An Alternative Modeling Framework for Aggregate Production Planning

Jakob Asmundsson, Reha Uzsoy, and Ronald L. Rardin
Laboratory for Extended Enterprises at Purdue
School of Industrial Engineering
1287 Grissom Hall
Purdue University
West Lafayette, In 47907-1287

November 4, 2002

1 Introduction

Models of production and distribution systems have been developed in the operations research and management science literature since the initial emergence of these fields. A fundamental issue in this area has always been the development of computationally tractable models that accurately reflect the operational dynamics of the systems under study. A fundamental dichotomy appears to have emerged between the mathematical programming models used for aggregate planning and the queuing and simulation models used for performance analysis. The queuing models have revealed that critical system performance measures, especially lead times, are affected by the workload on the system relative to its capacity, or its utilization. In particular, lead times increase nonlinearly in both mean and variance as the system utilization approaches 100%. On the other hand, the aggregate planning models have suffered from a fundamental circularity. In order to plan production in the face of time-varying demands, they often use fixed estimates of lead times in their planning calculations. However, the decisions made by these models determine the amount of work released into the facility in a given time period, which determines the utilization and, in turn, the lead times that will be realized.

In this research we build on a number of ideas developed previously by a number of authors (Graves 1986; Karmarkar 1989; Srinivasan et al. 1988) to propose a mathematical programming framework for modeling capacitated systems that accurately captures the nonlinear relationship between workload and lead times. We use the idea of clearing functions that define the throughput of a capacity-constrained resource as a function of the work in process inventory (WIP). A simplified version of the formulation that only handles single product systems is presented here. For details on an extension to multiple products as well as a more detailed discussion the reader is referred to Asmundsson et al. 2002.

2 Clearing Functions

From Little's Law, it is apparent that in order to keep lead time fixed when WIP is increased, throughput must be increased proportionally. Being able to keep the lead time fixed independent of workload implicitly assumes infinite capacity, so instead of throughput being proportional to WIP, the throughput-WIP curve (clearing function) levels off when throughput approaches the theoretical capacity of the system. This is illustrated in the figure below with a nonlinear clearing function.

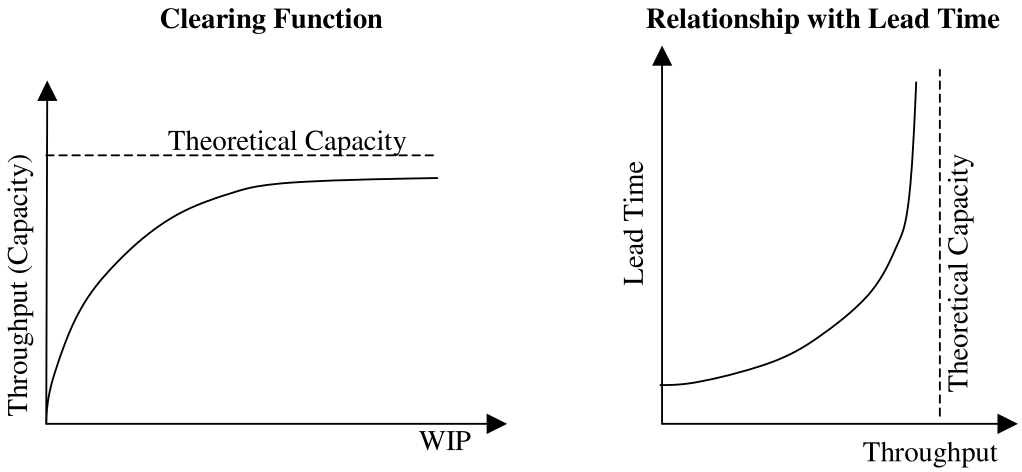

Figure 1 Relationships between throughput, WIP and lead time (Karmarkar 1989)

Since WIP and lead time are directly proportional due to Little's Law, the equivalent relationship between lead time and throughput can be drawn as well. This is shown in the figure on the right. Notice that as throughput approaches its theoretical capacity limit, the expected lead time goes to infinity. Hence by capturing the relationship between WIP and throughput, the lead time dynamics can be captured in a mathematical model.

3 Model Formulation

We present a simplified version of the formulation here for a single product and single stage system along the lines of Karmarkar (1989) and Srinivasan et al. (1988). The extension to a multi-stage system is straightforward, but the multi-product extension requires a more in-depth discussion. We consider a simple objective function that minimizes total system inventory for illustrative purposes, although a wide range of other objective functions can be used.

$$Minimize \quad \sum_{t=1}^{T} WIP_t + FGI_t$$

System inventory being composed of WIP and finished goods inventory (FGI), the formulation requires material balance equations for both inventory locations. WIP at the end of the period is equal to the beginning WIP, less the throughput (TP) during the period, plus the release of raw material (REL) during the period.

$$WIP_t = WIP_{t-1} - TP_t + REL_t \qquad t=1..T$$

Similarly, the finished goods inventory increases by the amount of throughput during the period, less the demand satisfied.

$$FGI_t = FGI_{t-1} + TP_t - Demand_t \qquad t=1..T$$

We are assuming that all demand can be satisfied on time, and no backlog can occur. This assumption can be relaxed in the usual manner.

The core of this formulation is the capacity constraint, which is of the form

$$TH_t \le f(WIP_{t-1}) \qquad t=1..T$$

Notice that in contrast to traditional linear programming models where the right hand side is a constant, the capacity is now some function of the beginning WIP level. For discussion on whether the beginning WIP should be used or some other measure of the average WIP during the period, the reader is referred to Karmarkar (1989) and Asmundsson et al. (2002). For now we use the beginning WIP levels as a measure of the congestion during the period, and therefore the effective capacity.

Let's now look closer at the clearing function in the right hand side. First, we have not discussed the analytical form of the function, nor how to derive it. The clearing function can be derived analytically or numerically from queuing analysis, or estimated from simulation models by sampling observations of throughput values for different WIP levels.

Second, the function is nonlinear, thus resulting in a nonlinear program. However since the clearing function is concave, it can be approximated with a convex hull of straight lines of the form $f(WIP_t) = \min_{c=1..C}\{\alpha_{ct} WIP_t + \beta_{ct}\}$. The capacity constraint can now be restated as

$$TH_t \le \alpha_{ct} WIP_t + \beta_{ct} \qquad t=1..T,\ c=1..C$$

The resulting formulation is a linear program that can be readily solved using conventional solvers such as those present in Xpress-MP.

4 Simple Example

To motivate the approach we present a simple example of a G/G/1 queuing system. We derived the clearing function analytically (Asmundsson et al. 2002) and approximated with five straight lines.

The implementation of the mathematical model as an Xpress-Mosel program is straightforward.

```
model "Aggregate Production Planning"
 uses "mmxprs"

 T:=30
 C:=5

 declarations
  Alpha:array(1..C) of real
  Beta:array(1..C) of real
  Demand:array(0..T) of real
  TP:array(0..T) of mpvar
  REL:array(0..T) of mpvar
  WIP:array(0..T) of mpvar
  FGI:array(1..T) of mpvar
  Initial_WIP:real
  Initial_FGI:real
 end-declarations

 initializations from "model.dat"
  Alpha Beta Demand Initial_WIP Initial_FGI
 end-initializations

 forall(t in 1..T)
  WIP(t)=if(t>1,WIP(t-1),Initial_WIP) - TP(t) + REL(t)

 forall(t in 1..T)
  FGI(t)=if(t>1,FGI(t-1),Initial_FGI) + TP(t) - Demand(t)

 forall(t in 1..T, c in 1..C)
  TP(t)<=Alpha(c)*WIP(t-1) + Beta(c)

 Total_Inventory:= sum(t in 1..T) (WIP(t) + FGI(t))

 minimize(Total_Inventory)

end-model
```

Solving the program using randomly generated demand yields the results summarized in the figure below.

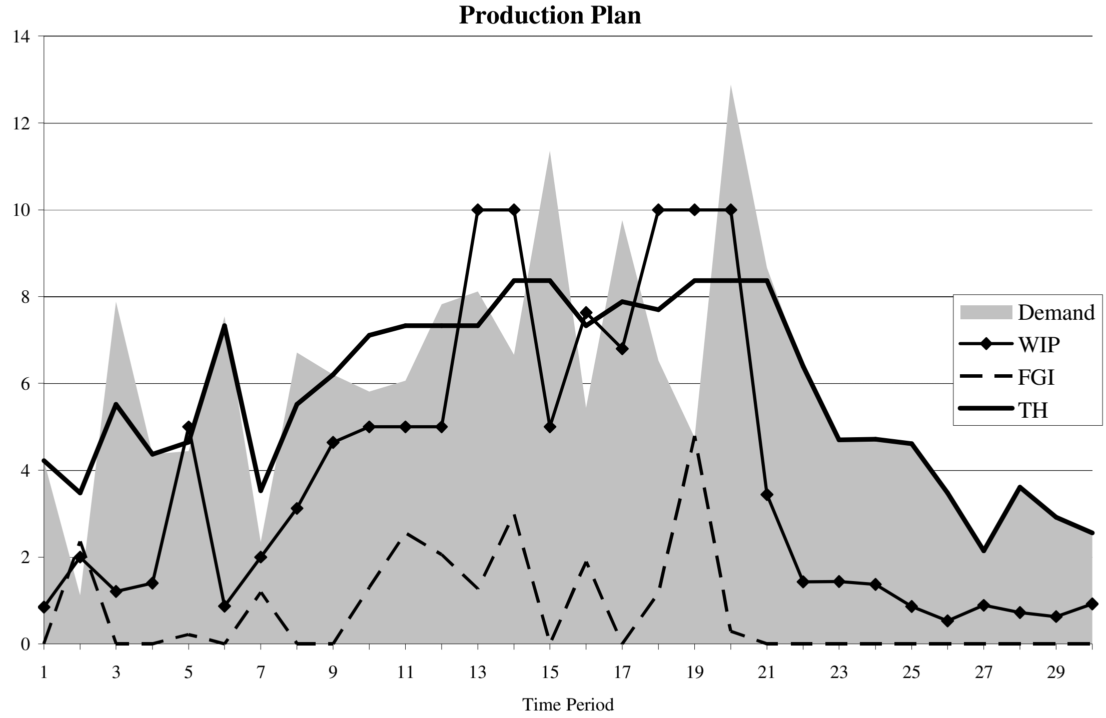

Figure 2 Production plan

Notice that in some cases, finished goods inventory is non-zero in periods where throughput is not very high. This is due to the fact that the marginal increase in throughput (capacity) is too low to warrant an increase in WIP and it is more economical to build ahead and keep FGI. In fact, the throughput in this example never even comes close to the theoretical capacity of 10 units. Notice that in the first 5 periods, the optimal solution has FGI, although all demand can be satisfied without holding any FGI.

By comparing the cumulative releases into the system to the cumulative throughput, the planned production lead time can be derived (Figure 3). Notice that the production lead time varies considerably, ranging from 0.2 to 1.2 periods over the planning horizon. These are significantly greater than the raw processing time, which is 0.1 periods.

An interesting question to ask here is: if one were to use the traditional linear programming models based on fixed lead time estimates, what lead time value should be used? Certainly not the raw processing time, since that would grossly underestimate the actual lead time. A consequence of that would be late material release and therefore demand would not be satisfied on time.

How about some measure of the average lead time? In this case material is released either too early or too late, resulting in unfilled demand in some periods and excess inventory in others.

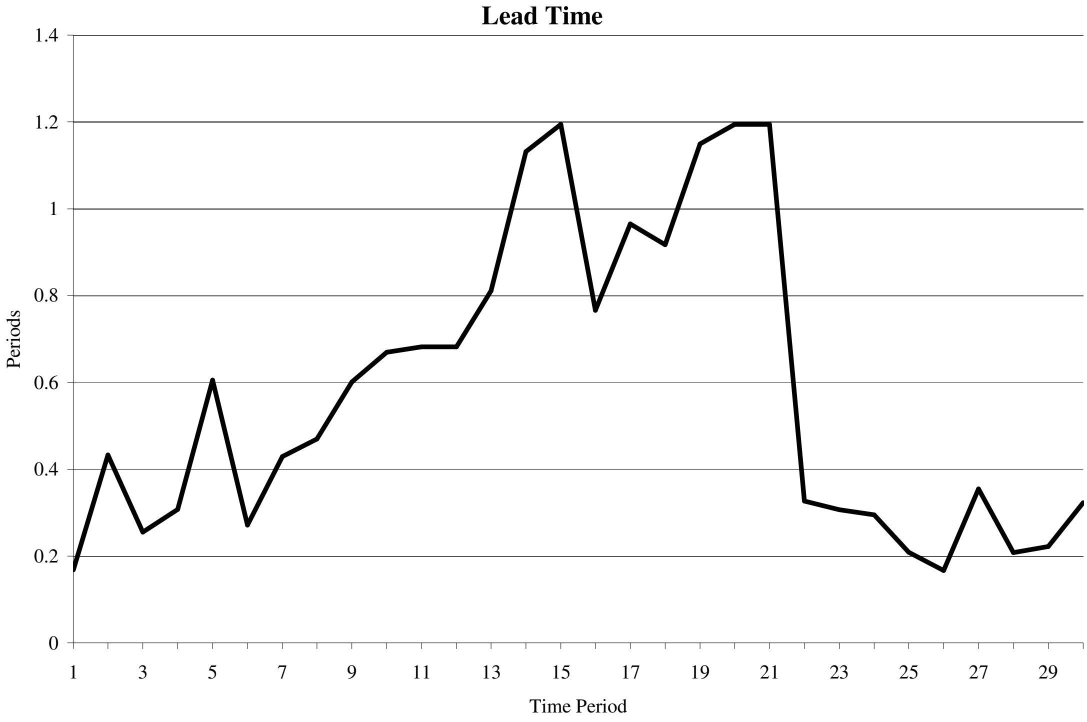

Figure 3 Production lead time

5 Conclusions

The goal of this research has been to develop a computationally tractable modeling framework that captures the nonlinear dynamics that exist between workload, lead-time and capacity. This is fundamental in modeling production networks accurately, whether it is composed of workstations within a single plant or nodes in a supply chain. Because lead times are captured implicitly in the formulation, the mathematical program develops the lead times that optimize system performance with respect to the chosen objective function. Concavity of the clearing function that is used to represent capacity allows us to develop a linear programming formulation for the planning problem that is relatively tractable and can be readily solved.

6 Acknowledgments

This research has been supported by The Laboratory for Extended Enterprises at Purdue (LEEAP), the UPS Foundation, and NSF Grants DMI-0075606 and DMI-0122207 through the Scalable Enterprise Systems Initiative. We also thank Dr. Alkis Vazacopoulos of Dash Optimization Inc. for supporting us through software donations. We used Xpress-IVE to model and solve the mathematical programs. The software allowed us to examine different scenarios and test our models against other modeling techniques with minimal programming efforts.

7 List of References

Asmundsson, J., R. L. Rardin, and R. Uzsoy (2002). "Tractable Nonlinear Capacity Models for Aggregate Production Planning", Research Report, Laboratory for Extended Enterprises at Purdue, Purdue university, West Lafayette, In 47907-1287.s

Graves, S. C. (1985). "A Tactical Planning Model for a Job Shop" *Operations Research*, 34, pp. 552-533.

Karmarkar, U. S. (1989). "Capacity Loading and Release Planning with Work-in-Progress (WIP) and Lead-times" *J. Mfg. Oper. Mgt.*, 2, pp. 105-123.

Medhi, J. (1991). *Stochastic Models in Queuing Theory*, Academic Press.

Srinivasan, A., M. Carey and T. E. Morton (1988). "Resource Pricing and Aggregate Scheduling in Manufacturing Systems" Unpublished paper, GSIA, Carnegie-Mellon University.