Web Based Certificate Verification using online Scheduling Algorithm for Parallel Batch Processing

Akshay Pawar, Ajit Thakare,
Niteen Gajbhare, Vaibhav Bhujbal
Sinhgad Institute of Technology,Lonavala
University of Pune
Lonavala, 410401

***Abstract*-This paper studies online scheduling problem on parallel machines that process computing jobs arrivingstochastically in a batch pattern. With the objective function asminimizing total completion time of weighted jobs or minimizingtotal latency of job completion, we propose 'Minimizing Max Weighed Processing Time Algorithm ' and its effective implementation in web based verification of documents. With the assumptionthat the processing time of a batch of documents is a constant,the jobs with a higher weight are processed first where weighting factor is 'time' and we assume that the latency of any job exceed a given. The main objective of this paper is toadjust thejob's waiting time to eventually minimize the value of objectivefunction.**

***Keywords*: batch processing, competitive ratio, online scheduling,objective function**

I. INTRODUCTION

The concept of batch scheduling has been the topic of researchsince 1990sand it is considered to be an optimal approach for computation of similar kinds of jobs. In case of single job processing, jobs of same kind are not processed simultaneously whereas the batch processing aggregatesmultiple jobs under certain conditions in a batch andsimultaneously process them under a batch schedulingalgorithm. Nowadays enterprises mainly use batchscheduling in the product lines such as semiconductorproduction, goods dispatch and delivery, steel casting, etc.Due to its wide application in various domains, the batchscheduling has attracted an increasing attention from researchcommunity.But still its effective implementation in Government Sector hasn't done yet.

There is extensive research on job scheduling on parallel machines. In traditional interval scheduling [1][2][3], jobs are processedin the intervals on the real time line,each job has to be processed on some machine, and every machine can process only one job at any time. There is muchliterature on real-time scheduling, where eachjob has to be processed on some machines during a time interval between its release time and due date or eachjob has to be processed during fixed interval between itsstart-time and end-time assuming a machine can processa single job at any time. However, to the bestof our knowledge, it is among the first such paper todiscuss the objective of minimizing the total completion timefor online real-time job scheduling. Previously some work has been done on job scheduling on a setof machines so as to maximize the total profit but in these works the cost of schedulingeach job is fixed. As pointed out further the cost ofscheduling each of the jobs depends on the other jobsscheduled on the same machine in the correspondingtime interval, thus it may change over time and amongdifferent machines; our real-time scheduling problem isdifferent from batch scheduling of conflicting jobs.

JiTian[5] (2011) studied on-line scheduling on *m* parallel-batch machines where the batch capacity is not bounded and the jobs belonged to m incompatible jobfamilies, and gave the different competitive ratios to thebatches with different job numbers. Jianfa Cao [6] studiedonline scheduling of equal length jobs with precedenceconstraints on m parallel batching machines and yielded acompetitive ratio of 2. The rest of paper is organized in thisway:

Research objectives are stated in Chapter II, two onlinescheduling algorithms presented in Chapter III, andsimulation results given in Chapter IV.Implementation of this algorithm in web based certificate verification is given in chapter V. In Chapter VI we drawpreliminary conclusions for this research.

VI. RESEARCH OBJECTIVE

In the batch scheduling domain, researchers tend to focus on two scheduling models: offline scheduling andonline scheduling. Offline scheduling assumes thatall scheduling information of the scheduled jobs are knownprior to the arrival of jobs, while online scheduling deals withthe scenarios in which job information is not known until itarrives for scheduling. Under the online scheduling schema the focus is on the jobs with varied arriving time or unknownarrival time, or whether there are more jobs to

come. Once thejob arrives and all job information is revealed, then the keyproblem becomes how to design an optimized scheduling
algorithm to yield appropriate job waiting times.
In our work presented in this paper, the online batch
scheduling model ,there are *m* processing machines withidentical work capability. The maximum number of jobs thatcan be processed simultaneously on a machine is assumed tobe a constant *B*, and the processing time of each machine isalso a constant *P*. So the goal is to group the jobs into batches and determine when to process the batches such that the valueof objective function is minimized. By means of the FBLPTalgorithm, we can easily prove that the competitive ratio isgreater than 2 when schedule the job without waiting forothers. Thus, our research interest becomes how to determine the appropriate wait time for each job, with the help of usingcompetitive ratio. We will first use the competitive ratio and weighing factor to adjust the waitingtime for each job, and then to minimize the objective functionthat includes weighed completion times. Second, we try tooptimize the objective function in completion latency byadjusting the job waiting time against its completion deadline and penalty factor if the job is delayed.

VII. ONLINE SCHEDULING ALGORITHM DESIGN

A. Competitive Ratio

The competitive ratio of algorithm A can be defined
asR_A=, $\sup_{\forall L}\{C_A(L)/C^*(L)\}$
where $C_A$andC^* refer tothe maximum completion time under online and offline,respectively, and *L* is the input job set. It is found that thevalue of competitive ratio is greater than 1. The competitive
ratio is used to assess the performance of the online
scheduling algorithms, with lower value indicates higherperformance. In the following work, we let
$\alpha=(\sqrt{5}-1)/2$and assume that the number of machines isappropriate that no large-scale batches will be generated andprocessed, but they must wait for the idle machine to becomeavailable. When the processing time is a constant *P* and thebatch size a constant *B*, the competitive ratio of online
scheduling algorithm for parallel processing machine is notgreater than1+ α . Zhang G.C [4] (2003) proved this with *P*= 1, and it is easy to expand the conclusion to the cases when
P is a constant but not equal to 1.

B. Minimizing Max Weighed Processing Time Algorithm

Under this scenario the objective function is designed tominimize the maximum weighed completion time, i.e.$F_{min}=\sum w_jC_j$
where the $w_j$is the weight of each job, $w_j=w_1,w_2$, ...w_i.....w_n,and $C_j$is the completion time of the *j-th*job.Design of the algorithm is elaborated as below and thealgorithmic flow chart is shown in Fig. 1:

Minimizing Max Weighed Processing Time Algorithm

1.Let the first job arrival time t = 0;

2. Compute the number of existing job U (t),
if U (t) < B (where B is the maximum number of jobs a processing machine can handle simultaneously), then go to Step 4;

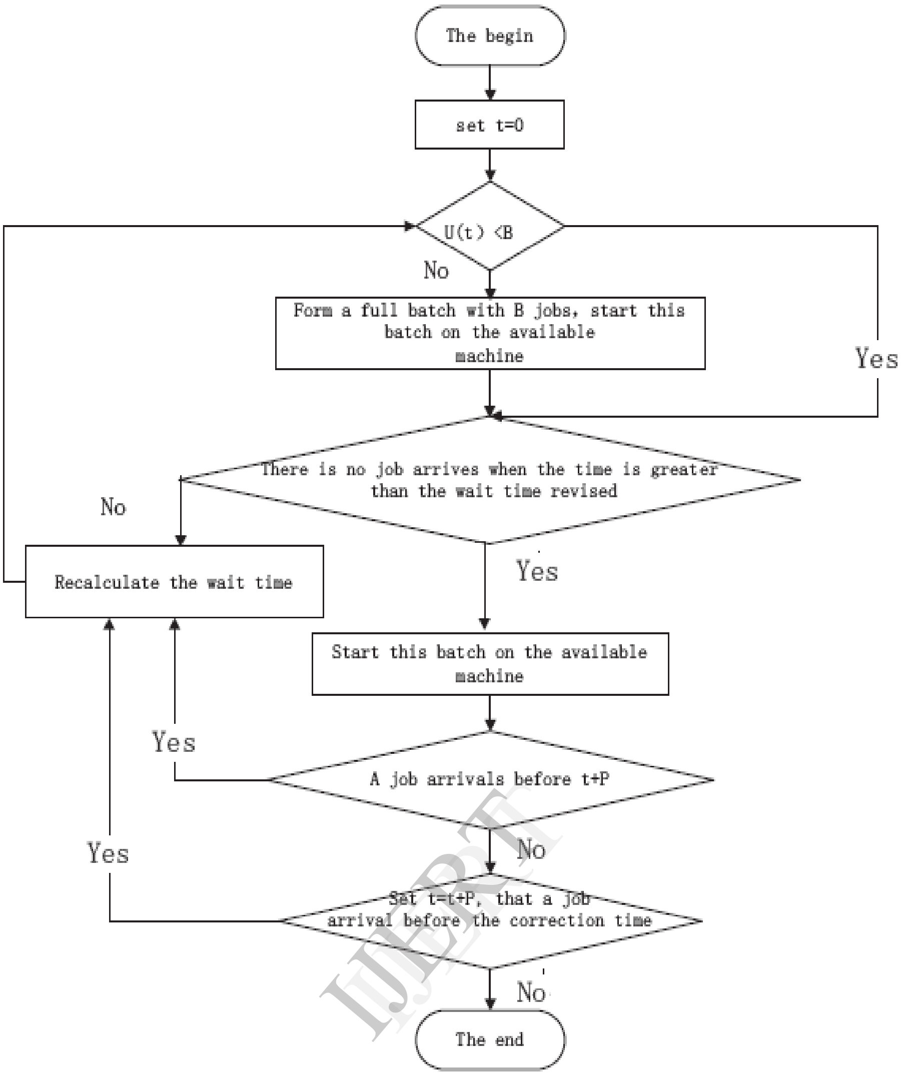
3. Form a full batch and look for idle machines for processing, and update t to the starting time, and then go back to Step 2;

4. If the number of existing jobs U (t) = 0, go to Step 6; otherwise let $\gamma_i(t)$ be the latest arrived job in the remaining jobs, then compute the job's maximum waiting time without considering the weight, $\delta_i(t) = (1+\beta)\gamma_i(t) +\beta$.
If $w_i / w_{i+1} > 1$ the waiting time is decreased and the (i+1)th job's waiting time becomes $\delta'_{i+1}(t) = \delta_{i+1}(t) - [(w_i - w_{i+1})/w_i] \mid \delta_i(t)- \delta_{i+1}(t)|$.
If the calculated wait time is less than the job's arrival time, then these (i+1) jobs shall be scheduled for processing as earliest as possible.
If the machines are available at the time t, then the batch processing shall be started and the earliest processing starting time shall be updated to t, then go back to Step 2;

5. Wait for the next arriving job, denoted by J_j, with the arrival time $\delta_{j-1}(t) = (1+\beta)\gamma_{j-1}(t) +\beta$ to come.
Adjust the job waiting time according to the weighing factor as described in Step 4. If during this period of time there is a new job J_h arrived, denote the arrival time as $\gamma_h(t)$, then recalculate and adjust J_h 's wait time and go to Step 2;

6. If there are more new jobs arriving, update its arrival time $\gamma_k(t)$ by the method stated in Step 4, and then go to Step 2; otherwise the scheduling algorithm goes to the end.

Figure 1 Minimizing Max Weighed Processing Time algorithm



IV. ALGORITHM SIMULATION

Processing Time algorithm

Assume the jobs arrive in a Poisson distribution, the number of processing machines *m* = 3, with each machine'scapability *B* = 100, and the processing time for a batch *P* =3.0. Based on the Minimizing

Max Weighed Processing Timealgorithm and with the given test parameters, we calculate thesuch items as waiting time (WT), modified waiting time (MWT), unmodified completion time (UM), modifiedcompletion time (MM), unmodified objective value (UOV),and modified objective value (MOV), and use them tocalculate and sum up the weighed processing time. Since the jobs arrives in a Poisson process in the simulation, so we design and conduct a few test cases, witheach case consists of 5 test runs, to get the statistic results ofmeasurements.

Figure 2 shows the results of two test cases for the Minimizing Max Weighed Processing Time algorithm, inwhich the simulation configuration and test parameters are thesame, except the job's arriving times are randomly generated through the Poisson process, which differ in the two cases.

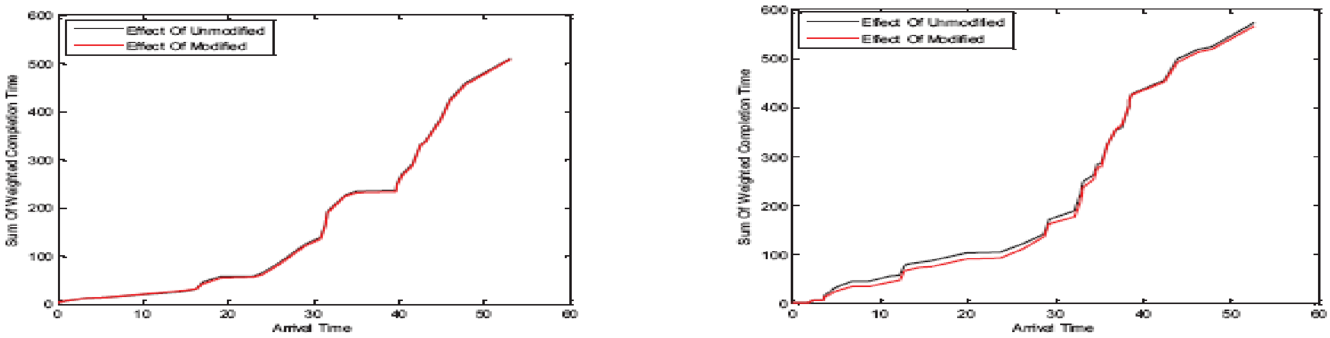
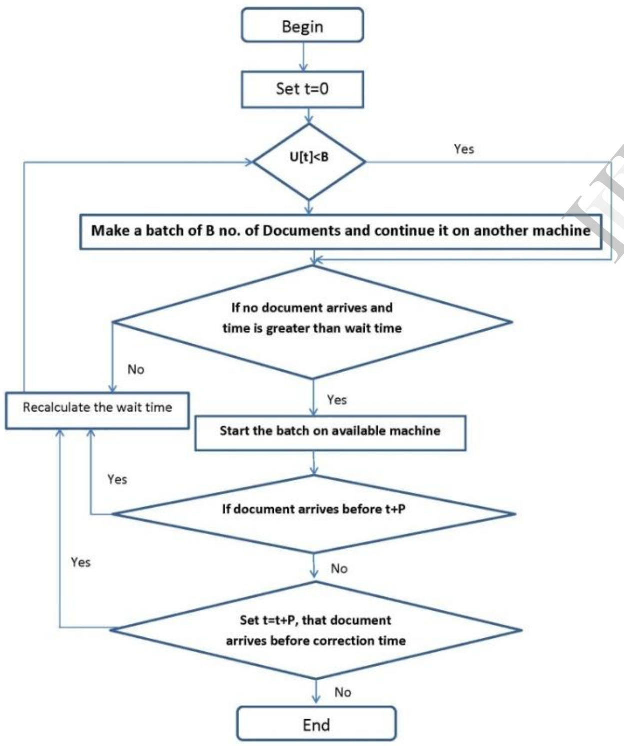


Figure 2 Simulation results for Minimizing Max Weighed Processing Time algorithm (P = 3.0)

V.ALGORITHM IMPLEMENTATION

Here, We have used this algorithm in web based certificate verification. The batch consists of 'm' no. of documents. Processing of documents is done in batch manner. The weighting factor is considered as 'Time'. That means batch of 'm' documents are processed within a particular time constraint. The implementation can be done by using following flow chart.



REFERENCES

1. lzakDuenyas and John J.Neale., Stochastic scheduling of a batchprocessing machine with incompatible job families. Annals ofOperations Reserch,70(1997)191-220.
2. Brucker P., Gladky A., Hoogeveen H., KovalyovM.Y.,PottsC.N.,TautenhahnT.and Van de Velde S.I: Scheduling a batchingmachine ,Journal of Scheduling ,1998,1:31-54.
3. Lee.C.Y. andUzsoy ,R.(1999) Minimizing completion time on a singlebatch processing machine with dynamic job arrivals International Journal of Production Research,40,764-775
4. ZHANG Yuzhong, CAO Zhigang, Parallel Batch Scheduling:A survey.Advances In Mathematics, 2008,37(4):392-403.
5. JiTian, T.C.E. Cheng,C.T. Ng and Jinjiang Yuan, Onlinescheduling on unbounded parallel-batch machines with incompatiblejob families , Theoretical Computer ScienceVolume 412, Issue 22, 13 May 2011, Pages 2380-2386.
6. Jianfa Cao, Jinjiang Yuan, Wenjie Li ,Hailin Bu, Online scheduling onbatching machines to minimise the total weighted completion time ofjobs with precedence constraints and identical processing times,International Journal of Systems Science, Jan2011, Vol. 42 Issue 1,p51-p55