Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

DISCRETE EVENT SIMULATION MODELING OF RESOURCE PLANNING AND SERVICE ORDER EXECUTION FOR SERVICE BUSINESSES

Young M. Lee
Lianjun An
Sugato Bagchi
Daniel Connors
Shubir Kapoor
Kaan Katircioglu

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

Wei Wang
Jing Xu

IBM China Research Lab
ZHONG GUANG CUN SOFTWARE PARK
Beijing, 100094, China

ABSTRACT

In this paper, we present a framework for developing discrete-event simulation models for resource-intensive service businesses. The models simulate interactions of activities of demand planning of service engagements, supply planning of human resources, attrition of resources, termination of resources and execution of service orders to estimate business performance of resource-intensive service businesses. The models estimate serviceability, costs, revenue, profit and quality of service businesses. The models are also used in evaluating effectiveness of various resource management analytics and policies. The framework is aided by an information meta-model, which componentizes modeling objects of service businesses and allows effective integration of the components.

1 INTRODUCTION

Service businesses, such as business consulting, call centers, technical services and IT outsourcing, are growing rapidly (Dietrich et al. 2006), and have become a significant portion of the U.S. and world economy. Process modeling and simulation have been used for many years to analyze the performance of manufacturing processes, supply chain processes and other business processes (see Bagchi et al. 1998 for example). However, it is difficult to apply those analytic tools developed for manufacturing and supply chain to service businesses, because service businesses are quite different from the other businesses and are also complex. Therefore, there have not been many structured modeling methods for analyzing the performance of service businesses. Unlike most supply chains, the service (human) resource is perishable, and the skill set of resources is diverse and inexact. The skill levels also change through training and engagement experiences. The resource requirements for service engagements are also complex and inexact; therefore, the bill-of-resources (BOR) is much more difficult to model than a bill-of-materials (BOM) in a manufacturing supply chain. Moreover, the availability of resources can also degrade over time due to attrition and termination of workforce resources. All these factors have made the modeling and simulation of service businesses very difficult. There have been only a handful of simulation studies done for service business, and they focus on a small subset of service businesses. Anderson et al. (1999) used system dynamics modeling to study correlation between capacity planning and backlog for a simplified mortgage approval process. Akkermans et al. (2000) also used system dynamics modeling to study upstream amplification of workload in a service supply chain (telecommunication business). However, there has not been any simulation model that describes the interactions of engagement/resource planning, execution of service orders, as well as resource management in a more detailed and discrete level for resource-intensive service businesses.

In this simulation framework, we model and simulate five major modules of resource-based service businesses; (1) demand planning of service engagements, (2) supply planning of human resources, (3) attrition of resources, (4) termination of resources and (5) execution of service engagement orders using discrete-event simulation (DES) methods. Each of these modules interacts with others through a notion of availability outlook, which is a view of resource availability with respect to specific skill sets, time periods, lines of business, locations, skill levels etc. One example of a resource availability outlook is the number of expert Java Programmers available, in weekly time buckets, in the next 52 weeks; for example, 10 people this week, 12 people next week, 20 people the week after etc. The resource availability outlook is often called the resource availability pool; but we prefer the term outlook here since we are interested in the availability for the current as well as the future periods. The supply planning module increments the resource availability in the availability outlook, while the module of service order execution allocates the available resources to specific service order engagements thus decrementing the availability accordingly. The modules of resource attrition and resource termination also decrement the availability. Therefore, the modules are tightly linked to one another. The performance of service businesses can depends more on the interaction of these five components than each individual one. For example, even if supply planning is done optimally, if the right resource is not allocated to the right engagement by a project manager, the business can suffer. Effectiveness of the availability management process in supply chain has been modeled and simulated by Lee (2007). That work models the effectiveness of availability (materials such as products and components) management on supply chain performance, but does not address service operations, which is different and more complex to model since service operations involve dealing with availability of human resources.

Another difference between service business and manufacturing supply chain is the fact that many more complicated policies are used in service business. For example, the resource acquisition part of the supply planning may have complicated policies on whether and how many resources should be hired or contracted or re-trained. For re-training itself, there can be many policies on who should be re-trained, and how long the training should be. Our simulation framework has a capability of plugging in various polices and analyzing the impact of the policies on the business performance.

The simulation framework can also interact with various optimization models. The supply planning could encompass several optimization problems, such as RCP (Resource Capacity Planning) and RAP (Resource Acquisition Planning). When these optimization models are available, the simulation framework can easily interact with the model, thus evaluating their effects on the overall service business performance.

2 MODELING APPROACH

We model five events that are triggered independently. The demand planning is typically a weekly or monthly event that produces the forecasted demand by engagement types. The supply planning is also typically carried out once a week or month, and uses the demand forecast, supply constraints and other criteria to convert expected future engagements to resources requirements. The outcome of supply planning is added to the resource availability outlook, which is a view of resources into the future by resource attributes such as skill, location and line of business. Resource attrition is a random event that describes the attrition of existing resources. The resource termination event is a human resource related event that can be triggered at certain frequencies to eliminate portions of resources for various reasons such as financial or efficiency. The resource attrition and termination events take away resources from the resource availability outlook. Finally, the engagement execution is an event of a service order arriving to a service business and being fulfilled using the service availability outlook information.

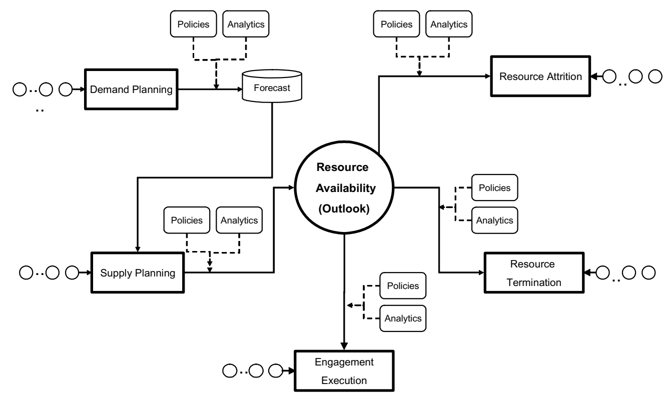

Figure 1: Overview of Simulation Model

We simulate five events (see Figure 1) being triggered and the corresponding activities being carried out. The events interact one another by using and updating information in the resource availability outlook. Each event could use various policies or analytics (to be explained later) in executing its tasks. The effective interaction of the events as well as the performance of individual events are manifested as business performance metrics such as revenue, profit, resource costs, service level as well quality of service business. Our models simulate various scenarios and estimate the performance metrics.

Our simulation models are developed using IBM's business process modeling/simulation software, WBM (http://www-306.ibm.com/software/integration/wbimodeler/) and SmartSCOR (Dong et al. 2006).

3 SIMULATION MODELING MODULES

3.1 Demand Planning of Service Engagements

Demand planning produces forecasted demand by engagement type, also called SPL (service product line), and by engagement start time period (usually in time buckets of weeks or months), duration and also size of engagement. The forecast can be based on historic demand patterns, expected business growth and engagement opportunities in the pipeline. Demand can be forecasted by simple business insights (we call them policies here) or can be computed by various forecasting algorithms such as time-series methods and/or methods based on partially known demand information and lead time. The demand forecast is later used by supply planning as is shown in Figure 2, and is explained in next section.

3.2 Supply Planning of Resources

Supply planning estimates the resource requirements based on demand forecasts and decides the level of resource that should be acquired. One of the activities in supply planning is Resource Capacity Planning (RCP), as shown in Figure 2. RCP breaks engagement forecasts into resource requirement through explosion of BOR (Bill-of-Resources). Resource requirements are computed by skill groups. Therefore, RCP converts engagement forecasts (grouped by SPL and by time periods) into resource requirements (grouped by skill groups and by time periods). RCP can also use information on supply constraints and by imploding the supply against the demand, it can determine how much of the demand can be fulfilled. RCP computation can be done deterministically when the demand forecast is given as a fixed number. When the demand forecast is given with uncertainty, RCP can also compute resource requirements that include resource buffers (safety resource) that protect against the demand variability. The output of RCP is gaps (shortages) and gluts (surpluses) of resource. The gaps and gluts are grouped by SPL and by time periods.

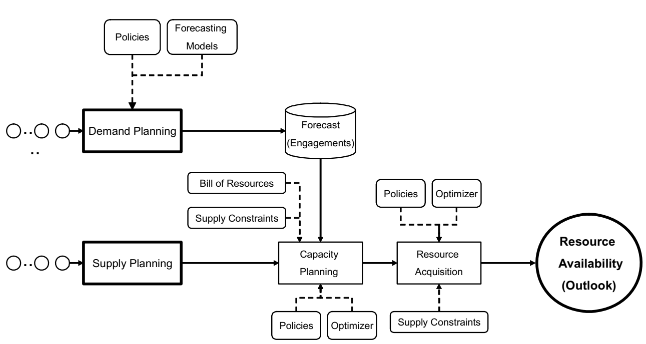

Figure 2: Demand Planning & Supply Planning

The gaps and gluts computed by RCP are used by the resource acquisition activity that decides how to acquire the additional resource. There are several ways to acquire resources; hiring additional resources, contracting temporary resource or retraining existing resources. Hiring new resources (employees) involves one time costs such as advertising, interviewing or relocation, and on-going costs of salary and benefits once the resources are hired. Hiring also takes time. Depending on the skill sets that are sought, it may take a few months to bring some resources and have them be ready to be used for engagements. Once resources are hired, it may not be easy to terminate the resource due to complicated labor agreements and expensive severance payments. Contracting resources from a resource agency may be easier and faster than hiring employees, but typically costs significantly higher per utilization. Re-training existing resources which are not being utilized is another option, but that can take substantially more time depending on which skills the resources are acquiring Engagement managers (or project managers), however, may be reluctant to train someone and prefer hiring or contracting the right resource with right skill sets as soon as possible. Therefore the resource acquisition task is complicated and needs to consider costs, time, and service level. This is in fact a rich optimization problem that can be solved analytically. However, in practice many service firms use rather simple policies to make resource acquisition decisions.

The simulation models that we are developing can simulate many different policies and also interace with optimization models during the simulation runs. The policies can be simple or complicated. We are currently modeling the following polices; (1) hire when needed and hold, (2) consider hiring first, contracting next, (3) consider contracting first, then hiring, (4) consider retraining first then contracting and then hiring etc. Each policy can also specify parameters such as decision threshold on lead time, costs, backlog etc. The resource acquisition optimizer can have an objective function of maximizing revenue (or profit) while satisfying constraints of service level agreement (SLA) and costs etc.

The output of resource acquisition is a plan which specifies which resources would be becoming available at which period and duration (grouped by skill sets and by time periods) The resource acquisition increments the resource availability outlook (see Figure 2).

3.3 Resource Attrition

Resources in a service firm are not stationary, but can be reduced in numbers as a result of resignations, retirements, or deaths. Attrition rates can depend on many factors such as the working environment, salary, age of employees, job market, types of skills etc. The influence of those factors on attrition is intuitively understandable, but the exact correlation is not easy to define. The causal relationship and feedback control mechanism are better suited for system dynamics (SD) approach (Forrester 1961, Sterman 2000). Such system dynamics modeling is not within scope of this paper. However, the resource attrition module of the simulation model, which is a DES model, can interact with a SD model to estimate the dynamics of resource attrition. The output of resource attrition decrements the resource availability outlook (see Figure 3).

3.4 Resource Termination

In resource-oriented service businesses, underutilized resources are directly related to the financial performance of the firms. Resources (employees) are paid their salaries and are provided benefits regardless of whether they are engaged in projects (except a few business, where employees are paid only when they are utilized). In fact, in consulting businesses, employees may be released if they are not utilized (also called 'on the bench') for a period of a few consecutive months. Service businesses usually monitor the bench resources and make periodic decision on whether, how many and which resource should be terminated. Our model simulates the termination event and activities.

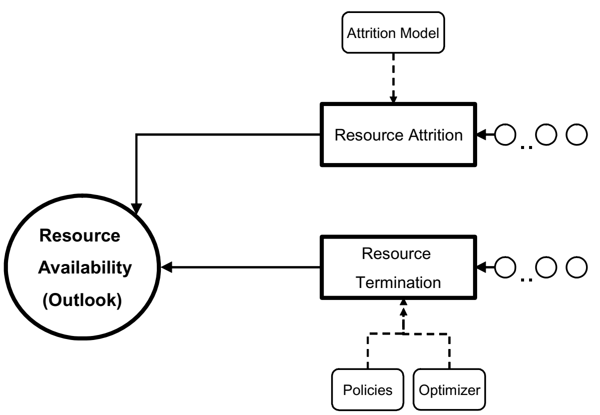

Figure 3: Resource Attrition & Resource Termination

The termination decision is based on many things such as the cost of termination (such as severance payment, union pressure), future prospect of the business, priority of skill sets and the morale of other employees. This decision problem can be formulated as an optimization problem. If such an optimization model is available, the simulation model can interact with it. The termination can also be decided by policies, which can also be simulated in our model. The output of the termination event decrements the resource availability outlook.

3.5 Engagement Execution of Service Orders

The most dynamic module of our simulation models is the event of engagement execution. In this module, we simulate service order arrival, allocating resources to it and fulfillment of the order as shown in Figure 4.

The arrival of service orders are modeled as a Poisson process. Upon arrival, each order can be characterized by assigning attributes such as service type (SPL for example), start date, duration, customer class, revenue, geographic location etc. These attributes are decided by randomly drawing from a distribution function (or histogram) derived from historic data with anticipated future adjustment.

Each service order can be exploded through a Bill-of-Resources (BOR) and the required resources can be allocated one at a time. This is actually a common practice, and project managers or program managers often allocate (or reserve) resources as soon as possible to ensure on-time fulfillment of a particular engagement. However, it may be more efficient to batch up service orders by time interval, such as weekly or daily, and then simultaneously allocate or reserve resources for several service orders thus creating opportunity for optimal allocation of resources. As shown in Figure 4, in the simulation model, the batching frequency, T_b, can be specified as a simulation parameter. Or, the simulation model can also be set up in such a way that each order is processed as soon as it arrives.

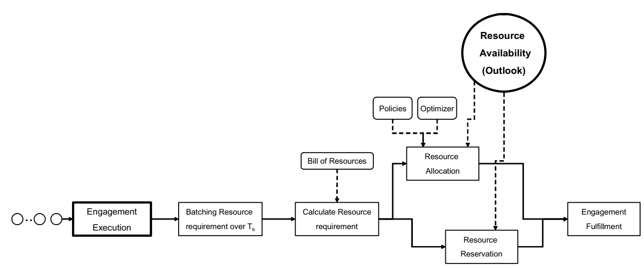

Figure 4: Engagement Execution of Service Order

The batched service orders are exploded together against the BOR, and appropriate resources are identified using the resource availability outlook. There are typically two ways of identifying resources for service orders. The first one is resource allocation. This is a firm allocation of a specific resources against a service order that is due to start within the lead time of the resource acquisition. If a specific resource is not available for this firm allocation, the fulfillment of this service order is either backlogged (start date is delayed) or goes away as lost sales opportunity. The second one is resource reservation against service orders. For service orders that are due to start beyond lead time (e.g, there is enough time to acquire the required resources), resources are loosely reserved rather than firmly allocated. This potentially allows other service orders access to the reserved resources for allocation if necessary. If there aren't enough resources to be reserved for certain service orders, the reservation reservation is computed and forwarded to the demand planning module, which will include this as a part of demand forecast, and the service order would be on-hold until after the next demand planning cycle, and re-scheduled. At each batching cycle, the reserved resources are converted to allocated resources and the service order is firmly scheduled.

Decisions on resource allocations and reservations may be made based on different polices. One example of such policies could be a first-come first served (FCFS), which is, in fact, a commonly used policy. Another example could be a policy with resource substitution option. For this policy, if the required (primary) resource is not available, alternative (secondary) resources are sought and allocated. Other more complicated ones could be a rationing policy, which tries to distribute resources across various classes of service orders such as customers, geographic locations or time periods. Others policies may be based on revenue, profit, or duration. The policy can also be designed to develop desirable skill sets for the future, and so the policy may prefer to allocate resource with less experience so that a firm can build up a resource pool with many highly skilled personnel for the future.

The engagement execution decrements the resource availability outlook when it allocates resources, and releases the resources back to the availability outlook with higher skill levels (though the engagement experience), when the engagement fulfillment is completed.

The performance metrics that the simulation model generates include revenue, profit, profile of bench resources, fulfillment delay, backlog, lost sales opportunity and resource utilization. When each engagement execution is simulated, the engagement is recorded with simulated information such as arrival time, requested start date, resource allocated, actual start date, completion date, resource costs, revenue, profit etc. At the end of a simulation run, covering for example for one year, the information of all the simulated engagements are summarized to compute the performance metrics above.

4 SIMULATION EXPERIMENTS & RESULTS

Using the framework we described above, we develop a simplified resource management simulation model, and conducted simulation experiments. This simulation model will later become a detailed model for analyzing many scenarios for many different business environments. In this paper, we report only our early simulation results to demonstrate the usefulness of the simulation model. More comprehensive results will be reported as they become available.

We report simulation results of two experiments. These experiments are for two different resource allocation policies. The first experiment is for a simple FCFS (First-Come First-Serve) resource allocation policy. For this policy, when an engagement comes in, the engagement manager identifies the resource requirement by exploding through the BOR, looks for the required resource in the availability outlook, and allocates the resources if they are found in the order of the engagement arrival. If all the required resources are not found, then the engagement is treated to be lost. From the simulation result, the benche resource level is plotted in Figure 5, and the revenue (cumulative) generated is plotted in Figure 7.

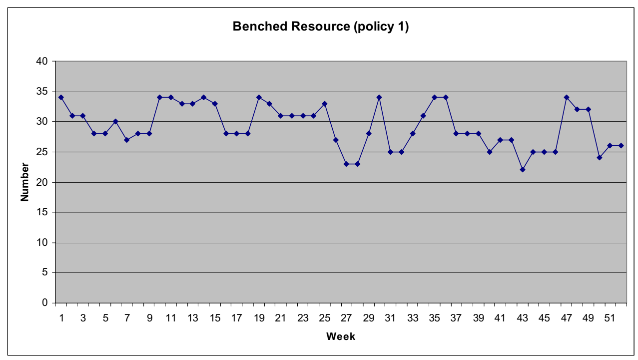

Figure 5: Bench Resource Profile from Policy1

The second experiment is for a resource allocation policy with resource substitution option. For this policy, the required resource is sought in the availability outlook in the order of engagement arrival. However, if the required (primary) resource is not available, alternative (secondary) resources are sought. The engagement is treated to be lost only if neither primary nor secondary is found. The bench resource is plotted in Figure 6. The bench resource level is substantially lower than the FCFS policy (policy 1, Figure 5). In this example, the average bench resource for the FCFS policy is 29 people, while that of the substitution policy is 26 people (10 % less). The reduction could mean a significant financial impact of a firm. The revenue generated from this policy is plotted in Figure 7 along with one from the FCFS policy. Substantially higher revenue is shown from the simulation. The total revenue generated from the policy 1 is \$8.2 million while that of the policy 2 is \$14 million (70% increase).

The purpose of the simulation results presented here is not to show the specific numerical figures representing business impact of the particular simulation setting, but rather to demonstrate that the simulation model can be very useful in simulating service businesses and estimating business performance metrics.

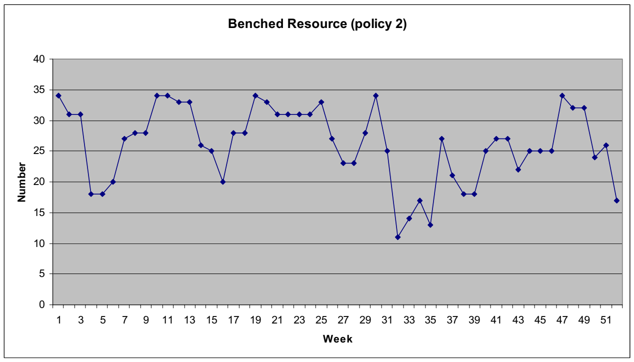

Figure 6: Bench Resource Profile from Policy2

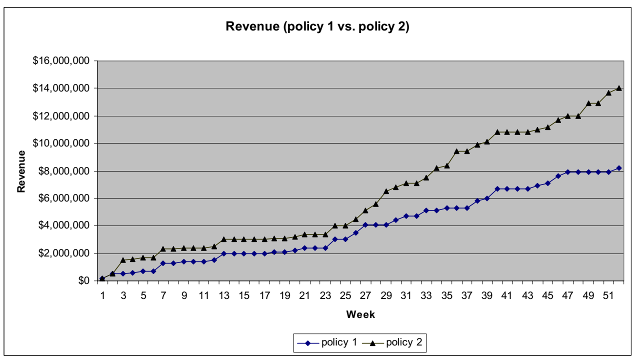

Figure 7: Revenue Profile for Policy1 & Policy2

5 CONCLUDING REMARKS

The management of (human) resources for resource-intensive service businesses are much more complicated than the management of resources (materials and equipment) for manufacturing or supply chain businesses. Therefore, a tool for modeling and simulating the service business to estimate the effectiveness of such businesses would be very useful. We summarize an approach of developing a simulation framework that can be used as a testbed for evaluating various components of the resource management processes, policies and analytics. We demonstrate the usefulness of the simulation tool by presenting sample simulation results. We hope to extend the research to come up with a comprehensive simulation framework that can be used to set up a simulation model for various resource-intensive service businesses so that the service businesses are better understood and improved.

ACKNOWLEDGMENTS

The authors would like to thank Steve Buckley, Dick Lam and Lea Deleris for many useful discussions and suggestions.

REFERENCES

Anderson, E., Morrice, D., 1999. A Simulation Model To Study The Dynamics in a Service-Oriented Supply Chain, In *Proceeding of the 1999 Winter Simulation Conference*.

Akkermans, H, Vos, B., 2000. Amplification in Service Supply Chains: An Exploratory Case Study from the Telecome Industry, *Production and Operations Management*, 2000.

Bagchi, S., Buckley, S., Ettl, N., Lin, G. 1998. Experience Using the IBM Supply Chain Simulator, In *Proceedings of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

Dietrich, B., Harrison, T., 2006. Servicing the Services, *OR/MS Today*, June 2006.

Dong, L., Ding, H., Ren, C., Wang, W., 2006. IBM SMARTSCOR - A SCOR Based Supply Chain Transformation Platform Through Simulation and Optimization Techniques, In *Proceedings of the 2006 Winter Simulation Conference*.

Forrester, J. W., 1961. Industry Dynamics, MIT Press, Cambridge, MA

Lee, Y. M. 2007. Analyzing the Effectiveness of Availability Management Process. In Trends in Supply Chain Design and Management: Technologies and Methodologies. Jung, H., Chen, F.F., Jeong, B., (eds.). Springer.

Sterman, J. D., 2000. Business Dynamics: System Thinking and Modeling for a Complex World, Irwin McGraw-Hill, Boston.

AUTHOR BIOGRAPHIES

YOUNG LEE is a member of the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His e-mail address is <ymlee@us.ibm.com>.

LIANJUN AN is a member of the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His e-mail address is <alianjun@us.ibm.com>.

SUGATO BAGCHI is a member in the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His research interests are in the use of simulation for operational decision-making in various industries..

DANIEL CONNORS is a member of the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His e-mail address is <dconnors@us.ibm.com>.

SHUBIR KAPOOR is a member of the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His e-mail address is < shubirk@us.ibm.com >.

KAAN KATIRCIOGLU is a member of the Mathematical Sciences department at the I.B.M. T.J. Watson Research Center in Yorktown Heights, NY. His e-mail address is < kaan@us.ibm.com >.

WEI WANG is a member of the I.B.M. China Research Lab in Beijing. His e-mail address is < wangwcrl@cn.ibm.com>.

JING XU is a member of the I.B.M. China Research Lab in Beijing. His e-mail address is < xujbj@cn.ibm.com>.