of the mechanism.

Initiator's Optimal Offering Strategy

In "Properties of the mechanism", we showed that the DSM mechanism provides the initiator with positive incentives to follow rules (I1)–(I4). Rule (I1) guarantees that at each round, the initiator offers between $L_{\min}$ and $L_{\max}$ agreements to the responders. However, the exact number of offers and the agreements that should be offered at each round are the initiator's decision-making variables. In this section, we derive the optimal offering strategy that the initiator should adopt to maximize its utility. For ease of presentation, we restrict attention to the case where outcomes (i) are independent (i.e. the feasibility of one outcome for a participant provides no information about the feasibility of other outcomes for it), and (ii) have equal privacy importance (i.e. $w_{I,k} = 1, \forall k$) and equal desired belief (i.e. $D_{Ii}(O_{t,k}) = D_{Ii}, \forall i, k$) from the initiator's point of view. However, the method can be simply generalized to the correlated outcomes with unequal privacy importance.

For any fixed offer size L, it is optimal for the initiator to offer its top L agreements that have not been discussed before. This is because when agreements are independent, previous discussions give no information about the feasibility of undebated agreements. Therefore, the chance of an agreement being feasible for all participants, as well as the privacy leakage of offering it, are the same across all agreements. In this case, the only factor that distinguishes the agreements for the initiator is their valuations $V_I(O_n^j)$. Therefore, when the offer size is fixed, it is most advantageous for the initiator to offer the agreements which have the maximum valuation for it. With this known, our objective in the rest of this section is to derive the optimal offer size L.

The optimal offer size depends on different factors; the most important being:

(a) The initiator's bargaining cost β_I: An initiator with a higher bargaining cost prefers to offer more agreements at each round to reduce the number of negotiation rounds;

(b) The initiator's privacy sensitivity θ_{Ii}, $i \in R_I$: A more privacy-sensitive initiator offers fewer agreements at each round to reduce its privacy leakage;

(c) The values of different agreements for the initiator $V_I(O_n^j)$, $j \in \mathcal{O}(t)$: If there is a big drop in valuation between the j^{th} and $j+1^{th}$ top agreements, the initiator prefers to keep agreement $j+1$ for the next round and first try the more valuable agreements;

(d) Number of participants $|A(t)|$: The chance of an agreement working for all agents is inversely proportional to the number of participants. Therefore, when the initiator is negotiating with a larger group of agents, it offers more outcomes at each round to increase the chance of reaching an agreement;

(e) The responders' strictness coefficients d_i, $i \in R_I$: Dealing with responders who are more strict in their preferences motivates the initiator to offer more agreements at each round to increase the chance of reaching an agreement.

(f) The initiator's budget $b_I(t)$: As discussed in "Properties of the mechanism", the initiator always chooses an offer size L such that the maximum budget needed for selecting a feasible agreement is below its available budget. The maximum budget needed for selecting a feasible agreement is increasing in terms of the offer size. Therefore, an initiator with a higher budget is able to offer more agreements at each round.

To find the initiator's optimal offering strategy, we formulate the problem as a Markov Decision Process (MDP). MDPs are the standard formalism for learning optimal sequential decision making in stochastic domains. The present problem is stochastic from the initiator's point of view, as it is not aware of the responder's preferences and hence their responses to the offers it makes.

In an MDP, the environment is modeled as a set of states and actions that can be performed to control the system's state. The effectiveness of an action can be measured by its impact on the system state and the instantaneous reward it provides. In our problem, we define a state s by the number of agreements that are feasible for the initiator and have not been discussed yet. We also define two additional states: *Succ* and *Fail*, where *Succ* (*Fail*) means that the agreement (disagreement) arises. These two states are terminal states or absorbing states that terminate a negotiation round. Let $A_s = \{1, \dots, \min\{s, \gamma(b_I)\}\}$ represent the action set at a non-terminal state $s \in \mathbb{Z}_{++}$, where $\gamma(b_I)$ is the maximum offer size that is affordable for an initiator with budget b_I. The action sets corresponding to terminal states are empty, i.e. $A_s = \emptyset$, for $s = Succ, Fail$.

By choosing L as the offering size in a state $s \in \mathbb{Z}_{++}$, the system makes a transition from s to a new state s', based on a probability transition function $P_L(s, s')$. If at least one of the offers is feasible for all the responders, the state transits to the *Succ* state. Otherwise, the state transits to the *Fail* state, if either no options remained, i.e., $L = s$, or the rules do not allow the initiator to move forward, i.e. $L < L_{\min}$. If neither of these conditions is satisfied, the state transits to state $s - L$. Thus, we have

$$P_L(s, s') = \begin{cases} 1 - (1 - \prod_{i \in R_I} (1 - d_i))^L, & \text{If} \quad s' = Succ, \\ (1 - \prod_{i \in R_I} (1 - d_i))^L, & \text{If} \quad s' = Fail, (L = s \text{ or } L < L_{\min}), \\ (1 - \prod_{i \in R_I} (1 - d_i))^L, & \text{If} \quad s' = s - L, L < s, L \geq L_{\min}, \\ 0, & \text{Otherwise}, \end{cases} \tag{23}$$

where $\prod_{i \in R_I} (1 - d_i)$ is the probability that a specific agreement is feasible for all responders.

The last major element of an MDP that remains to be defined is the reward function. The reward function $R_L(s, s')$ specifies the initiator's immediate reward when the state transits from s to s' as a result of an action L. Note that since the DSM mechanism is individually budget balanced (Lemma 4), the initiator's expected point income is zero irrespective of the offer size L. Therefore, the initiator's offering strategy at a negotiation $\mathcal{N}(t)$ has no effect on its negotiation power in the future negotiations. Thus, to choose the optimal offering strategy, the initiator should only focus on the instant profit it gains at the current negotiation. The initiator gains no positive reward if its action does not lead to a terminal state. Therefore, for $s' \neq Succ, Fail$, its reward only consists of the communication cost β_I, and the privacy leakage (9), i.e.

$$R_L(s, s') = -\beta_I - LC, s' \neq Succ, Fail, \tag{24}$$

where $C = \sum_{i \in R_I} \theta_{Ii} \min(d_I, 2(1 - D_{I,i}) - d_I)$ is the privacy loss of offering a single agreement. When the negotiation fails and disagreement happens, the initiator incurs an infinite cost. We model this fact by considering

$$R_L(s, Fail) = -K - \beta_I - LC, \tag{25}$$

where K is an arbitrarily large number. Parameter K has no other role in the results than to prevent the initiator from offering less than $L_{\min}$ agreements, when $s \geq L_{\min}$. Therefore, its exact value is not important.

Deriving the reward function for $s' = Succ$ is challenging. The reason is as follows. Once the agreement happens, the initiator's utility is computed according to (18) as a function of the final outcome $O_n^{j^*}$. The final outcome $O_n^{j^*}$ is the solution of the optimization problem (20), which should be solved by the initiator at the assessment stage. The optimal solution of (20) depends on the responders' scores and the payment function $r(.)$. Therefore, computing the probability that any specific outcome O_n^j becomes the solution of the optimization problem (20) is complex. Due to this complexity, the initiator often fails in practice to derive the reward function analytically, and hence finds itself faced with an MDP whose reward function is not known to it.

The most attractive way to solve MDP problems whose model is not completely known is reinforcement learning (RL) [32, 87]. In "Optimal RL-based offering strategy, we detail a RL technique to derive the optimal offering strategy for the initiator. This strategy gives a theoretical upper bound for the utility that the initiator can achieve in the negotiation. However, it is not very practical in dynamic settings like piMAP. The reason is that the RL techniques need a learning phase to calculate the unknown parameters of the model. However, in an incremental problem the parameters change continuously over time. Thus, there is no time for the initiator to first learn and then perform an action. In such problems, decisions must be made instantly and swiftly at the time they are called for. Considering this fact, in "Heuristic policy", we construct a heuristic policy with no need for a learning phase that performs very close to the optimal one. This algorithm is based on an approximate calculation of the reward function $R_L(s, s')$ for $s' = Succ$.

Optimal RL-Based Offering Strategy

Reinforcement learning is a general class of algorithms in the field of machine learning that aim to achieve an optimal policy when complete information about the environment is not available. RL techniques allow an agent to interact with the environment to gain knowledge about how to optimize its behavior.

One of the most popular RL methods is Q-learning where the agent estimates a Q value for every possible state-action pair (s, L), indicating the utility of performing action L in state s [94]. More specifically, assume the agent observes in m^{th} round of learning that action L executed at state s results in state s' and some immediate reward r. The whole observation is performed to update the Q value as follows:

$$Q(s, L) = (1 - \alpha_m)Q(s, L) + \alpha_m(r + \max_{L'} Q(s', L')), \tag{26}$$

where the learning rate α_m determines to what extent the old information will be overridden by the newly acquired information at round m. At any time, the Q values suggest a policy for choosing actions, namely the one which, in any state s, chooses action L that maximizes $Q(s, L)$. However, the following theorem defines a set of conditions under which the repeated application of this update equation eventually yields Q values that give rise to a policy that maximizes the expected cumulative reward. This theorem is proved in [94].

Theorem 5 *Let $m^i(s, L)$ denote the index of the ith time that action L is tried in state s. Given bounded rewards and learning rates* $0 \leq \alpha_m \leq 1$, *and*

$$\sum_{i=1}^{\infty} \alpha_{m^i(s,L)} = \infty, \quad \sum_{i=1}^{\infty} \left[\alpha_{m^i(s,L)}\right]^2 < \infty, \forall s, L, \tag{27}$$

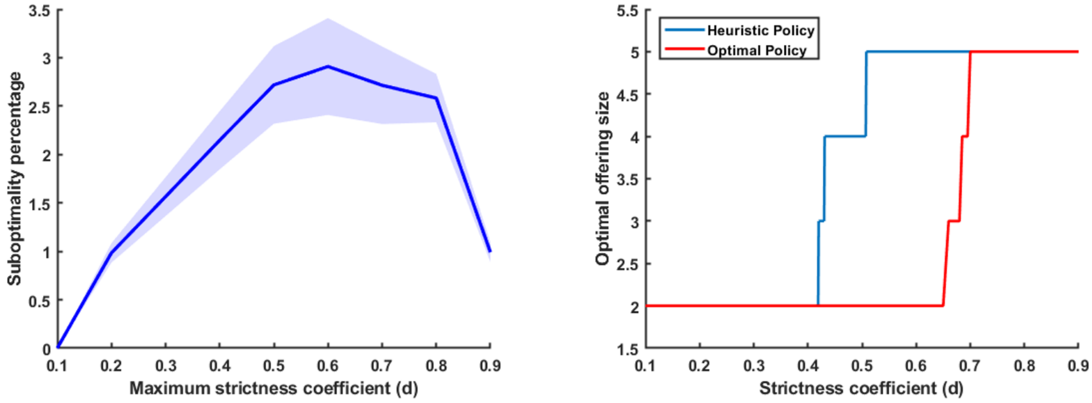
then the Q-learning solution converges to the optimum with probability 1.

Using the Q-learning algorithm, we derive the optimal offering strategy π^* for the initiator. However, RL techniques require a large amount of exploration of all actions and states for proper convergence to the optimal policy. In the present problem, the optimal policy depends on the negotiation specifications, such as the number of participants and the values of different agreements for the initiator. These parameters usually change from one negotiation to another. Therefore, in practice, the initiator cannot interact with a fixed environment repeatedly and learn the optimal policy.

To solve this issue, in "Heuristic policy", we propose a heuristic offering strategy that, as we shall demonstrate in "Numerical results", performs closely to the optimal policy π^*. This strategy can be derived analytically and does not need any exploration.

Heuristic Policy

A key component in the construction of the heuristic policy is to find a good approximation for the reward function of the



(a) Suboptimality of the heuristic policy.

(b) A case where heuristic policy finds a non-optimal solution. $|A(t)| = 6$, $\beta_I = 0.15$, $\theta_{Ii} = 0.02$, $d_i = d$, $V_I = (100, 95, 90, 75, 10)$, $L_{min} = 2$, $L_{max} = 5$.

Fig. 2 Accuracy of the heuristic policy

MDP proposed in "Initiator's optimal offering strategy". As previously noted, the reward function is hard to derive due to the complexity of calculating the probability $(P1)$ that any specific offer O_n^j becomes the solution of the optimization problem (20). In this section, we propose a low-complexity approximation for this probability and derive the reward function accordingly. Then, we solve the approximated MDP via a typical approach based on dynamic programming, to obtain a heuristic offering strategy for the initiator.

In more detail, we approximate $(P1)$ by assuming that all offers have an equal chance of being selected as the solution to problem (20). Therefore, the probability $(P1)$ can be approximated as

$$P1 \approx \frac{1-(1-\prod_{i\in R_I}(1-d_i))^L}{L}. \tag{28}$$

Equation (28) is not an equality, as in reality, the offers with higher valuations for the initiator have higher chances of being selected at the assessment stage. However, our assumption leads to a good approximation, because the optimal policy π^* selects the agreements with relatively close valuations for being offered at each round. This observation is our main motivation for this approximation and the main reason it leads to a good result.

Using (28), it can be shown that the reward function $R_L(s, Succ)$ for an initiator with L best agreements $O_n^1, \dots, O_n^L$, can be approximated as

$$R_L(s, Succ) \approx \frac{1}{L}\sum_{j=1}^{L} V_I(O_n^j) - \beta_I - LC, \tag{29}$$

when $L \le L_{\max}$. This reward is the average valuation of the initiator's L top offers minus the communication cost and the privacy loss.

Reward function (29) completes the definition of the approximated MDP. The initiator could solve this MDP analytically to derive a suboptimal offering strategy $\hat{\pi}$. In "Numerical results" we present performance results based on numerical simulations and show that $\hat{\pi}$ does indeed give results very close to the optimal offering strategy π^*.

Remark 1 As stated in "Strategic multi-agent negotiation over a communication graph", we assume that agents have a rich history of interactions with their neighbors and hence have accurate estimations of their strictness coefficients d_i, $i \in \mathcal{V}$. Therefore, the initiator is able to calculate the transition probabilities (23) accurately. Now, consider a situation where an initiator joins a network recently and does not have sufficient interactions with its neighbors to obtain an accurate belief about their strictness coefficients. In this case, the initiator can use its beliefs to derive an estimate, but not the precise, transition probabilities.

However, this does not prevent the initiator from using the Q-learning technique. The reason is that in Q-learning, the initiator does not use its prior information about the transition probabilities, but it learns both the transition probabilities and the rewards in a trial and error fashion. Therefore, Q-learning helps the newly-arrived agents to not only learn their optimal offering strategies, but also derive a good estimation of their neighbors' strictness coefficients. Once the estimation is complete, the initiator can switch to the heuristic policy to derive its future offering strategies with less complexity.

It is important to note that the desirable properties of the DSM mechanism discussed in "Properties of the mechanism" do not depend on the common knowledge assumption on the strictness coefficients. This assumption has not been used in the proofs and hence the DSM mechanism satisfies faithfulness, individual rationality, and individual budget balanced even if the agents have no precise estimation on the strictness coefficients.

Numerical Results

In "Properties of the mechanism", we proved that the DSM mechanism is faithful and gives sufficient incentives to the agents to follow its rules. In this section, we study through numerical simulations the performance of the DSM mechanism when agents follow the rules. We evaluate the mechanism based on a set of metrics including privacy leakage, speed of convergence, and expected social-welfare. This is done is "Evaluation of the DSM mechanism". In "Evaluation of the heuristic offering strategy", we study the performance of the heuristic offering strategy $\hat{\pi}$ proposed in "Heuristic policy" and show by numerical testing that the heuristic offering strategy $\hat{\pi}$ achieves almost the same performance compared to the optimal strategy π^*. This result allows us to consider $\hat{\pi}$ as the initiator's offering strategy for the performance evaluation of "Evaluation of the DSM mechanism".

Evaluation of the Heuristic Offering Strategy

In Fig. 2a, we compare the expected utilities the initiator can achieve by adopting the optimal offering strategy π^* and the heuristic offering strategy $\hat{\pi}$. The x-axis shows the maximum strictness coefficient d of all participants. For each fixed value of d, we generate 10000 negotiation instances with $\max_{i \in A(t)} d_i = d$, where the effective parameters such as the number of participants $|A(t)| \in \mathbb{Z}_{++}$, the number of possible agreements $o(t) \in \mathbb{Z}_{++}$, the initiator's bargaining cost $\beta_I \in \mathbb{R}_+$, and the initiator's privacy sensitivity $\theta_{Ii} \in \mathbb{R}_+$, $i \in R_I$, are selected randomly and uniformly from the corresponding intervals. The valuation vector of each participant i is constructed based on its strictness coefficient d_i. For each outcome $O_{t,k}$, $V_i(O_{t,k})$ takes the value $-\infty$ with probability d_i, meaning that the outcome $O_{t,k}$ is infeasible for agent i; otherwise the outcome is feasible for i and hence $V_i(O_{t,k})$ is uniformly selected from the set of all positive real numbers.

For each negotiation instance, we compute the suboptimality of the solution found by adopting the heuristic policy $\hat{\pi}$ to the solution found by π^*. Figure 2a shows the mean and standard deviation of the suboptimality gaps for different strictness coefficients d. The suboptimality gap is below 3.5% for all $d \in [0, 1]$, showing that the heuristic solutions generated are close to the optimum. Therefore, it is almost without loss of optimality for the initiator to adopt policy $\hat{\pi}$ for making the offers.

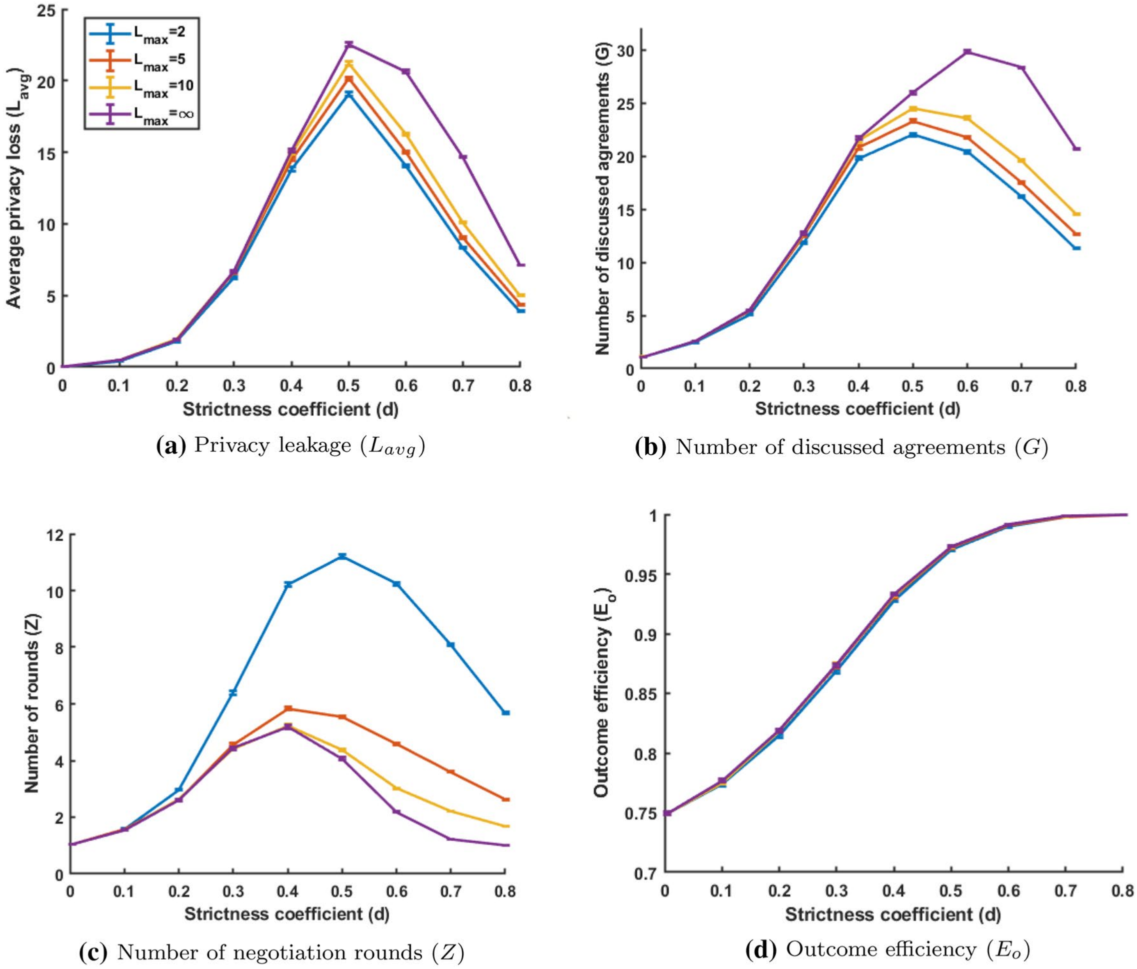
We can see from Fig. 2a that the average suboptimality gap varies from very small values at the ends of the spectrum to a maximum value 2.9% near the middle of the strictness range. The reason for this behavior is as follows. When d is close to 0, the probability that an arbitrary agreement satisfies the responders' constraints are close to 1. Therefore, both the optimal and heuristic policies recommend the initiator to be thrifty and offer only $L_{\min}$ agreements to the responders. When d approaches 1, there exists very few (if any) agreements that can satisfy all the responders' constraints. In this case, both policies advise the initiator to be generous and offer $L_{\max}$ agreements at each round, to expedite the search for the only feasible agreement(s). Therefore, the heuristic and optimal policies have similar behaviors at the ends of the spectrum. However, their behaviors can be different in the middle. The reason is as follows. As discussed in "Heuristic policy", the heuristic strategy $\hat{\pi}$ is based on estimating the reward function by using two low-complexity approximations. The error of these approximations may cause some delay or advance in following the optimal policy's pattern. Figure 2b shows a case where an advance occurs. In this case, the optimal offering strategy is of threshold type with three thresholds $d_1^* = 0.65$, $d_2^* = 0.68$, and $d_3^* = 0.695$. However, the heuristic policy estimates these thresholds as $\hat{d}_1 = 0.419$, $\hat{d}_2 = 0.43$, and $\hat{d}_3 = 0.507$, respectively. Such inaccuracies in estimating the optimal policy's pattern degrades the heuristic policy's performance in the middle of the strictness range. However, as we can see from Fig. 2a, the overall error caused by such inaccuracies is not very significant, which means that cases similar to what has been shown in Fig. 2b do not happen frequently.

We have discussed in "Heuristic policy" that deriving the heuristic policy is much less complex than deriving the optimal policy, since it does not need a learning process like reinforcement learning to calculate the unknown parameters of the models. Now, Fig. 2a shows that the performance of the heuristic policy is very close to the performance of the optimal policy and competes with it very well. Therefore, through the rest of the numerical investigations, we assume that the initiator always adopts the heuristic offering strategy.

Evaluation of the DSM Mechanism

In this section, we conduct empirical studies to evaluate our mechanism based on a set of metrics including privacy leakage, speed of convergence, and social-welfare. We also evaluate the role of parameters $L_{\min}$ and $L_{\max}$ on the performance of mechanism $\Gamma = (L_{\min}, L_{\max}, D, r(.))$.

For these experiments, we consider a piMAP where the needs for different negotiations arise over time. Similar to



(a) Privacy leakage (L_{avg})

(b) Number of discussed agreements (G)

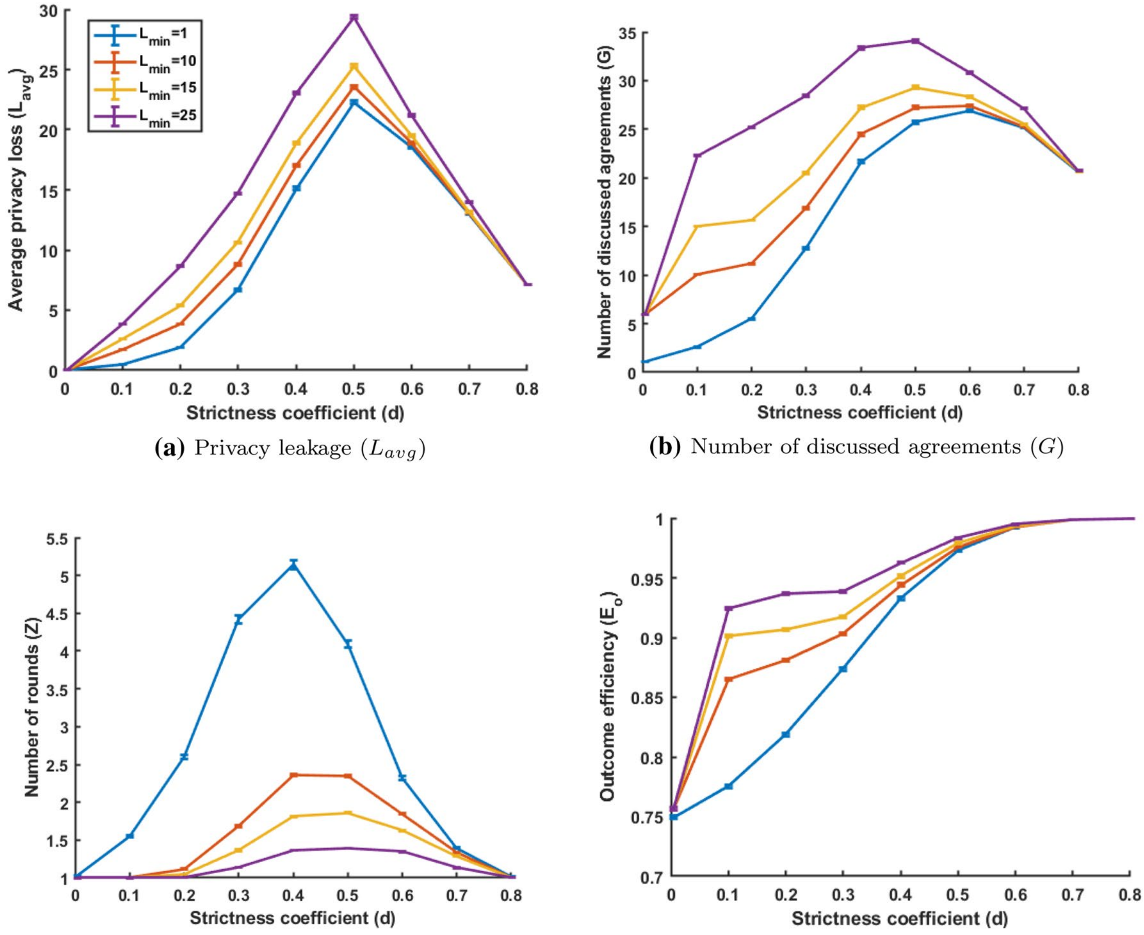
(c) Number of negotiation rounds (Z)

(d) Outcome efficiency (E_o)

Fig. 3 Performance evaluation: role of parameter $L_{\max}$

the settings used in [31, 93], we set the number of possible outcomes for each negotiation to 100 and the number of agents who participate in each negotiation is uniformly distributed from 5 to 10. To explore a broad range of reasonable negotiation strategies, we assume that the agents' bargaining costs and privacy sensitivities are drawn uniformly from [0.1, 10]. Bargaining cost β_i and privacy sensitivity θ_{ij} determine the weights that the convergence speed and privacy leakage, respectively, receive in agent i's utility function. The range 0.1 to 10 for these parameters means that the importance of convergence speed and privacy leakage for the agents is comparable to the importance of the final agreement's value. Different valuation functions with $V_i \in [0, 100]$ are tested in the experiments and all results are averaged over 10000 cases to ensure statistical robustness (p value less than 10^{-5}).

In Fig. 3, we consider a symmetric network where the agents have the same strictness coefficients d and study the role of parameter $L_{\max}$ on the performance of the mechanism. For this study, we have plotted (1) the agents' average privacy leakage (Fig. 3a), (2) the number of offers that are made by the initiator (Fig. 3b), (3) the number of negotiation rounds for the successful creation of an agreement (Fig. 3c), and (4) the outcome efficiency (Fig. 3d), when the agents employ a DSM mechanism with $L_{min} = 1, D = 3$, and $L_{\max} \in \{2, 5, 10, \infty\}$. In each figure, the error bars show the standard error of the mean, which depicts the accuracy of the results. Figure 3a shows the average privacy leakage for the agents who are interested in maximizing the others' uncertainties about their preferences, i.e. $D_{i,j} = 0.5$, for all i, j (see (5)). As the strictness coefficient d increases, i.e., agents are more strict in their preferences and hence fewer agreements can satisfy their requirements, the average privacy leakage first increases and then decreases. The intuition behind this result is twofold. First, according to (9)–(10), discussing each possible outcome induces $\min(d, 1-d)$ privacy loss to a responder and $(|A(t)| - 1)\min(d, 1-d)$ privacy loss to the initiator. Therefore, for each negotiation, the average privacy loss over all agents is



(a) Privacy leakage (L_{avg})

(b) Number of discussed agreements (G)

(c) Number of negotiation rounds (Z)

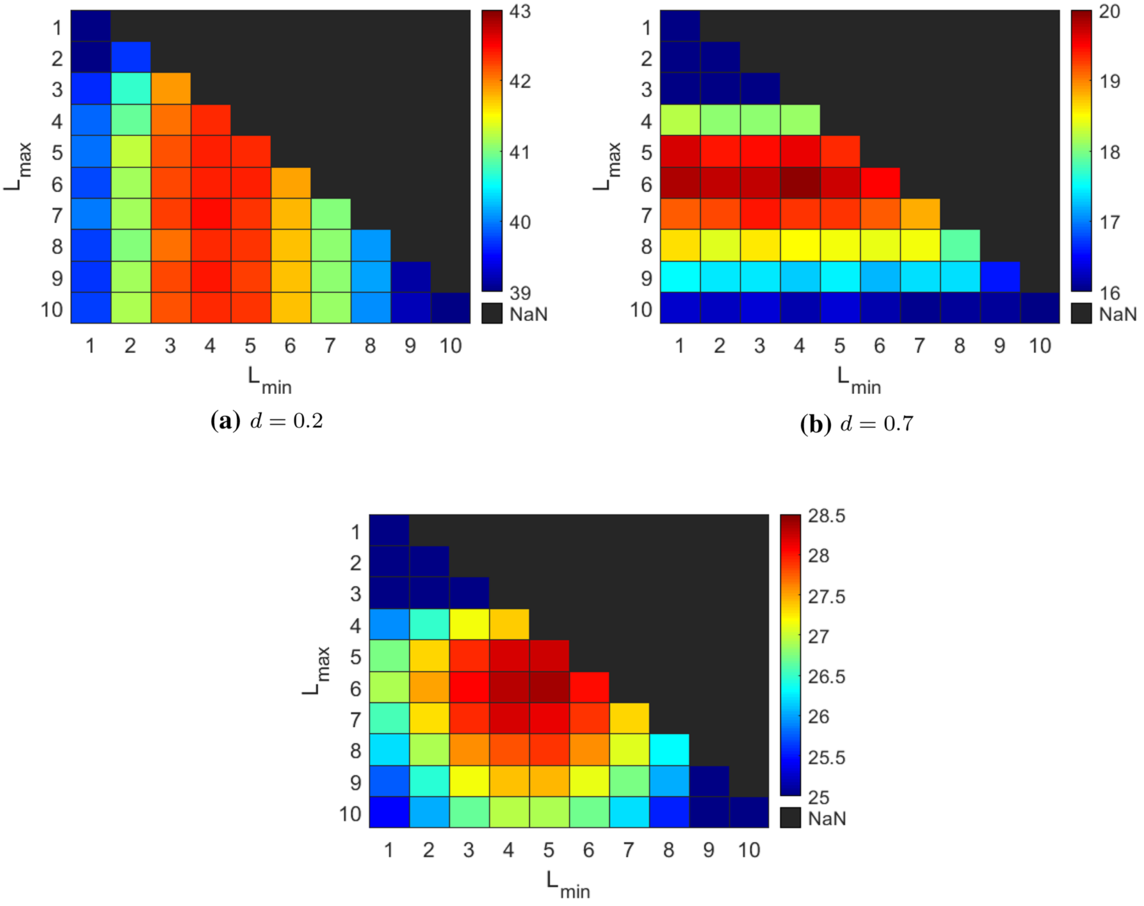
(d) Outcome efficiency (E_o)

Fig. 4 Performance evaluation: role of parameter $L_{\min}$

$$L_{avg} = \frac{1}{|A(t)|} \sum_{i,j \in A(t)} L_{i,j} = \frac{2(|A(t)| - 1)}{|A(t)|} \min(d, 1 - d)G, \tag{30}$$

where G is the number of discussed outcomes. Function $\min(d, 1 - d)$ is monotonically increasing on the interval [0, 0.5] and monotonically decreasing on the interval [0.5, 1]. Therefore, the same behavior is expected to be observed from the average privacy loss. The second intuition comes from the behavior of function G which is reported in Fig. 3b. We can see from this figure that the average number of discussed outcomes in a negotiation starts from 1 when $d = 0$, reaches a maximum at an intermediate value of d and then drops. Such a result comes from two opposite trends. On one hand, finding an agreement that is feasible for agents who have higher strictness coefficients requires more negotiation. On the other hand, a more intransigent initiator is satisfied with a smaller number of agreements and hence has fewer candidate agreements to offer. The tradeoff between these two effects describes a first increasing and then decreasing characteristics of both G and $\frac{1}{|A(t)|}\sum_{i,j\in A(t)} L_{i,j}$.

Figure 3a shows that increasing the upper bound L_{max} of the number of offers that are allowed to be made at each round, results in an increase in the privacy leakage. However, based on Fig. 3c, the increase of parameter $L_{\max}$ can promote the speed of convergence. For a larger value of $L_{\max}$, the initiator is allowed to make more offers at each round to speed up the negotiation. However, since the initiator is privacy-sensitive it does not unleash all its freedom to do so. Therefore, even in the absence of an upper bound (i.e. $L_{\max} = \infty$), the initiator does not offer all its candidate agreements in one negotiation round. For $L_{\max} = \infty$, the agents can reach an agreement at one round when d approaches 0 or 1, but it takes about 5.18 rounds on average for them to make an agreement when $d = 0.4$. This is because when $d = 0$, the agents are totally flexible and accept any agreement; therefore, the initiator makes only 1 offer and the responders accept that. When d approaches 1 the privacy loss is small



(a) $d = 0.2$

(b) $d = 0.7$

(c) Uniform d

Fig. 5 Performance evaluation: optimal parameters

(see Fig. 3a); therefore, the initiator is not worried about its privacy and makes all its offers at one round. However, for middle-range values of d in which the privacy leakage is significant, the initiator prefers to be prudent and offer agreements gradually, even if it is allowed to offer them en masse.

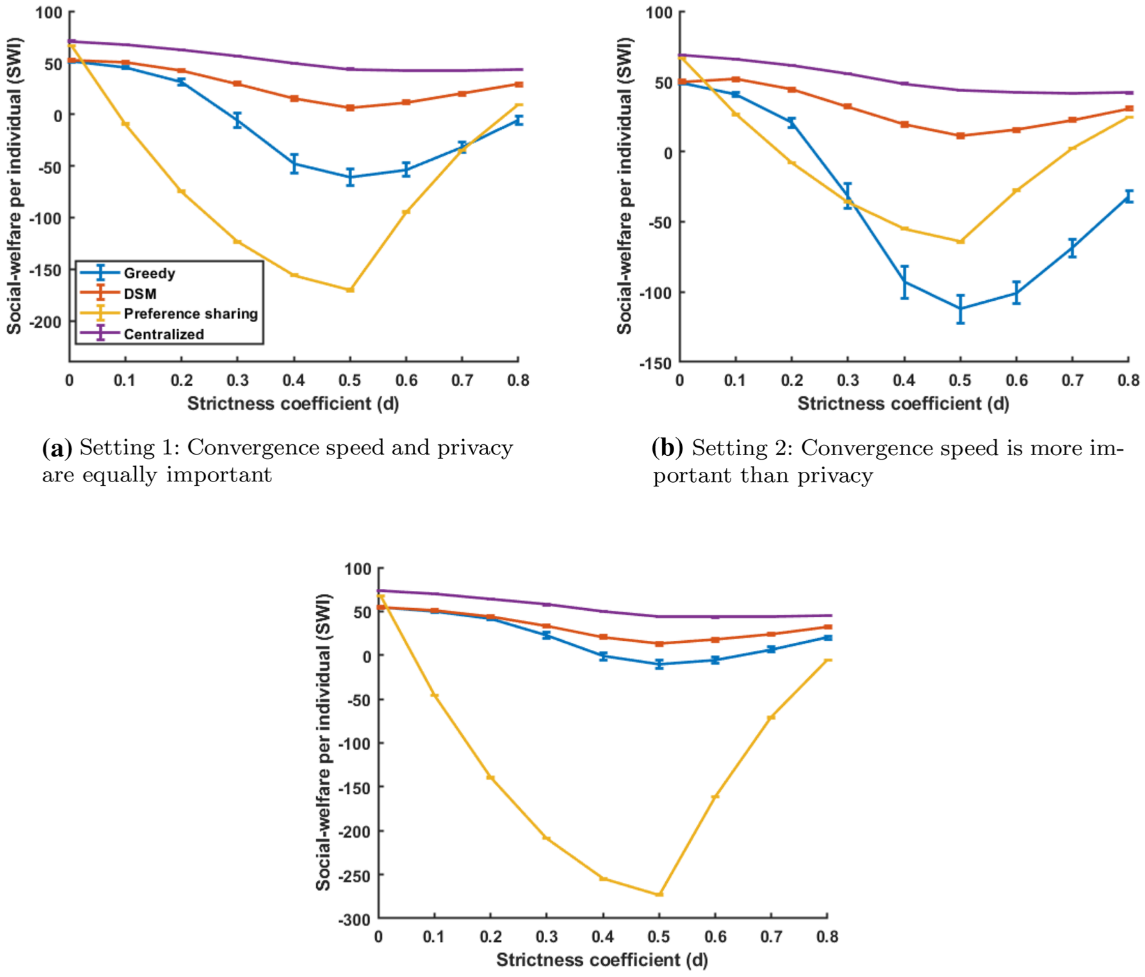
In Fig. 3d, we plot the outcome efficiency versus the strictness coefficient d for different upper bounds $L_{\max}$. The outcome efficiency of negotiation $\mathcal{N}(t)$ is defined as

$$E_o = \frac{\sum_{i \in A(t)} V_i(O_{t,k})}{\max_{k'} \sum_{i \in A(t)} V_i(O_{t,k'})}, \tag{31}$$

where $O_{t,k}$ is the negotiation outcome. The outcome efficiency is the ratio of the aggregate valuation of the selected outcome for the agents to the aggregate valuation of the best socially-accepted outcome. Figure 3d shows that irrespective of the upper bound's value $L_{\max}$, the outcome efficiency monotonically increases with d and reaches the limit 1 at a certain value of d. Based on the results above, we can conclude that the upper bound $L_{\max}$ can tune the tradeoff between convergence speed and privacy leakage, without affecting the outcome efficiency.

In Fig. 4, we study the role of parameter $L_{\min}$ on the performance of the DSM mechanism. For this study, we have fixed the upper bound $L_{\max}$ to be 30. This rather high value was chosen to provide a wide range of possible lower bounds $L_{\min}$. We can observe in Fig. 4, that increasing the lower bound of the number of offers the initiator is allowed to make increases the privacy loss (see Fig. 4a), decreases the number of negotiation rounds (see Fig. 4c), and enhances the outcome efficiency (see Fig. 4c). Therefore, the lower bound $L_{\min}$ is able to tune the relative importance of all three metrics.

Two important differences should be noted between the roles of the lower bound $L_{\min}$ and the upper bound $L_{\max}$. First, the upper bound $L_{\max}$ has no impact on the outcome efficiency; thus, $L_{\min}$ is the only parameter that can balance the tradeoff between outcome efficiency and the other two metrics. Second, the lower bound $L_{\min}$ has a more significant impact on the privacy leakage and convergence speed than the upper bound $L_{\max}$. For example, increasing $L_{\min}$



(a) Setting 1: Convergence speed and privacy are equally important

(b) Setting 2: Convergence speed is more important than privacy

(c) Setting 3: Privacy is more important than convergence speed

Fig. 6 Performance evaluation: social-welfare

from 1 to 25 increases the peak of average privacy leakage by 57%, while this increment is only about 18% when $L_{\max}$ rises from 2 to infinity. This is because when $L_{\max}$ increases, the initiator enlarges its offer sets only if it finds it profitable. However, when $L_{\min}$ increases, the initiator is obliged to make at least $L_{\min}$ offers at each round to keep the chance of continuing alive.

In Figs. 3 and 4, we have studied the role of $L_{\min}$ and $L_{\max}$ on some individual metrics and shown that these bounds may have opposite effects on each of the metrics. Now, we are going to investigate the role of these parameters on the overall performance of the mechanism, which is quantified by the social-welfare per individual (SWI) [17]:

SWI measures the average utility realized by the agents and acts as an indicator of the satisfaction level an algorithm provides to the network users. This metric is the normalization of the social-welfare [7] to the number of agents and is designed for evaluating the performance of algorithms over networks with different number of users [17].

In Fig. 5, we illustrate the social welfare per individual achieved by DSM with different L_{min} and L_{max}, for (1) a network of flexible agents with $d = 0.2$ (Fig. 5a), (2) a network of strict agents with $d = 0.7$ (Fig. 5b), and (3) a network of agents with uniformly selected strictness coefficients (Fig. 5c). We can see that when agents have low strictness coefficients, the lower bound $L_{\min}$ has a more significant

$$SWI = \frac{1}{|A(t)|}\sum_{i\in A(t)} U_i(.) = \frac{1}{|A(t)|}\sum_{i\in A(t)} [V_i(O_{t,k}) - \beta_i N - \sum_{\substack{j \in A(t)\\ j \neq i}} \theta_{ij} L_{i,j}(M_{i\to j})]. \tag{32}$$

effect on SWI than the upper bound $L_{\max}$. This observation is evident from the range of changes in each row and column of Fig. 5a. However, when agents become more strict in their preferences (Fig. 5b), the role of $L_{\max}$ becomes more significant. Intuitively, in negotiating with flexible agents, the initiator is not interested in making many offers itself. Therefore, as long as the upper bound is not very low, its changes do not affect the initiator's decision-making. However, when the initiator is negotiating with strict agents, it prefers to approach the upper bound to enhance the chance of reaching an agreement. Thus, the mechanism's performance is very dependent on the upper bound's value. We can observe that when $d = 0.2$, the best performance is achieved by a DSM mechanism with lower bound $L_{\min} \in \{3, 4, 5\}$. When $d = 0.7$, the optimal performance is obtained when $L_{\max} \in \{5, 6, 7\}$. When the negotiators' strictness coefficients are not fixed, but take random values within $[0, 1]$ with a uniform distribution (Fig. 5c), both lower and upper bounds have significant impacts on SWI and the optimal performance is achieved at $L_{max} = 6$ and $L_{min} = 5$.

So far, we have studied the roles of different parameters on the performance of our proposed mechanism and found their optimal values. In Fig. 6, we compare the performance of the DSM mechanism with the following benchmarks:

- Centralized mechanism (similar to [23, 33]): Suppose that there is a trusted central entity to whom all agents share their preferences without any concern about privacy leakage. Then, the central entity can choose the agreement that maximizes the social-welfare per individual. This solution provides the maximum theoretical SWI that could be achieved, if a central trusted entity existed. Hence, it serves as an upper bound for the performance of a mechanism.
- Greedy mechanism (similar to [45]): The initiator acts in a greedy fashion to improve its own satisfaction of the final outcome. It sorts the agreements from highest to lowest values and offers them one by one until an agreement that is feasible for all agents is found.
- Preference-sharing mechanism (similar to [103]): The initiator and the responders share all their preferences with each other and then the initiator chooses the agreement that maximizes the SWI. We can evaluate how much privacy leakage can be saved compared to this mechanism. The preference-sharing mechanism differs from the centralized solution in that the agents do not share their information with a trusted entity, but with each other.

We show the comparison in three different settings. In Fig. 6a, we considered a network where the range of parameters β_i and $\{\theta_{ij}\}$ are the same (Setting 1). Both of these parameters take random values in [0.1, 10] with a uniform distribution. This means that, on average, the agents give equal importance to both privacy and speed of convergence. In Fig. 6b, we consider a setting where agents are more concerned about the speed of convergence (Setting 2). To this end, we assume that the bargaining costs and privacy sensitivities are drawn uniformly from [5, 10] and [0.1, 5], respectively. Fig. 6c shows the other way around when β_is and $\{\theta_{ij}\}$s are drawn from [0.1, 5] and [5, 10], respectively, and hence the privacy has more importance (Setting 3). We can see that in all cases, the DSM mechanism is significantly superior to the preference-sharing mechanism, except when $d \leq 0.04$. The superiority comes mainly from better privacy protection. However, when d is almost 0, even before the start of negotiation, the agents are almost certain about the feasibility of all agreements for all negotiators. Therefore, revealing information does not reduce the agents' privacy. In this case, sharing the preferences has no negative impact on the agents' utilities and hence is superior to the DSM mechanism due to its higher convergence speed.

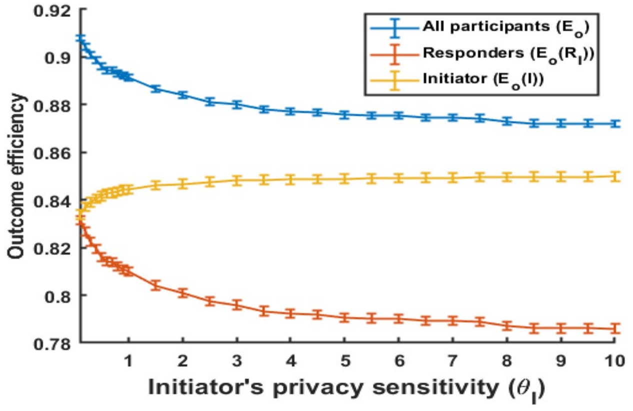


Fig. 7 Impact of initiator's privacy sensitivity on the final outcome

We can see from Fig. 6 that the DSM mechanism is always superior to the greedy algorithm, however the extent of its superiority depends on the relative importance of privacy and speed of convergence. The main advantage of the DSM mechanism over the greedy algorithm is the convergence speed. Therefore, the superiority is higher when agents are more concerned about speed (i.e., Fig. 6b) and is lower when speed is less important for the agents (Fig. 6c). We observe from Fig. 6 that DSM outperforms the existing distributed mechanisms by up to 67.5%, 75.2%, and 23.4%, in Settings 1–3, respectively. Moreover, the proposed mechanism achieves up to 75%, 73.2%, and 79% of the theoretical upper bound in Settings 1–3, respectively.

Finally, as the last experiment, we examined the impact of the initiator's privacy sensitivity on the negotiation process and the outcome. In Fig. 7, we plot the outcome efficiencies of (1) all the participants (blue line), (2) the responders (red line), and (3) the initiator (yellow line), versus the initiator's privacy sensitivity $\theta_{Ii} = \theta_I, \forall i \in R_I$. The outcome efficiency of a subgroup X of participants, which we denoted by $E_o(X)$, is defined in a similar way as in (31) replacing $A(t)$ by X. This metric indicates how close the selected outcome is to the ideal outcome for subgroup X. We can observe in Fig. 7 that the initiator's privacy sensitivity has inverse relations with E_o and $E_o(R_I)$, while it has a direct relation with its own outcome efficiency $E_o(I)$.

For interpreting this behavior, let's take a look at two extreme cases that could happen in the DSM mechanism with $L_{\min} = 1$ and $L_{\max} = \infty$. First, consider an initiator with a very high privacy sensitivity. This type of initiator is not willing to share much information with others and hence offers the options one by one. Since the initiator offers the options in order of its own satisfaction, the final outcome of this case would be the best option for the initiator that is feasible for all responders. In this case, the initiator's privacy awareness prevents it from taking socially-efficient decisions. Now consider the other end of the spectrum. A non-privacy concerned initiator offers all of the available options at one round and then selects the option that balances the tradeoff between its own satisfaction with the final outcome and the number of points it needs to spend to satisfy the responders. The number of points is inversely proportional to the responders' satisfaction from the final outcome. Therefore, a non-privacy concerned initiator takes the responders' satisfaction into consideration and selects an outcome that is better for the responders. The intuition behind these extreme cases justifies the behavior observed in Fig. 7. As the initiator's privacy sensitivity decreases, the extent to which the responders' preferences influence the initiator's decision making improves. Therefore, the final outcome shifts from the initiator's optimal one to the one that provides the same level of satisfaction to both the initiator and the responders.

Conclusions and Future Work

We have studied the distributed piMAP where a set of self-interested and privacy-preserving agents are required to make incremental agreements. Using an artificial currency, we developed a distributed multi-round negotiation mechanism which enables self-interested agents to reach a socially-desired agreement with limited information leakage. Through theoretical analysis, we proved that the DSM mechanism is (1) faithful, meaning that it gives sufficient incentives to the agents to follow the rules,(2) individually rational, meaning that it is dominant strategy for the agents to participate in the mechanism, and (3) individual budget balanced, meaning that each agent's expected payment is zero at all on- and off-equilibrium paths.

The DSM mechanism induces negotiation games with incomplete information among the agents. We derived the optimal negotiation strategies of both the initiator and the responders in such games. Specifically, we proposed a RL algorithm that allows the initiator to learn its optimal negotiation strategy, in a trial-and-error fashion. To avoid the complexity and time consuming nature of the RL algorithm, we also propose a simple heuristic strategy for the initiator in which the game's unknown parameters are not learned, but approximated by analytic expressions. We showed by numerical simulations that this heuristic strategy performs very closely to the optimal policy. Fixing the agents' strategies, we studied the performance of the proposed mechanism in terms of outcome efficiency, privacy leakage, and convergence speed, through comprehensive simulation experiments. In particular, we showed that the mechanism has two tuning parameters that can be used to adjust the tradeoff among the above-mentioned performance metrics. By setting these parameters suitably, the social welfare of the DSM mechanism significantly surpasses that of the best available distributed mechanisms by 67.5%.

To apply the DSM mechanism in real-life applications, we need an infrastructure that (1) provides a communication platform for message transmission among agents. Using this platform, the agents can send/receive offers to/from their neighbors. They can also give scores to the offers they receive; (2) assigns a fixed budget of convenience points to each agent upon arrival and then keeps track of its budget; (3) deducts the scoring cost from the agents' budgets; and (4) allows agents to convey their points to others. The initiator can use this feature to award the promised rewards to the responders. Using this infrastructure, the DSM mechanism can help the agents to reach socially-desirable agreements. The properties of the DSM mechanism discussed in the paper can guarantee its good performance in real-life applications. First of all, the faithfulness of the mechanism ensures that even if the agents devote all their computing power to manipulate the negotiation process to their own benefits, they will not find any strategy better than loyalty to the rules. Second, the property of individual budget balance ensures that consecutive negotiations do not lead to the agents losing all their points and negotiation powers. Assigning a sufficiently high number of points to the agents upon arrival can guarantee that their budgets do not vanish, but just fluctuate around the initial budget. With such an infrastructure in place, our mechanism can be used in any real-world applications that require privacy sensitive negotiations (including those mentioned in "Introduction".)

As it stands, the DSM mechanism assumes that agents are non-malicious and do not build their strategies to oppose others or steal their information. Assuming the agents are good-natured, DSM is able to incentivize agents to follow the rules faithfully. However, the presence of malicious agents may make others more cautious and hence more reluctant to follow the rules. For future work, it would be interesting to design a faithful distributed mechanism for piMAPs with malicious agents. Further extensions to DSM include designing an algorithm for reviewing previous agreements if an agent's preference changes significantly. In the current version of DSM, it is assumed that the agents' preferences over different agreements do not change over time. However, in some cases, the agents may need to update their preferences based on the new information they receive over time. In such settings, an agent may need to renegotiate a pre-agreed decision. Therefore, designing a faithful renegotiation procedure would be of interest in future work.

Appendix

A Proof of Theorem 1

We prove this theorem by showing that for any fixed belief profile that satisfies (21), we can design a reward function $r(.)$ such that the constraints (a)–(d) are maintained. In order to satisfy constraints (c) and (d), we design

$$r(D-1, A, F, L) = 0, \quad \forall A, F, L \leq L_{\max}, \tag{33}$$

$$r(s, A, F, L) = \infty, \quad \forall s, A, F, L > L_{\max}. \tag{34}$$

Then, we split the design of reward function $r(s, A, F, L)$ into three cases based on the degree of flexibility F:

Case 1: $F \geq (A+1)^{D-2}$ and mod $(F, (A+1)^{D-2}) = 0$. These two conditions on the degree of flexibility F are satisfied if and only if the responder gives score $D-1$ to all offers that are feasible for it, when at least one such offer exists. In this case, no offer with a score between 1 and $D-2$ exists. Therefore, constraint a can be rewritten as

$$P(D-1, A, F, L)r(D-1, A, F, L) - C(A, F) = 0. \tag{35}$$

Based on (15), giving scores 0 and $D-1$ to offers are free of charge. Thus, we have $C(A, F) = 0$. Substituting this result and (33) into (35) shows that constraint (a) is satisfied. In this case, $r(s, A, F, L)$ is only well-defined for $s = D-1$. Therefore, condition (b) is trivially satisfied.

Case 2: $F \geq (A+1)^{D-2}$ and mod $(F, (A+1)^{D-2}) > 0$. In this case, the responder gives score $D-1$ to at least one, but not all, the feasible offers. In this case, substituting (33) in (21), we can rewrite constraint (a) as follows:

$$\sum_{s=1}^{D-2} P(s, A, F, L)r(s, A, F, L) - C(A, F) = 0, \quad \forall L \leq L_{\max}, \forall A \leq L. \tag{36}$$

For each $L \leq L_{\max}$ and $A \leq L$, equation (36) is a multi-variable linear equation with positive coefficients and positive sum (see (22)), which always has a non-negative solution that satisfies condition (b).

Case 3: $0 < F < (A+1)^{D-2}$. This case happens when the responder gives score $D-1$ to none of the offers. In this case, constraint (a) does not put any restriction on the reward function. However, to satisfy constraint (e), we design $r(s, A, F, L) = r(s', A', F', L)$, where $\mathbf{s}' = (\mathbf{s} + c)\,\mathrm{sign}\,(\mathbf{s})$ and $c = D - 1 - \max\,(\mathbf{s})$. The score vector $\mathbf{s}'$ is defined so as to assign score $D-1$ to at least one offer. Therefore, $F' \geq (A'+1)^{D-2}$ and hence the reward function $r(s', A', F', L)$ is designed as per Cases 1 and 2. The reward function corresponding to F' is decreasing in terms of s'. Therefore, this property is inherited by the reward function corresponding to F.

Discussions made above show that we can design the reward function $r(.)$ such that it satisfies constraints (a)–(e).

B Proof of Theorem 2

We prove the initiator's loyalty to the rules (I1)–(I4).

(I1): $r(s, A, F, L) = \infty$, for $L > L_{\max}$. Therefore, if it offers more than $L_{\max}$ outcomes it cannot get responders' consent for any agreement. This feature forces it to observe the upper limit $L_{\max}$. In addition, in each round of the negotiation, there is a positive probability that none of the offers are feasible for all participants, and hence the need for moving to the next round arises. Therefore, to avoid disagreement, which has a value of negative infinity for the initiator, it prefers to propose at least $L_{\min}$ offers at each round, if possible, to preserve the chance of continuing the negotiation.

(I2): If the initiator offers L options to the responders when it doesn't have enough budget to select a feasible time slot for at least one tuple of the scores it might receive, it removes the chance of selecting some feasible options and hence increases the chance of disagreement. To avoid this situation, the initiator always chooses an offer size such that the maximum budget needed for selecting a feasible agreement is below its available budget.

(I3): There is no returning to the past options. Therefore, to prevent disagreement, the initiator prefers to select a feasible agreement as soon as it finds one. If the initiator finds more than one agreement that is feasible for everyone, it selects the solution of optimization problem (20) to maximize its utility.

(I4): The responders do not keep their commitment unless they receive the promised rewards. Therefore, the initiator awards the promised rewards to the responders to avoid breaking the agreement.

C Proof of Lemma 1

The agents are expected utility-maximizers. Therefore, each responder i's first priority is to reduce the chance of intolerable outcomes that drive its utility to negative infinity, i.e. $V_i(O_n^l) = \infty$. Responder i's utility becomes negative infinity when either (E1) the final agreement is infeasible for it, or (E2) disagreement arises. The disagreement arises only when no agreement receives non-zero scores from all responders. This is because, as we have discussed in rule (I2), the initiator makes its offers such that it can always afford to select a feasible agreement. In the following we show that truth-telling about the feasibility of offers is the best strategy for a responder to minimize $P(E1) + P(E2)$.

Let $\{O_n^1, \dots, O_n^L\}$ denote the initiator's offers in round n of the negotiation. We sort the offers based on the responder j's preferences and assume that offers $O_n^1, \dots, O_n^k$ are feasible and offers $O_n^{k+1}, \dots, O_n^L$ are infeasible for j. In the following two cases, we show that responder j has no incentive to either give a non-zero score to an infeasible agreement O_n^l, $l = k+1, \dots, L$, or give a zero score to a feasible agreement O_n^l, $l = 1, \dots, k$.

Case 1: We show that giving a non-zero score to an infeasible agreement O_n^l, $l = k+1, \dots, L$, increases $P(E1) + P(E2)$ and hence is not in agent j' favor. To show this, we partition the space of all possible preferences among the responders into three disjoint subspaces: (C1) There is at least one agreement, except $O_n^{k+1}, \dots, O_n^L$, that is feasible for all participants; (C2) There is no common feasible agreement when we exclude $O_n^{k+1}, \dots, O_n^L$, and responder j is the only responder who is against agreement O_n^l; (C3) There is no common feasible agreement when we exclude $O_n^{k+1}, \dots, O_n^L$, and agreement O_n^l is infeasible for at least two responders.

It can be seen that if responder j tells the truth about the feasibility of the offers, we have $P(E1) = 0$ and $P(E2) = P(C2) + P(C3)$. If responder j gives a non-zero score to agreement O_n^l which is infeasible for it, the probabilities change to $P'(E1) = P(C2) + \gamma P(C1)$ and $P'(E2) = P(C3)$, where $0 < \gamma \leq 1$ is the probability that O_n^l will be selected in case $(C2)$ when responder j gives a non-zero score to agreement O_n^l. Therefore, we have $P'(E1) + P'(E2) = P(C2) + P(C3) + \gamma P(C1) > P(E1) + P(E2)$ and hence a responder always gives score 0 to an agreement that is infeasible for it.

Case 2: Now, we show that giving a zero score to a feasible agreement O_n^l, $l = 1, \dots, k$, increases $P(E1) + P(E2)$ and hence is not in agent j' favor. In this case, we partition the space into the following disjoint subspaces: (D1) There is at least one agreement, except O_n^l and $O_n^{k+1}, \dots, O_n^L$, that is feasible for all participants; (D2) There is no common feasible agreement when we exclude $O_n^{k+1}, \dots, O_n^L$, and agreement O_n^l is feasible for all responders; (D3) There is no common feasible agreement when we exclude $O_n^{k+1}, \dots, O_n^L$, and agreement O_n^l is infeasible for at least one responder.

If responder j tells the truth about the feasibility of the offers, we have $P(E1) = 0$ and $P(E2) = P(D3)$. However, if it gives a zero score to agreement O_n^l which is feasible for it, we have $P'(E1) = 0$ and $P'(E2) = P(D2) + P(D3)$. Therefore, we have $P'(E1) + P'(E2) > P(E1) + P(E2)$, which proves optimality of giving a non-zero score to a feasible agreement. This completes the proof of Lemma 1.

D Proof of Lemma 2

Suppose a responder i finds it optimal to give score vector $\mathbf{s}_{i,n}$ to offers of round n, where $A_{i,n} > 0$ and $H_{i,n}^{D-1} = 0$. Now, we show that the responder's utility increases if it chooses score vector $\mathbf{s}'_{i,n} = (\mathbf{s}_{i,n} + c)\,\mathrm{sign}\,(\mathbf{s}_{i,n})$, where $c = D - 1 - \max_j s_{i,n}^j$. This contradicts the optimality of score vector $\mathbf{s}_{i,n}$.

The privacy leakages corresponding to $\mathbf{s}_{i,n}$ and $\mathbf{s}'_{i,n}$ are the same. Moreover, according to constraint (d), the reward function is invariant to shifting of the scores. The initiator's decision about the final outcome is based on the rewards it needs to pay. Therefore, if responder i changes its score from $\mathbf{s}_{i,n}$ to $\mathbf{s}'_{i,n}$ nothing except its cost is impacted. According to (15) and (16), the cost function is decreasing in terms of the scores. Therefore, the cost of giving score $\mathbf{s}'_{i,n}$ is less than the cost of scoring $\mathbf{s}_{i,n}$. Therefore, score vector $\mathbf{s}'_{i,n}$ can achieve a similar performance to $\mathbf{s}_{i,n}$, but with a lower cost. Therefore, score vector $\mathbf{s}_{i,n}$ cannot be an optimal response for responder i.

E Proof of Lemma 3

We proved in Lemma 1 that the agents announce the feasibility of the agreements truthfully. Therefore, all the rational scores a responder i would give to the offered agreements at round n have the same $A_{i,n}$. For $A_{i,n} = 0$, the problem becomes trivial as the only rational choice responder i has is to give score 0 to all offers. Therefore, in the rest of the proof, we focus on the case where $A_{i,n} > 0$.

Based on Lemma 2, responder i gives a score $D - 1$ to at least one of its feasible agreements. Therefore we have $H_{i,n}^{D-1} > 0$ and hence $F \geq (A+1)^{D-2}$. Now, using (21), we derive the expected point income of responder i at round n of the negotiation, when it gives a rational score vector $\mathbf{s}_{i,n}$ to offers as follows:

$$\mathbf{E}[B_{i,n}] = \sum_{s=1}^{D-1} P(s, A_{i,n}, F_{i,n}, L_n) r(s, A_{i,n}, F_{i,n}, L_n) - C(A_{i,n}, F_{i,n}) = 0, \tag{37}$$

where the second equality holds due to (21). This completes the proof of Lemma 3.

F Proof of Lemma 4

We prove this lemma by contradiction. Consider two agreements O_n^1 and O_n^2 where $V_i(O_n^1) > V_i(O_n^2)$. Suppose that in round n of the mechanism, agent i is asked to score agreements $\{O_n^1, O_n^2, \dots, O_n^{L_n}\}$, and it gives a higher score to O_n^2 than O_n^1, i.e. $s_{i,n}^2 > s_{i,n}^1$. Now, we construct another score vector $\mathbf{s}'_{i,n}$ from $\mathbf{s}_{i,n}$ by exchanging its first and second elements (i.e. $\mathbf{s}'_{i,n} = (s_{i,n}^2, s_{i,n}^1, s_{i,n}^3, \dots, s_{i,n}^{L_n})$) and show that responder i could achieve a higher expected utility if it gave score $\mathbf{s}'_{i,n}$ to the offers. Showing this fact contradicts the rational behavior of the responder and hence proves Lemma 4.

Let

$$\mathbf{E}[U_{i,n}(\mathbf{s}_{i,n})] = \sum_{j=1}^{L_n} (H_{i,n}^{s_{i,n}^j})^{-1} P(s_{i,n}^j, A_{i,n}, F_{i,n}, L_n) V_i(O_n^j) - \beta_i - \theta_{iI} L_{i,I}(\mathbf{s}_{i,n}), \tag{38}$$

denote the expected instantaneous utility agent i gets at round n of the mechanism, when it gives score $\mathbf{s}_{i,n}$ to the offers. Function $P(s_{i,n}^j, A_{i,n}, F_{i,n}, L_n)$ indicates the probability that one of the offers that receives score $s_{i,n}^j$ from agent i will be selected by the initiator. $H_{i,n}^{s_{i,n}^j}$ is the number of offers that receive score $s_{i,n}^j$ from agent i at round n. Therefore, the probability that any specific offer O_n^j is being selected is $(H_{i,n}^{s_{i,n}^j})^{-1} P(s_{i,n}^j, A_{i,n}, F_{i,n}, L_n)$. If offer O_n^j is selected, agent i gets value $V_i(O_n^j)$ while it incurs the bargaining cost β_i and the privacy loss $\theta_{iI} L_{i,I}(\mathbf{s}_{i,n})$.

If responder i changes its scores to $\mathbf{s}'_{i,n}$, the frequencies $H_{i,n}^s$, $\forall s$, the availability level $A_{i,n}$, and the degree of flexibility $F_{i,n}$ remain unchanged. Therefore, we can write its expected instantaneous utility as follows:

$$\begin{aligned}\mathbf{E}[U_{i,n}(\mathbf{s}'_{i,n})] \\ &= (H_{i,n}^{s_{i,n}^2})^{-1} P(s_{i,n}^2, A_{i,n}, F_{i,n}, L_n) V_i(O_n^1) + (H_{i,n}^{s_{i,n}^1})^{-1} P(s_{i,n}^1, A_{i,n}, F_{i,n}, L_n) V_i(O_n^2) \\ &\quad + \sum_{j=3}^{L_n} (H_{i,n}^{s_{i,n}^j})^{-1} P(s_{i,n}^j, A_{i,n}, F_{i,n}, L_n) V_i(O_n^j) - \beta_i - \theta_{iI} L_{i,I}(\mathbf{s}'_{i,n}).\end{aligned} \tag{39}$$

The privacy leakages corresponding to $\mathbf{s}_{i,n}$ and $\mathbf{s}'_{i,n}$ are the same. Therefore, we have:

$$\mathbf{E}[U_{i,n}(\mathbf{s}'_{i,n})] - \mathbf{E}[U_{i,n}(\mathbf{s}_{i,n})] = (K^{s_{i,n}^2} - K^{s_{i,n}^1})(V_i(O_n^1) - V_i(O_n^2)), \tag{40}$$

where $K^s := (H_{i,n}^s)^{-1} P(s, A_{i,n}, F_{i,n}, L_n)$ is the probability that a specific agreement that got score s from responder i is selected at round n. This probability is increasing in terms of s. Therefore, $s_{i,n}^2 > s_{i,n}^1$ implies that $K^{s_{i,n}^2} > K^{s_{i,n}^1}$. Using this result, we can conclude from (40) that $\mathbf{E}[U_{i,n}(\mathbf{s}'_{i,n})] > \mathbf{E}[U_{i,n}(\mathbf{s}_{i,n})]$ which contradicts the rational behavior of the agent and shows that responder i achieves more utility if it scores the offered agreements in an ordering consistent with its valuation function.

G Proof of Lemma 5

Suppose that at round n, the initiator offers agreements $\{O_n^1, \dots, O_n^{L_n}\}$ to the responders. We assume that the numbering of these agreements are according to agent i's preferences. That is,

$$V_i(O_n^1) \geq V_i(O_n^2) \geq \dots \geq V_i(O_n^{L_n}). \tag{41}$$

We focus on agreements $\{O_n^1, \dots, O_n^{A_{i,n}}\}$ that are feasible for responder i. Responder i gives a score between 1 and $D - 1$ to each of these agreements.

Suppose that agent i strictly prefers outcome O_n^k to O_n^{k+1}, i.e. $V_i(O_n^k) > V_i(O_n^{k+1})$, but it gives a similar score to both of them, i.e. $s_{i,n}^k = s_{i,n}^{k+1}$. We show by contradiction that this scoring cannot be optimal for responder i, if $H_{i,n}^1 = 0$ and the agent has not run out of budget.

Suppose that $\mathbf{s}_{i,n}$ is responder i's optimal scoring and $H_{i,n}^1 = 0$. We will prove that agent i could achieve a higher utility if it changes the scores to $\mathbf{s}'_{i,n}$, where

$$s_{i,n}^{\prime j} = \begin{cases} s_{i,n}^j - 1, & \text{If} k + 1 \leq j \leq A_{i,n}, \\ s_{i,n}^j, & \text{Otherwise.} \end{cases} \tag{42}$$

We define the selection vector $K_{i,n} = (K_{i,n}^1, \dots, K_{i,n}^{A_{i,n}})$, where $K_{i,n}^j = (H_{i,n}^{s_{i,n}^j})^{-1} P(s_{i,n}^j, A_{i,n}, F_{i,n})$ is the probability that responder i assigns to the selection of agreement O_n^j at round n. We denote the selection vector corresponding to rating $\mathbf{r}'_{i,n}$ by $K'_{i,n}$. Using this new notation, we write the expected instantaneous utility of responder i at round n when it gives scores $\mathbf{s}_{i,n}$ and $\mathbf{s}'_{i,n}$ to the offers, as

$$\begin{aligned}\mathbf{E}[U_i] = \sum_{j=1}^{A_{i,n}} K_{i,n}^j [V_i(O_n^j) + r(s_{i,n}^j, A_{i,n}, F_{i,n}, L_n) - C(A_{i,n}, F_{i,n})] \\ - \beta_i - \theta_{iI} L_{i,I}(\mathbf{s}_{i,n}),\end{aligned} \tag{43}$$

and

$$\begin{aligned}\mathbf{E}[U'_i] = \sum_{j=1}^{A_{i,n}} K'^{j}_{i,n}[V_i(O^j_n) + r(s'^{j}_{i,n}, A_{i,n}, F'_{i,n}, L_n) - C(A_{i,n}, F'_{i,n})] \\ - \beta_i - \theta_{iI} L_{i,I}(\mathbf{s}'_{i,n}),\end{aligned} \tag{44}$$

respectively. Score vectors $\mathbf{s}_{i,n}$ and $\mathbf{s}'_{i,n}$ deliver the same message about feasibility and infeasibility of the offered agreements for responder i. Therefore, they cause an equal amount of privacy leakage, i.e. $L_{i,I}(\mathbf{s}_{i,n}) = L_{i,I}(\mathbf{s}'_{i,n})$.

In Lemma 3, we proved that as long as the responder gives score $D - 1$ to at least one of the feasible offers, its expected point income will be zero. The score vector $\mathbf{s}_{i,n}$ is assumed to be optimal. Therefore, according to Lemma 2, it gives score $D - 1$ to at least one of the agreements. The mapping (42) keeps the top score fixed. Therefore, the condition is satisfied for $\mathbf{s}'_{i,n}$ as well. Thus, we have

$$\begin{aligned}&\sum_{j=1}^{A_{i,n}} K^{j}_{i,n}[r(s^{j}_{i,n}, A_{i,n}, F_{i,n}, L_n) - C(A_{i,n}, F_{i,n})] \\ &\quad = \sum_{j=1}^{A_{i,n}} K'^{j}_{i,n}[r(s'^{j}_{i,n}, A_{i,n}, F'_{i,n}, L_n) - C(A_{i,n}, F'_{i,n})] = 0.\end{aligned} \tag{45}$$

Therefore, we have

$$\mathbf{E}[U'_i] - \mathbf{E}[U_i] = \sum_{j=1}^{A_{i,n}} K'^{j}_{i,n} V_i(O^j_n) - \sum_{j=1}^{A_{i,n}} K^{j}_{i,n} V_i(O^j_n) = \mathbf{E}_{K'}[V_i] - \mathbf{E}_K[V_i]. \tag{46}$$

Using the properties of the payment function, it can be shown that the selection vectors satisfy the following condition:

$$\sum_{j=1}^{l} K'^{j}_{i,n} \geq \sum_{j=1}^{l} K^{j}_{i,n}, \tag{47}$$

for all $l \leq A_{i,n}$. This means that the selection vector $K'_{i,n}$ has first-order stochastic dominance over $K_{i,n}$. Now, we can use the first-order stochastic ranking theorem stated below. This theorem is proved in the literature.

Theorem 6 *If u is strictly increasing, and cumulative F first-order stochastically dominates cumulative $G \neq F$, then* $\mathbf{E}_F[u(x)] > \mathbf{E}_G[u(x)]$.

Using this theorem, we can conclude that $\mathbf{E}_{K'}[V_i] > \mathbf{E}_K[V_i]$. Substituting this in (46) we have $\mathbf{E}[U'_i] > \mathbf{E}[U_i]$ which contradicts the optimality of scoring $\mathbf{s}_{i,n}$. Therefore, as long as the number of satisfaction levels and the responders' budget allow, it is optimal for the responders to give unequal scores to agreements with unequal values.

Funding This study was funded by Samsung Electronics R&D Institute UK (SRUK).

Declarations

Conflict of interest There is no other conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abernethy JD, Cummings R, Kumar B, Taggart S, Morgenstern JH. Learning auctions with robust incentive guarantees. Adv Neural Inf Process Syst. 2019;32:11591–601.
2. BenHassine A, Ho TB. An agent-based approach to solve dynamic meeting scheduling problems with preferences. Eng Appl Artif Intell. 2007;20:857–73.
3. Bhargava M, Majumdar D, Sen A. Incentive-compatible voting rules with positively correlated beliefs. Theoretical Economics. 2015;10(3):867–85.
4. Bichler, M.: A roadmap to auction-based negotiation protocols for electronic commerce. In: Proceedings of the 33rd annual Hawaii international conference on system sciences; 2000.
5. Binmore K, Vulkan N. Applying game theory to automated negotiation. Netnomics. 1999;1:1–9.
6. Bogliolo A, Polidori P, Aldini A, Moreira W, Mendes P, Yildiz M, Ballester C, Seigneur J. Virtual currency and reputation-based cooperation incentives in user-centric networks. In: 8th international wireless communications and mobile computing conference (IWCMC); 2012, p. 895–900.
7. Borgers T, Krahmer D, Strausz R. An introduction to the theory of mechanism design. Oxford: Oxford University Press; 2015.
8. Brito I, Meseguer P. Privacy in distributed meeting scheduling. In: Proceedings of the 11th conference on artificial intelligence research and development; 2008, p. 118–27.
9. Brooks RR. Distributed sensor networks: a multiagent perspective. Int J Distrib Sens Netw. 2008;4(3):285.
10. Budish E, Cachon G, Kessler J, Othman A. Course match: a large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. Oper Res. 2017;65(2):314–36.
11. Bui HH, Venkatesh S, Kieronska D. A multi-agent incremental negotiation scheme for meetings scheduling. In: Proceedings of Third Australian and New Zealand conference on intelligent information systems. ANZIIS-95; 1995, p. 175–180.
12. Buttyan L, Hubaux JP. Nuglets: a virtual currency to stimulate cooperation in self-organized mobile ad hoc networks. Technical report; 2001.

13. Chen Z, Wu T, Deng Y, Zhang C. An ant-based algorithm to solve distributed constraint optimization problems. In: AAAI; 2018.
14. Chun HW, Wong RY. N* an agent-based negotiation algorithm for dynamic scheduling and rescheduling. Adv Eng Inform. 2003;17(1):1–22.
15. Crawford E, Veloso M. Mechanism design for multi-agent meeting scheduling including time preferences, availability, and value of presence. In: Proceedings. IEEE/WIC/ACM international conference on intelligent agent technology, 2004. (IAT 2004); 2004, p. 253–259.
16. Crawford E, Veloso M. Negotiation in semi-cooperative agreement problems. In: 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, vol. 2; 2008, p. 252–258.
17. Creedy J, Scutella R. The role of the unit of analysis in tax policy return evaluations of inequality and social welfare. Australian J Labour Econ (AJLE). 2004;7:89–108.
18. Davin J. Hierarchical variable ordering for multiagent agreement problems. In: In AAMAS; 2006, p. 1433–1435.
19. Defago X, Hassine A, Ho T. Agent-based approach to dynamic meeting scheduling problems. In: AAMAS; 2004, p. 1132–1139.
20. Dongjun L, Spong MW. Agreement with non-uniform information delays. In: American control conference (ACC); 2006.
21. Doshi P, Matsui T, Silaghi M, Yokoo M, Zanker M. Distributed private constraint optimization. In: Proceedings of the international conference on intelligent agent technology (IAT); 2008, p. 277–81.
22. Ephrati E, Zlotkin G, Rosenschein JS. A nonmanipulable meeting scheduling system. In: International workshop on distributed artificial intelligence; 1994.
23. Faltings B. A budget-balanced, incentive-compatible scheme for social choice. In: Faratin P, Rodriguez-Aguilar JA, editors. Agent-mediated electronic commerce VI. Theories for and engineering of distributed mechanisms and systems. Berlin: Springer; 2005. p. 30–43.
24. Faltings B, Leaute T, Petcu A. Privacy guarantees through distributed constraint satisfaction. In: Web intelligence and intelligent agent technology; 2008, p. 350–8.
25. Farhadi F, Golestani SJ, Teneketzis D. A surrogate optimization-based mechanism for resource allocation and routing in networks with strategic agents. IEEE Trans Autom Control. 2019;64:464–79.
26. Farhadi F, Jennings NR. A faithful mechanism for privacy-sensitive distributed constraint satisfaction problems. In: Bassiliades N, Chalkiadakis G, de Jonge D, editors. Multi-agent systems and agreement technologies. Cham: Springer International Publishing; 2020. p. 143–58.
27. Farhadi F, Tavafoghi H, Teneketzis D, Golestani SJ. An efficient dynamic allocation mechanism for security in networks of interdependent strategic agents. Dyn Games Appl. 2019;9:914–41.
28. Feigenbaum J, Papadimitriou C, Sami R, Shenker S. A bgp-based mechanism for lowest-cost routing. Distrib Comput. 2002;18:173–82.
29. Feigenbaum J, Shenker S. Distributed algorithmic mechanism design: recent results and future directions. In: Proceedings of the 6th international workshop on discrete algorithms and methods for mobile computing and communications, DIALM'02, p. 1–13. Association for Computing Machinery, New York, NY, USA; 2002.
30. Franzin MS, Rossi F, Freuder EC, Wallace R. Multi-agent constraint systems with preferences: efficiency, solution quality, and privacy loss. Comput Intell. 2004;20:264–86.
31. Freuder EC, Minca M, Wallace RJ. Privacy/efficiency tradeoffs in distributed meeting scheduling by constraint-based agents. In: Proceedings of IJCAI DCR; 2001.
32. Geibel P. Reinforcement learning for mdps with constraints. In: Machine learning: ECML; 2006, p. 646–53.
33. Gershkov A, Moldovanu B, Shi X. Optimal voting rules. Rev Econ Stud. 2016;84:688–717.
34. Gershman A, Meisels A, Zivan R. Asynchronous forward bounding for distributed cops. J Artif Intell Res. 2009;34:61–88.
35. Gorokh A, Banerjee S, Iyer K. Near-efficient allocation using artificial currency in repeated settings. In: Proceedings of the 12th international conference on web and internet economics, WINE '16.
36. Gorokh A, Banerjee S, Iyer K. From monetary to non-monetary mechanism design via artificial currencies. SSRN Electron J 2017;563–4
37. Greenstadt R, Grosz B, Smith M. Ssdpop: improving the privacy of dcop with secret sharing. In: AAMAS'07—Proceedings of the 6th international joint conference on autonomous agents and multiagent systems, proceedings of the international conference on autonomous agents; 2007, p. 1098–100.
38. Greenstadt R, Pearce JP, Tambe M. Analysis of privacy loss in distributed constraint optimization. In: National conference on artificial intelligence; 2006. AAAI Press, New York, p. 647–53.
39. Grinshpoun T. When you say (dcop) privacy, what do you mean?—categorization of dcop privacy and insights on internal constraint privacy. In: Proceedings of the 4th international conference on agents and artificial intelligence—vol. 2: ICAART; 2012, p. 380–86.
40. Grinshpoun T, Tassa T. P-syncbb: a privacy preserving branch and bound dcop algorithm. J Artif Intell Res. 2016;57:621–60.
41. Hirayama K, Yokoo M. Distributed partial constraint satisfaction problem. In: Principles and practice of constraint programming. New York: Springer; 1997. p. 222–36.
42. Hoang K, Hou P, Fioretto F, Yeoh W, Zivan R, Yokoo, M.: Infinite-horizon proactive dynamic dcops. In: Proceedings of the 16th conference on autonomous agents and multiagent systems (AAMAS); 2017, pp 212–220
43. Huang Z, Kannan S. The exponential mechanism for social welfare: private, truthful, and nearly optimal. In: IEEE 53rd annual symposium on foundations of computer science; 2012, p. 140–9
44. Jennings N, Jackson A. Agent-based meeting scheduling: a design and implementation. Electron Lett. 1995;31:350–2.
45. Jung T, Li XY, Han J. A framework for optimization in big data: privacy-preserving multi-agent greedy algorithm. In: Wang Y, Xiong H, Argamon S, Li X, Li J, editors. Big data computing and communications. New York: Springer International Publishing; 2015. p. 88–102.
46. Kekulluoglu D, Kokciyan N, Yolum P. Preserving privacy as social responsibility in online social networks. ACM Trans Internet Technol. 2018;18(4):1–22.
47. Lakkaraju K, Gasser L. A unified framework for multi-agent agreement. In: AAMAS; 2007, p. 1–3.
48. Leaute T, Faltings B. Privacy-preserving multi-agent constraint satisfaction. Int Conf Comput Sci Eng. 2009;3:17–25.
49. Leaute T, Faltings B. Protecting privacy through distributed computation in multi-agent decision making. J Artif Intell Res. 2014;47:649–95.
50. Leu SS, Son PVH, Nhung PTH. Hybrid Bayesian fuzzy-game model for improving the negotiation effectiveness of construction material procurement. J Comput Civ Eng. 2015;29:1–12.
51. Leung S, Lui E. Bayesian mechanism design with efficiency, privacy, and approximate truthfulness. In: Goldberg PW, editor. Internet Netw Econ. Berlin: Springer; 2012. p. 58–71.
52. Litov O, Meisels A. Forward bounding on pseudo-trees for dcops and adcops. Artif Intell. 2017;252:83–99.
53. Macarthur, K.S., Stranders, R., Ramchurn, S.D., Jennings, N.R.: A distributed anytime algorithm for dynamic task allocation in

multi-agent systems. In: Proceedings of the 25th AAAI conference on artificial intelligence, AAAI'11; 2011, p. 701–6
54. Maheswaran R, Pearce J, Bowring E, Varakantham P, Tambe M. Privacy loss in distributed constraint reasoning: a quantitative framework for analysis and its applications. Auton Agent Multi-Agent Syst. 2006;13:27–60.
55. Mahmud S, Choudhury M, Khan MM, Tran-Thanh L, Jennings NR: Aed: an anytime evolutionary dcop algorithm. In: Proceedings of the 19th international conference on autonomous agents and multiagent systems, AAMAS'20, p. 825–33. International foundation for autonomous agents and multiagent systems, Richland, SC (2020)
56. Majumdar D, Sen A. Ordinally Bayesian incentive compatible voting rules. Econometrica. 2004;72:523–40.
57. Meisels A, Lavee O: Using additional information in discsps search. In: DCR (2004)
58. Mezzetti C. Mechanism design with interdependent valuations: efficiency. Econometrica. 2004;72(5):1617–26. https://doi.org/10.1111/j.1468-0262.2004.00546.x.
59. Mhanna S, Chapman AC, Verbic G. A faithful and tractable distributed mechanism for residential electricity pricing. IEEE Trans Power Syst. 2018;33(4):4238–52.
60. Mhanna S, Verbic G, Chapman AC. A faithful distributed mechanism for demand response aggregation. IEEE Trans Smart Grid. 2016;7(3):1743–53.
61. Modi PJ, Shen WM, Tambe M, Yokoo M. Adopt: asynchronous distributed constraint optimization with quality guarantees. Artif Intell. 2005;161:149–80.
62. Modi, P.J., Veloso, M.: Bumping strategies for the multiagent agreement problem. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS '05; 2005, p. 390–6. Association for Computing Machinery, New York, NY, USA.
63. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. Technical report; 2009
64. Nguyen DT, Yeoh W, Lau HC, Zivan R. Distributed gibbs: a linear-space sampling-based dcop algorithm. J Artif Intell Res. 2019;64(1):705–48.
65. Nisim K, Xiao D. Mechanism design and differential privacy. New York: Springer; 2016.
66. Nissim K, Orlandi C, Smorodinsky R. Privacy-aware mechanism design. In: Proceedings of the 13th ACM conference on electronic commerce, EC'12; 2012, p. 774–789. Association for Computing Machinery
67. Okamoto S, Zivan R, Nahon A. Distributed breakout: beyond satisfaction. In: IJCAI; 2016, p. 447–53.
68. Singer Y. Budget feasible mechanisms In: 2010 IEEE 51st Annual Symposium on Foundations of Computer Science. 2010. p. 765–774. https://doi.org/10.1109/FOCS.2010.78.
69. Parkes DC, Kalagnanam JR, Eso M. Achieving budget-balance with vickrey-based payment schemes in exchanges. In: 17th IJCAI; 2001
70. Parkes DC, Shneidman J. Distributed implementations of vickrey-clarke-groves mechanisms. In: AAMAS; 2004, p. 261–8
71. Pascal C, Panescu D. On applying discsp for scheduling in holonic systems. In: 20th international conference on system theory, control and computing; 2016, p. 423–8.
72. Pecorino P. Negotiation games: applying game theory to bargaining and arbitration. Manag Decis Econ. 2004;25(3):175–6.
73. Peng D, Yang S, Wu F, Chen G, Tang S, Luo T. Resisting three-dimensional manipulations in distributed wireless spectrum auctions. In: 2015 IEEE conference on computer communications (INFOCOM); 2015, p. 2056–64.
74. Petcu A, Faltings B. A scalable method for multiagent constraint optimization. In: Proceedings of the 19th international joint conference on artificial intelligence, IJCAI'05;2005, p. 266–71. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
75. Petcu A, Faltings B.:Superstabilizing, fault-containing distributed combinatorial optimization. In: Proceedings of the 20th national conference on artificial intelligence—Vol. 1, AAAI'05; 2005, p. 449–454. AAAI Press.
76. Petcu A, Faltings B, Parkes DC. M-dpop: faithful distributed implementation of efficient social choice problems. J Artif Intell Res. 2008;32:705–55.
77. Prendergast C. The allocation of food to food banks. EAI Endorsed Trans Serious Games. 2016;3:e4.
78. Procaccia AD, Tennenholtz M. Approximate mechanism design without money. In: Proceedings of the 10th ACM conference on electronic commerce (EC); 2009.
79. Pultowicz P. Multi-agent negotiation and optimization in decentralized logistics. Ph.D. thesis, University of Vienna; 2017.
80. Rogers A, Farinelli A, Stranders R, Jennings N. Bounded approximate decentralised coordination via the max-sum algorithm. Artif Intell. 2011;175:730–59.
81. Rubinstein A. Perfect equilibrium in a bargaining model. Econometrica. 1982;50:97–109.
82. Sen S, Durfee EH. A formal study of distributed meeting scheduling. Group Decis Negot. 1998;7:55–68.
83. Shneidman J, Parkes DC. Specification faithfulness in networks with rational nodes. In: Proceedings of the twenty-third annual ACM symposium on principles of distributed computing. Association for Computing Machinery; 2004, p. 88–97.
84. Silaghi M, Faltings B. A comparison of distributed constraint satisfaction approaches with respect to privacy. In: DCR; 2002.
85. Silaghi MC, Mitra D. Distributed constraint satisfaction and optimization with privacy enforcement. In: Intelligent agent technology; 2004, p. 531–5.
86. Singh DK, Mazumdar BD. Agent mediated negotiation in e-commerce: a review. Int J Modern Trends Eng Res. 2017;9:207–16.
87. Strehl AL, Li L, Littman ML. Reinforcement learning in finite mdps: Pac analysis. J Mach Learn Res. 2009;10:2413–44.
88. Such JM, Rovatsos M. Privacy policy negotiation in social media. ACM Trans Auton Adapt Syst. 2016;11(1):1–29. https://doi.org/10.1145/2821512.
89. Sultanik E, Lass R, Regli W. Dynamic configuration of agent organizations. In: Proceedings of the 21st international joint conference on artificial intelligence—Vol. 1, IJCAI 2009; 2009, p. 305–11.
90. Tanaka T, Farokhi F, Langbort C. A faithful distributed implementation of dual decomposition and average consensus algorithms. In: 52nd IEEE conference on decision and control; 2003, p. 2985–90.
91. Tanaka T, Farokhi F, Langbort C. Faithful implementations of distributed algorithms and control laws. IEEE Trans Control Netw Syst. 2017;4(2):191–201.
92. Vincent DR. Bidding off the wall: why reserve prices are kept secret. Discussion Papers 838, Northwestern University, Center for Mathematical Studies in Economics and Management Science; 1989.
93. Wallace RJ, Freuder EC. Constraint-based reasoning and privacy/efficiency tradeoffs in multi-agent problem solving. Artif Intell. 2005;161:209–27.
94. Watkins C, Dayan P. Technical note: Q-learning. Mach Learn. 1992;8:646–53.
95. Wei K, Chen YF, Smith A, Vo B. Whopay: a scalable and anonymous payment system for peer-to-peer environments. Technical report; 2005.
96. Yeoh W, Felner A, Koenig S. Bnb-adopt: an asynchronous branch-and-bound dcop algorithm. J Artif Intell Res JAIR. 2008;38:85–133.

97. Yeoh W, Varakantham P, Sun X, Koenig S. Incremental dcop search algorithms for solving dynamic dcops. In: The 10th international conference on autonomous agents and multiagent systems, AAMAS'11; 2011, p. 1069–70.

98. Yokoo M. Protocol/mechanism design for cooperation/competition. In: 3rd International joint conference on autonomous agents and multiagent systems; 2004, p. 3–7.

99. Yokoo M, Suzuki K, Hirayama K. Secure distributed constraint satisfaction: reaching agreement without revealing private information. Artif Intell. 2005;161:229–45.

100. Zhang C, Fan Y, Wang L, Chen H. Consensus of multi-agent systems by distributed self-triggered control. In: 33rd youth academic annual conference of Chinese association of automation (YAC); 2018, p. 270–5.

101. Zhu R, Shin KG. Differentially private and strategy-proof spectrum auction with approximate revenue maximization. In: 2015 IEEE conference on computer communications (INFOCOM); 2015, p. 918–26.

102. Zivan R, Peled H. Max/min-sum distributed constraint optimization through value propagation on an alternating dag. In: Proceedings of the 11th international conference on autonomous agents and multiagent systems—Vol. 1, AAMAS '12. International Foundation for Autonomous Agents and Multiagent Systems; 2012, p. 265–72.

103. Zou J, Meir R, Parkes D. Strategic voting behavior in doodle polls. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing; 2015, p. 464–72.

104. Zunino A, Campo M. Chronos: a multi-agent system for distributed automatic meeting scheduling. Expert Syst Appl. 2009;36:7011–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.