Journal of Imaging Science and Technology® 54(6): 060401–060401-9, 2010.

Productivity Analysis of Print Service Providers

Jun Zeng, I-Jong Lin▲, Eric Hoarau▲ and Gary Dispoto▲
Hewlett-Packard Laboratories, Palo Alto, California 94304-1100, USA
E-mail: jun.zeng@hp.com

Abstract. *Digital transformation of commercial print production brings new opportunities including exploitation of embedded sensing and computing power that enable real-time communication during various phases of the production chain and dynamic reconfiguration of production flow. Simulation based modeling can help to exploit these opportunities at both strategic and operational levels. In this article we report our ongoing work on simulating an end-to-end print production process. We draw a close parallel between print production design and electronic design automation, and model print production system as a network of interconnected, distinct processes. We describe our simulation framework, simulation validation against queuing network theory, preliminary simulation results, and path to policy design and analysis. © 2010 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2010.54.6.060401]

INTRODUCTION

Commercial print has annual retail sales over $700 billion. The print production process has not yet been treated as a standard manufacturing activity due to its own unique characteristics: from the product's perspective, it simultaneously demands both product diversity and mass production; from investment's perspective, it is both capital-intensive and labor-intensive. The highly variable and dynamic job mix with highly personalized customer requirements results in many different combinations of equipments, resources, print shop configurations and business philosophy.[1] Many print shops rely on the mental models of one or a few skilled masters to tackle the complexities of the print production systems. Such artisan (or craftsman) decision-making practice has been identified as one of the key reasons that the productivity growth of the print industry as whole is falling far behind other manufacturing industries.[2]

Commercial print is dominated by analog offset printing. The emergence of digital print production provides both challenge and opportunity to print production industry. Today, digital print is in its early stage, has only 3% of the market in volume, 10% in revenue, but with an impressive 10% annual growth rate. The near-zero setup cost of the digital print pushes the product diversity to a new extreme: every page can be different. This has enabled a new business paradigm: print-on-demand or publishing-on-demand. The production agility enabled by digital print makes possible the integration of pervasive sensing and monitoring, real-time computing power, embedded actuation and reconfigurable workflow architecture, thus the transformation of the print production into an autonomous system formed by networked computing and physical components and moving the "lights-out" print factory closer toward reality.[3] The secular trend[4] is with the digital print: the job run-length is getting shorter; the significant setup cost of the plate creation is making analog press less economical for growing number of jobs. Additionally, the digital print uniquely enables a new array of print services, such as transactional print and targeted advertising (transpromotional print).

To attack the low productivity problem associated with current print industry, and to best uncover the full potential of the digital print promise, we propose a holistic, model-based approach that analyzes the design and management of print production as an integrated system, accounting for the performance, efficiency, stability, and sustainability as organic system attributes. This article summarizes our ongoing effort toward this research direction. We map the analysis of print production system as an electronic design problem. We adapt an open-source electronic design automation (EDA) toolkit as our modeling platform. We model print production as a heterogeneous, concurrent, integrated system.

In next section, we describe the modeling problem. Following that, we outline our approach. Next we demonstrate system simulation results, and address validation issues. Having discussed the simulation-based evaluative models, we subsequently discuss the generative models and the applications to runtime policy design. We conclude the article by enumerating future work. We also note that the effectiveness of model-based analysis relies on winning acceptance and adoption from decision-makers. This is an equally or even more critical challenge that modelers and analysts must strive to meet.

PROBLEM STATEMENT

Commercial print service providers (PSP) process jobs (print requests) with given resources; the service level objectives (SLO) are usually dominated by the on-time delivery constraint or the due date. Resources include machines (e.g., Raster Imaging Process servers, print presses, and finishers) and labors; collectively they are referred to as servers. Multiple servers may perform the same task (e.g., black-and-white print) with different capabilities (e.g., number of pages per minute), availabilities (e.g., the length of waiting queue),

▲IS&T member

Received Nov. 27, 2009; accepted for publication Aug. 10, 2010; published online Nov. 5, 2010.
1062-3701/2010/54(6)/060401/9/$20.00.

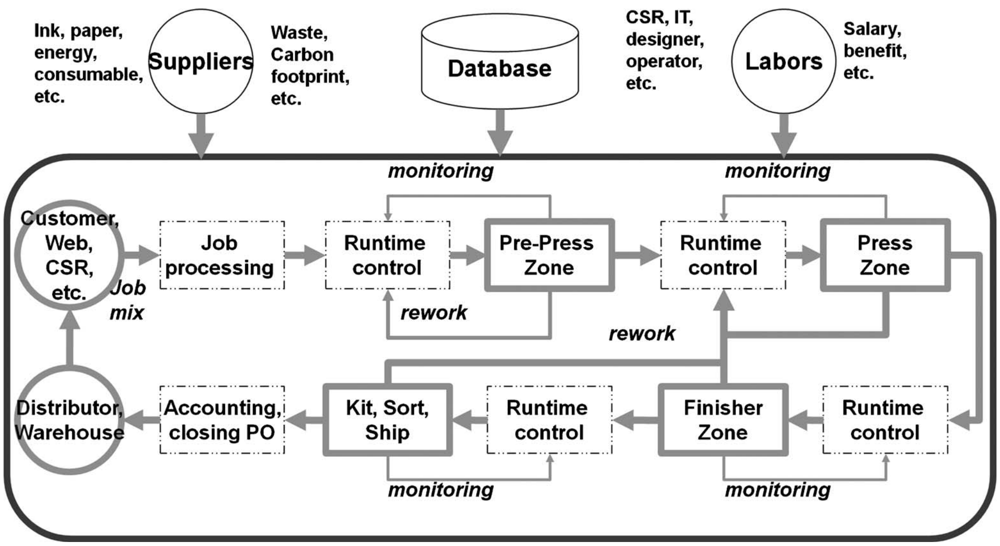

Figure 1. Depiction of end-to-end process logic of commercial print operation.

and status (e.g., type of substrate already loaded). Each job may need to visit different servers sequentially or in parallel for completion. Given the existing resources and client base (thus the job mix), the business objective of the PSP is to drive up the throughput to dilute the fixed-cost and maximize the profit while guarantees the quality of service.[1]

Figure 1 describes a typical print workflow:

(a) *Job acquisition.* Jobs (customers' requests for prints) are submitted through the storefront. This includes customers' physical visits, and, increasingly, jobs submitted electronically through the internet (web-to-print, for instance, www.blurb.com, www.snapfish.com). Once accepted, a unique job ticket is assigned. A job ticket usually contains a job number, number of copies, due dates, prices and payments, client information, inks and presses, finishing, and shipping methods.

(b) *Prepress.* The submitted electronic files (e.g., PDFs) are examined for the correctness (preflight), edited for color quality and accuracy, imposed, and flattened and half-tuned (raster imaging process) to create the bitmaps. In case of the offset process, plates are created (computer-to-plate).

(c) *Press.* Printer-ready electronic files are sent to the printers to produce physical copies. Different printers are used to produce book blocks and the book covers.

(d) *Postpress.* In terms of book printing by web press, the printed web is first singulated into printed pages. Printed pages are then folded, collated, and bound into book blocks. The book blocks are joined by book covers, and then dust jackets.

(e) *Distribution.* The finished books are sorted, labeled, and shipped. The purchase order (PO) is concluded.

Jobs arrive at indeterminate pace; the possible involvement of customers in proofing cycles can lead to long bid times. Other uncertainties associated with the jobs include job size, job type (content), thus workflow and equipments involved, and urgency (due dates). Variation associated with resources includes equipment capability, performance, stability, and costs of capital depreciation, material, energy, rent, insurance, labor, etc. The variability in run-time policy includes job prioritization, and timing, form and route associated with job release, possible monitoring and feedback, and scrap rate and quality assurance.

OUR APPROACH

We would like to draw a close parallel between the analysis of print production and electronic chip design (EDA) problems, in particular, that of laboratories-on-a-chip (LoC). LoC is a new system-on-chip architecture fabricated by micro electromechanical systems (MEMS) technology, with primary focus on high-throughput, massively parallel life science applications. A large number of independent assay operations can be carried out concurrently within this fingernail-sized chip; each involves a diverse set of sample operations including multistep sample preparations, assay, and detections. Similar to print production, a general-purpose LoC such as digital microfluidics also anticipates content diversity of assay requests (jobs). The LoC synthesis techniques aim to optimize the management of resources and runtime policies for incoming jobs. This includes scheduling assay operations, binding assay operations to available resources and temporal intervals accounting for resource sharing and resource constraints, and module placement and route (layout).[5] In print production systems, we face the same classes of design issues, about which, through both formal methods and heuristics, EDA provides a solid framework built on past 30 years of aggressive research and development.

We chose an open-source EDA toolkit, PTOLEMY,[6] as our modeling framework for print production systems. PTOLEMY is a Java-based, actor-oriented modeling framework for concurrent, real-time, embedded systems. Com-

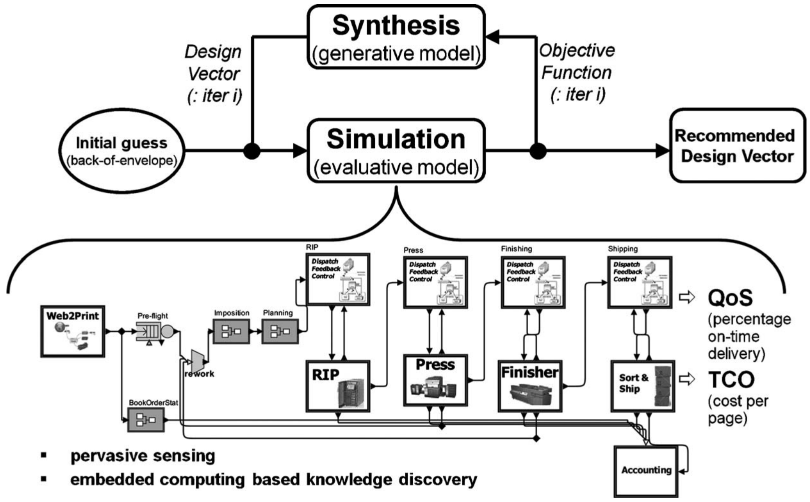

Figure 2. Sketch illustrates our modeling infrastructure for print production system. It is composed of both simulation and synthesis layers. The simulation integrates both devices models (components with frames) and devices themselves (represented by frameless components).

pared to object-oriented design practice, actor-oriented design emphasizes the concurrency and communication among components. PTOLEMY implements a set of well-defined models of computation (for instance, continuous time, discrete event, finite state machine) that govern the component interactions. It provides a hierarchical component assembly design environment that enables the use of heterogeneous mixtures of models of computation (e.g., hybrid and mixed-signal models). The print production system and control involves compute, logical and physical components, for which PTOLEMY's ability of blending different computational models provides the necessary simulation infrastructure support. In addition, we are already taking advantage of PTOLEMY's extensive customizability that only an open-source toolkit can offer.

We envision that the advanced state of our simulation infrastructure (Figure 2) will integrate both the system components themselves and models of other components that are otherwise too difficult or too expensive to be included. This simulation paradigm incorporates both hardware-in-the-loop (HIL) and software-in-the-loop (SIL). The accuracy of the system simulation relies on the precision of the component models it uses. Therefore, whenever possible, the components themselves are preferred over their models in simulation. In addition, integrating the real components in the simulation enables direct testing of the decision parameters. This approach benefits us with the agility and cost-effectiveness that simulations bring and the higher level of fidelity that experiments provide.

Alternative to PTOLEMY, standard discrete-event simulation packages (for instance, WITNESS software provided by Lanner Group, http://www.lanner.com/en/witness.cfm) can also perform workflow simulations. However, they do not usually provide sufficient programming accessibility and flexibility essential to the HIL and SIL paradigm. For this, an open source software platform provides unique benefit. PTOLEMY's strength in embedded systems design, i.e., designing and producing embedded software together with the systems within which it is embedded, is an excellent fit. We have successfully constructed a PTOLEMY model that integrates real-time raster imaging process as a component executed across intranet.

Simulation provides performance evaluation of a system design specified by a given set of design parameters. It is effective in discriminating among different design alternatives. However, simulation alone does not explicitly generate design or decision parameters. For this, we develop an additional layer of models, hereafter referred to as the generative models. In contrast, simulations are also referred to as the evaluative models. The generative models produce a set of design parameters that are favored (optimal) according to the design objectives formalized by the objective function (e.g., minimum makespan). The generative models are implemented on top of the simulations. For many design problems in print production process, simulation is the only feasible means to map design parameters and objective functions.

EDA provides a theoretical foundation, suite of classical techniques, and well-established design tools that can be used for print production design. However, the unique characteristics of print production process require these EDA techniques be tailored for print production applications. One example is the substantiation of human factors as a particular server class. Our research is aimed to provide interface between EDA and commercial print, and to synthesize the EDA methodology and print production application.

Research on productivity improvement is gaining accelerated attention in print industry in recent years. Most notably is the LDP solution by Xerox.[7] LDP is Xerox's simulation-based service solution offered through Xerox Managed Services (XMS) that helps to enhance print shop productivity. While applauding and encouraged by LDP's achievement, we would like to draw a distinction between our approaches. We approach print production as a content-driven cyber-physical system; we anticipate that the perva-

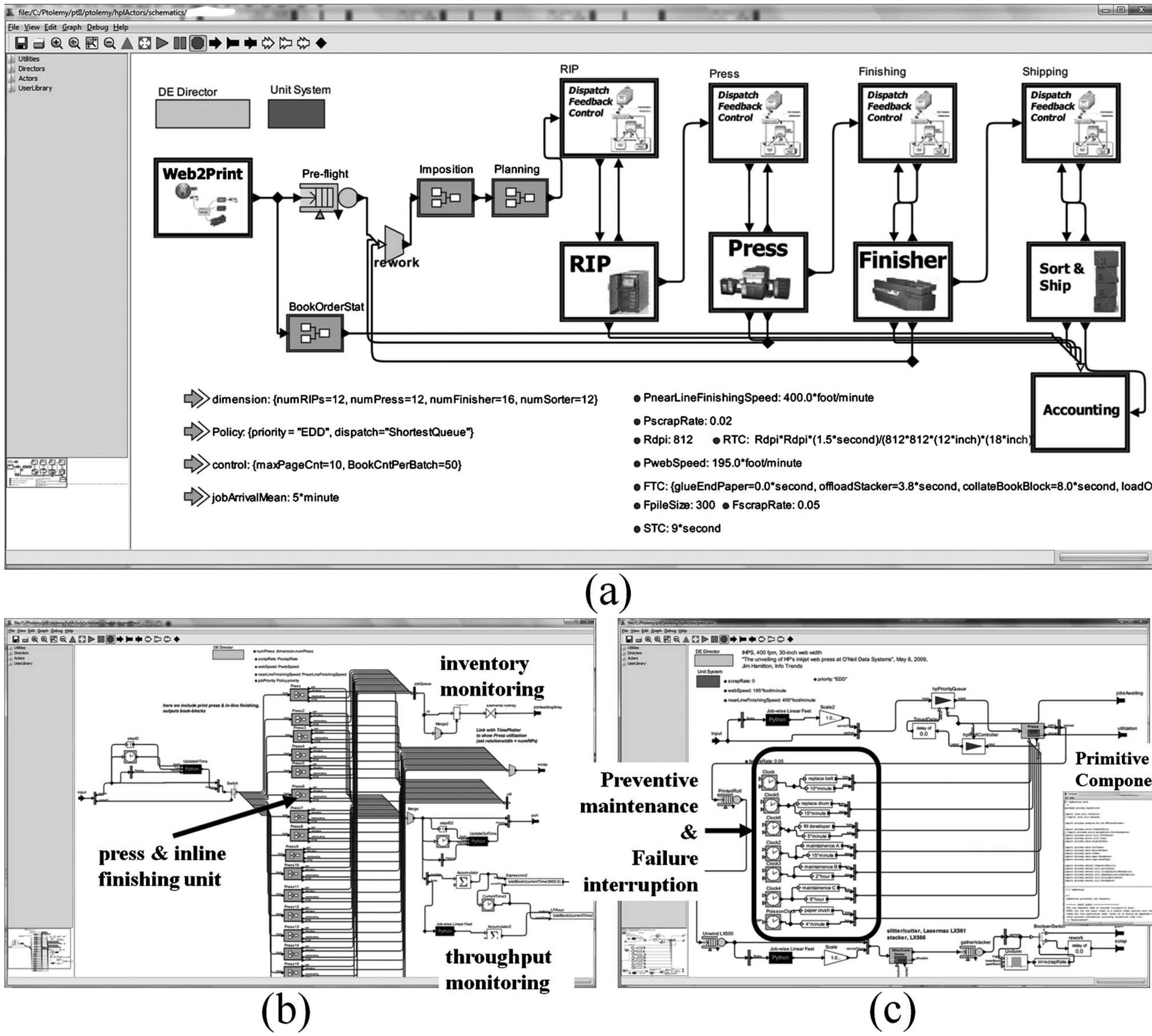

Figure 3. An explosive view of the simulation model hierarchy of end-to-end print production workflow. (a) Top level schematics. (b) Schematics of the press zone implements the Press composite actor shown in (a). An array of presses generate printed pages. (c) Schematics simulates a single press.

sive sensing and computing-based knowledge discovery will enable additional design space and flexibility.

PERFORMANCE SIMULATION

Figure 3 illustrates the implementation of the simulation model of print production system using PTOLEMY. The simulation starts from web-to-print, where the print requests (jobs) are collected. Multiple event generators are used that trigger print requests concurrently, simulating the presence of multiple storefronts. The print requests are simulated as Poisson processes with a prescribed mean interarrival rate. Once a print request is received, a job is generated. A job is a freely extendible record (or *struct*) containing the job arrival time, due, shipping address, array of book records, etc. Each book record corresponds to a unique book, containing number of pages, book sizes, paper types, and number of copies ordered. The job description is inherently stochastic; job generation involves random number generators according to prescribed statistical distributions. Figure 4(a) shows the statistical makeup of the jobs used in simulation. It includes the distribution of number of books per job, number of pages per book, and the due duration of print requests. The statistical makeup needs to be extracted from past job history and used as a simulation input. In this particular case, these statistical distributions shown in Fig. 4(a) are extracted from real job history data provided by a partner print service provider. This record form of job description allows programming flexibility and potentially straightforward transformation into JDF. This job record travels through all stages of print production process. Information may be extracted, updated, added, and subtracted onto the job record. The use of the job record resembles the job ticket in print shops.

When jobs arrive at each process stage, they can be ad-

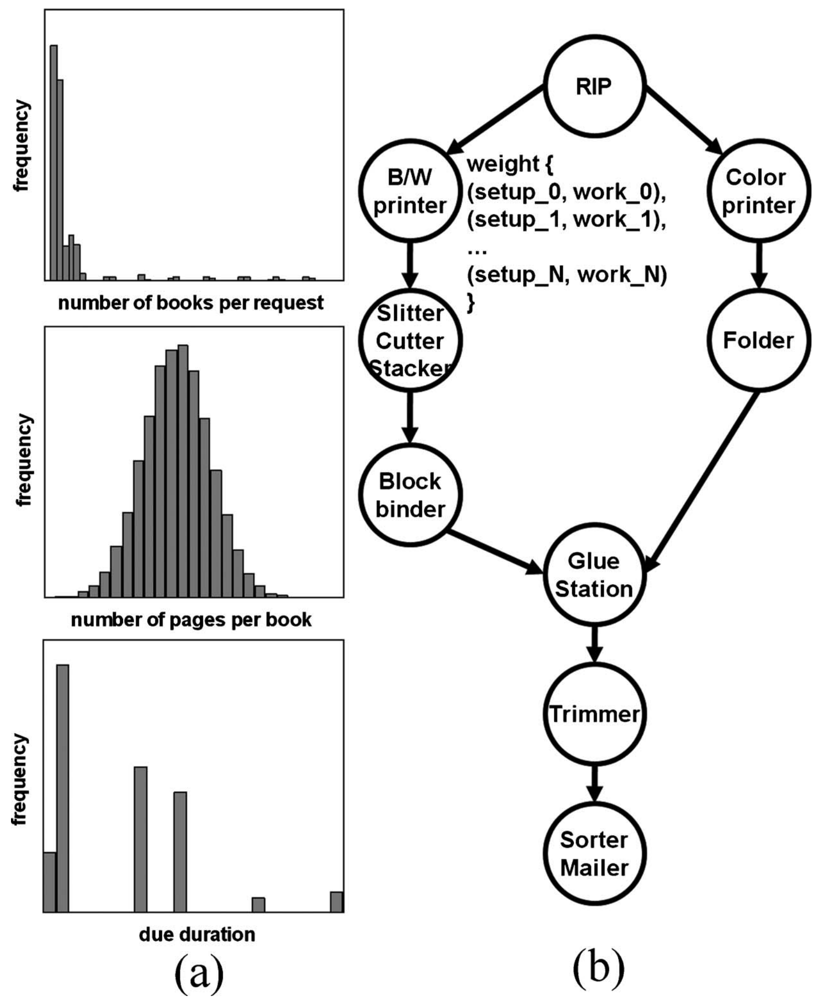

Figure 4. Representation of jobs. (a) Statistical job pattern obtained from a PSP, extracted from 5000 jobs (totally 150 000 books), from top to bottom, the distribution of number of books per job, number of pages per book, and the due duration of requests.(b) Direct acyclic graph documenting a single job.

ministered on a first-come-first-serve (also known as first-in-first-out) basis. To ensure on-time delivery and maximize productivity, other heuristic job priority policies have been exploited, for instance, earliest-due-date, shortest-processing-time, minimum-slack-time (slack time is defined as the difference between the due date and the sum of the remaining work and the current time), and so on. Usually there are arrays of machines that can perform the same task with various capabilities (e.g., service time and capacity), as illustrated in Fig. 3(b). The prioritized jobs need to be assigned to particular machines out of the pool of machines that can provide the same service (e.g., black-and-white print). Round-Robin with fixed-size quantum (e.g., total page count) can be used. Alternatively, machine priority policies can be implemented, for instance, shortest-queue, shortest-responsive-time, and so on. In addition, to minimize the machine setup cost batching is used. The batch size needs to be engineered otherwise it may have detrimental effect on on-time delivery since urgent jobs may be delayed.[8] Problems of this kind, constrained by the uncertainty and heterogeneority of the content and the availability of resources, have been proven NP-complete. Best heuristic practices are print shop specific depending on job contents and resource constraints.

The implementation of the machines accounts for the interruptions due to regular maintenance and unexpected machine failure, as illustrated in Fig. 3(c). Figures 5(a) and 5(b) show example simulation results of the machine utilizations and the inventory. They can be used to provide information about the effectiveness of certain policies that prioritize jobs and bind jobs with machines.

At the sort-and-ship stage, each book is sorted according to its shipping address. Once all the books for particular print request are collected, we calculate quality-of-service (QoS), defined as percentage of on-time delivery of completed jobs, and total-cost-of-ownership (TCO), defined as cost per page. The primary reference we use on cost is Ref. 9. Example calculation results of QoS and TCO are shown in Fig. 5(d).

The fidelity of the simulation model is measured by how closely it resembles reality. The acceptance and adoption of model-based analysis relies on demonstrating not only the usefulness of the model but also its fidelity. Validation, that is, the goodness of the assumptions made in the

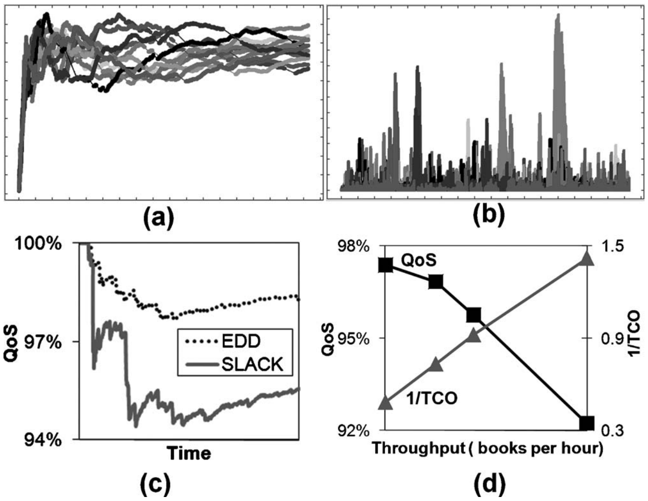

Figure 5. Example simulation results. The top row shows transient performance of 16 machines within the same process stage, (a) showing the utilization and (b) showing the inventory. The bottom row shows the system level metric. (c) shows the impact of different job prioritization algorithms. (d) illustrates the relationship between the objective of maximizing QoS and the objective of minimizing TCO. The horizontal axis is throughput. The increase of throughput reduces TCO but may impact on-time delivery negatively as it may increase the response time.

model about the reality, and verification, that is, the correctness of the model implementation, are a considerable portion of the model development effort. In particular, the validation is of higher complexity, drawing on both domain knowledge and modeling expertise, and lasting throughout the model development process. Depending on particular problems of study and the accessibility and makeup of the experimental data, we usually infer simulation fidelity from three possible sources, (a) analytical results, (b) experimental measurements, and (c) intuitions of domain experts. Analytical results provide not only validation but also means for verification, however, only much simplified models of real-life problems can be solved analytically. Agreement with experimental measurements from real-life system provide highest confidence in the simulations; however, the accessibility of the "right" experimental measurements is a challenge. The interpretation and extrapolation of the experimental data is best carried out coupling with the intuition of domain experts.

Below we discuss a validation study using analytical results derived from queuing network theory. Consider the following process, as illustrated in Figure 6(a), the fulfillment of a PO includes N sequential steps. POs arrive with a mean arrival rate of λ. At each step (indexed by j), multiple machines, number of M_j, each denoted by (i,j), can perform the same function with various capabilities (characterized by mean service rate $\mu_{i,j}$). The probability that a job is assigned to one of the M_j machines is determined such that, statistically, jobs receive equal response time (system time) regardless which machine it is assigned to. Upon completion of the last step, defect jobs, of scrap rate p, are routed to the initial step to repeat the fulfillment process. Both job arrival and machine service are assumed to be Poisson processes.

Even though the process described above is a great simplification of the real print production process, it embodies its essence and involves several key components. Most importantly, this simplified process can be solved analytically according to queuing network theory.[10] Therefore, it is chosen as a validation case for the simulations.

The runtime policy is defined such that, for any j step, the responsive time mean is the same for any one of the M_j machines. The responsive time mean W_j can be calculated, as

$$W_j = \frac{M_j}{\sum_i \mu_{i,j} - \lambda/(1-p)}. \tag{1}$$

Therefore, the probability that an incoming job arriving at step j to be assigned to machine (i,j) is

$$P_{i,j} = \frac{\mu_{i,j} - 1/W_j}{\lambda/(1-p)}. \tag{2}$$

The utilization of machine (i,j) is

$$\rho_{i,j} = \frac{\lambda P_{i,j}}{(1-p)\mu_{i,j}}. \tag{3}$$

The length of the waiting queue is

$$Q_{i,j} = \frac{\rho_{i,j}^2}{1-\rho_{i,j}} \tag{4}$$

Simulation results are plotted together with the analytical results shown in Figs. 6(b)–6(e). The simulated mean values are calculated using sub-sampling method (also known as method of batch means). Good agreement between the simulation results and that of queuing network theory are observed.

POLICY DESIGN

Jobs

The input to the PSP is the job requests from the print buyers; let **J** represent the entire job set yet to be released to the factory floor. The job pattern describes the statistical attributes including job arrival (tempo, can be modeled as Poisson process characterized by a mean arrival rate λ) and payload (spatial). The payload describes the work needs to be fulfilled including number of copies ordered, dimension for each copy and the delivery parameters. The job pattern can be acquired through field study; they depend on PSP's client makeup. Fig. 4(a) shows a typical job payload obtained from a large PSP indicating most jobs are of short run-length and with tight due dates. A job, denoted by j_n, can be documented using a sequencing graph [direct acyclic graph, or DAG, as shown in Fig. 4(b)]: the job is partitioned into set of tasks, defined as work can be completed by one server on the factory floor. The tasks are the nodes of the DAG, associated weight records costs, and the directed edges

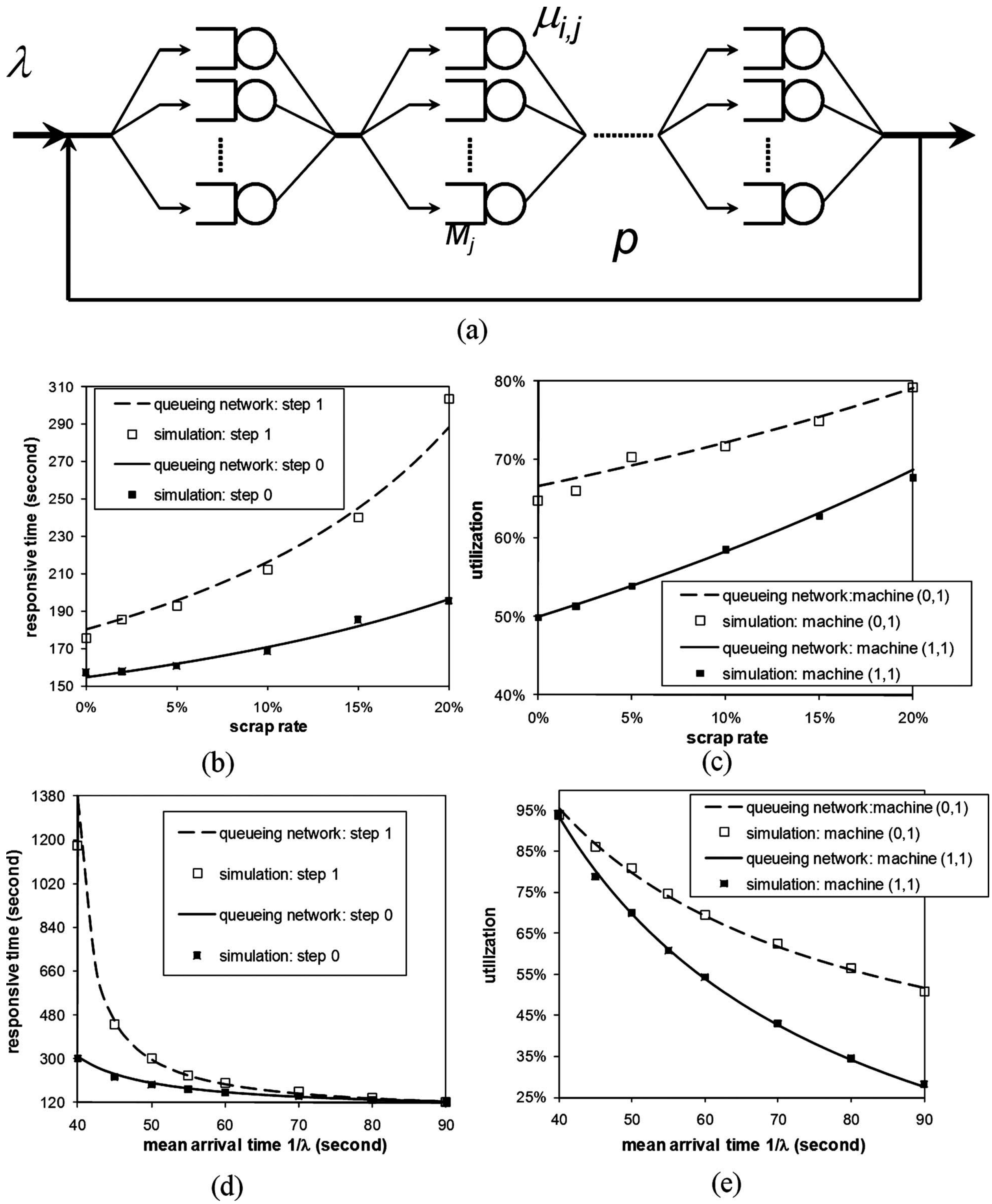

Figure 6. Simulation validation by queuing network theory. (a) illustrates the validation problem. (b)–(e) shows the comparison between the simulation results and that of queuing network theory. In this particular simulation, $N=2$, $M_0=3$, and $M_1=2$. $1/m_{i,j}=\{1, 1.5, 2; 1, 1.5\}$ min. In (b) and (c), the mean interarrival time $1/l$ is *1* minute and scrap rate varies from 0% to 20%. In (d) and (e), the scrap rate is 5% and $1/l$ varies from 40 to 90 s. Good agreement between simulation and queuing network theory is observed.

indicate the precedence constraints. $T^{(n)}$ denotes the task set associated with j_n ; $E^{(n)}$ denotes the associated edge set,

$$E^{(n)} = \{e|e = (u,v),\ u,\ v \in T^{(n)},\ u \neq v,\ start(u) + do(u) \leq start(v)\}, \quad (5)$$

where functions *start(.)* and *do(.)* return the time to start and time taken to perform a task u. Let **T** denote the complete task set on **J** and **E** the complete edge set on **T**, the SLO can be written as

$$start[last(E^{(n)})] + do[last(E^{(n)})] \leq due^{(n)}, \quad \forall\ n, j_n \in \mathbf{J} \quad (6)$$

where function *last(.)* returns the last task of job j_n. Function *do(.)* is a function of both task and the assigned server to fulfill this task, it includes possible setup cost which usually depends on the server status, i.e., the last task this server just completed.

Resources

The resource is characterized by its performance, make-ready effort, cost, consumables, waste, and exception patterns which includes both the nondeterministic (e.g., paper break, machine mechanical failure) and deterministic (e.g., regular maintenance, scheduled ink refill, worker shift) interruptions, denoted as **M**. A compatibility variable c can be defined to classify resources according to the tasks they can fulfill: $c(t,m)=1$ when resource m can fulfill task t; otherwise $c(t,m)=0$.

Policy

Many jobs are processed simultaneously on the factory floor, competing for the resource pool to meet their associated SLO. The policy problem includes: (a) determine which job has the priority to receive available resources (job sequencing); (b) determine which server has the priority to receive a new job (resource binding). Solutions of job sequencing and resource binding provide the job batching procedure. Faulty (scrap) jobs are rerouted to necessary upstream to repeat the fulfillment process (re-entrant). To formulate the policy, we define a binding variable $b{:}b(t,m)=1$ when task t is assigned to resource m and $c(t,m)=1$; otherwise $b(t,m)=0$. Task set $T_{(m)}$ includes all the tasks assigned to resource m, i.e., $T_{(m)}=\{t|b(t,m)=1, \forall\, t \in T\}$ for each resource m. We further define a permutation variable $p(m,t,j){:}p(m,t,j)=1$ when $b(t,m)=1$ and t is scheduled the j-th to be fulfilled by m; otherwise $p(m,t,j)=0$. By definition,

$$\sum_j p(m,t,j) = 1,\quad \forall\, t \in T_{(m)},\quad \forall\, m \in M;\quad \sum_{t \in T_{(m)}} p(m,t,j) = 1,\quad \forall\, j \in [1, size(T_{(m)})],\quad \forall\, m \in M. \quad (7)$$

Additionally, resource m can fulfill only one task at a time, that is

$$start(u) + do(u) \leq start(v)\ \ or\ \ start(v) + do(v) \leq start(u),\quad \forall\, u,\ v \in T_{(m)}\ \ and\ \ u \neq v,\quad \forall\, m \in M. \quad (8)$$

Objective

Find the runtime policy that achieves business objectives, for instance, minimizes the makespan and simultaneously guarantees on-time delivery,

$$\min E\left[\frac{1}{T'}\int_0^{T'} (\max^{j_n \in J}\{start[last(E^{(n)})] + do[last(E^{(n)})]\})dt'\right] \quad (9)$$

where function $E(.)$ returns the expectation, T' denotes the time duration of interest.

The job profile can be extracted from past job submission history. It is the input for system analysis and design. The resource and run-time policy domains project the design space.

SOLUTION APPROACH

Step 0: A subset of high-level questions may find the queuing network model acceptable. In this case, an explicit mixed-integer linear programming model can be formulated through mean value analysis, which will generate the policy solutions. *Otherwise,*

Step 1: A handful of heuristic procedures are implemented in today's PSP operation, for instance, EDD (earliest due date) as shown in Fig. 5(c). A collection of popular heuristic solutions are already implemented in the print workflow simulator; they form the heuristic solution space. Enumerative procedure is applied to scan through this heuristic solution space and search for policy solutions meeting performance expectation. *Otherwise,*

Step 2.0: Iterative procedure is applied over the print workflow simulator to map the performance metric space and the decision variable space, and derive the policy solution recommendation.[11,12]

Step 2.1: Relaxation schemes are applied to reduce and simplify the obtained policy solution for practical implementation.

FUTURE WORK

Simulation based modeling can help to optimize the print production system at both strategic and operational level. It can help to determine optimal workflow, predict production bottlenecks, guide capacity planning, and incubate different print production paradigms without incurring material expenses. This article reports our ongoing work on simulating a print production system adopting EDA tools and methodology, and preliminary simulation results.

Many efforts will need to be made on multiple fronts in order to develop this simulation platform into a computer-aided design tool for digital print production, including

(1) Implementing more new component models and provide complete component library for print production process. This includes not only models of various production machines, but also models of workforces, product tracking mechanisms (e.g.,

wireless barcode readers that are broadly implemented in print factories today), and decision making logic.

(2) Validating the simulation solutions experimentally. We have engaged with one of the largest digital print service providers in the United States as our research partner to facilitate field validation and reality checking.

(3) Further developing the generative model. Our plan is to repurpose the suite of EDA synthesis technologies developed over past 30 years to address the design automation and optimization problems involved in digital print. For instance, the policy design problem outlined above describes one particular use case of this generative model. Another example is the print production factory layout design which closely resembles the EDA geometrical synthesis that provides the optimal solutions for module placement and routing.

(4) Investigating different print production paradigms. Print service providers have started to embrace the lean methodology. A common practice today for the print service providers is to engage with external lean consultants to implement the lean practice. These lean-oriented changes may affect the operations policy, for instance, "batch of one;" they may affect the factory layout, for instance, incorporating more conveyor that will incur additional material expenses. Many of these changes have resulted in great success—higher throughput, faster turnover rate, better customer satisfaction score, business expansion without incurring additional headcount, and so forth. However, some of these changes have failed after business interruption owing to material costs. We plan to deploy this simulation platform to investigate different print production practices, to provide quantitative assessment on, for instance, the effectiveness of "batch of one," and also to illustrate the use of the modeling as the "virtual sandbox" that print service providers can use to first test out changes and verify the practicality and effectiveness before implementing them on the factory floor.

The effectiveness of model-based analysis relies on winning the acceptance from the decision-makers. This article summarizes our first step in demonstrating the fidelity and usefulness of simulation and modeling, and prompting the model-based design approach.

ACKNOWLEDGMENTS

Contribution from Moshe Abergel (HP/Indigo), Ron Askeland (HP/ICE), Nathan Guhthrie (HP/DDO), Johan Lammens (HP/LFP), Edward Lee (University of California, Berkeley), Jim Rowson (HP/CTO), and Ronen Yancu (HP/Indigo) is gratefully acknowledged.

REFERENCES

[1] *Handbook of Print Media: Technologies and Production Methods*, edited by H. Kipphan (Springer, Berlin, 2001).

[2] J. Uribe, *Print Productivity: A System Dynamics Approach*, M.Sc. thesis, Rochester Institute of Technology, 2007.

[3] I.-J. Lin, J. Zeng, E. Hoarau, and G. Dispoto, "Proposal for Next-Generation Commercial Print Infrastructure: Gutenberg-Landa TCP/IP", *Proc. IS&T's NIP25: Int'l. Conference on Digital Printing Technologies and Digital Fabrication 2009* (IS&T, Springfield, VA, 2009).

[4] According to http://www.businessdictionary.com/definition/secular-trend.html, the secular trend refers to economy or market trend associated with some characteristic or phenomenon which is not cyclical or seasonal but exists over a relatively long period. This revision has included this definition in the reference section.

[5] K. Chakrabarty and F. Su, *Digital Microfluidic Biochips: Synthesis, Testing, and Reconfiguration techniques* (CRC Press, Boca Raton, FL, 2006).

[6] C. Brooks, E. A. Lee, X. Liu, S. Neuendorffer, Y. Zhao, and H. Zheng, *Heterogeneous Concurrent Modeling and Design in Java* (University of California, Berkeley, CA, 2008).

[7] S. Rai, C. B. Duke, V. Lowe, C. Quan-Trotter, and T. Scheermesser, "LDP Lean Document Production—OR-enhanced productivity improvements for the print industry", Interfaces **39**(1), 69–90 (2009).

[8] W. Hopp and M. Spearman, *Factory Physics*, 3rd ed. (McGraw-Hill/Irwin, New York, 2007).

[9] *Budgeted Hourly Cost Study: Digital Prepress Operations; Sheetfed Press Operations; Web Press Operations; Bindery, Finishing & Mailing Operations* (NAPL, Paramus, NJ, 2008).

[10] R. W. Wolff, *Stochastic Modeling and The Theory of Queues* (Prentice Hall, New York, 1989).

[11] M. C. Fu, "Optimization by simulation: a review", Ann. Operat. Res. **53**, 199–247 (1994).

[12] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 3rd ed. (Springer, Berlin, 2008).