



ELSEVIER

Available at

www.ElsevierComputerScience.com

POWERED BY SCIENCE @ DIRECT®

Electronic Notes in
Theoretical Computer
Science

Electronic Notes in Theoretical Computer Science 91 (2004) 148–157

www.elsevier.com/locate/entcs

Aggressive Online Deadline Scheduling

Tak-Wah Lam^{1,2}

*Department of Computer Science
University of Hong Kong
Hong Kong*

Tsuen-Wan Johnny Ngan³

*Department of Computer Science
Rice University
Houston, TX, U.S.A.*

Kar-Keung To² and Prudence Wai-Ha Wong²

*Department of Computer Science
University of Hong Kong
Hong Kong*

Abstract

This paper is concerned with online algorithms for scheduling jobs with deadlines on a single processor. It has been known for long that unless the system is underloaded, no online scheduling algorithm can be 1-competitive, i.e., matching the performance of the optimal offline algorithm. Nevertheless, recent work has revealed that some online algorithms using a moderately faster processor (or extra processors) can significantly improve the competitiveness [10] or even be 1-competitive [16,12]. This paper takes a further step to investigate online scheduling algorithms with an even higher performance guarantee (i.e., better than 1-competitive algorithms) and in particular, presents an extra-resource analysis of the earliest-deadline-first strategy (EDF) with respect to such a higher performance guarantee.

Keywords: online algorithms, extra-resource analysis, firm deadline scheduling, earliest deadline first

¹ This research was supported in part by Hong Kong RGC Grant HKU-7024/01E.

² Email: {[twlam](mailto:twlam@cs.hku.hk), [kkto](mailto:kkto@cs.hku.hk), [whwong](mailto:whwong@cs.hku.hk)}@cs.hku.hk

³ Email: twngan@cs.rice.edu

1 Introduction

This paper is concerned with online algorithms for scheduling jobs with deadlines on a processor that allows preemption (see, e.g., [2,14,18,15,6]). A typical example is the earliest deadline first (EDF) algorithm, which has been widely used in many real-time systems (see [19] for a survey of EDF). We measure such online algorithms by the total work (or value) of jobs that can be completed by their deadlines. An online algorithm is said to be $\frac{1}{c}$ -competitive for some fraction $\frac{1}{c} \leq 1$ if, for any job sequence, it guarantees to achieve at least $\frac{1}{c}$ of the total work (or value) obtained by any offline algorithm. Notice that a 1-competitive algorithm matches the performance of the optimal offline algorithm. It has been known for long that when the system is underloaded, EDF is actually 1-competitive [7]; however, in general, EDF as well as any online algorithm is not 1-competitive [8]; indeed, the best possible algorithm can match at most $\frac{1}{4}$ of the total work obtained by any offline algorithm [14]. (In general, if our aim is to maximize the total value, the best possible online algorithm is $1/(\sqrt{\lambda} + 1)^2$ -competitive [14], where λ is the importance ratio, which is the ratio of the largest possible value per unit of work to the smallest possible value per unit of work.)

In recent years, there have been many studies on how to obtain better performance guarantee of online schedulers using a faster processor or other extra resources; such studies are often referred to as the resource augmentation or extra-resource analysis (e.g., [3,5,6,9,10,11,12,15,16,17]). Intuitively, allowing the online scheduler to use a processor faster than the offline scheduler provides a way to compensate the online scheduler for the lack of future information. To simplify our discussion, we first discuss the results on scheduling that aims at maximizing the total work. Kalyanasundaram and Pruhs [10] were the first to show that a moderately faster processor can guarantee a $\frac{1}{c}$ -competitive algorithm, where c is a constant that can be made arbitrarily close to one. Later Lam et al. [16] further showed that EDF supplemented with admission control (denoted by EDF-AC) is indeed 1-competitive when using a speed-2 processor⁴, and more recently Koo et al. [12] showed that 1-competitiveness can also be achieved if the online scheduler exploits only two unit-speed processors.

Knowing that a faster processor can allow an online scheduler to match the optimal offline algorithm, one would naturally demand a performance better than the optimal offline algorithm. Roughly speaking, it is desirable to have a k -competitive algorithm for some $k > 1$. Yet, from a technical viewpoint, such a guarantee is not possible because there always exist job sequences

⁴ A speed- x processor can process x units of work in one unit of time.

for which no online algorithm can outperform the optimal offline algorithm; for example, a job sequence may be underloaded (i.e., it can be completed entirely by an offline algorithm) or has only a short overloaded period. The above observation motivates us to investigate a less demanding notion called k -aggressive algorithms, which can be perceived as follows. Ideally, we want an online algorithm to complete all jobs whenever possible; when it cannot complete all, there must be some conflicting jobs and we require the algorithm to guarantee its performance on such jobs is at least k times as good as that of the optimal algorithm.

Definition 1.1 An online algorithm A is said to be k -aggressive if A can satisfy the following requirements. Given any job sequence I , A partitions the jobs into two categories, called overloaded jobs and underloaded jobs, such that

- with respect to overloaded jobs, the work (or value) attained by A is at least k times that of any offline algorithm (using a unit-speed processor); and
- for underloaded jobs, A completes all of them by their deadlines.

By definition, for any $c \leq 1$, an algorithm is c -competitive if and only if it is c -aggressive. In other words, aggressiveness can be regarded as an extension of competitiveness. As mentioned before, EDF-AC using a two times faster processor is 1-competitive [16], and thus 1-aggressive. It is however non-trivial whether a higher degree of aggressiveness can be achieved. In this paper we answer in the affirmative that EDF-AC can be k -aggressive for any $k \geq 1$ when using a speed- $(k+1)$ processor.

Our aggressive analysis of EDF-AC is indeed tight as we can show that using a speed- $(k+1)$ processor, EDF-AC is not $(k+\epsilon)$ -aggressive for any $\epsilon > 0$. Furthermore, if the speed factor is reduced to be slightly less than $k+1$, then EDF-AC is not k -aggressive. Such lower bound results can be generalized to any online algorithms that decides at release time, i.e., the decision of committing to a job to meet its deadline is made immediately when the job is released.

Our aggressive analysis of EDF-AC can be extended to the case when the objective is to maximize the total value instead of the total work of jobs completed. In particular, we show that a simple extension of EDF-AC can be k -aggressive for any $k \geq 1$ when using a speed- $((2k+1)\lceil \log \lambda \rceil)$ processor, where λ is the importance ratio.

The remainder of this paper is organized as follows: Section 2 presents the aggressive analysis of EDF-AC when the concern is the total work, or equivalently, $\lambda = 1$. Section 3 extends the analysis to the case of general λ (i.e., jobs with non-uniform value densities). Section 4 shows that our analysis

is tight. Before leaving this section, we give a precise definitions of the deadline scheduling problem (which is often referred to as the firm deadline scheduling problem in the literature) and EDF-AC.

Problem definition: Consider a single processor system. Jobs are released in an unpredictable fashion, each requesting a certain amount of work (processing time). The work, deadline, and value of a job are known only when the job is released. Deadlines are firm in the sense that completing a job after its deadline gives zero value. Notice that a system may be overloaded, in which case there is no way to schedule every job released to meet the deadline. The aim of a scheduler is to maximize the total value (or work) of jobs meeting the deadlines. Preemption is allowed at no cost (i.e., a preempted job can be restarted from the point of preemption at any time). The design of a good scheduler is further complicated by the fact that jobs may have different value densities, i.e., different ratios of value to processing time. The importance ratio λ of a system is defined as the ratio of the largest possible value density to the smallest possible value density. When $\lambda = 1$, all jobs have the same value density. The problem of maximizing the total work is actually a special case of the problem of maximizing the total value where $\lambda = 1$.

To simplify our argument, we assume that jobs have distinct release times and deadlines since ties can be broken consistently, say, by job identification numbers. Furthermore, we assume that the first job is released at time 0.

EDF and EDF-AC: In this paper, EDF refers to the strategy of scheduling the job with the earliest deadline. Note that the current job will be preempted when a new job with an earlier deadline is released. EDF is often supplemented with some kind of admission control to avoid excessive preemption when the system is overloaded. In the following EDF-AC denotes EDF enhanced with the following simple form of admission control. Upon release, a job must pass a test to get admitted for EDF scheduling. The test simply checks whether the new job together with the previously admitted jobs can all be completed by their deadlines using EDF. Once a job is admitted, EDF-AC guarantees to complete it.

2 The Aggressiveness of EDF-AC

In this section we show that when jobs have uniform value density (or equivalently, when we want to maximize the total work), EDF-AC, when using a speed- $(k + 1)$ processor, is k -aggressive, where k is any number at least one.

Theorem 2.1 *EDF-AC, when using a speed- $(k + 1)$ processor, is k -aggressive for any $k \geq 1$.*

We prove Theorem 2.1 by contradiction. Assume that EDF-AC, using a speed- $(k + 1)$ processor, is not k -aggressive for some job sequence. That is, there exists some job sequence that does not admit a categorization into overloaded and underloaded jobs such that EDF-AC completes all underloaded jobs, and for overloaded jobs, EDF-AC achieves a total work at least k times that of the optimal offline algorithm. Let I be such a job sequence containing the fewest jobs. We will first show that I must satisfy certain properties (see Lemmas 2.2 and 2.4). Then we prove that EDF-AC is k -aggressive for any job sequence which possesses those properties. In other words, EDF-AC is k -aggressive for I , and we obtain a contradiction to complete the proof of Theorem 2.1.

Lemma 2.2 *When scheduling the job set I , EDF-AC is busy over exactly one continuous period.*

Proof. Assume that EDF-AC is busy over two or more disjoint periods. Let t_ℓ be the start time of the last busy period. Divide I into two parts I_1 and I_2 , one for the jobs released before t_ℓ and one for the rest. If we schedule I_1 and I_2 by EDF-AC separately, the schedules produced are disjoint and their union is identical to the schedule of I . Since I is the smallest counter-example showing that EDF-AC is not k -aggressive, jobs in I_1 (I_2 , respectively) can be categorized into overloaded jobs called O_1 and underloaded jobs U_1 (O_2 and U_2 , respectively) such that the k -aggressive property holds (see the definition in the introduction). Consider the following categorization of I . Jobs in $O = O_1 \cup O_2$ are considered overloaded, and jobs in $U = U_1 \cup U_2$ underloaded. EDF-AC, when scheduling I , completes all jobs in U in time, while the amount of work scheduled by EDF-AC for overloaded jobs would still exceed that of any offline algorithm by a factor of k . In other words, I should be k -aggressive, contradicting our assumption on I . \square

The second property of I is related to a concept called repudiation, defined as follows.

Definition 2.3 Consider the moment when EDF-AC fails to admit a newly released job J . That is, when J is released, it is found that using EDF to schedule J together with all jobs admitted previously would cause some job to miss the deadline. Any such job is said to *repudiate* J . Note that it is possible that a job repudiates itself.

Lemma 2.4 *Let J_ℓ be the job in I with the latest deadline. In the course of scheduling I there is at least one time when J_ℓ repudiates a job.*

Proof. Suppose to the contrary that J_ℓ never repudiates any job. Note that a job can only repudiate another job if the latter has an earlier deadline. Since

J_ℓ does not repudiate itself, no job can repudiate it. Hence J_ℓ is admitted by EDF-AC at its release time. Consider any moment after J_ℓ is admitted. Any newly released job, if rejected by EDF-AC, must be repudiated by a job other than J_ℓ . Thus if we remove J_ℓ from I , EDF-AC will not admit more jobs. Since J_ℓ is the latest deadline job, removing it does not affect the EDF-AC schedule of any other job in I . Notice that $I - \{J_\ell\}$ contains one less job than I , so EDF-AC is k -aggressive for $I - \{J_\ell\}$. In other words, jobs in $I - \{J_\ell\}$ can be categorized into overloaded jobs O and underloaded jobs U in such a way that the k -aggressive property holds. Consider the following categorization of I : jobs in O are overloaded, and jobs in $U \cup \{J_\ell\}$ are underloaded. When EDF-AC schedules I , the underloaded jobs are all scheduled, while the amount of work scheduled by EDF-AC for overloaded jobs would still exceed that of any offline algorithm by a factor of k . In other words, I should be k -aggressive, contradicting the definition of I . \square

With the above two properties of I , we are ready to prove that EDF-AC is k -aggressive for I .

Proof of Theorem 2.1. By Lemmas 2.2 and 2.4, we can assume that I contains only one busy period, and J_ℓ , the job with the latest deadline in I , repudiates another job J at some time t . Let $p(J)$ and $d(J)$ denote the processing time and deadline of job J , respectively. Below we show that I can satisfy the requirement for k -aggressiveness, in particular, by assigning all jobs of I to be overloaded. Since all jobs have deadline no later than $d(J_\ell)$, the offline algorithm can only complete work amounting to $d(J_\ell)$. We want to show that EDF-AC can complete at least $k d(J_\ell)$ units of work.

Recall that J is repudiated by J_ℓ at time t . By definition, at time t , if we attempt to use EDF to schedule J together with all the previous jobs admitted by EDF-AC, J_ℓ can complete only after its deadline (i.e., $d(J_\ell)$). Note that a speed- $(k + 1)$ processor takes $p(J)/(k + 1)$ units of time to process J . That means, at time t , all the previous jobs admitted before must be able to keep the processor busy up to $d(J_\ell) - p(J)/(k + 1)$. Thus, EDF-AC is busy during the whole period from 0 to $d(J_\ell) - p(J)/(k + 1)$. The length of this interval is at least $d(J_\ell) - d(J)/(k + 1) > d(J_\ell) - d(J_\ell)/(k + 1) = k d(J_\ell)/(k + 1)$. Using a speed- $(k + 1)$ processor, EDF-AC can complete at least $k d(J_\ell)$ units of work, thus satisfying the k -aggressive requirement. This contradicts the fact that EDF-AC is not k -aggressive for I . \square

Notice the proof above implicitly shows how to categorize jobs into overloaded and underloaded jobs so that EDF-AC satisfies the requirement of k -aggressiveness. Details are as follows. Consider the schedule produced by EDF-AC. It may contain multiple busy periods, in which case we will con-

sider each busy period independently, each time categorizing those jobs released during one busy period as follows. We will find the latest deadline job J among the jobs released during one busy period. If the job J repudiates another job during its scheduling, all jobs in the period are categorized as overloaded, and the categorization completes. Otherwise, J is categorized as underloaded, and is removed from consideration. Once we remove J , the scheduling of the remaining jobs may again contain multiple busy periods. They are categorized recursively using the same strategy. The proof above shows that the k -aggressive property holds for this categorization.

3 General value density

It is more difficult to guarantee aggressiveness when jobs have vastly different value densities. Indeed, the aggressiveness guarantee of EDF-AC deteriorates quickly as the importance ratio (i.e., the ratio between the largest value density and smallest value density of jobs within the job set) increases. Using the proof techniques of Section 2, we can show that EDF-AC is k -aggressive when using a speed- $(k\lambda + 1)$ processor, where $\lambda > 1$ is the importance ratio. Luckily, we can use EDF-AC as a building block to devise an improved algorithm, called λ -EDF-AC, that requires only a speed- $((2k + 1)\lceil \log \lambda \rceil)$ processor.

We assume the value densities to be between 1 and λ . The λ -EDF-AC algorithm splits the processor into $\lceil \log \lambda \rceil$ speed- $(2k + 1)$ virtual processors. Each of them runs EDF-AC for some of the jobs. In particular, the i -th processor will schedule jobs with value density between $[2^{i-1}, 2^i)$, and jobs of value density λ is also scheduled by the $\lceil \log \lambda \rceil$ -th processor even if λ is a power of 2.

Theorem 3.1 *λ -EDF-AC is k -aggressive when each virtual processor is speed- $(2k + 1)$, i.e., when the speed of the single processor is speed- $((2k + 1)\lceil \log \lambda \rceil)$.*

Proof. Notice that the scheduling of jobs for each virtual processor is k -aggressive. In other words, the jobs S_i given to the i -th processor can be categorized as overloaded jobs O_i and underloaded jobs U_i , such that EDF-AC in the i -th virtual processor completes all jobs in U_i , and attains at least k times the work attained by any offline algorithm for O_i . Hence $U = \bigcup U_i$ and $O = \bigcup O_i$ defines a categorization for the set of jobs to show that λ -EDF-AC is k -aggressive: All jobs in U are completed by λ -EDF-AC. For jobs in O , λ -EDF-AC attains a value at least k times the sum of values that an offline algorithms can attain for each O_i , where the sum is at least the maximum value obtained by an offline algorithm for the job set O . \square

4 Lower Bound for Aggressiveness

An algorithm is said to *decide at release time* if the decision of committing to a job to meet its deadline is made immediately when the job is released. Note that EDF-AC is an algorithm that decides at release time. For algorithms of this type, we can show that the above speed and aggressiveness are tight.

Theorem 4.1 *Let \mathcal{A} be an online algorithm using a speed- $(k+1)$ processor for some positive integer k . If \mathcal{A} decides at release time, then \mathcal{A} is not $(k+\varepsilon)$ -aggressive for any $\varepsilon > 0$.*

Proof. Consider the following sequence S of $k+2$ jobs. The first job J_0 has release time 0, processing time $\varepsilon/2$, and deadline $\varepsilon/2$. All the remaining $k+1$ jobs have release time, processing time, and deadline be δ , 1, and $1+\delta$, respectively, where $\delta < \varepsilon/(2k+2)$.

\mathcal{A} must commit to the first job positively, as it is the only available job at that time. Then it can commit to at most k late jobs positively, and has to reject at least one late job J . (This is because $(k+1+\varepsilon/2)/(k+1) = 1 + \varepsilon/(2k+2) > 1 + \delta$.) Since an aggressive algorithm must complete all underloaded jobs, J must be categorized as overloaded. An optimal offline algorithm can choose to work only on J . Being a $(k+\varepsilon)$ -aggressive algorithm, \mathcal{A} must thus complete a set of overloaded jobs with total work being at least $k+\varepsilon$. But as we have seen, \mathcal{A} must reject at least one job with processing time 1, and thus completes only work amounting to $k + \varepsilon/2 < k + \varepsilon$. \square

Theorem 4.2 *Let \mathcal{A} be an online algorithm using a speed- $(k+1-\varepsilon)$ processor for some positive integer k and some $\varepsilon > 0$. If \mathcal{A} decides at release time, then \mathcal{A} is not k -aggressive.*

Proof. Consider the following sequence S of $k+2$ jobs. The first $k+1$ jobs have release time 0, processing time $1 - \varepsilon/3k$, and deadline 1. The last job has release time $\varepsilon/4k$, processing time $1 - \varepsilon/4k$ and deadline 1. \mathcal{A} cannot commit to all the first $k+1$ jobs, since it can only complete $k+1-\varepsilon$ work in 1 unit of time, while the total work of the $k+1$ jobs is $k+1 - (k+1)\varepsilon/3k$ which is larger. Thus \mathcal{A} rejects at least one of these $k+1$ jobs, which must thus be considered as overloaded job. Since an offline algorithm may work on that job, to be k -aggressive \mathcal{A} must commit to all the remaining k jobs. As a result, the last job (of slightly larger work) must be rejected, since to complete it means the algorithm completes work amounting to $k - \varepsilon/2 + 1 - \varepsilon/4k$, which is more than the maximum we have mentioned earlier. So the last job must be categorized as overloaded. The offline algorithm may opt to complete the job, so to be k -aggressive the algorithm must complete work amounting to $k - \varepsilon/4$, which is not the case (\mathcal{A} only completes $k - \varepsilon/3$ work). \square

5 Concluding remarks and future work

While we show that EDF-AC using a speed- $(k + 1)$ processor is k -aggressive, such a virtue is definitely not limited to EDF-AC. It can be shown that the EDF-Plus algorithm, introduced in [12], also has this property.

There are several open problems related to aggressiveness. It is non-trivial to us whether there exists any online algorithm that is k -aggressive when given extra unit-speed processors. The special case when $k = 1$ has been resolved in [12]. Furthermore, we would like to study aggressiveness in the context of multiprocessor scheduling.

References

- [1] Baruah, S., Overload tolerance for single-processor workloads. In *IEEE Symposium on Real time technology and application*, pages 2–11, 1998.
- [2] Baruah, S., G. Koren, B. Mishra, A. Raghunathan, L. Rosier and D. Shasha, On-line scheduling in the presence of overload. In *Proc. 1991 IEEE Real-Time Systems Symposium*, pages 101–110, 1991.
- [3] Berman, P. and C. Coulston, Speed is more powerful than clairvoyance. In *Proc. 6th SWAT*, pages 255–263, 1998.
- [4] Borodin, A. and R. El-Yaniv, *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.
- [5] Brehob, M., E. Torng and P. Uthaisombut, Applying extra-resource analysis to load balancing. *Journal of Scheduling*, 3(5):273–288, 2000.
- [6] Chrobak, M., L. Epstein, J. Noga, J. Sgall, R. van Stee, T. Tichý and N. Vakhania, Preemptive scheduling in overloaded systems. In *Proc. ICALP*, pages 800–811, 2002.
- [7] Dertouzos, M. L., Control robotics: the procedural control of physical processes. In *Proc. IFIP Congress*, pages 807–813, 1974.
- [8] Dertouzos, M. L. and A. K. L. Mok, Multiprocessor On-Line Scheduling of Hard-Real-Time Tasks *IEEE Transactions on Software Engineering*, 15(12): 1497–1506, 1989.
- [9] Edmonds, J., Scheduling in the dark. In *Proc. STOC*, pages 179–188, 1999.
- [10] Kalyanasundaram, B. and K. R. Pruhs, Speed is as powerful as clairvoyance. *J. ACM*, 47(4):617–643, 2000.
- [11] Kalyanasundaram, B. and K. R. Pruhs, Maximizing job completions online. In *Proc. ESA*, pages 235–246, 1998.
- [12] Koo, C. Y., T. W. Lam, T. W. Ngan and K. K. To, Extra Processors versus Future Information in Optimal Deadline Scheduling. In *Proc. SPAA*, pages 133–142, 2002.
- [13] Koren, G. and D. Shasha, MOCA: A multiprocessor on-line competitive algorithm real-time system scheduling. *Theoretical Computer Science*, 128:75–97, 1994.
- [14] Koren, G. and D. Shasha, D^{over} : An optimal on-line scheduling algorithm for overloaded uniprocessor real-time systems. *SIAM J. Comput.*, 24(2):318–339, 1995.
- [15] Lam, T. W. and K. K. To, Trade-offs between speed and processor in hard-deadline scheduling. In *Proc. SODA*, pages 623–632, 1999.

- [16] Lam, T. W. and K. K. To, Performance Guarantee for Online Deadline Scheduling in the Presence of Overload. In *Proc. SODA*, pages 755–764, 2001.
- [17] Phillips, C. A, C. Stein, E. Torng and J. Wein, Optimal time-critical scheduling via resource augmentation. In *Proc. STOC*, pages 140–149, 1997.
- [18] Sgall, J., On-line scheduling — a survey. In A. Fiat and G. Woeginger, editors, *On-line Algorithms: The State of the Art*, pages 196–231. Lecture Notes in Computer Science, Springer Verlag, 1998.
- [19] Stankovic, J. A., M. Spuri, K. Ramamritham and G. C. Buttazzo, *Deadline scheduling for real-time systems: EDF and related algorithms*. Kluwer Academic Publishers, 1998.