Average-Case Approximation Ratio
of Scheduling without Payments

Jie Zhang

Electronics and Computer Science
University of Southampton, United Kingdom
jie.zhang@soton.ac.uk

Abstract

Apart from the principles and methodologies inherited from Economics and Game Theory, the studies in Algorithmic Mechanism Design typically employ the *worst-case analysis* and *approximation schemes* of Theoretical Computer Science. For instance, the *approximation ratio*, which is the canonical measure of evaluating how well an incentive-compatible mechanism approximately optimizes the objective, is defined in the worst-case sense. It compares the performance of the optimal mechanism against the performance of a truthful mechanism, for all possible inputs.

In this paper, we take the *average-case analysis* approach, and tackle one of the primary motivating problems in Algorithmic Mechanism Design – the scheduling problem (Nisan and Ronen 1999). One version of this problem which includes a verification component is studied by Koutsoupias (2014). It was shown that the problem has a tight approximation ratio bound of $(n + 1)/2$ for the single-task setting, where n is the number of machines. We show, however, when the costs of the machines to executing the task follow *any* independent and identical distribution, the *average-case approximation ratio* of the mechanism given in (Koutsoupias 2014) is upper bounded by a constant. This positive result asymptotically separates the average-case ratio from the worst-case ratio, and indicates that the optimal mechanism for the problem actually works well on average, although in the worst-case the expected cost of the mechanism is $\Theta(n)$ times that of the optimal cost.

Introduction

The field of Algorithmic Mechanism Design (Nisan and Ronen 1999; Nisan et al. 2007; Procaccia and Tennenholtz 2009) focuses on optimization problems where the input is provided by self-interested agents that participate in the mechanism by reporting their private information. These agents are utility maximizers so they may misreport their private information to the mechanism, if that results in a more favorable outcome to them. Given the reports from the agents as input, a *mechanism* is a function that maps the input to allocations and payments if monetary transfers are allowed. The goal of the mechanism designer is twofold. On the one hand, the objective is to motivate agents to always report truthfully, regardless of what strategies the other

agents follow; on the other hand, the aim is to optimize a specific objective function that measures the quality of the outcome, subject to a polynomial-time implementability constraint. However, these objectives are usually incompatible. Therefore, often we need to trade one objective to achieve the other. One standard approach is to maintain the *truthfulness* property of the mechanism, and approximately optimize the specific objective function (e.g., social welfare maximization, revenue maximization, or cost minimization). The *approximation ratio* is the canonical measure for evaluating the performance of a truthful mechanism towards this goal. It compares the performance of the truthful mechanism against the optimal mechanism which is not necessarily truthful, over all possible inputs.

The approximation ratio is defined in the worst-case sense which resembles the worst-case time complexity of the algorithms. These are strong but very pessimistic measures. On the one hand, if it is possible to obtain a small worst-case ratio, then it is a very solid guarantee on the performance of the mechanism no matter what input is provided. On the other hand, if it turns out to be a large value, then one can hardly judge the performance of the mechanism as it may perform well on most inputs and perform poorly on only a few inputs. To address this issue, Deng, Gao, and Zhang (2017) propose an alternative measure, the *average-case approximation ratio*, that compares the performance of the truthful mechanism against the optimal mechanism, averaged over all possible inputs when they follow a certain distribution. Although average-case analysis is usually more complex than the worst-case analysis, it provides a different measure.

In this paper, we study the problem of scheduling unrelated machines without money. Scheduling is one of the primary problems in algorithmic mechanism design. In the general setting, the problem is to schedule a set of tasks to n unrelated machines with private processing times, in order to minimize the makespan. The machines (alternatively, speaking of the agents in game theoretical settings) are rational and want to minimize their execution time. They may achieve this by misreporting their processing times to the mechanism. No monetary payments are allowed in this problem. The objective is to design truthful mechanisms with good approximation ratio. One important version of the problem is studied by (Koutsoupias 2014) in which the ma-

chines are bound by their declarations. More specifically, in case a machine declares a longer time than its actual time for a task and it is allocated the task, then in practice its processing time must be the declared value. This is in the spirit that machines have been observed during the execution of the task and cannot afford to be caught lying about the execution times (an unaffordable penalty would apply). One can consider this as a monitoring model of the problem and it is more complex than some other models. Koutsoupias (2014) devises a truthful-in-expectation mechanism which achieves the tight bound of $\frac{n+1}{2}$ for one task, and generalizes to $\frac{n(n+1)}{2}$ for multiple tasks when the objective is minimizing makespan and $\frac{n+1}{2}$ when the objective is minimizing social cost. We note that, the tight bound instance is obtained when the ratio of the minimum value of the processing times against the maximum value of the processing times approaches 0. Obviously this instance is very unlikely to occur in practice. Therefore, it would be interesting to understand how well the optimal mechanism given in (Koutsoupias 2014) performs on average, when the instances are chosen from a certain distribution.

Our contribution

This paper provides a fresh view of the performance of the mechanism developed by Koutsoupias (2014). In particular, we show the following:

- The average-case approximation ratio of the mechanism devised by Koutsoupias (2014) is upper bounded by a constant, no matter which distribution is given.

In contrast, the worst-case approximation ratio of the mechanism shown in (Koutsoupias 2014) is $\frac{n+1}{2}$ which is asymptotically different.

A major criticism of the average-case analysis is that the results usually depend on the assumptions about the distribution of inputs, and these are not guaranteed to hold in practice. Even for the same mechanism, when it is applied to different application areas, the real-world distribution may vary. For example, Deng, Gao, and Zhang (2017) show that their average-case result holds for a uniform distribution; the positive results in the Bayesian analysis of auctions usually need to assume that the hazard rate function is monotone non-decreasing. In this paper, we develop a powerful proof that works for any i.i.d distribution. This is extraordinary in the average-case analysis literature, even in the more established space such as the analysis of time complexity of algorithms. Our results imply that, although the worst-case approximation ratios of the known mechanisms are quite negative, the average-case approximation ratios are rather positive.

Related work

The field of Algorithmic Mechanism Design was initiated by Nisan and Ronen (1999) and is further enriched by Procaccia and Tennenholtz (2009) to approximate mechanism design without money. For a more detailed investigation, we refer the reader to Nisan et al. (2007).

The scheduling problem has been extensively studied. However, after nearly two decades, we still have little progress in resolving this challenge. The known approximation ratio upper bounds are rather negative, and there is a large gap between the lower bounds and upper bounds. For the model presented by (Nisan and Ronen 1999) where payments are allowed to facilitate designing truthful mechanisms, the best known upper bound is given in their original paper and is achieved by allocating each task independently using the classical VCG mechanism, while the best known lower bound is 2.61 (Koutsoupias and Vidali 2013). Ashlagi, Dobzinski, and Lavi (2012) prove that the upper bound of n is tight for anonymous mechanisms. For randomized mechanisms, the best known upper bound is $\frac{n+1}{2}$ shown by (Mu'alem and Schapira 2007). For the special case of related machines, where the private information of each machine is a single value, Archer and Tardos (2001) give an randomized 3-approximation mechanism. Lavi and Swamy (2009) show a constant approximation ratio for the special case that the processing times of each task can take one of two fixed values. Yu (2009) generalizes this result to two-range-values, while together with (Lu and Yu 2008) and (Lu 2009), they show constant bounds for the case of two machines. For the setting that payments are not allowed, Koutsoupias (2014) first considers the setting that the machines are bound by their declarations. This is influenced by the notion of impositions that appears in (Fotakis and Tzamos 2013) for the facility location problems, as well as the notion of verification that appears in (Auletta et al. 2009). Penna and Ventre (2014) present a general construction of collusion-resistant mechanisms with verification that return optimal solutions for a wide class of mechanism design problems, including the scheduling problem. The mechanism presented in (Koutsoupias 2014) has a tight approximation ratio bound of $\frac{n+1}{2}$ for the single-task setting; by running the mechanism independently on multiple tasks a tight bound of $\frac{n+1}{2}$ can be achieved for social cost minimization and an upper bound of $\frac{n(n+1)}{2}$ can be achieved for the makespan minimization. Kovács, Meyer, and Ventre (2015) further apply mechanism design with monitoring techniques to the truthful RAM allocation problem. There are some works on characterizing truthful mechanisms in scheduling problems, such as (Kovács and Vidali 2015), as well as scheduling with uncertain execution time, such as (Conitzer and Vidali 2014).

In (Deng, Gao, and Zhang 2017), the authors propose to study the average-case and smoothed approximation ratios, and conduct these analyses on the one-sided matching problem. They show that, although the asymptotically best truthful mechanism for the problem is random priority and its worst-case approximation ratio is bounded by $\Theta(\sqrt{n})$, random priority has a constant average-case approximation ratio when the inputs follow a uniform distribution, and it has a constant smoothed approximation ratio.

Notably, the analysis of average-case approximation ratio takes a similar but fundamentally different approach to the Bayesian analysis. In the Bayesian auction design literature (Chawla and Sivan 2014; Hartline and Lucier 2010), the focus is on how well a truthful mechanism can approximately maximize the expected revenue, when instances are taken

from the entire input space. More specifically, the dominant approach in the study of Bayesian auction design is the *ratio of expectations*. A more detailed comparison of the two metrics will be given in the next section after their definitions are given.

Preliminaries

In the problem of scheduling unrelated machines without payment, there are a set of self-interested machines (alternatively, speaking of self-interested agents in game theoretical settings) and a set of tasks. The general setting comprises n machines and m tasks. In this paper we consider the setting of a single task. The machines are lazy and prefer not to execute any tasks. There are no monetary tools to incentivise machines to execute tasks. Machine i needs time (or cost) t_i to execute the task, $i \in [n]$. These t_i's are independent to each other. There could be two different objectives. One is to allocate the task to machines so that the makespan is minimized; the other is to allocate the task so that the social cost is minimized. The *makespan* is the total length of the schedule, and the *social cost* is the sum of all agents' costs. In the single-task setting, these two objectives are identical. Obviously, allocating the task to the machine with minimum execution time is the optimal solution. However, the mechanism has no access to the values t_i. Instead, each machine reports an execution time $\tilde{t}_i$ to the mechanism, where $\tilde{t}_i$ is not necessarily equal to $t_i, \forall i \in [n]$. A mechanism is a (possibly randomized) algorithm which computes an allocation based on the declarations $\tilde{t}_i$ of the machines. Denote the output of the mechanism by $\mathbf{p} = (p_i)_{i \in [n]}$, where p_i is an indicator variable in deterministic mechanisms and is the probability of machine i getting allocated to execute the task in randomized mechanisms. We follow the standard literature and consider the case that machines are bound by their reports. That is, the cost of machine i for the task is $\max\{t_i, \tilde{t}_i\}$. So in case a machine i declares $\tilde{t}_i \geq t_i$ and it is allocated the task, then its actual cost is the declared value $\tilde{t}_i$ and not t_i. This is in the spirit that machines are being observed during the execution of the task and cannot afford to be caught lying about the execution times (a high penalty would apply). Therefore, the expected cost of machine i is $c_i = c_i(t_i, \tilde{\mathbf{t}}) = p_i(\tilde{\mathbf{t}}) \max(t_i, \tilde{t}_i)$. In approximate mechanism design, we restrict our interest to the class of *truthful mechanisms*. A mechanism is *truthful* if for any report of other agents, the expected cost c_i of agent i is minimized when $\tilde{t}_i = t_i$. We note that this weak notion of truthfulness, truthful-in-expectation, enables us to consider a richer class of mechanisms than universal truthfulness. However, as mentioned in the Introduction, even with this rich class of truthful mechanisms, the performance of these mechanisms is still very limited in terms of approximation ratio. The canonical measure of efficiency of a truthful mechanism M is the *worst-case approximation ratio*,

$$r_{\text{worst}}(M) = \sup_{\mathbf{t} \in \mathcal{T}} \frac{SC_M(\mathbf{t})}{SC_{\text{OPT}}(\mathbf{t})},$$

where $SC_{\text{OPT}}(\mathbf{t}) = \min_{\mathbf{p} \in \mathcal{P}} \sum_{i=1}^{n} c_i$ is the optimal social cost which is essentially the minimum t_i, for all $i \in [n]$;

$SC_M(\mathbf{t})$ is the social cost of the mechanism M on the input $\mathbf{t}$; and $\mathcal{T}$ is the input space. This ratio compares the social cost of the truthful mechanism M against the social cost of the optimal mechanism OPT over all possible inputs $\mathbf{t}$.

In (Koutsoupias 2014), the author devises the following randomized mechanism.

Mechanism M: Given the input $\mathbf{t} = (t_1, \ldots, t_n)$, without loss of generality, let the values of t_i's be in ascending order $0 < t_1 \leq t_2 \leq \cdots \leq t_n$. Then the allocation probabilities are

$$p_1 = \frac{1}{t_1} \int_0^{t_1} \prod_{i=2}^{n} \left(1 - \frac{y}{t_i}\right) dy,$$

$$p_k = \frac{1}{t_1 t_k} \int_0^{t_1} \int_0^{y} \prod_{\substack{i=2,\ldots,n \\ i \neq k}} \left(1 - \frac{x}{t_i}\right) dx dy, \text{ for } k \neq 1.$$

Note that this is a *symmetric* mechanism, so it suffices to describe it when $0 < t_1 \leq t_2 \leq \cdots \leq t_n$. It is shown in (Koutsoupias 2014) that this mechanism is truthful and achieves an approximation ratio tight bound of $\frac{n+1}{2}$.

Analogously to the definition of the *average-case approximation ratio* of mechanisms for social welfare maximization in (Deng, Gao, and Zhang 2017), we define it for social cost minimization as follows:

$$r_{\text{average}}(M) = \mathbb{E}_{\mathbf{t} \sim D} \left[\frac{SC_M(\mathbf{t})}{SC_{\text{OPT}}(\mathbf{t})} \right],$$

where the input $\mathbf{t} = (t_1, \ldots, t_n)$ is chosen from a distribution D. Hence, the metric we study in this paper is *the expectation of the ratio*.

Comparison with the Bayesian Approach

In Bayesian mechanism design (Chawla and Sivan 2014; Hartline and Lucier 2010), there is also a prior distribution from which the agent types come from. However, the objective is to characterize the maximum ratio (for some given distribution of the agent types) of the expected social welfare (or social cost) of a truthful mechanism over the expected social welfare (or social cost) of the optimal mechanism. So, the metric in the study of the Bayesian approach is *the ratio of expectations*. That is, the objective is to characterize the ratio r in the following formula,

$$r \cdot \mathbb{E}\left[SC_{\text{OPT}}(\mathbf{t})\right] \leq \mathbb{E}\left[SC_M(\mathbf{t})\right].$$

Therefore, in Bayesian approach, the optimal mechanism is in respect of the entire input space. In other words, it outputs the optimal solution in expectation, when the inputs run over the entire prior distribution. In contrast, in the analysis of average-case approximation ratio, the optimal mechanism is in respect of each individual input instance. So, if we interpret the optimal mechanism as an adversary to the truthful mechanism, this adversary can take different actions in each round.

In light of this difference, also due to the fact that the expectation of the ratio is a nonlinear function of the two random variables, the analysis of average-case approximation ratio introduces more technical challenges. In some specific scenarios, a constant average-case approximation ratio

would imply a constant approximation ratio under Bayesian approach.

Average-case Approximation Ratio

In this section we show that the average-case approximation ratio of the mechanism M is upper bounded by a constant, when the inputs t_i's follow any independent and identical distribution $D[t_{\min}, \infty)$, where $t_{\min}$ is the minimum processing time for the task.

Let h be a constant, and denote event $A = \{t_{\frac{n}{2}} \leq h \cdot t_{\min}\}$. So, it corresponds to the case that the $\frac{n}{2}$-th order statistic of the inputs t_i's is less than or equal to $h \cdot t_{\min}$. Firstly, we show that if A is true, then the social cost of the mechanism M is upper bounded by a constant times t_1.

Lemma 1 *For any constant $h > 0$, given that event A holds, we have*

$$SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1.$$

Proof: The expected cost of the mechanism M is

$$SC_{\mathrm{M}}(\mathbf{t}) = \sum_{i=1}^{n} p_i \cdot t_i = \int_0^{t_1} \prod_{i=2}^{n} \left(1 - \frac{y}{t_i}\right) dy$$
$$+ \sum_{k=2}^{n} \frac{1}{t_1} \int_0^{t_1} \int_0^{y} \prod_{\substack{i=2,\ldots,n \\ i \neq k}} \left(1 - \frac{x}{t_i}\right) dx dy$$

Since $1 - \frac{y}{t_i} \leq 1$ for any $y \in [0, t_1]$, $i = 2, \ldots, n$, we can simply bound the first term by

$$\int_0^{t_1} \prod_{i=2}^{n} \left(1 - \frac{y}{t_i}\right) dy \leq \int_0^{t_1} 1 dy = t_1$$

Because event A holds, i.e., $t_{\frac{n}{2}} \leq h \cdot t_{\min}$, we have $\prod_{\substack{i=2,\ldots,n \\ i \neq k}} \left(1 - \frac{x}{t_i}\right) \leq \left(1 - \frac{x}{h \cdot t_{\min}}\right)^{\frac{n}{2}-2} \cdot 1^{\frac{n}{2}}, \forall k = 2, \ldots, n$. So we can bound the second term as follows.

$$\sum_{k=2}^{n} \frac{1}{t_1} \int_0^{t_1} \int_0^{y} \prod_{\substack{i=2,\ldots,n \\ i \neq k}} \left(1 - \frac{x}{t_i}\right) dx dy$$
$$\leq \sum_{k=2}^{n} \frac{1}{t_1} \int_0^{t_1} \int_0^{y} \left(1 - \frac{x}{h \cdot t_{\min}}\right)^{\frac{n}{2}-2} \cdot 1^{\frac{n}{2}} dx dy$$
$$\leq \frac{n-1}{t_1} \int_0^{t_1} \int_0^{y} \left(1 - \frac{x}{h \cdot t_1}\right)^{\frac{n}{2}-2} dx dy$$
$$= \frac{n-1}{t_1} \int_0^{t_1} \left(\frac{2ht_1}{n-2} - \frac{2ht_1}{n-2}\left(1 - \frac{y}{ht_1}\right)^{\frac{n}{2}-1}\right) dy$$
$$= \frac{n-1}{t_1} \left[\frac{2ht_1^2}{n-2} + \frac{4h^2t_1^2}{n(n-2)}\left(\left(1 - \frac{1}{h}\right)^{\frac{n}{2}} - 1\right)\right]$$
$$\leq \frac{n-1}{n-2} \cdot 2ht_1$$
$$< 2h \cdot t_1$$

So, $SC_{\mathrm{M}}(\mathbf{t}) = \sum_{i=1}^{n} p_i \cdot t_i \leq (2h+1)t_1$. $\square$
Since $SC_{\mathrm{OPT}}(\mathbf{t}) = t_1$, we get the following Corollary.

Corollary 1 *When event A holds, we have*

$$\mathbb{E}_{\mathbf{t} \sim D} \left[\frac{SC_{\mathrm{M}}(\mathbf{t})}{SC_{\mathrm{OPT}}(\mathbf{t})}\right] \leq 2h + 1.$$

Obviously, Lemma 1 and Corollary 1 hold regardless of the distribution.

Secondly, we show that there exists a constant h such that event A occurs with a large probability. Intuitively, the larger h is, the higher probability that event A occurs. We will need the following Lemma to find such an h.

Lemma 2 *For any $n > 1$, we have $\binom{n}{n/2} \leq \frac{e}{\pi\sqrt{n}} \cdot 2^n$, where e is the base of the natural logarithm.*

Proof: According to the estimation by (Robbins 1955),

$$n! = \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+r(n)},$$

where $\frac{1}{12n+1} < r(n) < \frac{1}{12n}$. Here we only need a looser bound to prove our lemma, i.e.,

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n} < n! < \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+\frac{1}{12n}} < n^{n+\frac{1}{2}}e^{-n+1}.$$

We have

$$\binom{n}{n/2} = \frac{n!}{(\frac{n}{2})!(\frac{n}{2})!} \leq \frac{e\sqrt{n}(\frac{n}{e})^n}{\left(\sqrt{2\pi\frac{n}{2}}\left(\frac{n}{2e}\right)^{\frac{n}{2}}\right)^2} = \frac{e}{\pi\sqrt{n}} \cdot 2^n$$

$\square$

Next we show that event A can occur with a large probability, with a properly choosing h.

Lemma 3 *For any $n > 1$, there exists a constant h, such that $F(ht_{\min}) \geq \frac{11}{12}$, and we have*

$$\Pr[A] \geq 1 - \frac{e}{2\pi} \cdot \frac{1}{n}$$

Proof: Since event $A = \{t_{\frac{n}{2}} \leq h \cdot t_{\min}\}$, the probability that event A occurs can be calculated by

$$\Pr[A] = \Pr[t_{\frac{n}{2}} \leq h \cdot t_{\min}]$$
$$= \sum_{k=\frac{n}{2}}^{n} \binom{n}{k} (F(ht_{\min}))^k (1 - F(ht_{\min}))^{n-k}$$
$$= 1 - \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{k} (F(ht_{\min}))^k (1 - F(ht_{\min}))^{n-k}$$
$$(1)$$

Since $(1 - F(ht_{\min}))^{n-k} \leq (1 - F(ht_{\min}))^{n/2}$, $\binom{n}{k} \leq \binom{n}{n/2}$, $k = 0, \cdots, \frac{n}{2} - 1$, and $(F(ht_{\min}))^k < 1$, we get

$$(1) \geq 1 - \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{n/2} (1 - F(ht_{\min}))^{\frac{n}{2}}$$
$$= 1 - \frac{n}{2}\binom{n}{n/2} (1 - F(ht_{\min}))^{\frac{n}{2}}$$
$$\geq 1 - \frac{n}{2} \cdot \frac{e}{\pi\sqrt{n}} \cdot 2^n \cdot (1 - F(ht_{\min}))^{\frac{n}{2}} \quad (2)$$

By choosing h such that $F(ht_{\min}) \geq \frac{11}{12}$, we get

$$(2) \geq 1 - \frac{n}{2} \cdot \frac{e}{\pi\sqrt{n}} \cdot 2^n \cdot \left(\frac{1}{12}\right)^{\frac{n}{2}}$$

$$= 1 - \frac{e}{2\pi} \cdot \sqrt{n} \cdot 3^{-\frac{n}{2}}$$

$$\geq 1 - \frac{e}{2\pi} \cdot \frac{1}{n}$$

The last inequality is due to $3^n \geq n^3, \forall n > 1$. Therefore, $\Pr[A] \geq 1 - \frac{e}{2\pi} \cdot \frac{1}{n}$. $\qquad\square$

We can then bound the probability of the case that the expected cost of the mechanism M is larger than $(2h+1)t_1$.

Lemma 4 *When event* A *holds, and for the choice of h in the above lemma, we have*

$$\Pr\left[SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1\right] < \frac{e}{2\pi} \cdot \frac{1}{n}.$$

Proof: According to Lemma 1, event A implies $SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1$, so we have

$$\Pr[A] \leq \Pr[SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1]$$

Hence,

$$1 - \Pr[A] \geq \Pr[SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1]$$

According to Lemma 3, we have

$$1 - \Pr[A] \leq \frac{e}{2\pi} \cdot \frac{1}{n}$$

So,

$$\Pr\left[SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1\right] < \frac{e}{2\pi} \cdot \frac{1}{n}$$

$\qquad\square$

We have established necessary building blocks. By carefully choosing the parameter h, we can partition the valuation space into two sets: $\{SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1\}$ and $\{SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1\}$. Last, we will use Corollary 1 and Lemma 4 to prove our main result. Essentially, Corollary 1 upper bounds the expected approximation ratio of the first case and Lemma 4 upper bounds the probability of the second case occurring. Note that in any case, the worst-case ratio is upper bounded by $\frac{n+1}{2}$ according to (Koutsoupias 2014). By adding them up together, we obtain our upper bound.

Theorem 1 *For any distribution on $[t_{\min}, +\infty)$, and a constant h such that $F(ht_{\min}) \geq \frac{11}{12}$, the average-case approximation ratio of the mechanism* M *is upper bounded by $2h + 1.33$. That is,*

$$r_{average} = \mathbb{E}_{\mathbf{t}\sim D}\left[\frac{SC_{\mathrm{M}}(\mathbf{t})}{SC_{\mathrm{OPT}}(\mathbf{t})}\right] < 2h + 1.33$$

Proof: It is easy to see that the above two sets are collectively exhaustive and mutually exclusive, and Lemma 4 holds. So we have

$$r_{average} = \mathbb{E}_{\mathbf{t}\sim D}\left[\frac{SC_{\mathrm{M}}(\mathbf{t})}{SC_{\mathrm{OPT}}(\mathbf{t})}\right]$$

$$\leq \Pr\left[SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1\right] \cdot \mathbb{E}\left[\frac{SC_{\mathrm{M}}(\mathbf{t})}{SC_{\mathrm{OPT}}(\mathbf{t})}\right] +$$

$$\Pr\left[SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1\right] \cdot \mathbb{E}\left[\frac{SC_{\mathrm{M}}(\mathbf{t})}{SC_{\mathrm{OPT}}(\mathbf{t})}\right]$$

$$\leq \Pr\left[SC_{\mathrm{M}}(\mathbf{t}) \leq (2h+1)t_1\right] \cdot (2h+1) +$$

$$\Pr\left[SC_{\mathrm{M}}(\mathbf{t}) > (2h+1)t_1\right] \cdot \frac{n+1}{2}$$

$$\leq 1 \cdot (2h+1) + \frac{e}{2\pi} \cdot \frac{1}{n} \cdot \frac{n+1}{2}$$

$$= 2h + 1 + \frac{e}{4\pi} \cdot \frac{n+1}{n}$$

$$\leq 2h + 1 + \frac{3e}{8\pi}$$

$$< 2h + 1.33$$

Therefore, the average-case approximation ratio of the mechanism M is upper bounded by $2h + 1.33$. $\qquad\square$

In hindsight, when the costs of the machines t_i's follow any heavy-tailed distribution, no matter how heavy is the tail, the mechanism M has a constant average-case approximation ratio bound. However, this was not intuitively clear beforehand, as the social cost of the mechanism M depends on how often the inputs contain large t_i and how big they are.

In the following, we give a few examples of the distribution to show the choice of h and the constant upper bounds for these distributions.

Example 1: Pareto Distribution

The Pareto distribution is a power law distribution that is widely used in the description of social, scientific, geophysical, actuarial, and many other types of observable phenomena. According to the influential studies by (Arlitt and Williamson 1996) and (Reed and Jorgensen 2004) as well as the references therein, the distributions of the file size (of web server workload and of Internet traffic which uses the TCP protocol) match well with the Pareto distribution.

That is, for a random variable T chosen from this Pareto distribution, the probability that T is smaller than a value t, is given by

$$F(t) = \Pr(T < t) = \begin{cases} 1 - \left(\frac{t_{\min}}{t}\right)^\alpha & t \geq t_{\min} \\ 0 & t < t_{\min} \end{cases}$$

where $\alpha > 0$ is the tail index of the distribution.

Note that in the proof of Theorem 1, the only place we need to deal with the particular distribution is Lemma 3. So, by handling the constant h for the Pareto distribution, we obtain the following result.

Theorem 2 *For the Pareto distribution, let $h = 12^{\frac{1}{\alpha}}$. The average-case approximation ratio of the mechanism* M *is upper bounded by $2 \cdot 12^{\frac{1}{\alpha}} + 1.33$.*

Proof: For the Pareto distribution, let $h = 12^{\frac{1}{\alpha}}$. In Lemma 3, we would have

$$\Pr[A] = \Pr[t_{\frac{n}{2}} \leq h \cdot t_{\min}]$$

$$= \sum_{k=\frac{n}{2}}^{n} \binom{n}{k} \left(F(ht_{\min})\right)^k \left(1 - F(ht_{\min})\right)^{n-k}$$

$$= \sum_{k=\frac{n}{2}}^{n} \binom{n}{k} \left(1 - \frac{1}{h^\alpha}\right)^k \left(\frac{1}{h^\alpha}\right)^{n-k}$$

$$= 1 - \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{k} \left(1 - \frac{1}{h^\alpha}\right)^k \left(\frac{1}{h^\alpha}\right)^{n-k}$$

$$\geq 1 - \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{n/2} \left(\frac{1}{h^\alpha}\right)^{\frac{n}{2}}$$

$$= 1 - \frac{n}{2} \binom{n}{n/2} \left(\frac{1}{h^\alpha}\right)^{\frac{n}{2}}$$

$$\geq 1 - \frac{n}{2} \cdot \frac{e}{\pi\sqrt{n}} \cdot 2^n \left(\frac{1}{h^\alpha}\right)^{\frac{n}{2}}$$

$$= 1 - \frac{n}{2} \cdot \frac{e}{\pi\sqrt{n}} \cdot 2^n \left(\frac{1}{12}\right)^{\frac{n}{2}}$$

$$= 1 - \frac{e}{2\pi} \cdot \sqrt{n} \cdot 3^{-\frac{n}{2}}$$

$$\geq 1 - \frac{e}{2\pi} \cdot \frac{1}{n}$$

The rest of the proof follows the proof in Theorem 1. $\square$

Example 2: Exponential Distribution

We then consider the case that machines' costs t_i's are independent variables and follow a truncated Exponential distribution $D[t_{\min}, \infty)$. That is, for a random variable T chosen from this Exponential distribution, the probability that T is smaller than a value t, is given by

$$F(t) = \Pr(T < t) = \begin{cases} 1 - \frac{1}{e^{\lambda t}} & t \geq t_{\min} \\ 0 & t < t_{\min} \end{cases}$$

where $\lambda > 0$ is the tail index of the distribution.

Theorem 3 *For the Exponential distribution, let* $h = \frac{1}{\lambda t_{\min}} \ln 12$. *The average-case approximation ratio of the mechanism* M *is upper bounded by* $2 \cdot \frac{1}{\lambda t_{\min}} \ln 12 + 1.33$.

The proof is omitted here.

Example 3: Log-logistic Distribution

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails. It is used in networking to model the transmission times of data. The cumulative distribution function is

$$F(t; \alpha, \beta) = \frac{t^\beta}{\alpha^\beta + t^\beta}$$

where $\alpha > 0$ is a scale parameter and $\beta > 0$ is a shape parameter. For simplicity, we take $\alpha = 1$ and have the following result.

Theorem 4 *For the Log-logistic distribution, let* $h = \frac{1}{t_{\min}} \cdot e^{\frac{\ln 11}{\beta}}$. *The average-case approximation ratio of the mechanism* M *is upper bounded by* $2 \cdot \frac{1}{t_{\min}} \cdot e^{\frac{\ln 11}{\beta}} + 1.33$.

Conclusion and Future Work

In this paper, we extend the worst-case approximation ratio analysis for the scheduling problem studied in (Koutsoupias 2014) to the average-case approximation ratio analysis. We show that, when the costs of the machines are independent and identically distributed, the average-case approximation ratios of the optimal mechanism M have constant bounds, which is asymptotically better than the worst-case approximation ratios. Our results offer some relief for applying the mechanism M in practice, apart from the fact that in the worst case the expected cost of the mechanism is $\Theta(n)$ times of what the optimal cost is.

A lot of problems remain open. Firstly, as we employ the worst-case analysis as a framework for comparing truthful mechanisms, the average-case analysis is also a framework for doing so. Although the mechanism M in (Koutsoupias 2014) is the best mechanism for the problem in terms of the worst-case ratio, it is not clear whether there is any other mechanism could be better than M in terms of average-case ratio. In any attempts to comparing M and other mechanisms, the specific distribution that the inputs follow matters, as we have pointed out that the assumptions on the distribution are vital in the average-case analysis so the comparison has to be done on the same distribution. There may not be a mechanism that is universally best for any distribution.

Secondly, it would be interesting to show some lower bounds for the average-case ratio of any truthful mechanism, but one should expect some much more involved arguments than their worst-case lower bound counterparts, and again, very likely different distributions need to be handled differently.

One might query the smoothed analysis of the mechanism M. We should note that, unlike the random priority mechanism studied in (Deng, Gao, and Zhang 2017) that has a constant smoothed approximation ratio, the smoothed ratio of the mechanism M would not be asymptotically different from the worst-case ratio. To see this, the tight bound example in the worst-case analysis of the mechanism M is obtained when the ratio of the minimum value of the processing times against the maximum value of the processing times approaches 0, i.e., $t_1/t_n \to 0$. Obviously, any small perturbation around these inputs would not change the nature of this fact.

Many more approximate mechanism design problems deserve average-case analysis to understand the nature of the problems and the performance of the mechanisms.

References

Archer, A., and Tardos, É. 2001. Truthful mechanisms for one-parameter agents. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, 482–491. IEEE Computer Society.

Arlitt, M. F., and Williamson, C. L. 1996. Web server workload characterization: The search for invariants. In *Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, 126–137.

Ashlagi, I.; Dobzinski, S.; and Lavi, R. 2012. Optimal lower bounds for anonymous scheduling mechanisms. *Math. Oper. Res.* 37(2):244–258.

Auletta, V.; Prisco, R. D.; Penna, P.; and Persiano, G. 2009. The power of verification for one-parameter agents. *J. Comput. Syst. Sci.* 75(3):190–211.

Chawla, S., and Sivan, B. 2014. Bayesian algorithmic mechanism design. *SIGecom Exchanges* 13(1):5–49.

Conitzer, V., and Vidali, A. 2014. Mechanism design for scheduling with uncertain execution time. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 623–629.

Deng, X.; Gao, Y.; and Zhang, J. 2017. Smoothed and average-case approximation ratios of mechanisms: Beyond the worst-case analysis. In *42nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2017*, 16:1–16:15.

Fotakis, D., and Tzamos, C. 2013. Winner-imposing strategyproof mechanisms for multiple facility location games. *Theor. Comput. Sci.* 472:90–103.

Hartline, J. D., and Lucier, B. 2010. Bayesian algorithmic mechanism design. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010*, 301–310.

Koutsoupias, E., and Vidali, A. 2013. A lower bound of $1+\varphi$ for truthful scheduling mechanisms. *Algorithmica* 66(1):211–223.

Koutsoupias, E. 2014. Scheduling without payments. *Theory Comput. Syst.* 54(3):375–387.

Kovács, A., and Vidali, A. 2015. A characterization of n-player strongly monotone scheduling mechanisms. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI*, 568–574.

Kovács, A.; Meyer, U.; and Ventre, C. 2015. Mechanisms with monitoring for truthful RAM allocation. In *Web and Internet Economics - 11th International Conference, WINE 2015, Amsterdam, The Netherlands, December 9-12, 2015, Proceedings*, 398–412.

Lavi, R., and Swamy, C. 2009. Truthful mechanism design for multidimensional scheduling via cycle monotonicity. *Games and Economic Behavior* 67(1):99–124.

Lu, P., and Yu, C. 2008. An improved randomized truthful mechanism for scheduling unrelated machines. In Albers, S., and Weil, P., eds., *STACS 2008, 25th Annual Symposium on Theoretical Aspects of Computer Science, Bordeaux, France, February 21-23, 2008, Proceedings*, volume 1 of *LIPIcs*, 527–538. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

Lu, P. 2009. On 2-player randomized mechanisms for scheduling. In Leonardi, S., ed., *Internet and Network Economics, 5th International Workshop, WINE 2009, Rome, Italy, December 14-18, 2009. Proceedings*, volume 5929 of *Lecture Notes in Computer Science*, 30–41. Springer.

Mu'alem, A., and Schapira, M. 2007. Setting lower bounds on truthfulness: extended abstract. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, 1143–1152.

Nisan, N., and Ronen, A. 1999. Algorithmic mechanism design (extended abstract). In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing (STOC)*, 129–140.

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V. 2007. *Algorithmic Game Theory*. New York, NY, USA: Cambridge University Press.

Penna, P., and Ventre, C. 2014. Optimal collusion-resistant mechanisms with verification. *Games and Economic Behavior* 86:491–509.

Procaccia, A. D., and Tennenholtz, M. 2009. Approximate mechanism design without money. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, 177–186. ACM.

Reed, W. J., and Jorgensen, M. 2004. The double pareto-lognormal distribution – a new parametric model for size distributions. *Communications in Statistics Theory and Methods* 33(8):1733–1753.

Robbins, H. 1955. A remark on stirling's formula. *Amer. Math. Montly* 62:26–29.

Yu, C. 2009. Truthful mechanisms for two-range-values variant of unrelated scheduling. *Theor. Comput. Sci.* 410(21-23):2196–2206.