Simulation Based Analysis of Stochastic Online Machine Scheduling Problems

Gang Chen

Department of Industrial and Systems Engineering
University of Florida
Gainesville, Florida 32611, U.S.A.

Zuo-Jun Max Shen

Department of Industrial and Systems Engineering
University of Florida
Gainesville, Florida 32611, U.S.A.

ABSTRACT

In the stochastic online scheduling environment, jobs with unknown release times and weights arrive over time. Upon arrival, the information on the weight of the job is revealed but the processing requirement remains unknown until the job is finished. With the objective of minimizing the total weighted completion time, various asymptotically optimal algorithms have been proposed for the online single machine problem, uniform parallel machine problem and flow shop problem. We perform extensive and sophisticated simulation studies on these stochastic online scheduling problems and show that two generic nondelay algorithms converge very fast. Our simulation results suggest that some known asymptotic optimality results may be extendable to the stochastic online open shop and job shop problems. The simulation results also suggest that, compared with the total weighted completion time metric, the total weighted flow time metric and total weighted stretch metric are more sensitive and may be better performance measures.

1 INTRODUCTION

In the stochastic online scheduling environment a set of jobs $N = \{1, 2, \cdots n\}$ arrive over time and must be processed nonpreemptively on one or more of m machines. The release time and weight of every job $j \in N$ remain unknown until job j arrives. In addition, the processing requirement of every job $j \in N$ is a random variable whose actual value remains unknown until job j is finished. The processing time of job j is equal to its processing requirement divided by the speed of the machine on which it is processed. In this paper we will study five different stochastic online scheduling problems using three different performance measures. Specifically, we study the single machine problem, the uniform parallel machine problem, the flow shop problem, the open shop problem and the job shop problem in the stochastic online environment. The performance metrics that we use are the total weighted completion time, total weighted flow time and total weighted stretch. The flow time of a job is defined to be the difference of its release time and its completion time. The stretch of a job is the ratio of its flow time to its total processing requirement. Our work is motivated by the fact that, although various asymptotically optimal algorithms have been proposed for the online single machine problem, uniform parallel machine problem and flow shop problem with the objective of minimizing the total weighted completion time, little has been done to evaluate the speed of convergence. In addition, up to this date, no asymptotic performance result or evaluation is known for the total weighted flow time metric or the total weight stretch metric, which are two of the most natural metrics for online scheduling problems. We also compare the sensitivity of the three performance metrics by evaluating the performances of several algorithms under these three metrics.

First we give a brief description of the five problems we study. In the single machine problem, all the jobs must be processed, one at a time, on a single machine with a unit speed. In the uniform parallel machine problem, there are m machines each with a constant speed $s_i > 0$ ($i \in \{1, 2, \cdots m\}$) and each job j ($j \in \{1, 2, \cdots n\}$) has to be processed on one of the m machines. The shop problems, namely the flow shop, open shop and job shop, are similar in the sense that each of the m machines $i \in \{1, 2, \cdots , m\}$ has a unit speed and each job $j \in N$ must visit each of the m machines. The differences are that in the flow shop every job must visit the machines 1,2,$\cdots$,m in that same order, while in the open shop the sequence in which each job visits the m machines is arbitrary and in the job shop, each job has to visit the machines in a prespecified order which can be different from job to job. With the objective of minimizing the total weighted completion time, the stochastic online single machine problem, uniform parallel machine problem, flow shop problem, open shop problem and job shop problem can be denoted, in standard scheduling notation (see, e.g., Graham et al., 1979), by $1|x_j \sim stoch, r_j|\sum w_j C_j$, $Qm|x_j \sim stoch, r_j|\sum w_j C_j$, $Fm|x_{ji} \sim stoch, r_j|\sum w_j C_j$, $Om|x_{ji} \sim stoch, r_j|\sum w_j C_j$, and $Jm|x_{ji} \sim stoch, r_j|\sum w_j C_j$, respectively, where x_j is the processing requirement of job j. Note that in the shop problems $x_j = \sum_{i=1}^{m} x_{ji}$, where x_{ji} is the processing requirement of job j on machine i. The deterministic variant of these problems, in which the exact processing requirement of every job j is known upon job j's arrival at time r_j, are denoted respectively by $1|r_j|\sum w_j C_j$, $Qm|r_j|\sum w_j C_j$, $Fm|, r_j|\sum w_j C_j$, $Om|r_j|\sum w_j C_j$, and $Jm|r_j|\sum w_j C_j$. Let f_j and R_j denote, resp., the flow time and stretch of job $j \in N$, then the total weighted flow time metric and total weighted stretch metric can be denoted, resp., by $\sum w_j f_j$ and $\sum w_j R_j$. In this paper we say that a job is *waiting* at time t if it is released but not being processed at time t. We say that a job is *in the system* at time t if it has been released but not finished by time t. We say that a certain amount of processing requirement is *waiting* at time t if it has been released but not finished by time t. Furthermore, throughout this paper we assume that all jobs have bounded weights. We also assume that the machine(s) in each problem has (have) adequate capacity. That is, in the long run the mean rate at which jobs arrive is strictly less than the mean rate at which the machine(s) is (are) capable of processing. We justify this assumption by observing that, with this assumption being unsatisfied, the number of jobs that are waiting for processing will keep increasing and will eventually approach infinity in any feasible schedule. This means that, after a certain period of time, there will always be an extremely large number of jobs waiting for processing and the vast majority of job information is known whenever a decision is to be made. Such kind of a problem bears more characteristics of an offline problem than an online problem and should be regarded more appropriately as an offline problem and thus will not be considered in this paper.

The rest of this paper is organized as follows. In Section 2 we briefly review related results in the literature. In Section 3 we present our simulation studies using **Arena**, through which we demonstrate that two generic nondelay algorithms converge very fast to the optimal solutions under the total weighted completion time metric. Our simulation results also suggest that the same asymptotic optimality results proved by Chen and Shen (2003) for the stochastic online single machine problem, uniform parallel machine problem and flow shop problem are possibly extendable to the stochastic online open shop problem and job shop problem. The simulation results also suggest that, compared with the total weighted completion time metric, the total weighted flow time metric and total weighted stretch metric are more sensitive to changes in the system parameters and different algorithms and may be better performance measures to be used in the online scheduling environment. We finally draw our conclusions and suggest future research directions in Section 4.

2 LITERATURE REVIEW

In this section we briefly review some related results in the literature. Asymptotic performance analysis evaluates the performance of an algorithm on large size instances. Up to this date most known asymptotic results for online scheduling problems are with the total weighted completion time metric. Kaminsky and Simchi-Levi (2001a) study the single machine problem $1|r_j|\sum C_j$ and show that the Shortest Processing Time among Available jobs (SPTA) rule is asymptotically optimal for this problem. Building on the results of Goemans (1997) and Goemans et al. (1999), Chou et al. (1999) show that a generalized version of SPTA rule, the WSPRA, is asymptotically optimal for the weighted version of the single machine problem $1|r_j|\sum w_j C_j$ and the uniform parallel machine problem $Qm|r_j|\sum w_j C_j$ with bounded weights and processing requirements. In this heuristic, whenever a machine is available, the job with the largest ratio w_j/x_j among all the waiting jobs is selected to be processed next. If there is no job waiting, then the machine remains idle until the next job arrives. Chou et al. derive an upper bound on the maximum delay that any amount of work can incur in the WSPRA schedule, relative to the LP relaxation presented by Goemans (1997). They then derive from this bound the asymptotic optimality of the WSPR algorithm for the single machine and uniform parallel machine problems. Chou (2001) also extends this result to the stochastic version of the single machine problem $1|x_j \sim stoch, r_j|E[\sum w_j C_j]$, where the metric is to minimize the expected total weighted completion times, $E[\sum w_j C_j]$. They prove that the Weighted Shortest Expected Processing Time (WSEPT) rule is asymptotically optimal for $1|x_j \sim stoch, r_j|E[\sum w_j C_j]$ as long as the job weights and processing requirements are bounded and the processing requirements are independently distributed with known mean values.

For shop problems, Kaminsky and Simchi-Levi (2001b) study the flow shop problem $Fm||\sum C_j$ and show that the SPT rule is asymptotically optimal as long as the job processing requirements are continuously, independently, and identically distributed (i.i.d.). Kaminsky and Simchi-Levi (1999) and Xia et al. (2000) study the more general flow shop problem $Fm||\sum w_j C_j$. They use probabilistic analysis to characterize the effectiveness of the WSPT rule and show that the WSPT rule is asymptotically optimal for $Fm||\sum w_j C_j$ under some mild probabilistic assumptions on the distributions of job processing times and weights. Building on the results of Chou et al. (1999), Liu and Simchi-Levi (1999) present two online heuristics and one semi-online heuristic which are asymptotically optimal for the flow shop problem $Fm|r_j|\sum w_j C_j$ with bounded job processing requirements and bounded weights. All of these heuristics are closely related to the WSPRA algorithm.

Very recently, building on Liu and Simchi-Levi (1999), Chen and Shen (2003) show that any *nondelay* algorithm, i.e., algorithm that keeps the machines busy as long as there is at least one job available for processing (Pinedo, 1995), is asymptotically optimal for the stochastic online single machine problem $1|x_j \sim stoch, r_j|\sum w_j C_j$ and the stochastic online flow shop problem $Fm|x_{ji} \sim stoch, r_j|\sum w_j C_j$ with the assumptions that job weights are bounded, machine capacity is adequate and processing requirements are i.i.d. across the machines and jobs. For the stochastic online uniform parallel machine problem $Qm|x_j \sim stoch, r_j|\sum w_j C_j$, Chen and Shen (2003) characterize the relationship between any two nondelay schedules and show that any nondelay algorithm is asymptotically optimal with the additional assumption that all job processing requirements are bounded.

For the total weighted flow time metric, no asymptotic result is known up to this date. A few results are available in the domain of competitive analysis. The interested readers are referred to Chekuri and Khanna (2002) and Chekuri et al. (2001). For the total weighted stretch metric, no analytical result is known so far and the special case of total unweighted stretch have been analyzed, in the competitive analysis domain, by Muthukrishnan et al. (1999) and Bender et al. (2002)

3 SIMULATION STUDY

3.1 The Total Weighted Completion Time Metric

Although it is known that any nondelay algorithm is asymptotically optimal for $1|x_j \sim stoch, r_j|\sum w_j C_j$, $Qm|x_j \sim stoch, r_j|\sum w_j C_j$ and $Fm|x_{ji} \sim stoch, r_j|\sum w_j C_j$ as long as machine capacity is adequate and some mild conditions on the weights and processing requirements hold, the rate of convergence of a randomly selected nondelay algorithm to the optimal solution is still an open question. In this section, we perform extensive simulation studies on the performances of two simple nondelay heuristics denoted by *random* and $FCFS$, respectively. In *random*, whenever a machine is available we randomly select one of the jobs that are waiting to be processed next. In $FCFS$, jobs are processed in the order of their arrivals.

Ideally, we would like to report the performance of *random* and $FCFS$ relative to the optimal offline solutions. However, since all these three problems are NP-hard even in the deterministic setting, finding the optimal offline solutions is prohibitively expensive. To overcome this difficulty, we compare the performances of *random* and $FCFS$ to some lower bounds (see section 3.1.1) of the optimal offline solutions of these problems. Furthermore, we note that any nondelay algorithm for a stochastic online scheduling problem is also applicable to its deterministic counterpart but not vice versa. Thus by simulation we can evaluate the performances of *random* and $FCFS$ by comparing it with the $WSPRA$ based algorithms, which have been shown, by Chou et al. (1999) and Liu and Simchi-Levi (1999), resp., to be asymptotically optimal for the deterministic online problems $1|r_j|\sum w_j C_j$, $Qm|r_j|\sum w_j C_j$ and $Fm|r_j|\sum w_j C_j$.

3.1.1 The Lower Bounds

3.1.1.1 The Uniform Parallel Machine Problem

First, we present a preemptive single machine relaxation for the uniform parallel machine problem $Qm|r_j|\sum w_j C_j$. The preemptive single machine scheduling problem, referred to as problem P_m^1, is constructed as follows. To every instance I of the problem $Qm|r_j|\sum w_j C_j$, we associate an instance I_m^1 of the single machine problem $1|r_j, pmtn|\sum w_j C_j$, with the same number of jobs and same job characteristics as I and a machine speed $s^m = \sum_{i=1}^m s_i$. Chou et al. (1999) shows that the optimal solution of I_m^1, denoted by Z_m^{1*}, is a lower bound of optimal solution of I, denoted by Z_m^*.

Although Z_m^{1*} provides a lower bound for $Qm|r_j|\sum w_j C_j$, it remains prohibitively expensive to solve since the problem $1|r_j, pmtn|\sum w_j C_j$ is also NP-hard. To resolve this difficulty, we use the **mean busy date** relaxation of general scheduling problems introduced by Goemans (1997) to generate lower bounds for Z_m^{1*}. Given a feasible schedule ξ, the mean busy date M_j of a job j is defined to be the average of the time instants at which job j is being processing. It can be calculated by the formula: $M_j = \frac{1}{x_j}\int_0^T s_j(t)tdt$, where $s_j(t)$ is the speed at which job j is being processed at time t and T is the time horizon of ξ. When the speed functions $s_j(t)(j \in \{1,2,\cdots,n\})$ are piecewise constant on t, the mean busy date M_j can be expressed as the weighted average of the midpoints of the time intervals during which job j is processed at constant speed. Goemans (1997) (also see Chou et al. 1999) shows that for any given schedule ξ of $1|r_j,pmtn|\sum w_jC_j$, $\sum_{j=1}^n w_j(M_j + x_j/2)$ is a lower bound of $\sum_j^n w_jC_j$. Hence, let $1|r_j,pmtn|\sum w_jM_j$ denote the *mean busy date relaxation* of the problem $1|r_j,pmtn|\sum w_jC_j$ and let $M_j^*(j \in 1,2,\cdot,n)$ denote the optimal solution to $1|r_j,pmtn|\sum w_jM_j$, then $\sum_{j=1}^n w_j(M_j^* + \frac{1}{2}x_j)$ provides a lower bound of $1|r_j,pmtn|\sum w_jC_j$ and, consequently, of $Qm|r_j|\sum w_jC_j$. Furthermore, Goemans (1997) shows that the following greedy algorithm, denoted by **LP**, optimally solves problem $1|r_j,pmtn|\sum w_jM_j$: At the completion time and the release time of any job, consider all the jobs currently in the system and the one with the largest ratio of w_j/x_j is selected to start (or resume) processing immediately, even if this forces the preemption of a currently in-process job. If no job is available, the machine stay idle until at least one job arrives. Algorithm LP is easily implementable and will be used to compute lower bounds for the uniform parallel machine problem $Qm|x_j \sim stoch, r_j|\sum w_jC_j$.

3.1.1.2 The Shop Problems and Single Machine Problem

We present a single machine relaxation for the deterministic flow shop problem $Fm|r_j|\sum w_jC_j$, introduced by Liu and Simchi-Levi (1999). The *nonpreemptive* single machine problem, referred to as P_F^1, is constructed as follows. To every instance I of the problem $Fm|r_j|\sum w_jC_j$, we associate an instance I_F^1 of the single machine problem $1|r_j|\sum w_jC_j$, which has a machine speed of one, the same number of jobs with the same job characteristics as instance I except that each job j in I_F^1 has a processing requirement $x_j = x_{j1}$, namely the processing requirement of job j on the first machine in instance I. Let Z_F^* and Z_F^{1*} denote, resp., the optimal solution value of instance I and the optimal solution value of its single machine relaxation I_F^1. Liu and Simchi-Levi (1999) shows that Z_F^{1*} is a lower bound of Z_F^*

Since $1|r_j|\sum w_jC_j$ is also NP-hard, Z_F^{1*} remains prohibitively expensive to solve. We then use the relaxation introduced by Dyer and Wolsey (1990) for $1|r_j|\sum w_jC_j$ to generate the lower bounds for Z_F^{1*}. Dyer and Wolsey (1990) developed the following integer programming formulation for $1|r_j|\sum w_jC_j$:

$$\begin{aligned}
\textbf{Min} \quad & \sum_{j=1}^n w_jC_j \qquad (D') \\
\text{s.t.} \quad & \sum_{j=1}^n y_{jt} \le 1, \forall t = 1,2,\cdots,T; \\
& \sum_{t=1}^T y_{jt} = x_j, \forall j = 1,2,\cdots,n; \\
& \frac{x_j}{2} + \frac{1}{x_j}\sum_{t=1}^T (t-\frac{1}{2})y_{jt} = C_j, \forall j = 1,2,\cdots,n; \\
& y \in Y^*; \\
& y_{jt} \in \{0,1\}, \forall j = 1,2,\cdots,n; t = r_j+1,\cdots,T;
\end{aligned}$$

where T is the scheduling horizon, $y_{jt} = 1$ if job j is being processed in the time period $[t-1,t)$ and Y^* are constraints imposing that each job j is processed during x_j consecutive periods. Although in this formulation it is assumed that jobs have integral processing requirements and release times, it is easy to see that formulation D' is applicable as long as the processing requirements and release times take rational values, since the time unit in the formulation can be chosen to represent only a fraction of the actual unit time period, e.g., 1/3 minute. Now let problem (D) denote the linear programming problem obtained from (D') by dropping the constraints $y \in Y^*$ and the integrality constraints, then the optimal objective value of problem D is a lower bound of both the problem $1|r_j|\sum w_jC_j$ and the problem $Fm|r_j|\sum w_jC_j$. Dyer and Wolsey (1990) (see also Hall, et al. 1997) prove that the algorithm **LP**, described in section 3.1.1.1, optimally solves the problem D. Hence, we can use the algorithm LP to calculate lower bounds for the offline optimal solutions of both the single machine problem $1|x_j \sim stoch, r_j|\sum w_jC_j$ and the flow shop problem $Fm|x_{ji} \sim stoch, r_j|\sum w_jC_j$.

Furthermore, it is easy to verify that problem P_F^1 also provides valid lower bounds for the offline optimal solutions of $Om|x_{ji} \sim stoch, r_j|\sum w_jC_j$ and $Jm|x_{ji} \sim stoch, r_j|\sum w_jC_j$. Thus the algorithm LP can also be used to generate lower bounds for these two problems. We can further modify P_F^1 to obtain a tighter lower bound for the job shop problem, $Jm|x_{ji} \sim stoch, r_j|\sum w_jC_j$, as follows: to every instance I of the problem $Jm|r_j|\sum w_jC_j$, we associate an instance I_J^1 of the single machine problem $1|r_j|\sum w_jC_j$, which has a machine speed of one and the same number of jobs as instance I, with each job j has a processing requirement $x_j^{'} = x_{j1}$ and release time $r_j^{'} = r_j + \sum_{i\in\varphi_j} x_{ji}$, where φ_j is the set of machines on which job j has be processed before being processed on the first machine in instance I. In later discussions we will apply the algorithm LP to I_J^1 to obtain a lower bound of each instance of $Jm|x_{ji} \sim stoch, r_j|\sum w_jC_j$.

3.1.2 Experiment Design and Simulation Results

To recapitulate, we test the following 4 different algorithms for each of the five problems we consider:

1. *random*: In this heuristic, whenever a machine i is available, consider all the jobs that are waiting for machine i and randomly select one job to be processed next. If no job is waiting for it, machine i stays idle until at least one job becomes available.

2. $FCFS$: In this heuristic, whenever a machine i is available, consider all the jobs that are waiting for machine i and the one that has been waiting for the longest is selected to be processed next. If no job is waiting for it, machine i stays idle until at least one job becomes available. (Note: the waiting time of job j for machine i is defined to be the difference between the time when job j becomes available for machine i and the current time)

3. $WSPRA$: In this heuristic, whenever a machine i is available, consider all the jobs that are waiting for machine i and select a job with the largest ratio of w_j/x_j to be processed next. If there is no job waiting for if, machine i stays idle until at least one job becomes available. (Note: for the shop problems $x_j = \sum_{i=1}^m x_{ji}$)

4. LP: To simplify the notation, for the uniform parallel machine problem we define $s^m = \sum_{i=1}^m s_i$ and for the flow shop problem and single machine problem we define $s^m = 1$. We also define $x_{ji} = x_j (i \in \{1, 2, \cdots, m\})$ for the uniform parallel machine problem and $x_{j1} = x_j$ for the single machine problem. Now consider a preemptive single machine problem with machine speed s^m, weight w_j, job processing requirements x_{j1} and release time $r_j^{'} = r_j + \sum_{i\in\varphi_j} x_{ji}$, where φ_j is the set of machines which job j must visit before visiting the first machine. Whenever a job is released or completed, consider all the jobs that are currently in the system and the job with the largest ratio of w_j/x_{j1} is selected to start (or resume) processing immediately, even if this forces the preemption of a currently in-process job. If no job is available, the machine stays idle until at least one job arrives. When scheduling is finished the total weighted *mean busy date*, $\sum_{j=1}^n w_j(M_j + x_j/2)$, is calculated if it is the uniform parallel machine problem and the total weighted completion time, $\sum_{j=1}^n w_jC_j$, is calculated otherwise. (Note: in the single machine problem, uniform parallel machine problem, flow shop problem and open shop problem the set φ_j is always empty, $\forall j \in \{1, 2, \cdots, n\}$.)

As mentioned before, $WSPRA$ was proved, by Chou et al. (1999) and Liu and Simchi-Levi (1999), resp., to be asymptotically optimal for the deterministic online problems $1|r_j|\sum w_jC_j$, $Qm|r_j|\sum w_jC_j$ and $Fm|r_j|\sum w_jC_j$ with bounded job processing requirements and weights. Algorithm LP provides a lower bound for the optimal offline solution of each instance of the five stochastic online scheduling problems we study. We denote the total weighted completion times obtained by applying *random*, $FCFS$, $WSPRA$ and LP by, resp., wc_r,wc_F, wc_w and wc_l

We use three different distributions to generate data sets for testing. We first generate all the parameters, including processing requirements, inter-arrival times and weights, using uniform distributions. The simulation is then repeated using all the parameters generated from exponential distributions and then from empirical distributions. The parameters of the three distributions are randomly selected such that they have the same mean values. Specifically, if let ρ denote the mean value, then the corresponding uniform, exponential and empirical distributions are, resp., Uniform$(0, 2\rho]$, Exponential(ρ), and Empirical$(0.2, 0.6\rho, 0.5, 0.92\rho, 0.3, 1.4\rho)$, where in the empirical distribution 0.2, 0.5 and 0.3 are the probabilities that the random variable will take a value of, resp., 0.6ρ, 0.92ρ and 1.4ρ. (Thus we have the mean value equal to $0.2 \times 0.6\rho + 0.5 \times 0.92\rho + 0.3 \times 1.4\rho = \rho$ and, again, these parameters are randomly selected.) We use the **Arena** simulation package (see Kelton et al. 2002) to implement all of our simulation studies. The visual interface and modelling flexibility of Arena allow us to model the five problems, the four different algorithms and the three different performance measures easily. The number of simulation runs is chosen such that the half-width of each confidence interval of the weighted completion time is less than 10% of the average value.

We use $E[c_l^r]$, $E[c_l^F]$ and $E[c_l^w]$ to denote, resp., the average ratios wc_r/wc_l, wc_F/wc_l and wc_w/wc_l and use p, w and L to denote, resp., the mean values of the processing requirements, job weights and inter-arrival times. Due to limited space, we are unable to include all the simulation results in this paper. Selected simulation results of the five different scheduling problems are presented in Tables 1 through Table 6. Specifically, the simulation results for the stochastic online single machine, uniform parallel machine, flow shop, open shop and job shop problems with all the parameters generated from uniform distributions are presented, resp., in Table 1, Table 2, Table 4, Table 5 and Table 6. Table 3 shows the simulation results for the stochastic online uniform parallel machine problem with all the parameters generated from exponential and empirical distributions. Figure 1 and Figure 2 illustrate the relative performances of *random*, *FCFS* and *WSPRA* for the stochastic online uniform parallel machine problem and flow shop problem, resp., under the total weighted completion time metric.

From the simulation results we observe that when the total weighted completion time metric is used, both *random* and *FCFS* perform reasonablly well and are comparable to the WSPRA based algorithms. Specifically, we observe that

1. both *random* and *FCFS* converge fast to the optimal solution as n increases. The convergence becomes faster as the value of $\frac{s_m L}{p}$ increases or as the number of machines decreases;

2. the total weighted completion times obtained by applying *random* and *FCFS* are very close to the total weighted completion time obtained by applying *WSPRA*. This is true for all the five problems we test and even for problems of small sizes (less than 6% of difference);

3. The total weighted completion times obtained by applying *WSPRA*, *random* and *FCFS* get even closer as the ratio $\frac{s_m L}{p}$ increases or the number of machines increases;

From the simulation study, it appears that both *random* and *FCFS* are also asymptotically optimal for the stochastic online open shop and job shop problems (see Table 5 and Table 6), which suggest that the asymptotic optimality results proved by Chen and Shen (2003) for the stochastic single machine problem, uniform parallel machine problem and flow shop problem may be extendable to these two problems. Furthermore, the performances of *random* and *FCFS* are almost identical for all of the five problems studied. The total weighted completion time metric seems to be very insensitive to different scheduling algorithms and does not necessarily reveal valuable information about the goodness or badness of these algorithms.

Table 1: Simulation results for $1|x_j \sim stoch, r_j|\sum w_j C_j$.

n	p=10, w=10, Uniform Distri.					
	L=10.5			L=12		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.0333	1.0325	1.0024	1.0132	1.0131	1.0017
200	1.0226	1.0227	1.0014	1.0065	1.0065	1.0009
500	1.0128	1.0130	1.0007	1.0029	1.0029	1.0004
1000	1.0080	1.0080	1.0004	1.0014	1.0014	1.0002
3000	1.0032	1.0031	1.0001	1.0005	1.0005	1.0001

Table 2: Simulation results for $Qm|x_j \sim stoch, r_j|\sum w_j C_j$.
* $s_1 = 1.5, s_2 = 0.5$
$^{**}s_1 = 0.2, s_2 = 0.3, s_3 = 0.4, s_4 = 0.5, s_5 = 0.6$

n	p=10, w=10, Uniform Distri.					
	L=5.5, m=2 *			L=5.5, m=5 **		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.0469	1.0470	1.0257	1.0935	1.0946	1.0793
200	1.0275	1.0272	1.0132	1.0517	1.0515	1.0405
500	1.0117	1.0117	1.0054	1.0218	1.0219	1.0163
1000	1.0060	1.0060	1.0027	1.0109	1.0110	1.0082
3000	1.0020	1.0020	1.0009	1.0037	1.0037	1.0027

Table 3: Simulation results for $Qm|x_j \sim stoch, r_j|\sum w_j C_j$.

n	p=10, w=10, L=5.5, m=2, s_1=1.5, s_2=0.5					
	Exponential Distri.			Empirical Distri		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.0864	1.0854	1.0317	1.0241	1.0240	1.0219
200	1.0525	1.0529	1.0170	1.0124	1.0124	1.0111
500	1.0291	1.0292	1.0070	1.0050	1.0050	1.0045
1000	1.0163	1.0164	1.0036	1.0025	1.0025	1.0022
3000	1.0056	1.0056	1.0012	1.0008	1.0008	1.0007

Table 4: Simulation results for $Fm|x_{ji} \sim stoch, r_j|\sum w_j C_j$.

n	p=10, w=10, Uniform Distr.					
	L=11, m=3			L=11, m=5		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.1163	1.1163	1.0935	1.1991	1.1999	1.1771
200	1.0684	1.0682	1.0542	1.1208	1.1217	1.1071
500	1.0341	1.0342	1.0269	1.0599	1.0598	1.0519
1000	1.0180	1.0181	1.0144	1.0318	1.0318	1.0279
3000	1.0060	1.0060	1.0048	1.0110	1.0110	1.0097

Table 5: Simulation results for $Om|x_{ji} \sim stoch, r_j|\sum w_j C_j$.

n	p=10, w=10, Uniform Distri.					
	L=11, m=3			L=11, m=5		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.0837	1.0881	1.0585	1.1299	1.1361	1.1037
200	1.0473	1.0508	1.0307	1.0735	1.0754	1.0551
500	1.0213	1.0225	1.0126	1.0318	1.0325	1.0225
1000	1.0116	1.0123	1.0065	1.0159	1.0163	1.0113
3000	1.0042	1.0044	1.0022	1.0058	1.0059	1.0038

Table 6: Simulation results for $Jm|x_{ji} \sim stoch, r_j|\sum w_j C_j$.

n	p=10, w=10, Uniform Distr.					
	L=11, m=3			L=11, m=5		
	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$	$\mathbf{E}[\mathbf{c}_l^r]$	$\mathbf{E}[\mathbf{c}_l^F]$	$\mathbf{E}[\mathbf{c}_l^w]$
100	1.1187	1.1259	1.0613	1.2019	1.2039	1.1176
200	1.0749	1.0799	1.0343	1.1252	1.1295	1.0658
500	1.0329	1.0358	1.0141	1.0590	1.0634	1.0283
1000	1.0169	1.0186	1.0072	1.0339	1.0370	1.0150
3000	1.0060	1.0065	1.0025	1.0117	1.0130	1.0051

3.2 The Total Weighted Flow Time and Total Weighted Stretch Metric

Ideally, for the total weighted flow time metric and the total weighted stretch metric we should also compare the performances of *random*, *FCFS* and *WSPRA* to the lower bounds of the offline optimal solutions of the five stochastic online scheduling problems. The same techniques that we use to generate lower bounds for the total weighted completion time metric can also be used to generate valid lower bounds for the total weighted flow time metric, since the two objective functions differ only by an additive term,

Figure 1: Relative performances of *random*, *FCFS*, and *WSPRA* to *LP* for $Qm|x_j \sim stoch, r_j|\sum w_j C_j$

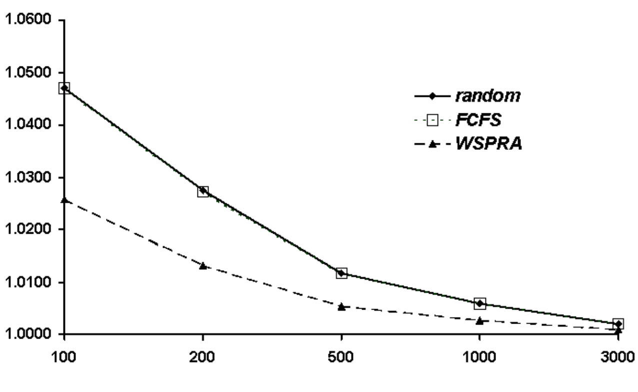
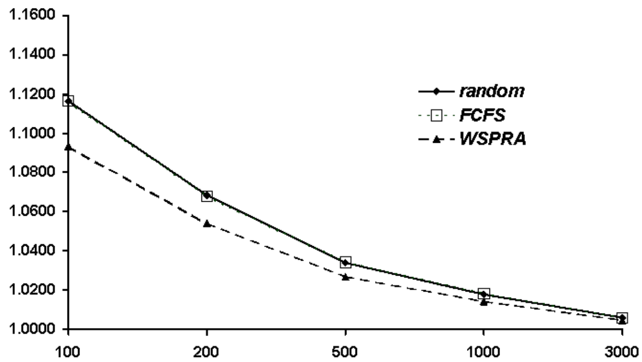


Figure 2: Relative performances of *random*, *FCFS*, and *WSPRA* to *LP* for $Fm|x_{ji} \sim stoch, r_j|\sum w_j C_j$



i.e., the total weighted release time, which is constant and independent of the schedule. However, these techniques do not provide asymptotically tight lower bounds for the total weighted flow time metric. For instance, the algorithm *LP* can be used to provide valid lower bound for the total weighted flow time metric but the gap between the lower bound that it provides and the optimal offline solution increases as the problem size increases. For the total weighted stretch metric, developing asymptotically tight bound is even harder. In fact, there has been an ongoing research in the literature to develop easily computable tight bounds for the total weighted flow time metric and total weighted stretch metric but so far no such algorithms has been reported. Hence we will compare the performances of *random* and *FCFS* against the performances of *WSPRA* under these two performance measures.

We use the same three distributions, i.e., the uniform, exponential and empirical distributions, to generate the processing requirements, inter-arrival times and weights. The number of simulation runs is chosen such that the half-width of the confidence interval of the weighted flow time is less than 10% of the average value. We denote the total weighted flow times obtained by applying *random*, *FCFS* and *WSPRA* by, resp., wf_r, wf_F and wf_w and the total weighted stretches obtained by applying *random*, *FCFS* and *WSPRA* by, resp., we_r, we_F and we_w. We use $E[f_w^r]$, $E[f_w^F]$, $E[e_w^r]$ and $E[e_w^F]$ to denote, resp., the average ratios wf_r/wf_w, wf_F/wf_w, we_r/we_w, and we_F/we_w. Selected simulation results of the five different scheduling problems are presented in Tables 7 through Table 9. Specifically, the simulation results for the stochastic online uniform parallel machine, flow shop, and job shop problems with all the parameters generated from uniform distributions are presented, resp., in Table 7, Table 8, and Table 9.

Table 7: Simulation results for $Qm|x_j \sim stoch, r_j|\sum w_j f_j$ and $Qm|x_j \sim stoch, r_j|\sum w_j R_j$.
* $s_1 = 1.5, s_2 = 0.5$
$^{**}s_1 = 0.2, s_2 = 0.3, s_3 = 0.4, s_4 = 0.5, s_5 = 0.6$

n	p=10, w=10, Uniform Distri.							
	L=5.5, m=2 *				L=5.5, m=5 **			
	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$
100	1.3790	1.3847	2.3972	2.4183	1.1291	1.1418	1.5329	2.0728
200	1.5175	1.5094	2.7128	2.5583	1.2008	1.2022	2.2423	2.2214
500	1.5448	1.5424	3.4875	2.8507	1.2375	1.2395	2.9026	2.3433
1000	1.5740	1.5708	2.1860	3.6154	1.2419	1.2471	2.5478	3.5289
3000	1.6027	1.6069	3.1682	3.4985	1.2537	1.2575	2.6455	2.9903

Table 8: Simulation results for $Fm|x_{ji} \sim stoch, r_j|\sum w_j f_j$ and $Fm|x_{ji} \sim stoch, r_j|\sum w_j R_j$.

n	p=10, w=10, Uniform Distri.							
	L=11, m=3				L=11, m=5			
	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$
100	1.1805	1.1804	1.2091	1.2074	1.1038	1.1058	1.1119	1.1151
200	1.1966	1.2000	1.2208	1.2284	1.1122	1.1162	1.1192	1.1243
500	1.2092	1.2179	1.2330	1.2450	1.1265	1.1253	1.1348	1.1330
1000	1.1987	1.2086	1.2217	1.2334	1.1180	1.1172	1.1262	1.1244
3000	1.2067	1.2062	1.2333	1.2307	1.1152	1.1147	1.1221	1.1213

Table 9: Simulation results for $Jm|x_{ji} \sim stoch, r_j|\sum w_j f_j$ and $Jm|x_{ji} \sim stoch, r_j|\sum w_j R_j$.

n	p=10, w=10, Uniform Distri.							
	L=11, m=3				L=11, m=5			
	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$	$\mathbf{E}[f_w^r]$	$\mathbf{E}[f_w^F]$	$\mathbf{E}[e_w^r]$	$\mathbf{E}[e_w^F]$
100	1.5032	1.5661	1.6547	1.7259	1.4543	1.4654	1.5314	1.5464
200	1.6605	1.7414	1.8304	1.9443	1.5882	1.6308	1.6779	1.7325
500	1.7487	1.8640	1.9372	2.0911	1.7220	1.8253	1.8257	1.9453
1000	1.7717	1.9015	1.9661	2.1362	1.8407	1.9803	1.9568	2.1160
3000	1.8198	1.9298	2.0349	2.1729	1.8666	2.0417	1.9941	2.1847

We observe that when the total weighted flow time metric or the total weighted stretch metric is used, both *random* and $FCFS$ perform badly compared with $WSPRA$. The gap tends to increase dramatically as the number of jobs increases or as the number of machines decreases. It is evident that the total weighted flow time metric and the total weighted stretch metric are much more sensitive to even slight changes in the system parameters and different algorithms than the total weighted completion time metric in the online scheduling environment. This might suggest that the total weighted flow time metric and the total weighted stretch metric are more appropriate metrics for many online scheduling problems. Despite the extremely limited effort that has been devoted so far to these two metrics in the online environment, the interest and potential are immense. Research on asymptotic results of algorithms under the total weighted flow time metric or the total weighted stretch metric in the online environment is important and yet challenging. One of the major challenges is to develop tight lower bounds for the total weighted completion time and total weighted stretch.

4 CONCLUDING REMARKS

In this paper we study five different stochastic online scheduling problems under three different performance measures. With the objective of minimizing the total weighted completion time, we show that two generic nondelay algorithms perform very well and converge to the optimal solutions very fast for all of these five problems. It appears that the asymptotic optimality results in Chen and Shen (2003) for the stochastic single machine problem, uniform parallel machine problem and flow shop problem may be extendable to the stochastic online open shop and job shop problems. The simulation results also suggest that, compared with the total weighted completion time metric, the total weighted flow time metric and total weighted stretch metric are more sensitive and may be better performance measures for scheduling in the online scheduling environment. A very interesting direction of future research would be to devise algorithms that are asymptotically optimal for the total weighted flow time metric or the total weighted stretch metrics. One of the major challenges is to develop tight lower bounds for the total weighted completion time and total weighted stretch.

REFERENCES

Bender, M., S. Muthukrishnan, and R. Rajaraman. 2002. Improved algorithms for stretch scheduling. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithm.*

Chen,G., and Z. J. Shen. 2003. Probabilistic Asymptotic performance of algorithms for stochastic online scheduling problems. Working Paper. Department of Industrial & Systems Engineering, University of Florida, Gainesville, Florida.

Chekuri, C., S. Khanna, and A. Zhu. 2001. Algorithms for weighted flow time. In *Proceedings of the 33rd ACM Symposium on Theory of Computing.*

Chekuri, C., and S. Khanna. 2002. Approximation schemes for preemptive weighted flow time. In *Proceedings of the 34th ACM Symposium on Theory of Computing.*

Chou, C.F., M. Queyranne, and D. Simchi-Levi. 1999. The asymptotic performance ratio of an online algorithm for uniform parallel machine scheduling with release dates. Working paper. Northwestern University.

Chou, C.F. 2001. Asymptotic performance analysis of machine scheduling problems with release dates. Ph.D Thesis. Northwestern University.

Dyer, M.E., and L.A. Wolsey. 1990. Formulating the single machine sequencing problem with release dates as a mixed integer program. *Discrete Applied Mathematics* 26, 255-270.

Goemans, M.X. 1997. Improved approximation algorithms for scheduling for scheduling with release dates. In *Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms*, 591-598.

Goemans, M.X., M. Queyranne, A.S. Schulz, M. Skutella, and Y. Wang. 1999. Single Machine Scheduling with Release Dates. Report 654. Fachbereich Mathematik, Technische Universität Berlin, Germany. Available at URL: http://www.math.tu-berlin.de/coga/publications/techreports/1999/Report-654-1999.html

Graham, R.L., E.L. Lawler, J.K. Lenstra, and A.H.G. Rinnooy Kan. 1979. Optimization and approximation in deterministic sequencing and scheduling: a survey. *Annals of Discrete mathematics* 5, 287-326.

Hall, L.A., A.S. Schulz, D.B. Shmoys, and J. Wein. 1997. Scheduling to minimize average completion time: offline and online approximation algorithms. *Mathematics of Operations Research* 22 (3), 513-544.

Kaminsky, P., and D. Simchi-Levi. 1999. Probabilistic Analysis and Practical Algorithms for the Flow Shop Weighted Completion Time Problem. *Operations Research* 46, 872-882.

Kaminsky, P., and D. Simchi-Levi. 2001a. Probabilistic Analysis of an On-line Algorithm for the Single Machine Mean Completion Time Problem With Release Dates. *Operations Research Letters* 21, 141-148.

Kaminsky, P., and D. Simchi-Levi. 2001b. The Asymptotic Optimality of the Shortest Processing Time Rule for the Flow Shop Completion Time Problem. *Operations Research* 49, 293-304.

Kelton, W.D., R.P. Sadowski, and D.A. Sadowski. 2002. *Simulation with Arena*, 2nd Edition. McGraw-Hill.

Liu H., and D. Simchi-Levi. 1999. On the asymptotic optimality of online algorithms for the flow shop problem with release dates. Working Paper, Northwestern University.

Muthukrishnan, S., R. Rajaraman, R. Shaheen, and J. Gehrke. 1999. Online scheduling to minimize average stretch. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science.*

Pinedo, M. 1995. *Scheduling: Theory, Algorithms and Systems.* Prentice Hall, Inc. Englewood Cliffs, New Jersey.

Xia, C.H, J.G. Shanthikumar, and P.W. Glynn. 2000. On the asymptotic optimality of the SPT rule for the flow shop average completion time problem. *Operations Research* 48 (4): 615-622.

AUTHOR BIOGRAPHIES

Gang Chen is a Ph.D. student in the Department of Industrial & Systems Engineering at the University of Florida. His areas of research include scheduling, algorithm analysis, simulation and logistics. He is a member of IIE and INFORMS. His e-mail address is `<cgang@ufl.edu>`.

Zuo-Jun Max Shen is an Assistant Professor in the Department of Industrial & Systems Engineering at the University of Florida. He received his Ph.D. from the department of Industrial Engineering and Management Sciences at Northwestern University. His areas of research include integrated supply chain design and management models and algorithm analysis. He is a member of INFORMS. His e-mail address is `<shen@ise.ufl.edu>`.