RC23934 (W0604-051) April 11, 2006
Other

IBM Research Report

Simulating Impact of Available-to-Promise Generation on Supply Chain Performance

Young M. Lee
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

SIMULATING IMPACT OF AVAILABLE-TO-PROMISE GENERATION ON SUPPLY CHAIN PERFORMANCE

Young M. Lee

IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598, U.S.A.

ABSTRACT

Availability management influences key supply chain performance metrics such as customer service level and inventory. The availability management process involves generating Available-to-Promise (ATP) quantities, scheduling customer orders against the ATP, and fulfilling the orders. ATP generation is a push-side of the availability management process, and it allocates expected availability into ATP quantities based on product types, demand classes, supply classes, and ATP time buckets as well as various availability management polices. This paper describes a simulation work done for IBM computer hardware business to evaluate how changes in ATP generation would impact supply chain performance. The simulation work played an important role in making strategic business decisions that impacted customer services and inventory cost.

1 INTRODUCTION

This work was motivated by supply chain processes of IBM's Computer Hardware businesses. In IBM, businesses are being managed as On-Demand business, where business strategies, policies and processes are continually evaluated and changed to meet increasingly demanding needs of customers. These changes are called "business transformations" in IBM. Various business transformation ideas are generated, evaluated and deployed to improve the effectiveness of the businesses especially in the area of supply chain. Availability Management Process (AMP) is one such area where transformation ideas are constantly evaluated and implemented. When a change in AMP is sought, the impact of such change has to be accurately assessed before they are implemented because the changes are typically expensive and time consuming to implement in large enterprises as IBM.

The availability management involves generating availability outlook, scheduling customer orders against the availability outlook, and fulfilling the orders. Generation of Availability Outlook is a push-side of the availability management process, and it allocates availability into ATP (Available-to-Promise) quantities based on various product and demand characteristics and planning time periods. Order Scheduling is a pull-side of availability management process, and it matches the customer orders against the Availability Outlook, determines when customer order can be shipped, and communicate the promised ship date to customers. Order fulfillment is executing the shipment of the order at the time of promised ship date. Even if an order is scheduled for shipment for a certain date based on the outlook of availability, the resources that are required to ship the product on the promised ship date may not actually available when the ship date comes. A key role for effective availability management process is to coordinate and balance the push-side and pull-side of ATP.

Ball at al. (2004) gave an overview of the push-side (Availability Planning) and pull-side (Availability Promising) of ATP with examples from Toshiba, Dell and Maxtor Corporation. They stressed the importance of coordinating the push and pull-side of availability management for supply chain performance by making good use of available resources. Although ATP functions has been available in several commercial ERP and Supply Chain software such as SAP's APO, i2's Rhythm, Oracle's ATP Server and Manugistics' SCPO modules etc. for several years (see Ball et al. 2000 for details), those ATP tools are mostly fast search engines for availability database, and they schedule customer orders without any sophisticated quantitative methods. Research on the quantitative side of ATP is still at an early stage, and there are only a limited number of analytic models developed in supporting ATP.

For the push-side of ATP, Ervolina and Dietrich (2000) developed an optimization model as the resource allocation tool, and described how the model is used for a complex Configured-to-Order (CTO) environment of the IBM Server business. They also stress how the push-side (Availability Promising) and pull-side (Availability Planning) have to be work together for the overall availability management performance.

For the pull-side of ATP, Chen et al. (2002) developed a Mixed-Integer Programming (MIP) optimization model for a process where order promising and fulfillment are handled in a predefined batching interval. Their model determines the committed order quantity for customer orders that arrive with requested delivery dates by simultaneously considering material availability, production capacity as well as material compatibility constraints. They also studied how the batching interval affects supply chain performance with different degree of resource availability. Moses et al. (2004) also developed a model that computes optimal promised ship date considering not only availability but also other order-specific characteristics and existing commitments to the previous scheduled orders. Pan et al. (2004) also developed a heuristics-based order promising model but with E-commerce environment in mind. They modeled a process where customer orders arrive via Internet and as earliest possible shipment dates are computed in real-time and is promised to customers.

All the previous work described above deal with either push-side of ATP or pull-side of ATP, but not together. There have not been any quantitative tool that looks at both the push and pull-side simultaneously as well as other dynamic factors in supply chain, and evaluates the effectiveness of the overall availability management process. Some of the work described above use simulation experiments to measure the effectiveness of their solutions, but their simulation work was only capable of simulating very specific supply chain environment, focusing only one aspect of ATP process.

In this paper, we describe a simulation work that evaluates how changes in ATP generation impact supply chain performance by simulating all three parts of the availability management (generating availability outlook, scheduling customer orders, and fulfilling the order).

The rest of paper is organized as follows. In section 2, we describe an availability management process in an IBM's hardware business which we conducted the simulation study for. In section 3, we describe the simulation study done for changes in ATP generation, its impacts and results. Section 4 provides conclusion and remarks.

2 AVAILABILITY MANAGEMENT PROCESS

In IBM hardware businesses, the availability management consists of three main tasks: (1) generating availability outlook, (2) scheduling customer orders against the availability outlook, and (3) fulfilling the orders. The business that we analyzed in this study is CCHW (Complex Configured Hardware) business, which manufactures rather expensive, server-type computers.

For the CCHW business, customers place orders in advance of their actual needs, often a few months in advance. Typically, CCHW customers place orders as early as 3 months before the requested delivery (due) dates, and early delivery and payment are not allowed. For this environment, products usually consist of a hierarchy of complex components, and require a longer supply planning. Many buyers in this environment purchase products based on a careful financial planning, and they typically know when they want to receive the products and make payment. Customer orders in this environment are typically highly skewed toward the end of quarter, e.g, only a small portion of orders are placed in the first week of a quarter, and the orders gradual increase, and finally as much as 60-70% of orders are placed in the last 2 weeks of a quarter.

Generation of Availability Outlook, is a push-side of the availability management process, and it pre-allocates ATP quantities, and prepare searchable availability database for promising future customer orders. For the CCHW business, the availability outlook is allocated by weekly buckets, and the availability is planned in much longer horizon, often a quarter (3 months) into the future. ATP quantity is also called Availability Outlook for this reason. The ATP quantity is typically generated based on product type, demand classes, supply classes, and outlook time buckets. The product type can be finished goods (FG) level for Make-to-Stock (MTS) business or components (Comp) level for Make-to-Order (MTO) or Configured-to-Order (CTO) business. Demand classes can be geographic sales locations, sales channels, customer priority, sensitivity to delivery dates, profitability and demand quantity. Supply classes can be degree of constraints and value of products. Outlook time buckets are typically in weekly buckets. Availability is pre-allocated into ATP bucket based on the dimension described above, and rolled-forward daily or weekly. The ATP quantity is determined based on the availability of components, finished goods, WIP (Work-In-Process), MPS (master production schedule), supplier commitment, and production capacity/flexibility. When customer orders arrive, ATP is searched in various ways according to scheduling polices to determine the ship (delivery) date that can be promised to customers.

Customer Order Scheduling is a pull-side of availability management, and it reacts to customer orders and determines ship date for the orders. The CCHW customers usually request orders to be shipped (or delivered) in specified future dates. And they would like to know whether the requested due date can be met or how long is the delay if the due (requested) date can't be met. Customer orders arrive with various information such as product types, the demand classes, customer classes and due dates. The order scheduler then searches through the availability outlook database, and identifies the availability that meets the characteristics. The scheduling can also be done by an ATP engine that uses certain algorithm to optimize the scheduling considering various resources, policies and constraints. The scheduler then reserves specific availability against each order, and decrements the availability according to the purchase quantity of the order. The ship date of the order is determined from the time bucket where the availability reserved, and it is promised to customers. Depending on the business environment, various rules and policies are applied in this order scheduling process. Examples are first-come-first-served policy, customer priority-based scheduling, and revenue (or profit)-based scheduling etc. In a constraint environment, certain ceiling can also be imposed to make sure the products are strategically distributed to various demand classes.

Order fulfillment is executing the shipment of the product at the time of promised ship date. Even if an order is scheduled with a specific promised ship date based on the availability outlook, the availability (ATP quantity) may not actually exist when the ship date comes. There are several reasons why the orders cannot be fulfilled at the promised date. One such reason is the quality of availability outlook generation. In CTO environment, availability outlook is often generated based on finished goods availability, which is estimated based on supplier commitment on components and forecasted configuration of the finished goods. Since the component availability changes often and there is certain error in configuration forecast, the components that are required to assemble a certain finished good may not be available when it is time ship the product to customer. Another source for the fulfillment problem is due to IT system that supports the availability management process. The order scheduling is done based on the availability outlook data in an IT system, which is typically refreshed periodically since it is very expensive to update the database in real time. The availability information kept in the IT system (system availability) are not always synchronized with the actual availability (physical availability). Due to the potentially inaccurate view of the availability, unrealistic ship date can be promised to customer. Therefore, for certain customer orders the necessary ATP quantity may not be there when the promised ship date arrives, thus creating dissatisfied customers. The impact of IT on the fulfillment is discussed in detail by Lee (2006). Therefore, a key role for effective availability management process is to coordinate and balance the push-side and pull-side of ATP as well as IT resources. In this paper, we studied how the push-side ATP would affect the overall availability management process.

3 SIMULATING IMPACT OF ATP GENERATION

For this study, we analyzed a situation where, one of IBM's hardware businesses was interested in managing availability based on new demand class, and they didn't know how the new demand class would impact their supply chain performance, specifically on their customer services and inventory cost. The business wanted to change from a demand class#1 representing 4 geographic demand regions to a new demand class#2 representing 8 new geographical demand regions. For this case, we developed a simulation model to evaluate the impact of the demand class change on supply chain performance. We modeled and simulated 4 different scenarios based on different ways of availability allocation and order scheduling as shown in Table 1.

Scenario 1 is the old (As-Is) availability management process, where availability outlook is allocated based on 19 Product Types, 4 Sources of Supply, 4 elements of Demand Class#1 and 13 Weekly buckets. When an order is generated, the order is assigned with attributes, e.g., a product type, a source of supply, a demand class and the customer requested ship date (also called due date). For the scenario 1, the simulation model tries to schedule each order by searching for availability for a specific product, a source of supply and a demand class, and then the weekly bucket that corresponds to the customer requested ship date. If no availability is found, the model goes back to earlier weekly buckets until it find the availability. If availability is still not found, the simulation model looks for available in later weeks until it finds the availability. If no availability is found in any of 13 weekly buckets, the order is considered backlogged. For this case study, we simulated more than 100,000 orders which represent customer orders for the business for a year. From the simulation, we estimated the customer services and inventory holding costs.

Table 1: Four Simulated Scenarios

	Allocation of ATP	**Constraint on Order Scheduling**
Scenario 1 (As-Is)	Product Type (19) Source of Supply (4) Demand Class1 (4) Weekly Buckets (13)	No constraint
Scenario 2 (To-Be 1)	Product Type (19) Source of Supply (4) Demand Class2 (8) Weekly Buckets (13)	No constraint
Scenario 3 (To-Be 2)	Product Type (19) Source of Supply a(4) Weekly Buckets (13)	Ceiling imposed by Product Type, Demand Class2 and Quarter
Scenario 4 (To-Be 3)	Product Type (19) Source of Supply (4) Weekly Buckets (13)	No constraint

Scenario 2 is the new (To-Be) availability management process that the business would like to evaluate. For this scenario, availability outlook is generated based on 19 Product Types, 4 Sources of Supply and 13 Weekly buckets. But, in addition, it is generated based on 8 elements of Demand Class#2, which represent new geographic demand regions.

Scenario 3 is another new (To-Be) availability management process that the business would like to evaluate. For this scenario, availability outlook is generated based on 19 Product Types (19), 4 Sources of Supply (4) and 13 Weekly buckets. It is not generated based on neither Demand Clas#1 nor Demand Class#2. However, in this case a constraint is imposed when scheduling order. The constraint is a ceiling, which is a maximum allowed quantity for scheduling a specific product type and a specific demand class#2. The ceiling is usually imposed with a predetermined flexibility, 2% etc.

Scenario 4 is another new (To-Be) availability management process that is similar to the scenario 3, but there isn't any ceiling imposed for the scheduling.

For some of key data used in the simulation model are as follows. Customer orders are highly skewed toward the end of 13 week period. The number of orders in the first week of the quarter starts with about 4% of quarterly volume, gradually increases, and for last two week of the quarter the number of weekly order goes up to about 15% of quarterly orders. In addition to the weekly skew of orders, the weekly demand itself has a variability. The variability of component supply is also modeled. The customer requested ship date (due date) is also skewed in that a large portion of orders arriving early part of the quarter request orders to be shipped latter part of the quarter, and the orders arriving in the latter part of the quarter request the orders to be shipped within a few weeks before the end of the current quarter.

One of the key performance metrics we wanted to measure for this study was scheduling delay. For this business, customer orders come with requested arrival dates (due date). Since the transportation lead time is known in advance based on the service level agreement with carriers, it is easy to figure out when the order should be shipped (requested ship date) so that the product arrives at customer's place on the requested arrival date. The scheduling delay here, therefore, is defined as the difference between scheduled ship date and requested ship date. The figures 1, 2, 3, 4 show the scheduling delays for the four scenarios for one product type. It is clear to see in the figure 1 and 2 that the scheduling delay gets worse when the demand class is changed from one that has less members (Demand Class#1) to one that has more members (Demand Class#2). This is obvious because when availability buckets are bigger it is easier to schedule orders against them than when the availability buckets are smaller. As it can be seen in the Figure 3, the scheduling delay is substantially reduced when the demand class is dropped from the availability allocation. However, the ceiling creates significant constraint in scheduling toward the end of quarter. Obviously when the ceiling is dropped (Figure 4) the scheduling delay at the end of quarter disappears. The scheduling delays for the four scenarios are summarized in Table 2.

Another the key performance metrics for this case study was inventory holding cost. We assumed here that the holding a product for one year costs 20% of the sales value. Table 3 compares inventory holding costs of the four scenarios. The scenario 2 would cost $2.827 million more than the scenario 2 (As-Is). However, the scenario 3 and 4 would generate a substantial saving as compared with the As-Is scenario, $3.730 million and $4.462 million respectively. According to the simulation results shown below, the scenario 3 and 4 appear to be good candidates for ATP generation methods, and the business is evaluating feasibility of implementing the scenarios.

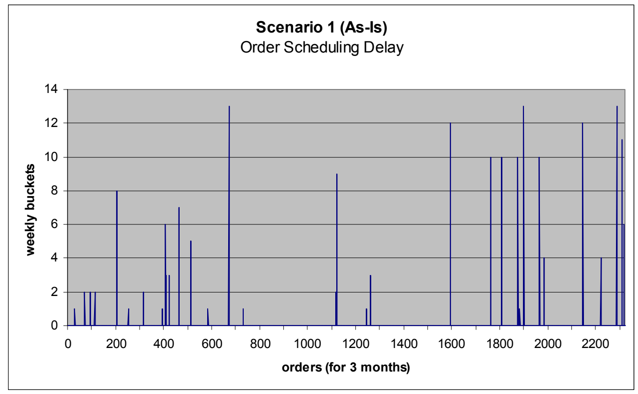

Figure 1: Order Scheduling Delay of Scenario 1 (As-Is)

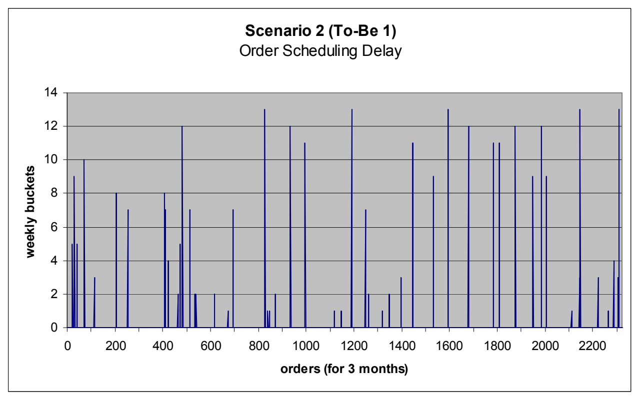

Figure 2: Order Scheduling Delay of Scenario 2 (To-Be 1)

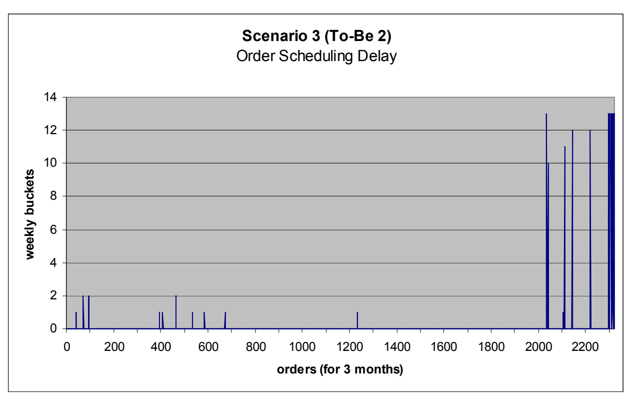

Figure 3: Order Scheduling Delay of Scenario 3 (To-Be 2)

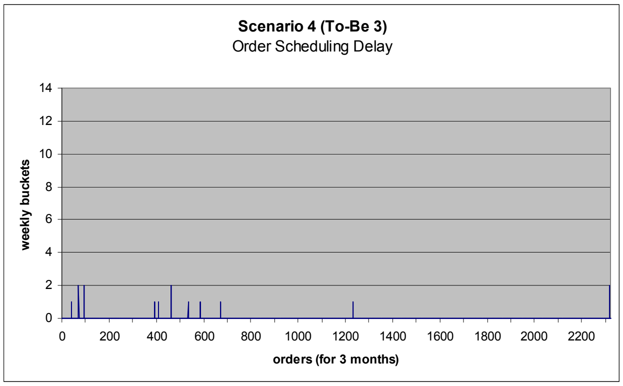

Figure 4: Order Scheduling Delay of Scenario 4 (To-Be 3)

Table 2: Order Scheduling Delay for 4 Scenarios

Order Scheduling Delay	Sce.1: As-Is	Sce.2: To-Be1	Sce.3: To-Be2	Sce.4: To-Be3
Week 0	72.10%	70.74%	78.25%	78.26%
Week 1	12.25%	11.57%	10.38%	10.42%
Week 2	4.64%	4.85%	2.73%	2.74%
Week 3	2.71%	2.99%	2.66%	2.70%
Week 4	2.87%	2.97%	3.03%	3.18%
Week 5	2.18%	2.04%	1.50%	1.61%
Week 6	1.33%	1.23%	0.57%	0.75%
Week 7	0.62%	0.78%	0.16%	0.19%
Week 8	0.14%	0.59%	0.03%	0.02%
Week 9	0.12%	0.28%	0.02%	0.01%
Week 10	0.12%	0.25%	0.05%	0.03%
Week 11	0.17%	0.33%	0.09%	0.02%
Week 12	0.23%	0.46%	0.11%	0.03%
> Week12	0.52%	0.95%	0.41%	0.04%

Table 3. Inventory Holding Costs for 4 Scenarios

	Sce.1: As-Is	Sce.2: To-Be1	Sce.3: To-Be2	Sce.4: To-Be3
Inventory Holding Cost	$13.135 million	$15.962 million	$9.405 million	$8.673 million
Inventory Holding Cost Saving (wrt As-Is)	--	-$2.827 million	$3.730 million	$4.462 million

4 CONCLUDING REMARKS

ATP generation, as a part of availability management process, directly influences key supply chain performance such as customer services and inventory. Simulation is a very useful tool to estimate how different ATP generation method would affect impact the supply chain performance. In this paper, we described a simulation work that was developed for IBM's computer hardware business to evaluate various alternatives in ATP generation method. The model simultaneously simulates the three main components of availability management process; generating availability outlook, scheduling customer orders and fulfilling the orders, as well as the effect of other dynamics in the supply chain. The simulation study has been useful in making important decision on ATP generation methods.

ACKNOWLEDGMENT

The author would like to thank Brad Howland, Joe DeMarco, and Barun Gupta of IBM Integrated Supply Chain (ISC) group for sharing their knowledge and experience in IBM's availability management processes.

REFERENCES

Ball, M.O., C-Y Chen, Z-Y. Zhao (2004), "Available To Promise", Handbook of Quantitative Analysis: Modeling in an-E-Business Era, eds. D. Simchi-Levi, D. Wu, Z.M. Shen, Kluwer, 446-483.

Chen, C-Y, Z. Zhao, M. Ball (2002), "A model for batch advanced available-to-promise" Production and Operations Management, 11-4, 424-440.

Ervolina, T., B. Dietrich (2001). "Moving Toward Dyanamic Available to Promise," in Supply Chain Management: Status and Future Research, S. Gass and A. Jones (eds) preprints produced by R. H. Smith School of Business, U. of MD and Manufacturing Eng. Lab., NIST.

Lee, Y. M. (2006), "Analyzing the Effectiveness of Availability Management Process", IBM Research Report, rc23903 (W0603-041).

Pan, Y., L. Shi (2004), "A Stochastic On-Line Model for Shipment Date Quoting with On-Line Delivery Gurantees", Proceedings of the 2004 Winter Simulation Conference, Eds. Ingalls, R.G., M.D. Rossetti, J. S. Smith, and B. A. Peters.

S. Moses, H. Grand, L. Gruenwald, S. Pulat (2004), "Real-time due-date promising by build-to-order environments, International Journal of Production Research, 42-20, 4353-4375.

YOUNG M. LEE works in the mathematical science department of IBM's T.J. Watson Research Center, U.S.A., in the areas of supply chain simulation and optimization. Prior to joining IBM, he had worked for BASF, where he had founded and managed the Mathematical Modeling Group, and led development of numerous optimization and simulation models for various logistics and manufacturing processes. Recently at IBM, he developed several complex simulation models that are instrumental in analyzing and improving business processes, supply chain and IT solutions. He has a B.S., a M.S., and a Ph.D. degree in Chemical Engineering from Columbia University in the City of New York. His research interest includes simulation and optimization of supply chain, manufacturing and business processes. His email address is <ymlee@us.ibm.com>.