Workload Control in Job Shops with Re-entrant Flows: An Assessment by Simulation

Matthias Thürer* and Mark Stevenson

Name: Professor Matthias Thürer
Institution: Jinan University
Address: Jinan University
              Huangpu Road, No 601
              510632 Guangzhou, PR China
E-mail: matthiasthurer@workloadcontrol.com

Name: Professor Mark Stevenson
Institution: Lancaster University
Address: Department of Management Science
              Lancaster University Management School
              Lancaster University
              LA1 4YX, U.K.
E-mail: m.stevenson@lancaster.ac.uk
Tel: 00 44 1524 593847

Corresponding Author: matthiasthurer@workloadcontrol.com

Workload Control in Job Shops with Re-entrant Flows: An Assessment by Simulation

Abstract

One of the key functions of Workload Control is order release. Jobs are not released immediately onto the shop floor – they are withheld and selectively released to create a mix of jobs that keeps work-in-process within limits and meets due dates. A recent implementation of Workload Control's release method highlighted an important issue thus far overlooked by research: how to accommodate re-entrant flows, whereby a station is visited multiple times by the same job. We present the first study to compare the performance of Workload Control both with and without re-entrant flows. Simulation results from a job shop model highlight two important aspects: (i) re-entrant flows increase variability in the work arriving at a station, leading to a direct detrimental effect on performance; and, (ii) re-entrant flows affect the release decision-making process, since the load contribution of all visits by a job to a station has to fit within the norm. Both aspects have implications for practice and our interpretation of previous research, since: (i) parameters given for work arriving may significantly differ from those realized; and, (ii) increased workload contributions at release mean that prior simulations may have been unstable, leading to some jobs never being released.

Keywords: *Order Release; Workload Control; Job Shop; Simulation; Re-entrant Flows.*

1. Introduction

Workload Control is a production planning and control concept that was developed for high-variety contexts, such as small and medium-sized make-to-order companies, which often have a job shop configuration (Zäpfel & Missbauer, 1993; Stevenson *et al.*, 2005). One of the key functions of Workload Control is order release, which decouples the shop floor from higher level planning. Jobs are not released onto the shop floor immediately but flow into a pre-shop pool from which they are released to meet due dates, while keeping work-in-process within limits or norms. This buffers the shop floor against variance in the incoming order stream (Melnyk & Ragatz, 1989; Thürer *et al.*, 2012a).

The Workload Control concept has been shown to significantly improve the performance of job shops both through simulation (e.g. Oosterman *et al.*, 2000; Fernandes & Carmo-Silva, 2010; Thürer *et al.*, 2012a, 2014a) and, on occasions, in practice (e.g. Hendry *et al.*, 2013). However, during a recent implementation of Workload Control's order release mechanism, we encountered an important real-life aspect: re-entrant flows, whereby a station or machine is visited more than once by the same job. The impact of this aspect of many real-life shops has been overlooked in previous research. Some studies in the Workload Control literature have allowed for re-entrant flows (e.g. Ragatz & Mabert, 1988; Ahmed & Fisher, 1992; Cigolini & Portioli-Staudacher, 2002) while others prohibit re-entrant flows (e.g. Melnyk *et al.*, 1989; Park & Bobrowski, 1989, Park & Salegna 1995; Land & Gaalman, 1998). But there has not been an investigation of the effect that the assumption made about re-entrant flows has on performance. A study either does or does not allow for re-entrant flows, but we could not identify any papers in the Workload Control literature that: (i) consider both options allowing for comparison; or, (ii) justify their decision concerning this aspect. Importantly, the implications for performance of the assumption made by a study on re-entrant flows could not be derived from the literature. It appears to be implicitly assumed that there is no performance effect.

Curiously, although many studies in the Workload Control literature published prior to 2003 accommodate re-entrant flows, only two such studies since 2003 could be identified (Fredendall *et al.*, 2010; Philipoom & Steele, 2011). If re-entrant flows do impact performance – and therefore the results and conclusions of a study – then this is a major shortcoming of recent work since re-entrant flows are commonly encountered in practice. For example, Fowler *et al.* (2002) reported that re-entrant flows are common practice in the semi-conductor industry (e.g. for

wafer-fabrication). Similarly, Lödding *et al*. (2003) reported on a case study where material flows were characterized by many re-entrant flows.

This study represents a move back from the practice of Workload Control and jobs shops to theory. Using a simulation model of a job shop – both with and without re-entrant flows – the paper explores, for the first time, how re-entrant flows affect the performance of the shop and of the Workload Control concept. This will not only provide important insights for future implementations of Workload Control in practice but also for the interpretation of results from previous research, which have either allowed for **or** prohibited re-entrant flows without any further explanation.

The remainder of this paper is structured as follows. In Section 2, we review how re-entrant flows have been considered in the Workload Control literature. The simulation model applied to evaluate the performance impact of re-entrant flows is then described in Section 3 before the results are presented, discussed and analyzed in Section 4. Finally, conclusions are drawn in Section 5, where managerial implications and future research directions are also outlined.

2. Literature Review

Re-entrant flows have received broad attention in the general operations management literature; a comprehensive review of the re-entrant scheduling problem is provided by Danping & Lee (2011). This literature typically assumes a given set of jobs that are optimized for a certain set of performance measures. But in the make-to-order context considered in our study, job arrivals follow a stochastic process and may occur at any time. This makes approaches based on linear programming (Hung & Leachman, 1996; Kim & Kim, 2001) or clearing functions (e.g. Pürgstaller & Missbauer, 2012) arguably not feasible. Meanwhile, other approaches, such as Constant Work-in-Process (ConWIP), lack the load balancing capabilities that are inherent to Workload Control and required in contexts with high processing time variability (Germs & Riezebos, 2010; Thürer *et al*., 2012a). Therefore, the review that follows focuses specifically on examining how re-entrant flows are treated in the Workload Control literature as this concept is designed for high-variety job shops.

2.1 Re-entrant *vs*. No Re-entrant Flows in the Workload Control Literature

Some of the Workload Control literature has allowed for re-entrant flows when evaluating the performance of order release methods, while other studies have prohibited re-entrant flows.

Below, we will briefly review the literature on both, beginning with the work that has allowed for re-entrant flows.

Within the Workload Control literature that allows for re-entrant flows, there are differences in terms of how this is modelled. Some of the literature allows for consecutive visits to a station, i.e. immediate re-entry, while other studies do not permit consecutive visits:

- *Literature that allows for consecutive visits*: The routing is purely random, for example, in Sabuncuoglu & Karapinar (1999, 2000). This means that a job could have two or more consecutive operations at the same station, where each operation is treated as being independent of the other(s). Meanwhile, authors such as Cigolini *et al*. (1998) and Cigolini & Portioli-Staudacher (2002) do not specify whether successive visits are prohibited – it is therefore assumed that they are possible. It is however questionable whether successive visits would be likely in practice. A successive visit means that an operation of a job is completed and the job moves back to the queue to await the second operation, potentially competing again with other jobs for the same resource. In practice, it seems more likely that both operations would normally be handled successively and thus would count as one operation. Of course, there will be exceptions, e.g. if sequence dependent set-up times justify such a procedure, but this was not considered in any of these studies.

- *Literature that prohibits consecutive visits*: Most of the literature that allows for re-entrant flows prohibits consecutive visits. For example, Bertrand & van Ooijen (1996, 2002) avoid consecutive visits since their routing is based on the probability that a job leaves the shop or moves to any of the other stations. This is similar to Ragatz & Mabert (1988) and Ahmed & Fisher (1992), while these authors also set a limit on the maximum number of stations in the routing of a job. No such limit exists for Bertrand & van Ooijen (1996, 2002), which allows for (at least hypothetically) infinite routings. Other authors who considered job shops with re-entrant flows without consecutive visits are Enns & Prongue-Costa (2002), Fredendall *et al*. (2010) and Philipoom & Steele (2011).

Re-entrant flows are prohibited altogether by authors such as Park & Bobrowski (1989), Melnyk & Ragatz (1989), Park & Salegna (1995), Land & Gaalman, (1998) and by most authors of the Workload Control research that has appeared since 2003. It is often not specified how exactly the random routing sequence without re-entrant flows is created. But one exception,

where the procedure is specified, is Fredendall & Melnyk (1995). The authors generated a random routing without replacement to avoid re-entrant flows.

Finally, some studies do not specify whether or not re-entrant flows were allowed in their studies (e.g. Philipoom *et al*. 1993; Perona & Portioli, 1998). Hence, this aspect of their simulation models remains unclear, making replication difficult.

2.2 Assessment of the Literature

To the best of our knowledge, no study has been produced that considers the effect on performance of re-entrant flows - either a study has allowed or has prohibited re-entrant flows, meaning the two have not been compared. While other shop characteristics, such as job release time distributions (e.g. Melnyk *et al*., 1994), routing direction (e.g. Oosterman *et al*., 2000; Thürer *et al*., 2012a), sequence dependent set-up times (e.g. Fernandes & Carmo-Silva 2010; Thürer *et al*., 2012b), due date tightness, and processing time variability (e.g. Thürer *et al*., 2015) have been widely assessed, the impact of this aspect of routing characteristics has been neglected. This is a major shortcoming since it questions the comparability of results between studies that do and do not allow for re-entrant flows. Moreover, most recent literature prohibits re-entrant flows although they are commonly encountered in practice. This leads to the following research question:

What is the impact of re-entrant flows on the performance of a job shop that is using the Workload Control (WLC) concept?

An exploratory study based on controlled simulation experiments will be used to provide an answer to this question. The simulation model and experimental factors used will be outlined next in Section 3.

3. Simulation Model

3.1 Overview of Modeled Shop and Job Characteristics

A simulation model of a randomly routed job shop (Conway *et al*., 1967) – later referred to as a pure job shop (Melnyk & Ragatz, 1989) – has been implemented in Python$^{©}$ using the SimPy$^{©}$ module. The shop contains six stations, where each station is a single, constant capacity resource. Our model is stochastic, whereby job routings, processing times, inter-arrival times and due dates

are stochastic (random) variables. The routing length of jobs varies uniformly from one to six operations. The routing length is first determined before the routing sequence is generated randomly according to the two types of job shops under consideration:

(i) *Job Shop with No Re-entrant Flows:* All stations have the same probability of being the next station in the routing of a job and no replacement occurs, i.e. a particular station is required at most once in the routing of a job; and,

(ii) *Job Shop with Re-entrant Flows:* All stations have the same probability of being the next station in the routing of a job except the last station visited for which the probability is zero, i.e. a station may be visited several times as long as it does not represent a consecutive visit. In other words, re-entry is allowed but a job cannot have two consecutive operations to the same station.

Operation processing times follow a truncated 2-Erlang distribution with a mean of 1 time unit after truncation and a maximum of 4 time units. The inter-arrival time of jobs follows an exponential distribution with a mean of 0.648, which – based on the average number of stations in the routing of a job – deliberately results in a utilization level of 90%. Due dates are set exogenously by adding a random allowance factor, uniformly distributed between 30 and 50 time units, to the job entry time. The minimum value will be sufficient to cover a minimum shop floor throughput time corresponding to the maximum processing time (4 time units) for the maximum number of possible operations (6) plus an arbitrarily set allowance for the waiting or queuing times of 6 time units. These settings have been chosen to facilitate comparisons with earlier studies on Workload Control (e.g. Melnyk & Ragatz, 1989; Oosterman *et al*., 2000; Thürer *et al*., 2012a, 2014b). While any individual job shop in practice will differ in many aspects from this stylized environment, it captures the typical job shop characteristics of high routing variability, processing time variability, and arrival variability.

Finally, Table 1 summarizes the simulated shop and job characteristics.

[Take in Table 1]

3.2 Order Release

There are many order release methods in the Workload Control literature; for examples, see the reviews by Wisner (1995), Land & Gaalman (1996), Bergamaschi *et al*. (1997), and Fredendall *et al*. (2010). In this paper, three different approaches will be considered:

- *LUMS COR* (Lancaster University Management School Corrected Order Release) because it has been identified as the best solution for Workload Control order release (Thürer *et al*., 2012a);

- *LUMS* because it has been widely applied in the prior literature (e.g. Hendry & Wong, 1994; Cigolini & Portioli-Staudacher, 2002; Thürer *et al*., 2010); and,

- *Continuous Release* as an alternative release method that recently emerged in the literature (e.g. Land *et al*., 2010; Fernandes & Carmo-Silva, 2011; Thürer *et al*., 2014b).

These approaches will be described in sections 3.2.1 to 3.2.3 before Section 3.2.4 describes how order release was implemented in the simulations.

3.2.1 LUMS COR

LUMS COR uses a *periodic* release procedure, executed at fixed intervals, to control and balance the shop floor workload. All jobs in the pool are sequenced according to planned release dates and then considered for release in turn. A job contributes to the workload of the station(s) in its routing, where the load contribution to a station in LUMS COR is calculated by dividing the processing time of the operation at a station by the station's position in the job's routing. This *corrected* aggregate load method (Oosterman *et al*., 2000) recognizes that an order's contribution to a station's direct load is limited to only the proportion of time that an order is at the station. The workload of each station is then compared against predetermined workload limits or norms. A job is released if the new workload at each station in the job's routing is below its workload norm; otherwise, the job is retained in the pre-shop pool and the next job in the sequence is considered. This procedure keeps the workload W_s^R released to a station s within a pre-established workload norm and creates a mix of jobs on the shop floor that balances the workload across resources. It can be formulated as follows:

(1) All jobs in the set of jobs *J* in the pre-shop pool are sorted according to their planned release date, given by their due date minus a constant operation throughput time for each operation in their routing.

(2) The job $j \in J$ with the earliest planned release date is considered for release first.

(3) Take R_j to be the ordered set of operations in the routing of job *j*. If job *j*'s processing time p_{ij} at the ith operation in its routing – corrected for station position *i* – together with the

workload W_s^R released to station s (corresponding to operation i) and yet to be completed fits

within the workload norm N_s^C at this station, that is $\dfrac{p_{ij}}{i} + W_s^R \le N_s^C \quad \forall i \in R_j$, then the job

is selected for release. That means it is removed from J, and its load contribution is included,

i.e. $W_s^R := W_s^R + \dfrac{p_{ij}}{i} \quad \forall i \in R_j$.

Otherwise, the job remains in the pool and its processing time does not contribute to the station load.

(4) If the set of jobs J in the pool contains any jobs that have not yet been considered for release, then return to Step 2 and consider the job with the next highest priority. Otherwise, the release procedure is complete and the selected jobs are released to the shop floor.

A released job contributes to W_s^R until its operation at this station is completed.

In addition to the above periodic release mechanism, LUMS COR incorporates a *continuous* workload trigger. If the load of any station falls to zero, the first job in the pool sequence with that station as the first in its routing is released irrespective of whether this would exceed the workload norms of any station. The continuous trigger avoids premature station idleness or starvation (see, e.g. Kanet, 1988; Land & Gaalman, 1998). When the continuous workload trigger releases a job, its workload contribution to a station is calculated using the same corrected aggregate load approach as used for the periodic release time element of LUMS COR.

3.2.2 LUMS

Early studies on Workload Control typically focused on limiting the so-called "classical" aggregate load (based on full processing times) of a station (e.g. Bertrand & Wortmann, 1981; Hendry & Kingsman, 1991). In other words, the workload contribution was not corrected. This method – with uncorrected workload contributions – will be called LUMS in our study. It was originally proposed in Hendry & Kingsman (1991). It is otherwise the same approach as LUMS COR, i.e. combining periodic and continuous release.

3.2.3 Continuous Release

Continuous Release (see, e.g. Land *et al*., 2010; Fernandes & Carmo-Silva, 2011; Thürer *et al*., 2014b) is a new breed of release methods, i.e. release methods that may release a job at any

moment in time. This method executes the periodic release element of LUMS COR continuously. In other words, LUMS COR's periodic element is executed whenever a new job arrives at the shop or an operation is complete (rather than at periodic time intervals).

3.2.4 Implementation of Order Release in the Simulation

As in previous simulation studies on Workload Control (e.g. Land & Gaalman, 1998; Fredendall *et al*., 2010; Thürer *et al*., 2012a), it is assumed that all jobs are accepted, materials are available and all necessary information, e.g. regarding shop floor routings and processing times, is known. Jobs flow into a pre-shop pool to await release according to LUMS COR, LUMS or Continuous Release. The time interval between releases for the periodic element of LUMS COR and LUMS is set to 4 time units. The constant allowance for the operation throughput time used for calculating planned release dates is set to 5 time units, based on preliminary simulation experiments. Finally, nine workload norm levels have been considered: from 4 (the maximum possible processing time) to 12 time units for LUMS COR and LUMS; and 7 to 15 time units for Continuous Release. Workload norm levels for LUMS – i.e. the classical aggregate load – are multiplied by 2.67 to reflect the average routing position.

As a baseline measure, experiments without controlled order release have also been executed, i.e. where jobs are released onto the shop floor immediately upon arrival. This is referred to as Immediate Release (IMM).

3.3 Shop Floor Dispatching

Three dispatching rules will be considered in this study for controlling the flow of jobs on the shop floor: (i) the Operation Due Date (ODD) rule, a time-based rule that considers the urgency of jobs; (ii) the Shortest Processing Time (SPT) rule, a load-based rule that has been previously shown to reduce throughput times in flow shops (e.g. Conway, 1967); and, (iii) the Modified Operation Due Date (MODD) rule, which combines the SPT and ODD rules (e.g. Baker & Kanet, 1983, Baker, 1984).

The calculation of the operation due date δ_{ji} for the ith operation of a job j follows Equation (1).

$$\delta_{ij} = \delta_j - (n_j - i) \cdot c \qquad i:1..n_j \tag{1}$$

The operation due date for the last operation in the routing of a job (with index n_j) is equal to the due date δ_j while the operation due date of each preceding operation is determined by successively subtracting a constant allowance c from the operation due date of the next operation. In this study, the allowance has been set to 5 time units based on preliminary simulation experiments that indicated it resulted in the best overall performance.

The ODD rule prioritizes jobs with the earliest operation due date. Meanwhile, the SPT rule selects the job with the shortest processing time from the queue. The MODD rule prioritizes jobs according to the lowest priority number, which is given by the maximum of the operation due date and earliest finish time, i.e. $max(\delta_{ij}, t+p_{ij})$ for an operation with processing time p_{ij}, where t refers to the time when the dispatching decision is made. The MODD rule shifts between a focus on ODDs to complete jobs on time and a focus on speeding up jobs – through SPT effects – during periods of high loads, i.e. when multiple jobs exceed their ODD (Land *et al*., 2015).

3.4 Experimental Design and Performance Measures

The experimental factors are: (i) the nine different norm levels for each release method; (ii) the three different release methods; (iii) the three different dispatching rules; and, (iv) the two different shop types (i.e. with and without re-entrant flows). A full factorial design was used with 162 cells, where each cell was replicated 100 times. Results were collected over 10,000 time units following a warm-up period of 3,000 time units. These parameters are in line with those used in previous studies that applied similar job shop models (e.g. Land, 2006; Thürer *et al*., 2012a) and allow us to obtain stable results while keeping the simulation run time to a reasonable level.

The principal performance measures considered in this study are as follows: *mean throughput time* – the mean of the completion date minus the release date across jobs; *mean lead time* – the mean of the completion date minus the pool entry date across jobs; *percentage tardy* – the percentage of jobs completed after the due date; and, *mean tardiness* – the conditional lateness, that is $T_j = \max(0, L_j)$, with L_j being the lateness of job j (i.e. the actual delivery date minus the due date of job j). To a degree, both the mean tardiness and the standard deviation of lateness measure the dispersion of lateness across jobs. In this study, we chose to measure the mean tardiness because the standard deviation of lateness is more sensitive to extreme values.

4. Results

To obtain a first indication of the relative impact of the experimental factors, statistical analysis has been conducted by applying ANOVA. The ANOVA is here based on a block design with the workload norm as the blocking factor, i.e. the nine levels of workload norm were treated as different systems. A block design allowed the main effect of the workload norm and both the main and interaction effects of our other three factors (re-entrant flow, Workload Control approach, and dispatching rule) to be captured. Results are presented in Table 2.

[Take in Table 2]

All main effects and most two-way interactions were shown to be statistically significant. The only two-way interactions that were not shown to be significant were between dispatching & re-entrant flows in terms of the throughput time and between both dispatching & re-entrant flows and release method & re-entrant flows in terms of the lead time. All three-way interactions where shown to be statistically significant.

Detailed performance results – to further assess the impact of re-entrant flows – will be presented next. Section 4.1 focuses on assessing the performance impact of re-entrant flows on LUMS COR under different dispatching rules. An in-depth analysis of the performance impact will then be presented in Section 4.2 before sections 4.3 and 4.4 examine the impact of re-entrant flows on LUMS and Continuous Release, respectively.

4.1 The Impact of Re-entrant Flows on LUMS COR

Figure 1a, 1b and 1c show the lead time, percentage tardy and mean tardiness results over the throughput time results, for ODD, SPT and MODD dispatching, respectively. Results are presented in the form of performance curves, where the left-hand starting point of the curves represents the tightest workload norm level (4 time units). The workload norm increases step-wise by moving from left to right in each graph, with each data point representing one norm level (from 4 to 12 time units); loosening the norms increases the workload levels and, as a result, the throughput times on the shop floor. In addition, and as a reference point, the results obtained when jobs are released immediately are also included. These results are referred to as IMM (IMMediate release) – see single point "X" (for no-re-entrant flows) and point "+" (for re-entrant flows) in the figures – and represent the outcome with no order release control. It is located to the right of the curves as it leads to the highest level of work-in-process and, consequently, the

11

longest throughput times on the shop floor. We will first discuss the general performance of dispatching and LUMS COR before we focus on the impact of re-entrant flows.

[Take in Figure 1]

The results for dispatching rule performance in isolation are given by the single points "X" and "+" towards the right-hand side of each figure (i.e. immediate release). As somewhat expected, SPT significantly reduces throughput and lead times compared to ODD, resulting in a lower percentage tardy. But this comes at the expense of a higher mean tardiness. Meanwhile, the MODD dispatching rule switches between ODD and SPT in periods of high load, resulting in the best tardiness performance (see also Land *et al*. (2015)).

The results for LUMS COR are given by the performance curve. Compared to immediate release, LUMS COR realizes significant performance improvements regardless of whether re-entrant flows are or are not allowed for all performance measures if ODD dispatching is used. If SPT dispatching is applied, LUMS COR reduces mean tardiness but has a negative impact on percentage tardy performance. Meanwhile, LUMS COR negatively impacts mean tardiness and the percentage tardy if MODD is applied; however, our simulations do not capture the expected behavioural effects of reduced work-in-process, such as improved quality and increased productivity (see, e.g. Schultz *et al*. 1999). In practice, it is expected that these benefits will offset the observed performance loss. In general, the tighter the workload norm, the fewer jobs are on the shop floor. This affects the selection possibilities of the dispatching rule and reduces performance impact and performance differences across rules (Ragatz & Mabert, 1988).

In terms of re-entrant flows, the following can be observed from our results:

- *Performance impact of re-entrant flows on the dispatching rule (i.e. immediate release)*: Re-entrant flows have a direct detrimental effect on performance under immediate release. This can be observed by comparing the single points "X" and "+".

- *Performance impact of re-entrant flows on LUMS COR*: The detrimental effect of re-entrant flows that was observed under immediate release remains largely unchanged if the release of jobs is controlled. In other words, the performance of LUMS COR is better if re-entrant flows are not permitted. However, the performance impact of LUMS COR in terms of the degree of improvement created relative to immediate release is not significantly affected by re-entrant flows.

12

It appears that whatever is responsible for the negative performance impact of re-entrant flows is already apparent under immediate release, i.e. without order release control. This will be explored further in the next subsection.

4.2 The Impact of Re-entrant Flows on LUMS COR: Performance Analysis

Re-entrant flows are expected to have two main effects. First, they change the distribution of work arrival to an individual station. Second, they increase the workload contributed by a job to a station at release, e.g. for a job with two operations at a station, both operations together have to fit within the norm. These two effects will be explored below in Section 4.1.1 and 4.1.2, respectively.

4.2.1 The Impact of Re-entrant Flows on the Distribution of Work Arrivals

To assess the effect of re-entrant flows on the distribution of work arrivals, we recorded the inter-arrival time of work and the work that arrives at an arbitrary station both with and without re-entrant flows. Results are presented in Figure 2a and Figure 2b, respectively. Two main (inter-related) phenomena can be observed from the figures:

(i)   The average inter-arrival time and the variability of the inter-arrival time increases (there is a shift to the right in the distribution) when re-entrant flows are permitted; and,

(ii) The average workload and variability in the workload that arrives at a station increases when re-entrant flows are permitted (again, there is a shift to the right in the distribution).

[Take in Figure 2]

Allowing for re-entrant flows has a significant impact on when and how much work arrives at an individual station; i.e. variability in the incoming workload increases when re-entrant flows are permitted. This explains the direct detrimental effect on performance that can be observed under immediate release. Another major implication of the distribution shift observed is that it prohibits the comparison of results between studies that allow for re-entrant flows and studies that do not. In fact, the input distribution specified in the former studies appears to be very different from the input distribution actually realized due to increases in variability of inter-arrival times and in the arriving workload.

13

4.2.2 The Impact of Re-entrant Flows on the Order Release Decision

Re-entrant flows are expected to have a major impact on the release decision. In our experimental settings for re-entrant flows (i.e. a maximum routing length of 6 and no consecutive visits), a job can visit a station up to 3 times, which means that the maximum workload contribution of a job using the corrected aggregate load is $4/1 + 4/3 + 4/5 = 6.133$ time units rather than 4 time units (the maximum processing time).

To further explore the effect of re-entrant flows, we recorded the performance of jobs that have the same station specifically twice in their routing (removing jobs with three visits in their routing), referred to as double visits, and jobs with the same station specifically three times in their routing, referred to as triple visits. The results are presented together with the results for all jobs (as a baseline measure) in Figure 3a, 3b and 3c for the lead time, percentage tardy and mean tardiness results over the throughput time results with ODD, SPT and MODD dispatching, respectively. In addition, Table 3 gives the percentage of jobs with double and triple visits to a station according to the routing length (i.e. for each class of jobs with the same number of operations in their routing).

[Take in Figure 3 and Table 3]

The results confirm the observations made on Figure 1 above. It appears that the major determining factor for the detrimental impact of re-entrant flows is the inter-arrival time and workload distribution of work arrivals evident under immediate release. There appears to be no detrimental effect on the pool waiting time – given by the lead time minus the throughput time – across jobs that visit a station either two or three times. Rather, performance differences are due to longer throughput times, which can be explained by the different routing lengths, e.g. jobs with 3 visits must have at least 5 operations in their routing (as consecutive visits to the same station are prohibited in our study). Therefore, it can be concluded that the positive effect of LUMS COR on performance is not jeopardized by re-entrant flows.

4.3 The Impact of Re-entrant Flows on LUMS

Figures 4a, 4b and 4c show the lead time, percentage tardy and mean tardiness results over the throughput time results for LUMS with ODD, SPT and MODD dispatching, respectively.

[Take in Figure 4]

14

The main difference between LUMS and LUMS COR is the use of the classical aggregate load instead of the corrected aggregate load. As suggested in Oosterman *et al*. (2000), the corrected aggregate load improves load balancing capabilities compared to the classical aggregate load. This effect can be observed by comparing the results in Figure 1 (LUMS COR) with those in Figure 4 (LUMS). While the overall pattern of performance remains the same, performance is in general worse with LUMS.

In terms of the impact of re-entrant flows, the same conclusions apply as for LUMS COR. We saw that re-entrant flows have two main effects. First, re-entrant flows influence the distribution of work arrivals to an individual station, which has a direct negative impact on dispatching rule performance (i.e. immediate release). Second, re-entrant flows impact the release decision, since all operations of a job at a certain station have to fit the norm. As a consequence, the maximum workload contribution for the classical aggregate approach, as used in LUMS, is 12 time units (3 visits multiplied by the maximum processing time per visit of 4 time units). This means that, for some of our experimental settings of the workload norm, there are some jobs that result in a workload contribution that is larger than the workload norm. These jobs can *never* be released by the periodic release time element. This effect does however not affect the stability of our simulations since jobs are eventually released by the continuous release time element of LUMS (as in LUMS COR), which releases jobs regardless of whether they violate the norm or not. It may however question the validity of simulations and, consequently, the results of previous studies that applied purely periodic release methods while allowing for re-entrant flows (e.g. Sabuncuoglu & Karapinar 1999, 2000; Cigolini & Portioli-Staudacher, 2002; Fredendall *et al*., 2010).

Similarly, care has to be taken with Continuous Release, a new breed of continuous release methods (i.e. release methods which may release a job at any moment in time) that use a workload-limiting upper bound (e.g. Land *et al*., 2010; Fernandes & Carmo-Silva, 2011; Thürer *et al*., 2014b). This will be discussed in the next subsection.

4.4 The Impact of Re-entrant Flows on Continuous Release

'Classical' continuous release methods based on re-order point logic release a new job whenever the workload falls below a certain limit regardless of a job's workload contribution (see, e.g. Melnyk & Ragatz, 1989). But under Continuous Release, a job is never released if its workload

contribution violates the norm. Thus, in shops with re-entrant flows, the minimum workload norm that can be used has to be higher than in shops where re-entrant flows are prohibited. In our simulation experiments, the minimum workload norm had to be higher than 6.133 time units. Figures 5a, 5b, and 5c show the lead time, percentage tardy, and mean tardiness results over the throughput time results for Continuous Release with ODD, SPT and MODD dispatching, respectively.

[Take in Figure 5]

The performance impact of Continuous Release follows a similar pattern to LUMS COR (Figure 1) and LUMS (Figure 4). Using an upper bound as part of a continuous release method, as in Continuous Release – which continuously attempts to fill up the workload norms – leads to SPT effects if many jobs are waiting to be released (Land *et al*., 2010). This is because, for a job to fit the norm, its workload contribution has to be equal to or smaller than the workload contribution freed up by the job completed. This switching behavior is similar to MODD dispatching and leads to significantly improved performance in terms of percentage tardy (compared to LUMS and LUMS COR). The switching behavior in Continuous Release is determined by the workload norm; the lower the norm, the more dominant SPT effects become. As a result, tighter norms lead to a strong increase in mean tardiness performance specifically for large jobs (Perona & Portioli, 1998, Land, 2006; Thürer *et al*., 2014b). It also provides an explanation as to why the best performance under Continuous Release is realized at larger throughput times (and thus work-in-process levels) than under LUMS and LUMSCOR.

Finally, in terms of the impact of re-entrant flows on Continuous Release, it can be observed that, as for LUMS COR and LUMS, the detrimental effect of re-entrant flows, which is observed under immediate release, remains largely unchanged if the release of jobs is controlled. In other words, the performance of all Workload Control release methods considered in our study is better if re-entrant flows are not permitted. However, the performance impact in terms of the degree of improvement created relative to immediate release is not significantly affected by re-entrant flows.

5. Conclusions

The Workload Control concept is a production planning and control approach specifically developed for job shops. But although its effectiveness has been demonstrated in simulation and practice, the impact of a major feature of many job shops has been overlooked: re-entrant flows, where jobs may visit a station more than once. The Workload Control literature has tended to either allow for re-entrant flows or to prohibit re-entrant flows but, to the best of our knowledge, the effect of this factor on performance has not been assessed. In response, we started by asking: *What is the impact of re-entrant flows on a job shop that is using the Workload Control (WLC) concept?*

Using an exploratory study based on controlled simulation experiments to compare a job shop under Workload Control that permits and does not permit re-entrant flows led us to identify two main effects. First, re-entrant flows increase the variability of work arrivals, as seen from an individual station on the shop floor. This leads to a direct detrimental effect on performance, regardless of whether order release is applied or not, i.e. if jobs are released immediately and control is only exercised by the dispatching rule. This effect questions the comparability of results between studies that allow for re-entrant flows and those that do not. Moreover, the distribution specified for the arrival of work differs from the distribution actually realized if re-entrant flows are permitted, which further complicates the interpretation of results. Second, re-entrant flows affect the release decision since the aggregate of the workload contribution of each visit has to fit within the norm. When norms are tight, some jobs may be at risk of never being released. This has a major impact on purely periodic release methods, as commonly applied in prior literature, and on continuous release methods that use an explicit workload limit. The only reason why this does not lead to unstable simulations (i.e. jobs never being released) in our experiments is that, under LUMS COR and LUMS, periodic release is combined with continuous release. The continuous release time element injects work to starving stations in-between periodic releases regardless of the workload contribution.

5.1 Managerial Implications

The major managerial implication that can be derived from this study is that Workload Control maintains its positive effects on performance, compared to immediate release, regardless of whether re-entrant flows are present or not. Hence, a job shop can benefit from Workload Control order release irrespective of whether it does or does not have re-entrant flows. The

negative relative performance impact observed for shops with re-entrant flows, compared to shops where re-entrant flows are prohibited, is due to effects that are evident under immediate release and thus extraneous to order release – an increase in the variability of work arrivals, as observed for an arbitrary station on the shop floor in our experiments.

5.2 Limitations and Future Research

A major limitation of our study is its limited set of environmental factors. This was justified to keep our study focused on the impact of re-entrant flows on different Workload Control methods. Future research should however consider broader experimental settings, e.g. evaluating factors such as processing time variability, due date tightness and sequence dependent set-up times. A further shortcoming of our study is that it has focused on the impact of re-entrant flows on an *order release* controlled job shop only. This was motivated by a recent implementation. But we have seen that re-entrant flows also have a major impact on shops without order release control, i.e. where jobs are released immediately and control is only exercised by the dispatching rule. A comparative study on what the assumption of re-entrant flows means, e.g. for the optimized scheduling literature, is therefore still missing. Future research could also consider how shops can avoid re-entrant flows and assess the impact of these solutions on performance.

References

Ahmed, I., and Fisher, W.W., 1992, Due date assignment, job order release and sequencing interaction in job shop scheduling, *Decision Sciences*, 23, 633 – 647.

Baker, K.R., 1984, Sequencing rules and due-date assignments in a job shop, *Management Science*, 30, 9, 1093-1104.

Baker, K.R., and Kanet, J.J., 1983, Job shop scheduling with modified operation due-dates, *Journal of Operations Management*, 4, 1, 11-22.

Bergamaschi, D., Cigolini, R., Perona, M., and Portioli, A., 1997, Order review and release strategies in a job shop environment: A review and a classification, *International Journal of Production Research*, 35, 2, 399 – 420.

Bertrand, J.W.M., and Wortmann, J.C., 1981, *Production control and information systems for component - manufacturing shops*, Elsevier Scientific Publishing Company, Amsterdam.

Bertrand, J.W.M., Van Ooijen, H.P.G., 1996, Integrating material coordination and capacity load smoothing in multi-product multi-phase production systems, *International Journal of Production Economics*, 46– 47, 1 – 12.

Bertrand, J.W.M., and Van Ooijen, H.P.G., 2002, Workload based order release and productivity: A missing link, *Production Planning and Control*, 13, 7, 665 - 678.

Cigolini, R., and Portioli-Staudacher, A., 2002, An experimental investigation on workload limiting methods with ORR policies in a job shop environment, *Production Planning and Control*, 13, 7, 602–613.

Cigolini, R., Perona, M., and Portioli, A., 1998, Comparison of Order Review and Release techniques in a dynamic and uncertain job shop environment, *International Journal of Production Research*, 36, 11, 2931-2951.

Conway, R., Maxwell, W.L., and Miller, L.W., 1967, *Theory of Scheduling*, Reading, MA: Addisson-Wesley.

Danping, L., and Lee, C. K. M., 2011, A review of the research methodology for the re-entrant scheduling problem, *International Journal of Production Research*, 49, 8, 2221-2242.

Enns, S.T., and Prongue Costa, M., 2002, The effectiveness of input control based on aggregate versus bottleneck workloads, *Production Planning and Control*, 13, 7, 614 - 624.

Fernandes, N.O., and Carmo-Silva, S., 2010, Order release in a workload controlled flow-shop with sequence-dependent set-up times, *International Journal of Production Research*, 49, 8, 2443 – 2454.

Fernandes, N.O., and Carmo-Silva, S., 2011, Workload Control under continuous order release, *International Journal of Production Economics,* 131, 257 – 262.

Fowler, J.W., Hogg, G.L., and Mason, S.J., 2002, Workload control in the semiconductor industry, *Production Planning and Control*, 13, 7, 568 – 578.

Fredendall, L., and Melnyk, S.A., 1995, Assessing the impact of reducing demand variance through improved planning on the performance of a dual resource constrained job shop, *International Journal of Production Research*, 33, 6, 1521 – 1534.

Fredendall, L.D., Ojha, D., and Patterson, J.W., 2010, Concerning the theory of workload control, *European Journal of Operational Research*, 201, 1, 99-111.

Germs, R., and Riezebos, J., 2010, Workload balancing capability of pull systems in MTO production, *International Journal of Production Research*, 48, 8, 2345-2360.

Hendry, L.C., and Kingsman, B.G., 1991, A decision support system for job release in make to order companies, *International Journal of Operations and Production Management*, 11, 6-16.

Hendry, L.C., and Wong, S.K., 1994, Alternative order release mechanisms: a comparison by simulation, *International Journal of Production Research*, 32, 12, 2827 – 2842.

Hendry, L.C., Huang, Y., and Stevenson, M., 2013, Workload control: Successful implementation taking a contingency-based view of production planning & control, *International Journal of Operations & Production Management*, 33, 1, 69-103.

Hung Y.F., and Leachman, R.C., 1996, A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations, *IEEE Transactions on Semiconductor Manufacturing*, 9, 2, 257–269

Kanet, J.J., 1988, Load-limited order release in job shop scheduling systems, *Journal of Operations Management*, 7, 3, 44 – 58.

Kim, B., and Kim, S., 2001, Extended model for a hybrid production planning approach, *International Journal of Production Economics*, 73, 165–173

Land, M.J., 2006, Parameters and sensitivity in workload control, *International Journal of Production Economics*, 104, 2, 625 – 638.

Land, M.J., and Gaalman, G.J.C, 1996, Workload control concepts in job shops: A critical assessment, *International Journal of Production Economics*, 46 – 47, 535 – 538.

Land, M.J., and Gaalman, G.J.C., 1998, The performance of workload control concepts in job shops: Improving the release method, *International Journal of Production Economics*, 56-57, 347-364.

Land, M.J., Su, N.P.B. and Gaalman, G..J.C, 2010, In search of the key to delivery improvement, *16th International Working Seminar on Production Economics, 1st – 5th March, Innsbruck, Austria, Conference Proceedings, 2, 297-308.*

Land, M.J., Stevenson, M., Thürer, M., and Gaalman, G.J.C., 2015; Job Shop Control: In Search of the Key to Delivery Improvements, *International Journal of Production Economics*, 168, 257-266.

Lödding, H., Yu, K.-W., and Wiendahl, H.-P., 2003, Decentralized WIP-oriented manufacturing control (DEWIP), *Production Planning & Control*, 14, 1, 42-54.

Melnyk, S.A., and Ragatz, G.L., 1989, Order review/release: research issues and perspectives, *International Journal of Production Research*, 27, 7, 1081 – 1096.

Melnyk, S. A., Denzler, D. R., Magnan, G. L., and Fredendall, L., 1994, An experimental model for investigating the sensitivity of job shop performance to job release time distribution parameters, *Production and Operations Management*, 3, 64 – 74.

Oosterman, B., Land, M.J., and Gaalman, G., 2000, The influence of shop characteristics on workload control, *International Journal of Production Economics*, 68, 1, 107-119.

Park, P.S., and Bobrowski, P.M., 1989, Job release and labour flexibility in a dual resource constraint job shop, *Journal of Operations Management*,8, 3, 230 – 249.

Park, P.S., and Salegna, G.J., 1995, Load smoothing with feedback in a bottleneck job shop, *International Journal of Production Research*, 33, 6, 1549 – 1568.

Perona, M., and Portioli, A., 1998, The impact of parameter setting in load oriented manufacturing control, *International Journal of Production Economics*, 55, 133 – 142.

20

Philipoom, P.R. and Steele, D.C., 2011, Shop floor control when tacit worker knowledge is important, *Decision Sciences*, 42, 3, 655-688.

Philipoom, P.R., Malhotra, M.K., and Jensen, J.B., 1993, An evaluation of capacity sensitive order review and release procedures in job shops, *Decision Sciences*, 24, 6, 1109-1133.

Pürgstaller, P., and Missbauer, H., 2012, Rule-based vs. optimisation-based order release in workload control: A simulation study of a MTO manufacturer, *International Journal of Production Economics*, 670 – 680.

Ragatz, G. L., and Mabert, V. A., 1988, An evaluation of order release mechanisms in a job shop environment, *Decision Sciences*, 19, 167 – 189.

Sabuncuoglu, I., and Karapinar, H.Y., 1999, Analysis of order review/release problems in production systems, *International Journal of Production Economics*, 62, 259-279.

Sabuncuoglu, I., Karapinar, H.Y., 2000, A load-based and due-date-oriented approach to order review/release in job shops, *Decision Sciences*, 31, 2, 413–447.

Schultz, K.L., Juran, D.C., and Boudreau, J.W., 1999, The Effects of Low Inventory on the Development of Productivity Norms, *Management Science*, 45, 12, 1664 – 1678.

Stevenson, M., Hendry, L.C., and Kingsman, B.G., 2005, A review of production planning and control: The applicability of key concepts to the make to order industry, *International Journal of Production Research*, 43, 5, 869 - 898.

Thürer, M., Silva, C., and Stevenson, M., 2010, Workload Control Release Mechanisms: From Practice back to Theory Building, *International Journal of Production Research*, 48, 12, 3593-3617.

Thürer, M., Stevenson, M., Silva, C., Land, M.J., and Fredendall, L.D., 2012a, Workload control (WLC) and order release: A lean solution for make-to-order companies, *Production & Operations Management*, 21, 5, 939-953.

Thürer, M., Silva, C., Stevenson, M., and Land, M., 2012b, Improving the Applicability of Workload Control (WLC): The Influence of Sequence Dependent Set-Up Times on Workload Controlled Job Shops, *International Journal of Production Research*, 50, 22, 6419-6430.

Thürer, M., Stevenson, M., Silva, C., Land, M.J., Fredendall, L.D., and Melnyk, S.A., 2014a, Lean control for make-to-order companies: Integrating customer enquiry management and order release, *Production & Operations Management*, 23, 3, 463-476.

Thürer, M., Qu, T., Stevenson, M., Maschek, T., and Godinho Filho, M., 2014b, Continuous Workload Control Order Release Revisited: An Assessment by Simulation, *International Journal of Production Research*, 52, 22, 6664-6680.

Thürer, M., Silva, C., Stevenson, M., and Land, M.J. 2014c, Controlled Order Release: A Performance Assessment in Job Shops with Sequence Dependent Set-up Times, *Production Planning & Control*; 25, 7, 603-615.

Thürer, M., Land, M.J., Stevenson, M., Fredendall, L.D., and Godinho Filho, M., 2015, Concerning Workload Control and Order Release: The Pre-Shop Pool Sequencing Decision, *Production & Operations Management,* 24, 7, 1179-1192.

Wisner, J. D., 1995, A review of the order release policy research, *International Journal of Operations and Production Management*, 15, 6, 25 – 40.

Zäpfel, G. and Missbauer, H., 1993, New concepts for production planning and control, *European Journal of Operational Research*, 67, 297-320.

Table 1: Summary of Simulated Shop and Job Characteristics

Shop Characteristics	Routing Variability	Random routing; re-entrant and no-re-entrant flows
	No. of Work Centers	6
	Interchange-ability of Work Centers	No interchange-ability
	Work Center Capacities	All equal
	Work Center Utilization Rate	90%
Job Characteristics	No. of Operations per Job	Discrete Uniform[1, 6]
	Operation Processing Times	Truncated 2–Erlang; (mean = 1; max = 4)
	Due Date Determination Procedure	Due Date = Entry Time + d; d U ~ [30, 50]
	Inter-Arrival Times	Exp. Distribution; mean = 0.648

Table 2: ANOVA Results

	Source of Variance	Sum of Squares	Degree of Freedom	Mean Squares	F-Ratio	p-Value
Throughput time	Norm	73315.37	8	9164.421	5705.34	0
	Release Method (RM)	39221.97	2	19610.99	12208.89	0
	Dispatching (D)	162080.9	2	81040.46	50452.01	0
	Re-entrant (RE)	3714.922	1	3714.922	2312.74	0
	RM x D	3058.226	4	764.5564	475.98	0
	D x RE	9.022029	4	2.255507	1.4	0.2297
	RM x RE	171.1091	2	85.55454	53.26	0
	D x RE x RM	433.5742	2	216.7871	134.96	0
	Residual	25980.1	16,174	1.606288		
Lead time	Norm	977.9502	8	122.2438	36.68	0
	Release Method (RM)	1790.234	2	895.1169	268.58	0
	Dispatching (D)	194309.7	2	97154.86	29150.89	0
	Re-entrant (RE)	13615.09	1	13615.09	4085.15	0
	RM x D	3913.608	4	978.402	293.57	0
	D x RE	7.273168	4	1.818292	0.55	0.7023
	RM x RE	12.54598	2	6.272989	1.88	0.1523
	D x RE x RM	808.7771	2	404.3885	121.34	0
	Residual	53905.13	16,174	3.332826		
Percentage tardy	Norm	0.298437	8	0.037305	55.64	0
	Release Method (RM)	1.859881	2	0.929941	1386.92	0
	Dispatching (D)	4.282537	2	2.141268	3193.51	0
	Re-entrant (RE)	1.720568	1	1.720568	2566.07	0
	RM x D	0.394556	4	0.098639	147.11	0
	D x RE	0.016626	4	0.004157	6.2	0.0001
	RM x RE	0.054361	2	0.027181	40.54	0
	D x RE x RM	0.398837	2	0.199418	297.41	0
	Residual	10.84478	16,174	0.000671		
Mean tardiness	Norm	290.2334	8	36.27917	119.82	0
	Release Method (RM)	56.61916	2	28.30958	93.5	0
	Dispatching (D)	6036.071	2	3018.036	9967.87	0
	Re-entrant (RE)	1026.156	1	1026.156	3389.15	0
	RM x D	400.4929	4	100.1232	330.68	0
	D x RE	11.2311	4	2.807774	9.27	0
	RM x RE	3.524876	2	1.762438	5.82	0.003
	D x RE x RM	34.01775	2	17.00888	56.18	0
	Residual	4897.106	16,174	0.302776		

Table 3: Percentage of Jobs that Visit a Station More than Once (Two or Three Times)

		RL1[a]	RL2	RL3	RL4	RL5	RL6
Percentage of Jobs with:	Double Visit	0.00%	0.00%	20.00%	52.00%	76.81%	81.96%
	Triple Visit	0.00%	0.00%	0.00%	0.00%	4.02%	14.18%

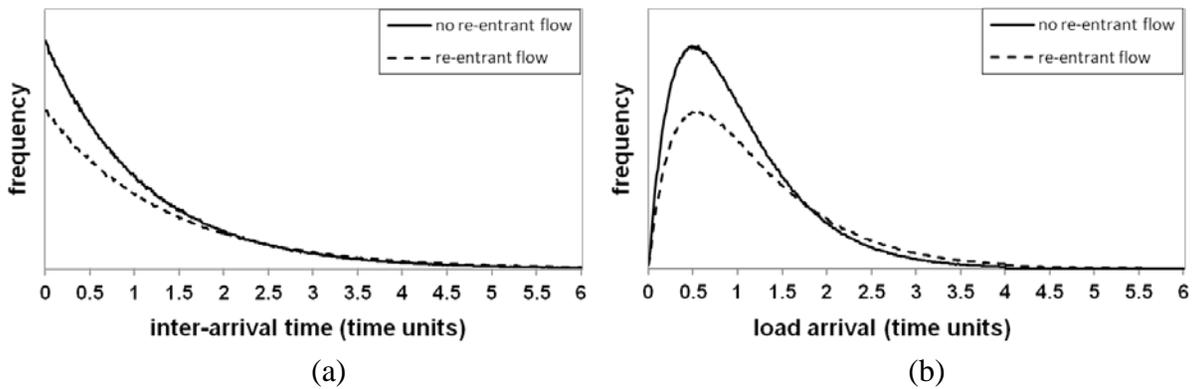[a]) Routing Length
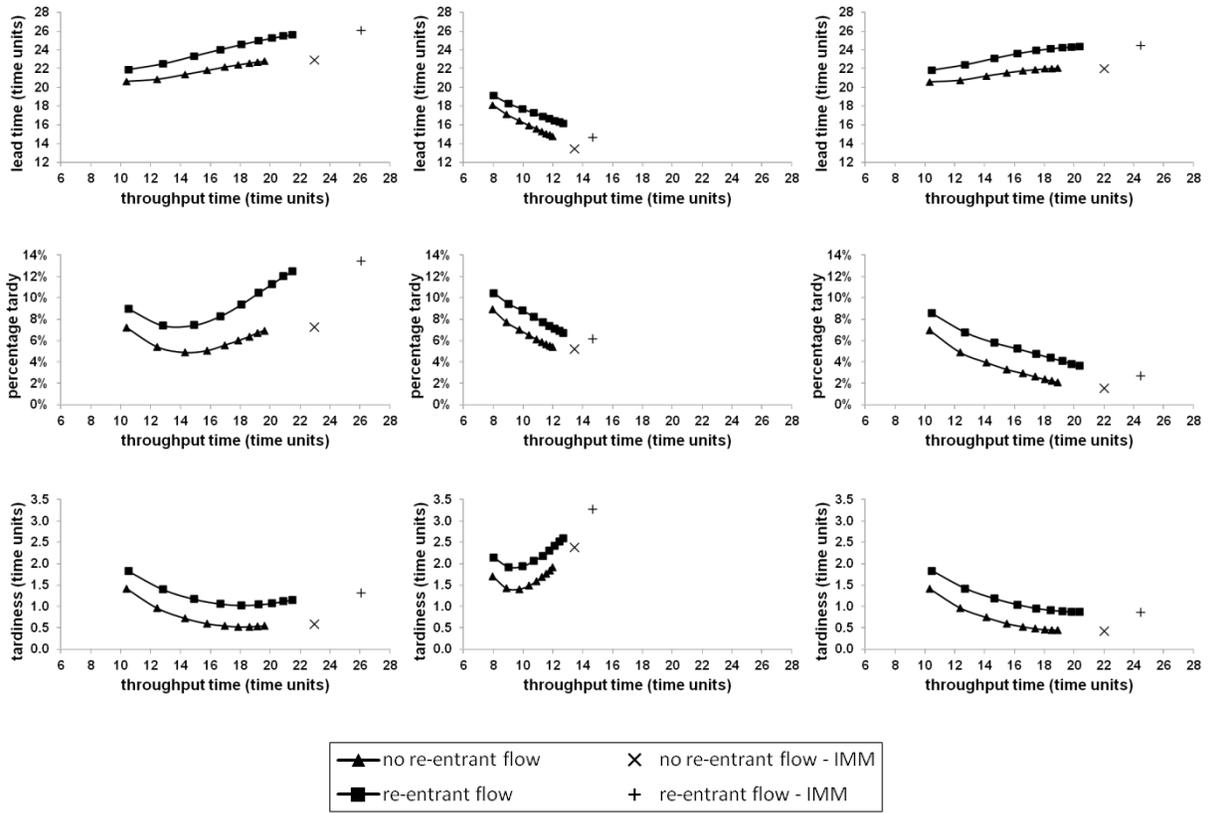
(a) ODD   (b) SPT   (c) MODD
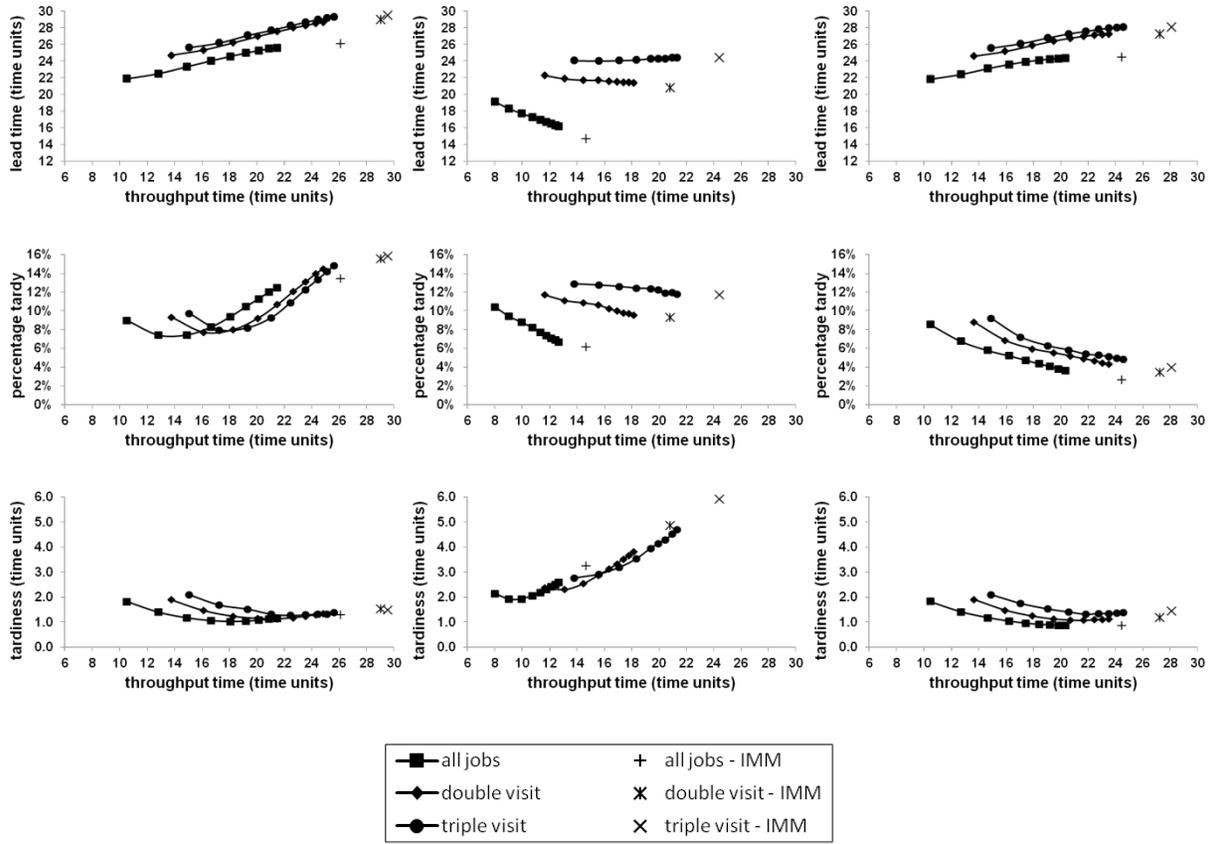
Figure 1: Performance Assessment under LUMS COR: (a) ODD Dispatching; (b) SPT Dispatching; and (c) MODD Dispatching

(a)   (b)

Figure 2: Change in the Distribution of Work Arrivals to an Arbitrary Station: (a) Inter-Arrival Time of Work; and (b) Workload per Arrival

25

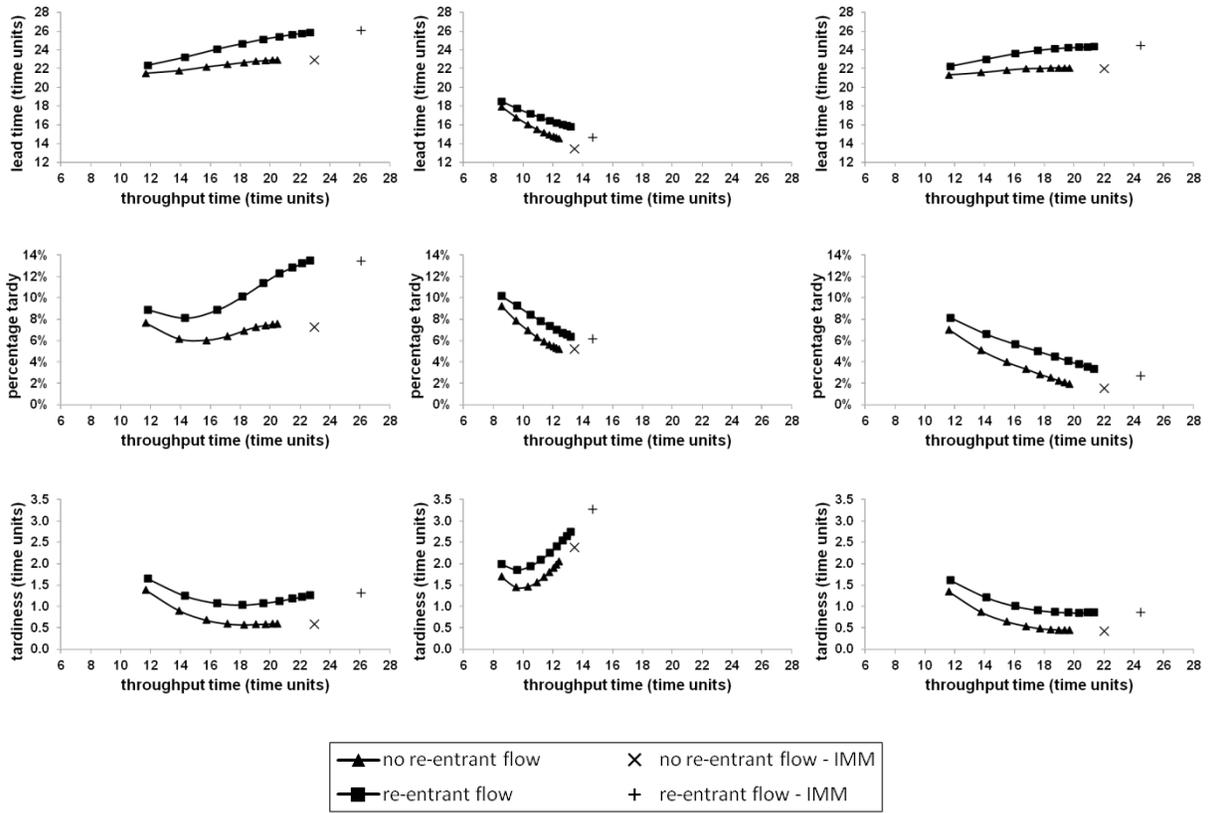(a) ODD   (b) SPT   (c) MODD

Figure 3: Performance Explosion – All Jobs, Jobs with Double and Triple Visits: (a) ODD Dispatching; (b) SPT Dispatching; and (c) MODD Dispatching
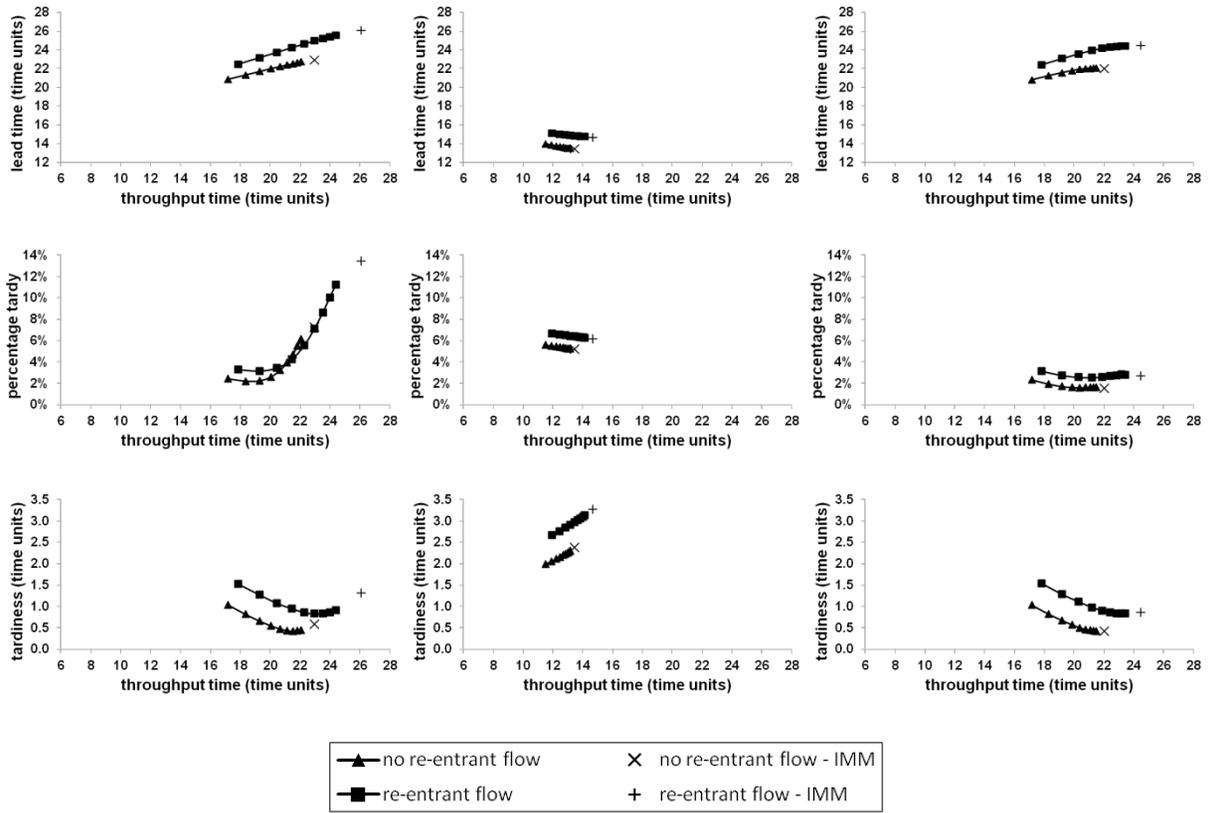
(a) ODD   (b) SPT   (c) MODD

Figure 4: Performance Assessment under LUMS: (a) ODD Dispatching; (b) SPT Dispatching; and (c) MODD Dispatching

(a) ODD   (b) SPT   (c) MODD

Figure 5: Performance Assessment under Continuous Release: (a) ODD Dispatching; (b) SPT Dispatching; and (c) MODD Dispatching