5 1

Introduction to Mechanism Design

1.1 Dominant strategies and Nash equilibria

In the previous lectures we have seen examples of games that admit several Nash equilibria. Moreover, some of these equilibria correspond to solutions that are far off the optimal ones. In particular, we have seen that

- For the "football or shopping" game,

$(\mathbf{2},\mathbf{1})$	$(0,0)$
$(0,0)$	$(\mathbf{1},\mathbf{2})$

  we could not predict the "behavior" of the agents. Indeed, it is not clear how the agents[1] can reach one of the two equilibria (shown in bold in the table of the payoffs) and which one of the two. Notice that, the best strategy for player 1 (i.e., the row) depends on the chosen strategy of player 2 (i.e., the chosen column), and vice versa. Intuitively, the two players should agree on some *joint* strategy.

- For the n links game, there exists a Nash equilibrium whose cost is $\Omega(\frac{\log n}{\log\log n})$ times the optimum.[2] Since we cannot predict which of the Nash equilibria is reached (if so), then it may be the case that the *worst* one is reached, thus a non-optimal configuration.

It is not clear whether a Nash equilibrium can be reached, even for those games that have *only one* such an equilibrium: the "matching pennies" game

$(1,-1)$	$(-1,1)$
$(-1,1)$	$(1,-1)$

[1] In the sequel we use the terms 'agent' and 'player' interchangeably.

[2] The proof of this lower bound is based on a generalization of the lower bound 3/2 for two identical machines. Consider n identical machines, n identical jobs and, for each job/agent, a job chooses one machine with uniform distribution.

Here, the unique equilibrium is reached when player 1 (resp., player 2) chooses a row (resp., column) with probability 1/2. However, how can the two players "agree" on this unique probability distribution? (consider the situation in which the game goes through several rounds and, at each round, one of the two players is allowed to change her/his probability distribution).

These two facts seem to denote the "weakness" of Nash equilibria. On one hand, there is no "preferred" Nash equilibrium (actually, it is not clear how such a configuration can be reached). On the other hand, even if we are guaranteed that a Nash equilibrium is reached, it may be a "very bad" one (i.e., the worst one, which for the m links game is non-optimal).

Let us now consider another example of a game with a *unique* Nash equilibrium: the "two prisoners' dilemma"

$(3,3)$	$(1,5)$
$(5,1)$	$(\mathbf{2},\mathbf{2})$

Besides the fact that $(2,2)$ is the unique equilibrium, there is a very important property in this game:

- Even though the payoff of player 1 depends on the strategy of the other player (i.e., the column), no matter what this strategy is like, for player 1 it is always better to choose the second row (recall that player 1 cannot choose the column). In other words, player 1 does not need to know what the other player is doing to maximize his/her payoff!

- Similarly, no matter which strategy (row) player 1 selects, for player 2 it is always better to choose the second column. Again, player 2 does not need to know what the other player is doing to maximize his/her payoff!

Intuitively, both players have a "universally" optimal strategy which guarantees that, no other strategy would increase his/her payoff. Consider a payoff table of the form

$(v_1(1,1), v_2(1,1))$	$(v_1(1,2), v_2(1,2))$
$(\mathbf{v_1(2,1), v_2(2,1)})$	$(v_1(2,2), v_2(2,2))$

such that

$$\forall j \in \{1,2\},\ v_1(2,j) \geq v_1(1,j)$$

(i.e., for player 1, row 2 is never worse than row 1)
and

$$\forall i \in \{1,2\},\ v_2(i,1) \geq v_2(i,2)$$

(i.e., for player 2, column 1 is never worse than column 2)
In this case, player 1 (resp., player 2) has no reason to choose a strategy different from row 2 (respectively, column 1).

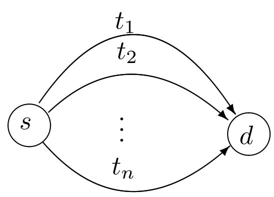
In general, we say that there exists a *dominant strategy* for players i if this strategy maximizes his/her payoff, for all possible strategies adopted by the other agents. Notice that, this is a property of the table of payoff, so an agent can actually verify whether such a strategy exists!

1.2 Payments inducing dominant strategies

We would like to reward each agent in such a way that

- every agent has a dominant strategy
- the dominant strategy is the "desired one"

We better explain the above two goals with an example. Consider the n links network in which each link is owned by a selfish agent i, $i = 1, 2, \ldots n$. Each link i is owned by an agent AG_i that knows the speed of that link, that is, the time t_i that a packet needs to traverse the link (i.e., $t_i = 1/speed\_link\_i$). This is a *private information*, in the sense that AG_i is the only one to know t_i (in particular, we do not know the value t_i). More importantly, since t_i is the time link i must be used in order to perform the transmission, t_i represents a *cost* for agent AG_i if his/her link is selected.



GOAL: We want to send one piece of traffic from s to t using the fastest of the n links, that is, the link minimizing t_i.

It seems necessary that, somehow, we obtain all correct values from the agents. It is then natural to ask the following:

QUESTION: Can we guarantee that each agent reports his/her speed correctly?

Observe that this is impossible in general: a *malicious* agent may always lie attempting to making us fail in our task! However selfish agents do not behave in this way: they simply try to maximize their own payoff or utility. This means that they lie only if there is a reason for that (namely, they get some benefit out of this). We then better address the following question:

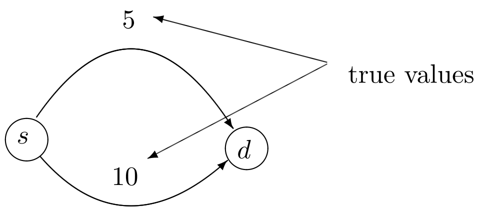
QUESTION: Can we guarantee that no agent i has an incentive in lying about his/her link speed, i.e., in reporting a value $s_i \neq t_i$?

We next consider several payments and see whether they provide a positive answer to the above question or not.

No payments. What happens if we do not provide any payment to the agents? The *utility* u_i of agent AG_i can be defined as the loss due to cost incurred by the agent, if his/her link is chosen:

$$u_i = \begin{cases} -t_i & \text{if link } i \text{ is chosen,} \\ 0 & \text{otherwise.} \end{cases}$$

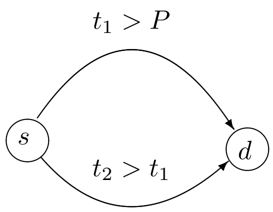
Consider the following example:



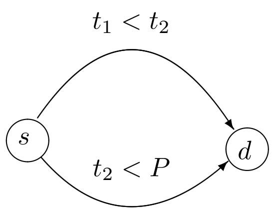
In this case, if agent AG_1 owning the link requiring 5 times unit would report a value $s_1 > 10$ then his/her utility would be 0 (the other link would be selected), while reporting $t_1 = 5$ would give $u_1 = -5$ (i.e., the cost he/she incurs when being selected).

Pay a fixed amount. We fix a value P and reward the agent reporting the largest speed an amount equal to P (i.e., this value is chosen in advance and independently of the agents declarations).

The following examples show that, doing so, we risk to overpay or underpay the agents:



we underpay the agent(s)

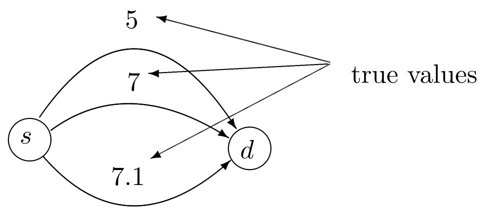


we overpay both agents

On the left, we are underpaying AG_1, so he/she would be better reporting $s_1 > t_2$, since the corresponding utility would be 0 against $u_1(t_1) = P - t_1 < 0$ when telling the truth. This is essentially the same problem as in the case of no payments. If, instead, we fix a value P which is sufficiently large to guarantee that $P \geq t_1$, then we risk to pay too much: since we do not know the values

t_i all we can do is to be sure that $P > \max_i t_i$ (assume we know the speed of the slowest existing link). Above on the right we have such an example: we are overpaying both agents, when selected; so AG_2 could report $s_2 < t_1 < t_2$ and obtain a positive utility equal to $u_2(s_2) = P - t_2 > 0 = u_2(t_2)$ (notice that the cost is always computed with respect to the true values).

Pay $P_i = s_i$. We choose the link with minimal reported s_i and pay to AG_i an amount $P_i = s_i$. Not surprisingly, this payments originate *speculation.* Consider the following example:



If agent AG_1 reports a value s_1 such that $t_1 = 5 < s_1 < t_2 = 7$, then her/his link is still selected and he/she gets a better payment: $P(t_1) = 5 < P(s_1)$. Therefore, also the utility would be better when reporting such an s_1: $u_1(t_1) = 5 - 5 < u_1(s_1) = s_1 - t_1$. So, AG_1 is clearly tempted to get as close to $t_2 = 7$ as possible!

This would not be harmful if *each agent would know the true/reported values of every other agent in advance*:[3] in this case, AG_1 may compute the minimum $s_i < t_2$ and report this value. If this happens, then we are still selecting the fastest link. However, this assumption is unreasonable: recall that t_i is a private information and AG_1 does not really know t_2. Also, communication between the agents cannot be assumed to be truthful (i.e., they may lie one to the other as well) or even possible.

So, all an agent AG_i knows is the following:

- If her/his link is the fastest one, according to the reported values s_i, then he/she could get something more by reporting s_1 just below the minimum s_j, with $j \neq i$.
- If her/his link is not the fastest one, according to the reported values s_i, then the best is to report $s_i = t_i$ (if he/she tries to be selected, the he/she would receive some payment $P_i \leq \min_{j \neq i}\{s_j\} < t_i$)

It is evident that agents are tempted to overbid (underbidding never pays off!): if AG_i is able to guess a "not too high" value $s_i > t_i$, then he/she might improve his/her utility. Suppose that all agents think that $s_i = t_i + a$ would be

[3]We consider a game in which every agent/player reports a value simultaneously with all other agents. This kind of games are usually termed *revelation games.*

a good choice, for some $a > 0$. In the example above, AG_2 would then declare $s_2 = 7+a$. Now, observe that the best value for AG_1 is no longer $s_1 < t_2 = 7$: if AG_1 reports a value $s_1 < s_2 = 7+a$, then she/he will be selected and rewarded an amount "close to" $7+a$. So, AG_1 might think that, since AG_2 will overbid by a factor a, then it is better for him/her to overbid by a factor $2a$. But also AG_2 may reasoning in the same way, thus leading AG_2 to overbid by a factor $3a$, and so forth...When will an agent stop doing this? Will all agents stop at the same value $k \cdot a$? Will all agents start with the same $a > 0$? This seems to be very unlikely!

Essentially what is not desirable in these payments is the fact that agents know that telling the truth is *not* the best strategy for them, but the best strategy *depends* on the other agents strategies (i.e., reported values). This makes the behavior of the agents "unpredictable": they may only guess a number trying to improve the utility. It then may be the case that, in the above example, AG_2 overbids by a factor $a = 1/2$ and agent AG_1 overbids by a factor $6a = 3$. In this case, we would fail since we would select the slowest link!

1.2.1 Vickrey-Clarke-Groves (VCG) payments

Though the payments considered above do not work, they have some good features:

- The "fixed P" is good since the amount of money agent AG_i receives does not depend on s_i. So, there is *no incentive to overbid.* overbid trying to increase his/her utility when selected.

- The payment "$P_i = s_i$" guarantees that we do not underpay an agent if he/she reports the true value (so he/she will not try to be excluded from the solution). Moreover, there is *no incentive to underbid*, since this always lead to a non-positive utility (in the above example, if AG_2 reports $s_2 \leq 5$, then she/he gets $u_2 = s_2 - 7 \leq -2$)

As mentioned above, the problem with the payments "$P_i = s_i$" is the fact that the highest payment that an agent AG_i can get depends on the other values s_js. One of the ideas of the VCG[4] payments is to compute this value and provide this (maximal) amount of money to the agent AG_i anyway.

Consider the following approach. Based on $s_1, s_2, \ldots, s_n$, do the following:

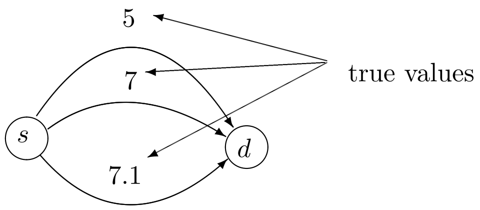
1. Choose the cheapest link i corresponding the cheapest value s_i;

2. Reward AG_i an amount equal to the 2^{nd} best/cheapest link, i.e, $P_i = \min_{j \neq i}\{s_j\}$.

Intuitively, this approach possesses both positive features of payments "fixed P" and "$P_i = s_i$": the payment P_i does not depend on s_i, and there is no incentive

[4]The name VCG is doe to the three fundamental works by Vickrey [3], Clarke [1] and Groves [2] on this area. These work contain all the main ideas we are going to discuss in this lecture.

to underbid since the rewarding is fixed to be the cost of the 2^{nd} cheapest link. Indeed, if an agent AG_j, owning a link of non-minimal cost, underbids pretending to be the cheapest one, then the payment he/she receives is no larger than his/her true cost t_j:



if AG_2 reports $s_2 \leq 5$, the he/she would receive $P_2 = 7$ and his/her link would be selected, thus a utility $u_2 = 7 - 7 = 0$; however, 0 is also the utility that he/she would obtain reporting $s_2 = t_2 = 7$, thus no reason to lie!

Also, the agent AG_i owning the truly cheapest link has no reason o underbid since her/his link would remain the cheapest and the payment does not depend on his/her declared value $s_i < t_i$. In the example above, if AG_1 reports $s_1 = 4$, her/his link is still selected and the payment remains $P_1 = 7$.

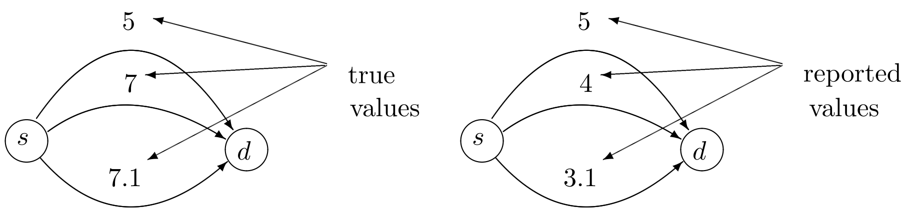
Finally, the most interesting feature of this payment scheme is the fact that it fixes the problem of the payment "$P_i = s_i$": now overbidding does not pay off! If, in the example above, AG_1 reports $s_1 > t_1 = 5$, then there are two possibilities:

1. link 1 is selected: AG_1 receives $P_1 = 7$, no matter what he/she declared.

2. link 1 is not selected: he/she receives no payment, thus a utility $u_1 = 0$.

It is clear that, overbidding does not increase the payment that he/she receives. However, if AG_1 overbids a too high value, he/she risks to loose the opportunity of being chosen (and having a positive utility $u_1 = 7 - 5$).

It is now natural to ask the following question: if an agent AG_j lies, is there an incentive for another agent AG_i to lie as well? Consider the following example:



According to the *reported* values, AG_1 is not owning the cheapest link. Does AG_1 have an incentive to lie? Observe that, the payment is fixed to be 4 (i.e., the 2^{nd} cheapest reported cost). So, if AG_1 tries to be selected (i.e., reports $s_1 < 3.1$) then he/she will obtain $u_1 = 4 - 5 < 0$.

More in general, the payment scheme described above guarantees the following two properties:

1. If we select link i, then $t_i \leq \min_{j \neq i}\{s_j\} = P_i$;
2. If we do not select link i, then $t_i \geq \min_{j \neq i}\{s_j\} = P_i$.

Since the value P_i does not depend on s_i, the two items above state that (i) if link i is the best, according to the reported values of the other agents, then there is an incentive for AG_1 in being selected (thus reporting $s_i = t_i$) since, in this case, the utility is positive; (ii) if link i is not the cheapest, according to the reported values of the other agents, then there is no incentive in being selected since this results in a non positive utility.

Formally, let $P_i(s_1, s_2, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n)$ denote the payment that each agent AG_i receives when the reported values are $s = (s_1, s_2, \ldots, s_n)$, with s_i being the value reported by AG_i. For every value x, let us define $(x, s_{-i}) := (s_1, s_2, \ldots, s_{i-1}, x, s_{i+1}, \ldots, s_n)$. The utility of AG_i when he/she reports s_i and the other agents reported values are $s_{-i} = (s_1, s_2, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ is then equal to

$$u_i(s_i, s_{-i}) = P_i(s_i, s_{-i}) + \begin{cases} -t_i & \text{if link } i \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

Recall that our goal is to guarantee that "truth-telling" is a dominant strategy for all agents, i.e., it is always the case that an agent cannot improve his/her utility by misreporting his/her true value t_i. This is actually what the above payments guarantee:

Theorem 1 *For all i, $i = 1, 2, \ldots, n$, it holds that*

$$\forall s_{-i} \forall s_i \quad u_i(t_i, s_{-i}) \geq u_i(s_i, s_{-i}).$$

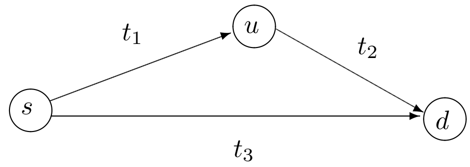
Proof. Exercise 16. □

In game-theoretic setting the strategy $\overline{s}_i = t_i$ is termed *dominant* for this game: for all possible strategy that the other agents choose, strategy $\overline{s}_i$ yields the maximum payoff that AG_i can obtain in this situation.

1.2.2 The shortest path problem

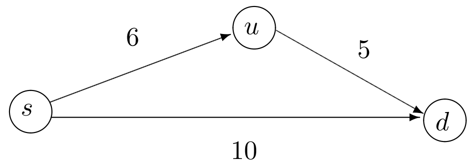
Consider now the following natural extension of the problem considered above: the nodes s and d are any two given nodes of a directed weighted graph; each edge i is owned by an agent and the corresponding weight is defined as in the previous problem. Assume also that every agent owns exactly one edge. [5] Here is an example:

[5] We make this assumption only for the sake of presentation/semplicity. However, this is not really necessary (see Exercise 20).



A first attempt: using the same payments for the n links problem

Consider a straightforward application of the payments for the n links problem to the example with three edges above: we select the cheapest path from s to d and we pay, to each agent owning a selected edge, the cost of the 2^{nd} cheapest path. Let us see what happens in the following instance (numbers represent the true costs/weights):



Let AG_1, AG_2 and AG_3 own the edge of cost 6, 5 and 10, respectively. If all agents are truth telling, then AG_1 and AG_2 have utility equal to 0 (we would select edge (s, d) and pay AG_3 only). However, if AG_1 reports a value $s_1 \leq 4$, then he/she would obtain something better: the payment would be 10 (now the 2^{nd} best path is the edge (s, d)) and AG_1 would be selected and receive that payment, thus a utility of $10 - 5 > 0$.

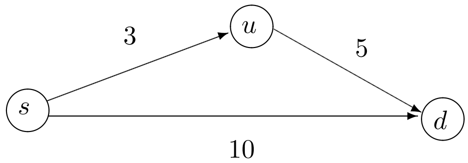
What is the problem here? It seems that we are paying too much, since a non-optimal edge is incentivated to get into the chosen path by lowering its cost.

n links revised: pay the "maximum speculation"

We briefly review the payments for the n parallel links, i.e., a special case of the shortest path problem. Consider the following instance:

Let us now see what would happen with payments "$P_i = s_i$". In particular, how much could AG_1, owning the link of cost 5, speculate? In order to receive some payment, he/she needs to be selected, thus implying that he/she must report some value $s_1 \leq 7$. This will also be the payment that, in this case, he/she will receive. So, 7 is the maximum he/she could: this is exactly the cost of the 2^{nd} cheapest link!

This seems to see another idea contained in the payments using the 2^{nd} cheapest link: pay the maximum amount that an agent could speculate if we were using the payments "$P_i = s_i$". Let us apply this idea to the following instance:



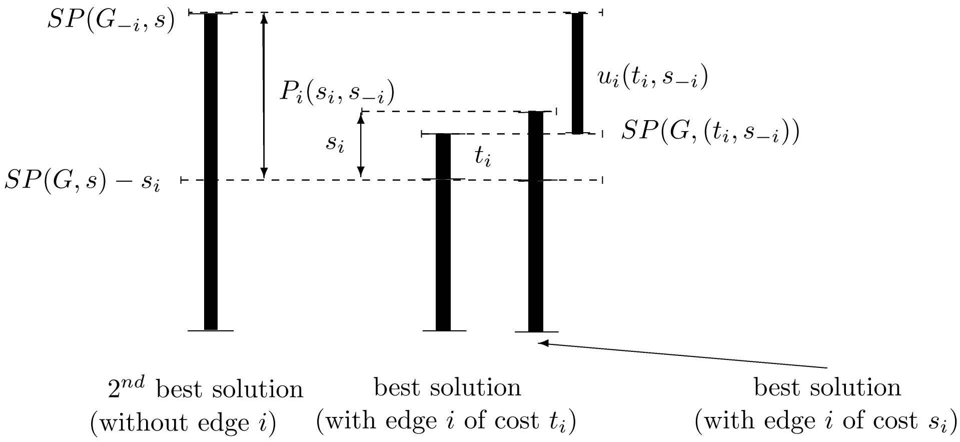
Assume all agents report the truth values, thus selecting the upper path. How much could get the agent AG_1 by lying about her cost, if $P_1 = s_1$? It must be $s_1 + 5 \leq 10$ (otherwise, we would select the other path). So, the payment P_1 should be equal to $10 - 5$. What does this value represent? The value 10 is the cost of the 2^{nd} cheapest path, while 5 is the cost of the chosen path (i.e., the cheapest one) *without counting edge of AG_1*.

VCG payments for the shortest path problem

Let us define $SP(G, w)$ the cost of the shortest path from s to d when the weight of edge i is equal to w_i, $w = (w_1, w_2, \ldots, w_n)$. Also let $SP(G_{-i}, w)$ denote the shortest path from s to d without using edge i. We then define the payment function of agent i as

$$P_i(G, (s_i, s_{-i})) := SP(G_{-i}, (s_i, s_{-i})) - (SP(G, (s_i, s_{-i})) - s_i). \tag{1.1}$$

Observe that the quantity $SP(G, (s_i, s_{-i})) - s_i$ corresponds to the cost of the shortest path, according to weights (s_i, s_{-i}), without counting s_i. Here is a pictorial explanation of the idea behind these payments for the shortest path problem:

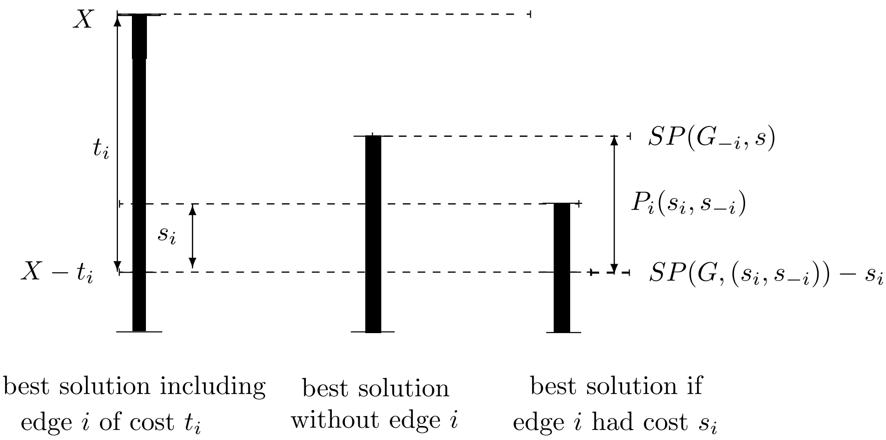


Observe that, the utility of an agent i whose edge is selected is equal to

$$\begin{aligned} P_i(t_i, s_{-i}) - t_i &= SP(G_{-i}, (s_i, s_{-i})) - (SP(G, (s_i, s_{-i})) - s_i) - t_i \quad (1.2) \\ &= SP(G_{-i}, (s_i, s_{-i})) - (SP(G, (t_i, s_{-i})). \quad (1.3) \end{aligned}$$

(The second equality es graphically shown above and easy to prove by considering the definition of $SP(\cdot)$.)
So, if agent i reports t_i, then he/she gets a non-negative utility. Moreover, if he/she reports a higher value $s_i > t_i$, such that edge i is still in the shortest path, then his/her payment does not change, thus also the utility does not change. Moreover, if s_i is "too high", so that the other solution becomes better, then the utility of agent i becomes 0 (the edge would not be selected anymore). Finally, lowering the reported cost to some $s_i < t_i$ does not change the payment nor the utility (the edge would be selected as well and payed the same amount).

The above considerations show that, if the shortest path with edge costs (t_i, s_{-i}), then AG_i has no incentive in reporting $s_i \neq t_i$. The next picture shows why the same holds when i is not in a shortest path for the weights (t_i, s_{-i}):



In this case, if AG_i reports $s_i < t_i$ so that his/her edge would be selected, then the payment he/she would get would not cover his/her cost t_i. Thus a non-positive utility, which cannot be better than reporting t_i and being not selected.

Using these arguments it is easy to show that the above payment functions guarantee that no agent owning an edge can improve his/her utility by misreporting his/her edge cost to the *mechanism* that selects the shortest path and pays the selected edges according to the payment functions in Eq. 1.1.

1.3 The VCG Theorem

The shortest path problem contains all ingredients to show the full power of VCG payments. Indeed, we did not use any particular property of the shortest path in order to prove that the payment functions in Eq. 1.1 work for this problem. In this section we formally define a class of optimization problem and a generalization of these payment functions such that, every problem in this class admits a so called *truthful mechanism*, that is, an algorithm A and a payment function $P = (P_1, P_2, \ldots, P_n)$ such that their combination $M = (A, P)$ incentivates the agents to report their true values t_is.

1.3.1 A general setting

We consider optimization problems defined as fourtuples $(\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$, where:

1. $\mathcal{I}$ is the set of *instances* of the problem;

2. $\mathsf{sol}(\cdot)$ is a function mapping every instance $I \in \mathcal{I}$ into a set $\mathsf{sol}(I)$ of *feasible solutions*;

3. $\mathsf{m}(\cdot)$ is the *measure* or *optimization function* mapping every $X \in \mathsf{sol}(I)$ into a non-negative real number $\mathsf{m}(X, I)$;

4. $\mathsf{goal} \in \{\min, \max\}$; the goal is to find an $X^* \in \mathsf{sol}(I)$ such that $\mathsf{m}(X^*, I) = \mathsf{opt}(I) := \mathsf{goal}_{X \in \mathsf{sol}(I)} \mathsf{m}(X, I)$. If $\mathsf{goal} = \min$ (respectively, $\mathsf{goal} = \max$) then Π is a *minimization* (respectively, *maximization*) problem.

We the consider a set of *selfish agents* that privately hold part of the input. In particular, for every $I \in \mathcal{I}$, we have that $I = (t, \sigma)$, where

1. $t = (t_1, t_2, \ldots, t_n)$ is the *private input*, and t_i is the *type* of agent AG_i (e.g., the cost of his/her edge);

2. σ is the *public part* of the input which is public knowledge (e.g., the nodes and the edges of a graph, in the case of the shortest path problem).

We consider agents that can report a value s_i in a set S_i, with $t_i \in S_i$. We also assume that the set of feasible solutions $\mathsf{sol}(I)$ does *not* depend on the agents type t. For every solution $X \in \mathsf{sol}(I)$, every agent AG_i has a *valuation function* $v_i(X, t_i)$: the function $v_i(\cdot, \cdot)$ is also public knowledge, but the type t_i is not. The value $v_i(X, t_i)$ represents a benefit (when positive) or a cost (when negative) for AG_i when a solution X is implemented (e.g., when a certain path containing his/her edge is selected). In our shortest path problem, the valuation functions are naturally defined as

$$v_i(X, t_i) = \begin{cases} -t_i & \text{if } X \text{ contains link } i, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2 (mechanism) *A mechanism for a problem* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$ *is a pair* $M = (A, P)$ *where, given an instance* $I = (t, \sigma)$*, (i)* A *is an algorithm computing a solution* $A(s, \sigma) \in \mathsf{sol}(I)$ *and (ii) each agent* AG_i *receives an amount of money equal to* $P_i(s_i, s_{-i})$*, with* $P = (P_1, P_2, \ldots, P_n)$*. In this case, the* utility *of* AG_i *is equal to* $u_i(s_i, s_{-i}) := v_i(A(s_i, s_{-i}), t_i) + P_i(s_i, s_{-i})$.

Definition 3 (truthful mechanism) *A mechanism $M = (A, P)$ for a problem $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$ is* truthful *if, for every agent AG_i, $i = 1, 2, \ldots, n$, the strategy $\overline{s}_i = t_i$ is dominant, that is,*

$$\forall s_{-i} \forall s_i \quad u_i(t_i, s_{-i}) \geq u_i(t_i, s_{-i}),$$

where $u_i(s_i, s_{-i}) = v_i(A(s_i, s_{-i}), t_i) + P_i(s_i, s_{-i})$.

Remark 4 *Observe that, if M is truthful, then we can assume that all agents will report their true type t_i. Therefore, algorithm A is provided with the correct input. So, if A is an exact (resp., c-approximate) algorithm for* Π *(without selfish agents), then the mechanism A guarantees that an exact (resp., c-approximate) solution.*

1.3.2 One more intuition on the shortest path problem: the agents "want to help" the algorithm

Let us step back to the shortest path problem for a second. Consider the utility function of agent AG_i when reporting s_i. With some abuse of notation, we let $SP((G, w')|w'')$ denote the cost of the shortest path for the instance (G, w') evaluated w.r.t. the instance (G, w''). By simply replacing t_i with s_i in Eq. 1.3 we obtain:

$$\begin{aligned} u_i(s_i, s_{-i}) &:= P_i(t_i, s_{-i}) - t_i = SP(G_{-i}, (s_i, s_{-i})) + && (1.4) \\ &- (SP(G, (s_i, s_{-i})) - s_i) - t_i && (1.5) \\ &= \underbrace{SP(G_{-i}, (s_i, s_{-i}))}_{\text{independent of } s_i} - SP((G, \underbrace{(s_i, s_{-i})}_{\substack{\text{provided}\\ \text{input}}})| \underbrace{(t_i, s_{-i})}_{\substack{\text{“correct”}\\ \text{input}}}). && (1.6) \end{aligned}$$

A few more words are needed. Clearly, the first quantity in Eq. 1.6 does not depend on s_i (by definition, edge i is not in the solution). Let us ask ourselves the following: how can AG_i maximize her/his utility? The only way is to minimize the quantity $SP((G(t_i, s_{-i}))| (t_i, s_{-i}))$, which we subtract. This quantity is defined as follows: a solution $X = A(G, (s_i, s_{-i}))$ is obtained by running a shortest path algorithm A on input (s_i, s_{-i}), and then the cost of X is computed w.r.t. the instance $(G, (t_i, s_{-i}))$. Since A finds a path of minimal cost when provided with the "correct " input (t_i, s_{-i}), the best that AG_i can do is to report $s_i = t_i$ so to "help" algorithm A in its job. So, the maximum value of $u_i(\cdot, s_{-i})$ is achieved for $s_i = t_i$.

This is yet another proof that the mechanism for the shortest path is truthful. There are two fundamental ingredients that we have used in the "proof":

1. In Eq. 1.5 we can "combine" the cost of a solution with the terms '$-s_i$' and 't_i': adding '$t_i - s_i$' to the cost of a solution yields the cost of the same solution on input (t_i, s_{-i});

2. We are using an algorithm A for the shortest path problem which returns a minimum-cost path when provided with the correct input. This is also fundamental for the mechanism in order to be truthful (see Exercise 17).

The first property yields a class of optimization functions for which the same idea works: the commonly termed *utilitarian* problems (see next section). The second property determines, for the class of utilitarian problems, which algorithms A can be turned into a truthful mechanism $M = (A, P)$: the optimal ones!

1.3.3 Truthful VCG mechanisms for Utilitarian Problems

As for the shortest path problem, utilitarian problems have the objective function which is the sum of all agents costs/valuations:

Definition 5 (utilitarian problem) *A minimization (resp., maximization) problem* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$ *is utilitarian if, for every instance* $I = (t, \sigma)$ *and for any* $X \in \mathsf{sol}(I)$*, it holds that*

$$\mathsf{m}(X, (t, \sigma)) = -\sum_{i=1}^{n} v_i(X, t_i)$$

(resp., $\mathsf{m}(X, (t, \sigma)) = \sum_{i=1}^{n} v_i(X, t_i)$*).*

Clearly, maximization utilitarian problems can be transformed into minimization ones and vice versa.

Example 6 (minimum spanning tree) *Consider the minimum spanning tree problem involving selfish agents owning one edge of a weighted graph each. The valuation function is defined analogously to the shortest path problem. It is easy to see that also this problem is a utilitarian optimization one.*

Definition 7 (VCG mechanism) *Let* A^* *be an optimal algorithm for a minimization utilitarian problem* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$*. Let*

$$\mathsf{m}(X, (\sigma, s_{-i})) := -\sum_{j \neq i}^{n} v_j(X, s_j)$$

and let

$$P_i^{VCG}(s_i, s_{-i}) := h_i(s_{-i}) - \mathsf{m}(A^*(\sigma, s), (\sigma, s_{-i})), \tag{1.7}$$

where $h_i(\cdot)$ *is any function independent of* s_i*. Then the resulting mechanism* $M = (A^*, P^{VCG})$ *is called* VCG mechanism *for* Π.

Theorem 8 (VCG theorem) *Any VCG mechanism for* $M = (A^*, P^{VCG})$ *a minimization utilitarian problem is truthful.*

The proof goes through a number of relatively easy steps/observations (see Exercise 18).

Observation 9 *If* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$ *is a minimization utilitarian problem, then*

$$\mathsf{m}(X, (\sigma, s_{-i})) - v_i(X, t_i) = \mathsf{m}(X, (\sigma, (t_i, s_{-i}))).$$

Lemma 10 *Let* $M = (A^*, P^{VCG})$ *be a VCG mechanism for a minimization utilitarian problem* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$*. Then, the utility function of agent* AG_i *is equal to*

$$u_i(s_i, s_{-i}) = h_i(s_{-i}) - \mathsf{m}(A^*(\sigma, (s_i, s_{-i})), (\sigma, (t_i, s_{-i}))).$$

Observation 11 *If* A^* *is an optimal algorithm for a minimization utilitarian problem* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$*, then the quantity*

$$\mathsf{m}(A^*(\sigma, (s_i, s_{-i})), (\sigma, (t_i, s_{-i})))$$

is minimized for $s_i = t_i$*.*

Now proving Theorem 8 is an easy quite simple (see Exercise 19).

1.3.4 Voluntary participation

The mechanism presented in Sect. 1.2.2 is a VCG mechanism in which $h_i(s_{-i})$ is the cost of the 2^{nd} best shortest path. If we look at the picture at pag. 10, we realize that this choice has a good feature: we always cover the costs of a truthful agent, that is, $u_i(t_i, s_{-i}) \geq 0$. This property is called *voluntary participation.*

Intuitively, we can achieve voluntary participation using VCG mechanisms whenever there exists an "alternative solution" X_{-i} that does not "involve" AG_i.

Definition 12 *Let* $\mathsf{sol}(\sigma_{-i})$ *denote the set of solutions* $X_{-i} \in \mathsf{sol}(\sigma)$ *such that, for every* t_i*,* $v_i(X_{-i}, t_i) = 0$*.*

Theorem 13 *Let* $\Pi = (\mathcal{I}, \mathsf{m}, \mathsf{sol}, \mathsf{goal})$ *be a minimization utilitarian problem* $M = (A^*, P^{VCG})$ *such that, for every* $I = (\sigma, t)$ *and for every* $i = 1, 2, \ldots, n$*,* $\mathsf{sol}(\sigma_{-i}) \neq \emptyset$*. Let* M *be the VCG mechanism for* Π *obtained by setting, in Def. 7,* $h_i(s_{-i}) = \min_{X \in \mathsf{sol}(\sigma_{-i})}\{\mathsf{m}(X, (\sigma, s_{-i}))\}$*. Then,* M *satisfies the voluntary participation constraint.*

Proof. Exercise 21. □

Exercises

Exercise 14 *Show that, for the n links problem, if we pay the cost of the 3^{rd} cheapest link (instead of the 2^{nd} one), then there is an instance for which some agent can improve her/his utility by reporting a false cost.*

Exercise 15 *Consider the following payment for the n links problem. We pay to the selected link i a quantity which is the average between the best (i.e., the cost t_i of the selected link) and the 2^{nd} best link:*

$$P_i = \frac{t_i + \min_{j \neq i}\{t_j\}}{2}$$

Show that, in this case, there is an instance for which some agent can improve her/his utility by reporting a false cost.

Exercise 16 *Give a proof from scratch of Theorem 1 , i.e., without making use of Theorem 8.*

Exercise 17 *Consider the mechanism for the shortest path problem and replace the algorithm A computing a shortest path with an algorithm A' which computes a 2^{nd} shortest path. Let $SP'(\cdot)$ be defined as $SP(\cdot)$ with the only difference that we now consider the cost of the 2^{nd} shortest path. Let us define payments $P'_i(\cdot)$ by replacing, in Eq. 1.1, $SP(\cdot)$ by $SP'(\cdot)$. Is the mechanism $M' = (A', P')$ truthful?*

Exercise 18 *Prove Observation 9, Lemma 10 and Observation 11.*

Exercise 19 *Prove Theorem 8. Also, define the payment functions for a utilitarian maximization problem.*

Exercise 20 *Using Theorem 8, show that the shortest path problem admits a truthful mechanism also when an agent owns more than one edge. Also, define the payment functions for this case.*

Exercise 21 *Prove Theorem 13.*

Bibliography

[1] E.H. Clarke. Multipart Pricing of Public Goods. *Public Choice*, pages 17–33, 1971.

[2] T. Groves. Incentive in Teams. *Econometrica*, 41:617–631, 1973.

[3] W. Vickrey. Counterspeculation, Auctions and Competitive Sealed Tenders. *Journal of Finance*, pages 8–37, 1961.