Implementation of Lean Document Production in the Printing Industry

SUDHENDU RAI*

Xerox Research Center Webster
MS 128-51E, 800 Phillips Road, Fairport, New York 14450, U.S.A.

(Received on January 01, 2012 and revised on September 21, 2012)

Abstract: A previous Interfaces (Edelman Finalist) paper by Rai et al. [12] described the LDP Lean Document Production toolkit and work process based on operations research techniques that demonstrated significant improvements in the productivity of small to mid-sized print shops. In this paper, the extensions of the methodology to improve productivity of very large print shops are described. The printing industry is segmented into four quadrants based on resource utilization and job variability. Job variability in one of the industry segments is extremely high and is characterized by fat-tail distributions. This paper describes the extensions required to handle the applications of LDP Lean Document Production to this market segment.

Keywords: manufacturing; performance/productivity; production scheduling; applications; statistics: data analysis; heavy-tail distributions

1. Introduction

It is well-known that the printing industry is large and fragmented. The North American Industrial Classification System (NAICS) code for “printing and related support activities” is 323. In 2005, the total value of print shipments corresponding to code 323 was $97.095 billion with an annual payroll of $24.893 billion [22]. The industry employed 642,300 employees with the payroll per employee of $38,753. An estimate of $100,000 in annual sales per employee is remarkably accurate in determining a commercial printer’s annual sales [21]. Changes over time in the numbers of small (1-9 employees), medium (10-49 employees), and large (50+) establishments provide a measure of industry dynamics. In Figure 1, we show the number of print shops grouped by the number of employees. The increase in the number of larger establishments and decline in the number of small and medium shops reveals that business is moving from small and medium sized shops to large shops.

![Figure 1: Commercial, Quick Printers by Employee Size, 2002 vs. 2004 vs. 2006. [21]](image-url)
The top business challenges faced by print and prepress firms are indicated in Figure 2. The top three, economic conditions, competition, and pricing, all are symptoms of the intense competitive dynamics of the industry. Since small and medium size shops are diminishing in number and large shops are growing, the role of Operations Research in the printing industry seems likely to increase. One of our major objectives in this paper is to present an extension of Operations Research methodologies to handle the fat-tailed job distributions characteristic of many of these increasingly dominant large shops.

Figure 2: Top Business Challenges faced by all Print and Prepress Firms in the Fall, of 2006. [21]

2. Market Segmentation

Because of the great diversities in print shops, we developed a categorization scheme that allows us to develop optimization tools and techniques for various segments of the print-shop market. Based on the experience from early engagements, we developed the print-shop market segmentation matrix shown in Figure 3. Print shops are characterized based on the job variability (characterized by job mix, job size distribution and job inter-arrival time distribution) and resource (equipment and manpower) utilization levels.

Figure 3: A Segmentation of Print Shops within the Printing Industry. FM designates Facilities Management. CRD designates Corporate Reprographics Department.
Segment I encompasses shops that operate individually to produce a few types of products on as-needed basis (e.g., store-front print shops for convenience document production). Segment II encompasses shops that are moderately sized (e.g., corporate reprographics departments), produce several different types of job types (typically less than forty), and are challenged with delivering high quality of service such as turnaround time and print quality at competitive costs. Segment III encompasses shops that typically specialize in a few different types of workflows to create products that are manufactured in high-volumes (e.g., large book manufacturers). Typically these shops are found to operate at higher levels of resource utilization than print shops found in segments I and II. XMS operates shops in segments I, II and III.

Segment IV contains shops that manufacture a wide array of documents often within a specific industry segment such as financial or healthcare, are large in size (e.g., over $50M of annual revenue), and exhibit leveraged economies of scale in production processes to achieve better utilization of resources than the shops in segments I and II. LDP consulting engagements are offered to facilities in this segment. During the course of the development of LDP Solutions print shops from each of the four segments were engaged. Most effort was, however, focused on print shops in segments I, II and III. Segment IV is the subject of ongoing research and development activity, the initial results of which are described below.

Description of the results of implementations of the LDP techniques to quadrants I, II and III of the market-segmentation matrix have been described in [12]. In this paper we describe the extension of the LDP Lean Document Production techniques to the more complex document factory consulting operations characteristic of Segment IV and of the results obtained in the field from utilizing these extensions.

3. The LDP (Lean Document Production) Methodology

The LDP methodology ([4],[6],[10],[12],[13],[14],[15],[16],[17],[20]) is used for production management to make the document production process more efficient in terms of high resource utilization, low waste, low Work In Process (WIP) and low inventory. It utilizes the concept of autonomous cells [13] combined with novel job-splitting, routing and event-driven scheduling, WIP management approaches using finite inter-process buffers [7], and labor cross-training to improve the productivity of print shops. This methodology has been described in some detail in a previous paper [12] to which we refer the reader for details.

4. Improving Complex Document Factory Operations

Shops in quadrant IV are characterized both by higher levels of variability and/or product variety and by higher levels of labor and equipment utilization (at least at bottlenecks) when compared to shops in other quadrants. They are most likely to be large production operations. These shops often employ industrial engineers who utilize manufacturing engineering tools and methods to improve productivity. They are customers for sophisticated optimization analyses produced by expert teams of operations research consultants or industrial engineers. Providing optimization analyses for these shops is a qualitatively different task than performing such analyses for the engagements described above for quadrants I, II and III.

In Figure 4, we show a bar chart description of a shop that processes 961 different product workflows. The process cycle efficiency (measured as the ratio of the value added time to the total process cycle time) of such large shops is often quite low (often less than 10%). While many factors contribute to the low efficiency of these shops, high product
variety can be an important one. Prior work [3] reports the potential of phase transitions in scheduling efficiency whereby the scheduling efficiency of shops drops dramatically when product workflow variety increases.

![Job Count by unique workflow](image)

**Figure 4**: Print shop showing the job count associated with each workflow type.

The coefficient of variation of the job-size distributions can be very high. Moreover, in some cases the distribution may be fat-tailed. Other sources of operational complexity arise from the sheer size of these shops. They can employ over 100 staff; have up to 100 pieces of equipment with sequence-dependent setup requirements; and process up to 1000 jobs per day with cycle time requirements that vary between a few hours to few days.

High variability as characterized by fat-tailed distributions in job sizes, job inter-arrival rate and job mix has been observed in - large transaction and commercial print shops (e.g., those with annual revenues near or above $100 million). Although fat-tail phenomena have been observed in Internet traffic [1] and computer workload on distributed processors [9] this behavior has not previously been reported in the manufacturing systems literature. In sections 7,8 and 9 we describe the extensions of LDP required to apply it successfully to the large shops in quadrant IV that exhibit fat-tailed job distributions. In the remainder of this section we describe the results of an initial exploration of the use of these methods in one particular large shop.

The job size distribution of this particular shop exhibited fat-tail behavior. Its equipments required up to six different types of setup depending on the sequence of jobs processed on them. The shop had a very high WIP level, typically WIP in excess of 3 million units on a given day. The LDP proposed productivity improvement workflow change was to separate small jobs from the large jobs and to dedicate autonomous cells to handle each class separately. Detailed simulation models were constructed to determine the structure of the autonomous cells and the corresponding scheduling policy. In Table 1 we show the results of implementing LDP on the production operations. This example highlights the fact that application of LDP Solutions results in significant productivity improvements for large, complex document production facilities as well as the smaller and/or simpler shops characteristic of quadrants I, II and III in Figure 3.
Table 1: Impact of implementation of LDP in a large production operation

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (one year)</td>
<td>178%</td>
</tr>
<tr>
<td>IRR- Internal Rate of Return</td>
<td>186%</td>
</tr>
<tr>
<td>Payback Period</td>
<td>11 months</td>
</tr>
<tr>
<td>Space Savings</td>
<td>15%</td>
</tr>
<tr>
<td>Product Travel Distance Reduction</td>
<td>75%</td>
</tr>
<tr>
<td>Throughput Time Savings</td>
<td>31%</td>
</tr>
<tr>
<td>Defects Per Million Reduction</td>
<td>55%</td>
</tr>
<tr>
<td>Productivity Improvement</td>
<td>12%</td>
</tr>
<tr>
<td>Sigma Metric Improvement</td>
<td>4.8 to 5.0</td>
</tr>
<tr>
<td>Yield Improvement</td>
<td>99.941 to 99.973</td>
</tr>
</tbody>
</table>

5. Shop and Scheduling Policy Design in the presence of Fat-tail Inputs

Large print shops are the first manufacturing systems reported to exhibit power-law distribution characteristics, and in many instances – a fat tail. An example is given in Figure 5.

Let $X$ be a random variable with cdf (cumulative density function) $F(x) = P[X \leq x]$ and complementary cdf (ccdf) $F_c(x) = P[X > x]$. We say here that a distribution $F(x)$ is fat-tailed if

$$F_c(x) \sim cx^{-\alpha}, 0 < \alpha < 2$$

In the limit of $x \to \infty$,

$$\lim_{x \to \infty} \frac{d \log F_c(x)}{d \log x} = -\alpha$$

The theory of stable distributions [5],[19] has been used to characterize the behavior of these distributions. A discussion of efficient methods for estimating the parameters and properties of these distributions can be found in [11].

Fat-tailed distributions behave quite differently from distributions such as normal distributions or exponential distributions that are most frequently used in characterizing manufacturing system behavior. Because their tails decline relatively slowly, the
probability of outlying observations is not negligible. Fat-tail distributions have infinite variance indicating high variability in the underlying process that generates these distributions. Because of these attributes, shop design in the presence of these distributions differs from that of most shops as described below.

A simple method to detect the presence of fat-tail distribution is to plot ln(CCDF) and ln(JobSize) as shown in Figure 6. As seen from equation (2) the negative value of the slope for large job sizes will be equal to $-\alpha$. The slope of curve in Figure 6 towards the end of the curve (i.e., for large job sizes) is -0.56 implying a fat-tail distribution.

Fat tailed distributions have many small jobs mixed with a few large jobs. Therefore, even though most of the job-sizes are small, the major contribution to sample mean or variance comes from the few large observations. In such distributions, the difference between means and medians is usually very high. e.g., for the print shop data shown in Figure 5, the mean = 17170 and median = 260.

![Figure 6: Log(CCDF) and Log(JobSize) plot of job Size data from print shop](image)

6. Calculation of Moments and Steady State

To analyze the behavior of the sample mean, we are concerned with the convergence properties of sums of random variables. The normal starting point for such discussions would be the Central Limit Theorem (CLT). Unfortunately, the CLT applies only to sums of random variables with finite variance, and so does not apply in this case. In the place of the CLT we instead have limit theorems for fat-tailed random variables first formulated by Lévy [5],[19].

To introduce these results we need to define the notation $A \xrightarrow{d} B$ which means that the random variable $A$ converges in the distribution to $B$ (roughly, has distribution $B$ for large $n$). Then the usual CLT can be stated as: for $X_i$ i.i.d and drawn from some distribution $F$ with mean $\mu$ and variance $\sigma^2<\infty$, define

$$A_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad (3)$$

$$Z_n = n^{-1/2} (A_n - \mu) \quad (4)$$

then

$$Z_n \xrightarrow{d} N(0, \sigma^2) \quad (5)$$
in which $N(0, \sigma^2)$ is a Normal distribution.

However, if $X_i$ are i.i.d. and drawn from some distribution $F$ that is fat-tailed with tail index $1 < \alpha < 2$, then if we define

$$Z_n = n^{1-1/\alpha} (A_n - \mu)$$

we find that

$$Z_n \xrightarrow{d} S_\alpha$$

where $S_\alpha$ is an $\alpha$–Stable distribution [19].

The consequence of equation (6) is that as $\alpha$ gets smaller, the convergence of the sample mean to the population mean becomes very slow – much slower than in the case of distributions with finite variance. In Figure 7 we show the sample mean calculations performed on samples drawn from two distributions – one being normal and another that is fat-tailed. It shows how the mean of the sample drawn from normal distribution converges at 100 data-points but the mean of samples drawn from a fat-tail job size distribution of a large print shop has not converged even when the sample size is 2000. Simulations performed with fat-tailed inputs [2] converge slowly to steady state and show high variability at steady state. The practical consequence is that it becomes difficult for such print shops to predict performance within tight bounds.

7. Scheduling

A prior paper [8] discusses various scheduling policies when task-size distribution is fat-tailed. They consider four task-assignment policies for such distributed server systems: Round Robin; Random; Size-Based, in which all tasks within a given size range are assigned to a given host; and Dynamic-Least-Work-Remaining, in which a task is assigned to a host with least outstanding work. They reported that when the task-sizes are not highly variable, the Dynamic-Least-Work-Remaining policy is preferable. However, when the task-size distribution is fat-tailed and exhibits high variability, then the Size-Based policy is the best choice.

![Figure 7: Mean calculations on samples drawn from a normal distribution and a fat-tail job-size distribution.](image-url)
insertion into envelopes. A setup in the form of roll-changeover is incurred each time the underlying pre-printed form changes. Similarly, a new setup is needed when requirements on the inserting machines change. The combined effect of sequence-dependent setups and fat-tail job size distribution creates difficulties in achieving high throughput in these shops.

In Figure 8 we show the scheduling architecture that we studied for creating cells and scheduling jobs in large print shops with fat-tailed inputs. In step 1, the incoming job stream is separated into two sets using a job-size threshold parameter – one that contains small jobs and the other that are jobs above the threshold. Autonomous cells are designed to process the small jobs in their entirety. The small jobs are further sorted and grouped based on common attributes to reduce setup time on machines. The larger jobs are split into two sets – one that requires less frequent setup and the other that requires more frequent setups. Multiple strategies are used to perform this partitioning. In one strategy, the number of unique setup requirements is determined separately for multiple production steps (for example, printer setup and insertion setup) for each job. Each job is assigned a number that is the maximum of the setup numbers for each processing step. A threshold value is picked. Jobs that have assigned number that is less than the threshold are put in the low-setup pool, and the rest are put in a high-setup pool. The threshold value is iterated upon using multiple simulation studies to determine the value that optimizes the print shops performance objective (for example, number of late jobs, resource utilization, makespan). Autonomous cells are then designed for each job stream. The key motivation behind this strategy is to separate jobs that require high setups from the ones that require low setup and to design different types of autonomous production cells to handle the two types.

8. Heuristic Solution

High-variability in job size distribution can lead to high queue waiting times. In Figure 9 we show a G/G/m queue where job arrivals are routed to a set of identical printers. G/G/m queue refers to a system of m parallel machines with general (denoted by G) inter-arrival
and processing time distributions. The waiting time in the queue is directly proportional to the sum of squares of coefficient of variation of the inter-arrival time distribution and job size distribution respectively.

Splitting jobs into smaller batches has the effect of lowering the coefficient of variation of the job size distribution but increasing the coefficient of variation of the inter-arrival time distribution. If the effect of batch-splitting leads to an overall reduction in the value of \( (c_v^2 + c_v^2)/2 \), there will be a reduction in the queue waiting times at each printer.

**Figure 9:** Queue waiting time in a G/G/m queue [23] with job arrivals routed to a set of identical printers

### 9. An Example

In Figure 10 we show the inter-arrival time distribution and the job size distribution from a print shop that has high variability in job size distribution. The coefficient of variation (CV) of the inter-arrival time distribution is 2.7 and the coefficient of variation (CV) of the job-size distribution is 5.6.

**Figure 10:** Job size distribution and inter-arrival time distribution of black-and-white jobs from a print shop. The coefficient of variation is indicated by CV.

Table 2 indicates the effect of batch splitting on shop performance. While the CV of the job size distribution gets reduced significantly, the CV of the split-job distribution
experiences only a moderate increase. The net result is that the value of the term \( \frac{c^2_e + c^2_a}{2} \) is reduced from 19.3 to 6.2 (i.e., a reduction by a factor of 3.1).

This example illustrates the fact that the result of splitting very large jobs (in the fat-tail distribution) into smaller batches prior to routing these to autonomous cells has the effect of reducing queue waiting times. Further discussion of a batch-splitting heuristics in the presence of high-variability job size inputs can be found in [18]. In addition to performing batch-splitting prior to routing jobs to cells, the batches that are processed within a cell can be further split according to policies discussed in [12] to minimize work-in-progress and improve flow and cycle time.

| Table 2: Changes in job size and inter-arrival distributions after batch splitting. |
|---------------------------------|----------------|----------------|
| **Job Size** | Original | After Splitting |
| Mean      | 6722     | 4403           |
| Stddev    | 31017   | 9698           |
| CV        | 5.6     | 2.2            |
| **Inter-Arrival Time** | Original | After Splitting |
| Mean      | 0.69    | 0.66           |
| Stddev    | 1.87    | 1.83           |
| CV        | 2.70    | 2.76           |

Efficient sequencing policies to assign production priority to jobs queued up at the various cells are implemented to optimize desired production requirements like minimizing the percentage of late jobs or improving utilization. The interaction of these variables and their resulting impact on cell design, scheduling policies and performance can be quite complex. Extensive simulation models were built that allow iteration and optimization of the parameters of the architecture shown in Figure 8.

The solution described by the architecture diagram in Figure 8 is a function of the job size, setup requirements, and related characteristics of the incoming job stream. As the customers’ job mix changes, the cells may require re-adjustment and optimization. While the proposed architecture delivers superior performance to a functionally organized print shop, it needs ongoing maintenance and optimization as the job mix changes over time.

10. Summary

This paper describes a four quadrant framework (Figure 3) for characterizing print shops in the printing industry based on the workflow complexity and utilization of resources. While an earlier paper [12] described the application of a productivity improvement solution namely, LDP Lean Document Production to three of the four quadrants (Quadrant I, II and III), the focus of this paper is on large shops (Quadrant IV) that exhibit high levels of workflow complexity and typically run at higher levels of utilization than shops in other quadrants. The formalism of heavy tail distributions is used to characterize the extreme levels of variability observed in job size. Methods of quantifying and characterizing the variability are presented. A production architecture (Figure 8) is described along with scheduling policy heuristics that improve the productivity of such shops. A simulation-based optimization framework is utilized to evaluate and optimize the various parameters associated with the architecture and scheduling policies. Data analytics from a large print shop where this heavy tailed job size distribution was observed is described. The implementation of the LDP Lean Document Production solution in a large
transaction shop is presented (Table 1) and is shown to deliver substantial productivity benefits.

References


Sudhendu Rai


Sudhendu Rai is a Xerox Fellow, Program Manager and a certified Lean Six Sigma Black Belt at the Xerox Research Center in Webster, N.Y. He received his Ph.D. from MIT in 1993, M.S. from Caltech in 1989, and B.Tech. from IIT, Kanpur (India) in 1988—all in Mechanical Engineering. He is the lead inventor of the LDP Lean Document Production® Solution that was selected as a finalist in the Franz Edelman (worldwide) competition sponsored by INFORMS. This work has also been selected as one of the three finalists for the Service Innovation Practice Award sponsored by SRII (Service Research and Innovation Institute). LDP also won the first place at the IIE Lean Best Practices Award in 2011. Starting in 1998, he led a team that developed the algorithms, software toolkit to support the initial offering and a training curriculum to train Xerox Global Services consultants. He has personally led and implemented process improvement initiatives in dozens of small and large print shops spanning multiple industry segments. He holds 37 patents (with 40 additional pending) and has published more than 25 technical papers in conference proceedings and technical journals. He is a senior member of IEEE. He is a recipient of the Xerox Excellence in Science and Technology Award and was selected as a finalist for the Rochester Engineer of the Year award in 2007.