Research Center
Finance & Information Management

Fraunhofer
FIT

Project Group
Business & Information
Systems Engineering

Integrating Business Partners on Demand: The Effect on Capacity Planning for Cost Driven Support Processes

by

Christoph Dorsch, Björn Häckel

Europäische Union
„Investition in Ihre Zukunft"
Europäischer Fonds für
regionale Entwicklung

in: Proceedings of the 45th Hawaii International Conference on System Sciences,
Maui, 2012, pp. 4796-4805

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: F.-v.-Schiller-Str. 2a, 95444 Bayreuth
Phone: +49 921 55-4710 (Fax: -844710)

Integrating Business Partners On Demand: The Effect on Capacity Planning for Cost Driven Support Processes

Christoph Dorsch
University of Augsburg, FIM Research Center
christoph.dorsch@wiwi.uni-augsburg.de

Björn Häckel
University of Augsburg, FIM Research Center
bjoern.haeckel@wiwi.uni-augsburg.de

Abstract

Capacity planning for business processes is still a major challenge. Technology enabling on demand integration of business partners handling peak load at short notice may help. We model a three-stage supply chain with a service provider receiving demand from its customers usually executed by an in-house unit. Alternatively the demand can be routed to external business partners if these offer capacity and a preferable execution. Then the capacity planning problem of the service provider is examined: In what way does this "on demand integration capability" affect the optimal level of in-house capacity? The model consists of two separate queuing systems (in-house unit and external business partners) along with their relevant cost functions. Furthermore a routing algorithm is developed. It evaluates both systems for every incoming order and decides about routing to the preferable one. To derive results a discrete-event simulation is necessary performed within a case study of the securities trading and settlement process. Three insights are gained: There are situations where on demand integration leads to reduced optimal capacity. Furthermore the risk of allocating an inappropriate amount of capacity can be reduced as well as the total operating costs of the business process.

1. Introduction

As every business activity and every business model is implemented by specific business processes [6], a value oriented management [16] of business processes is essential for a company's long-term success. Especially the ex ante planning of capacity (e.g. in terms of personnel or IT-capacity) to be assigned to a specific business process is a widely recognized challenge. This holds true for companies which offer manly IT driven services, too. While state-of-the-art IT-platforms are widely scalable, necessary manual interventions require resources which have to be scheduled ex ante [5]. Determining an appropriate level of capacity ex ante is a crucial task, especially for business processes characterized by highly volatile demand, non adjustable capacity in short-term and time critical execution due to customer needs [1, 5]. Seeking to optimize the level of assigned capacity for such business processes, a company is faced with a trade-off [13, 19]: Assigning a high level of capacity allows the buffering of temporarily peaks in customer demand but results in idle costs in time frames of low demand. Assigning less capacity avoids idle costs but results in rising waiting times in time frames of high demand. If the execution is time critical rising waiting times go along with waiting costs, e. g. caused by the violation of service level agreements (SLA). Thus, a company faced with this trade off aims at minimizing the total operating costs (consisting of idle and waiting costs) of a business process by choosing an appropriate level of capacity [4]. This is especially the case for a cost-driven support process, where costs are a central competitive factor.

The trade-off described may be mitigated at least to a certain extent by the opportunities arising with new developments in information technology. With the growing diffusion of service-oriented infrastructures suitable for the integration of web services as well as corresponding description languages (e. g. WSDL) or standards for data exchange (e. g. XML, EDIFACT) a simple and fast integration of business partners has become considerably easier. Especially for highly standardized IT-driven support processes even the on demand integration of external service providers meanwhile is a feasible alternative. Thereby on demand integration means that new business relations can be established (nearly) without any loss of time by building up links fast and cheap. This on demand integration of external service providers is closely related to the concept of "Business Process as a Service (BPaaS)". BPaaS recently has gained high attention going along with the increasing market penetration of innovations in IT like Cloud Computing: In a survey of Vehlow and Golkowsy [20] already 27% of the providers of cloud computing services stated to offer BPaaS. Within the concept of BPaaS services for standardized business activities such as e. g importing high data volumes out of databases are also offered as services tailored to specific industries as e. g. life insurance origination fulfillment services.

The possible on demand integration of external service providers may offer substantial benefits for companies. As with help of on demand integration companies are able to react more quickly on peaks in customer demand by routing excessive demand to external service providers, they may reduce the risk of choosing an inappropriate level of capacity.

Considering business processes with highly volatile demand, however, this potential economic benefit might be lowered, as external service providers in this case usually will not commit a SLA at reasonable costs for the on demand supply of capacities: To avoid high contractual penalties caused by the violation of SLA commitments for on demand supply an external service provider would have to provide high reserve capacities causing significant fixed costs [14]. These high fixed costs of external service providers will usually result in prohibitive high SLA costs for the buyer. Consequently, a company that wants to route excessive demand to external service providers by buying not SLA backed capacities on demand, faces the risk of only getting served as soon as capacity is available. That might cause delays for external routed demand. Thus, when optimizing the assigned capacity for a specific business process a company has to consider these possible delays and to weigh the risk of waiting times at the external service provider's market against those waiting times of in-house execution.

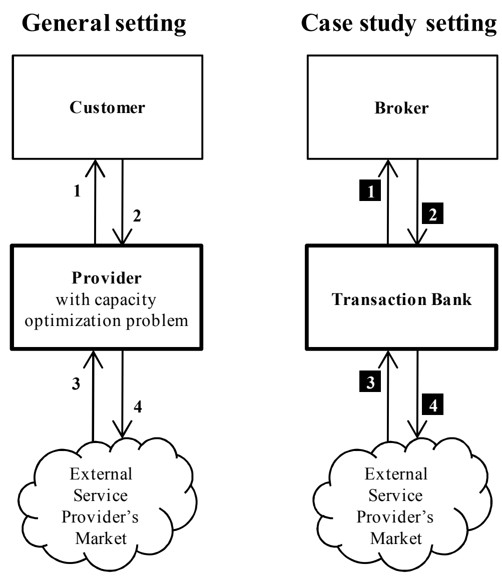
As the weighting of these effects is not a trivial task due to the stochastic nature of the capacity planning problem, in this paper we focus on the following central research question:

In what way does on demand integration capability affect the optimal level of in-house capacity assigned to a specific business process?

For answering this research question, we investigate the setting of a three-stage supply chain outlined in the left column of figure 1. Therefore we first analyze related literature and then develop an optimization model using queuing theory. The model consists of two separate queuing systems (in-house unit and external business partners) along with relevant cost functions. Furthermore a routing algorithm is developed. It evaluates both systems for every incoming order and decides about routing to the preferable one. To derive results a discrete-event simulation is necessary performed within a case study of the securities trading and settlement process (right column of figure 1). Finally, we summarize the key findings of the paper and give an outlook on prospective further research.

2. Literature Review

We review literature relevant for answering our main research question, namely to analyze the impacts of on demand integration capability on in-house capacity planning. As on demand integration capability represents a certain kind of flexibility in capacity planning (the flexibility to route excessive demand to external service providers), we will in particular take a closer look on literature dealing with aspects of flexibility in the context of capacity planning. Due to our quantitative optimization approach we will focus on quantitative oriented literature.



1 provides a service *with* committed service level agreement (SLA)

2 uses the provided service characterized by volatile demand

3 provides a support process "on demand" (*without* committed SLA) when capacity is available (in addition to the in-house unit of the provider with capacity optimization problem)

4 uses the support process if external service providers enable a preferable execution

1 provides the whole securities trading and settlement process *with* committed SLA

2 routes all securities orders to be executed (characterized by volatile arrival rates)

3 provides the necessary manual interventions on orders "on demand" (*without* committed SLA) when capacity is available

4 routes peak loads if manual interventions on orders at the external service provider's market are less expensive than within the corresponding in-house unit

Figure 1. Model setting of the three stage supply chain considered

Effects of flexibility with respect to capacity planning have especially gained high attention in the areas of production management and supply chain management (SCM). So a range of papers addresses the problem of optimal mixing dedicated and flexible manufacturing capacities. For this purpose the paper of Bassamboo et al. [3] studies the basic problem of capacity and flexible technology with a newsvendor network model. The authors consider a multiproduct firm and deal with the question, if different products should share resources or if the firm should establish dedicated resources for some of them. Tomlin and Wang [18] connect the mix-flexibility and dual-sourcing literature by studying unreliable supply chains that produce multiple products. Like in Bassamboo et al. the authors consider a firm that can invest in product-dedicated resources and totally flexible resources. Netessine et al. [15] determine the optimal mix of the different types of capacity considering the effects of increasing demand correlation. Analyzing the optimal mix of different kinds of capacity, however, none of these papers considers a possible on demand integration of external capacity and its effects on in-house capacity. Instead, the analysis in these papers is restricted to the in-house capacities of the firm.

Though, there are several papers taking into account aspects of flexibility in external supplier markets. In this context Tomlin [17] analyzes the effect of volume flexibility of suppliers on the sourcing strategy of a firm. For this, the paper studies a single-product setting in which a firm can source either from an unreliable but cheaper or from a reliable but more expensive supplier. Furthermore the reliable supplier may possess volume flexibility. The author shows, that contingent rerouting may constitute a possible tactic if a supplier can ramp up its processing capacity, that is, if it has volume flexibility. Dong and Durbin [9] study markets for surplus components, which allow manufacturers with excess component inventory to sell to firms with a shortage. The paper is motivated by recent developments in internet commerce, which have the potential to greatly increase the efficiency of such markets. Dong and Durbin derive conditions on demand uncertainty that determine whether a surplus market will increase or decrease supplier profits. Another paper dealing with flexibility of supplier markets is that of Lee and Whang [12]. Within this paper the authors investigate the impacts of a secondary market, where resellers can buy and sell excess inventories. For this, the authors develop a two-period model with a single manufacturer and many resellers. The authors derive optimal decisions for the resellers regarding their ordering policies and analyze the effects of the secondary market both on the sales of the manufacturer and the supply chain performance.

The last-named papers are closely related to our approach regarding the basic idea of a market, where firms with a shortage of capacity or inventory can buy available overcapacities or excess inventories from other firms. However, there are two fundamental differences between our approach and those outlined:

First, the papers mentioned above focus on physical and thus storable products. Hence, the named papers are more concerned with the possible trading of (physical) excess inventories and its implications on capacity planning. Thereby a time critical delivery time of products is not considered. In contrast, we focus on cost-driven business processes with a time critical execution and thus investigate the potential benefits of on demand usage of external service providers offering non-storable services.

Second, considering this on demand context a company has to decide for each incoming order whether to route it internally or externally. This requires a routing algorithm weighting costs of the internal and external execution for each order. As current literature does not consider the flexibility of on demand capacity, approaches for continuous order routing were not focused so far. Thus, the development of an algorithm for continuous order routing constitutes a further distinct difference to existing capacity planning literature. Thereby the proposed algorithm in particular takes into account the risk of time delays when routing demand externally.

To sum up, the effects of on demand integration capability on capacity planning for cost-driven business processes are only sparsely considered in literature. The objective of this paper is to contribute to the closure of this research gap.

3. Model formulation

To model the three-stage supply chain, we first specify the underlying optimization problem resulting from the volatile arrival rates of time-critical orders. Then we describe the in-house unit executing the necessary support process and the external service provider’s market along with the on demand integration capability. Finally we complete the model with the relevant cost-functions and an order routing algorithm necessary to analyze the model.

3.1. Capacity optimization problem

The provider considered offers an IT driven service to its customers. Each incoming order triggers the corresponding business process and the provider executes all necessary activities. This business process is widely automated using a specialized IT-platform. Nevertheless there are some activities which require manual interventions. These manual activities are executed by an in-house unit with limited execution capacity. As the IT-platform used for the business process is scalable for all possible amount of demand, only the manual interventions within the in-house unit might be a bottleneck executing the business process. The arrival rate (λ), i. e. the number of time-critical orders sent from the customers per unit time is random. Based on historical data and contractual agreements respectively the statistical distribution of λ can be approximated. The planning horizon considered is finite and divided into equidistant time units.

The provider has to decide ex ante about the number of orders (y), the in-house unit can execute simultaneously ("capacity"), which minimizes the total operating costs (c) for the business process:

$$\min_{y} c(\lambda, y)$$

Focusing IT driven business processes does not imply full automation where only (scalable) IT resources (e. g. computing power, network connectivity, storage capacity) are necessary. Manual intervention to enrich data, (double) check documents and other in-/outputs or make decisions not reducible to automated rules require employees, workplaces etc. Taking this into account it becomes clear why the decision has to take place ex ante: Even within the same organization it would be difficult to shift this kind of capacity to and from other tasks respectively following the volatile arrival rates (even without inducing analogous problems to these other tasks) as this might be possible with fast scalable IT resources.

3.2. Modeling approach

Concerning the characteristics of the optimization problem, the in-house unit as well as the external service provider's market will be modeled as two separate queuing systems. Therefore we extend the basic assumptions of queuing theory as e. g. described in Gross et al. [10] by parameters and functions necessary to specify the relevant trade-offs described above. The resulting service system for orders requiring manual intervention is outlined in figure 2.

It is obvious to use queuing systems to model this capacity optimization problem of a business process as its main characteristics like random demand and multiple service stations are inherent. Therefore we follow other publications, e. g. Braunwarth et al. [6], Chen and Nunez [7] or Hlupic and Robinson [11] which relied on queuing systems to model and examine different issues on business processes. Furthermore using queuing systems a wide range of possible settings of the underlying business process (e. g. limited business hours, day and night operation, overtime work, alternating processing times) can be modeled within the focused optimization problem.

3.3. Order execution and service levels

The execution of an order starts with the arrival of an order unless all units of capacity within the in-house unit are busy. Otherwise each incoming order lines up in an infinite waiting queue. The queued orders will be executed immediately after free capacity is available according to the first in/first out principle. The time frame the order stays in the queue in front of the in-house unit is called waiting time. The execution time of one order is the time between the beginning of the first activity and the end of the last activity of the business process. One order uses at least one unit of capacity for this time frame. Waiting and execution time in total are called processing time. Free units of capacity are idle or can be used to accelerate the execution of orders by assigning more than one unit of capacity to an order. A service level s is guaranteed to our business partners regarding the processing time. Any order which does not keep up to the service level agreed causes costs c_g per order.

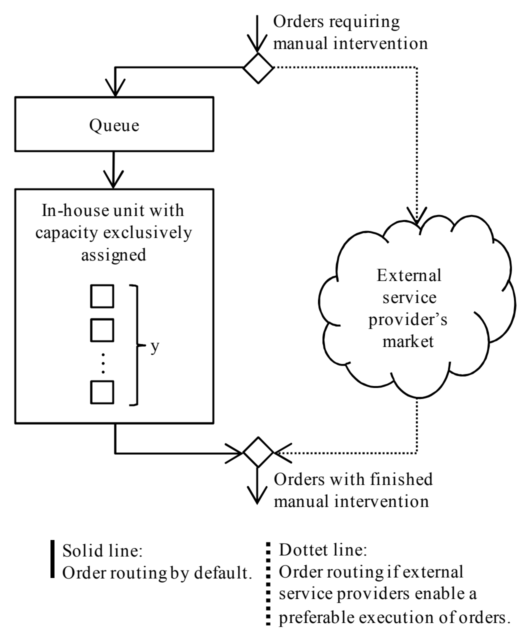


Figure 2. Service system with two queuing systems (in-house unit and external service provider's market)

For the optimization problem regarded in this paper a possible service level is a maximum processing time with monetary compensation for each time unit the order exceeds this limit. Another one is a fixed penalty for all orders of a given time frame which are not executed ahead of a final deadline.

3.4. External service provider's market

All manual interventions the in-house unit executes on orders are offered by external service providers as standardized services. Therefore these external service providers can be used to execute the manual interventions on orders in place of the in-house unit.

A set of external service providers which is basically able to execute the necessary manual interventions (necessary skills available, provider and employees trustworthy, tasks executed carefully etc.) is identified. Regarding the availability of capacity to execute orders no service level is agreed. An order can be executed externally only if capacities of one or more external service providers are (temporarily) underutilized.

External service providers are used in addition to the in-house unit, e. g. to execute peak loads resulting from volatile demand. As outlined above, capacities are not booked in advance on the external service provider's market as therefore usually fixed costs arise. Also no SLA is contracted due to high fees usually accompanying a pay-per-use agreement for volatile demand [2, 8], especially for resources not easily scalable. In fact excess capacity is used which cannot be employed otherwise. Orders are routed "on demand" to the external service provider's market, i. e. whenever the execution of orders on this market is preferable to the execution within the in-house unit (see section 3.7 for details).

Therefore the set of external service providers has to be evaluated constantly concerning their availability of capacity, execution costs etc. For this purpose, we assume: The provider's IT-platform allows a continuous evaluation and integration of external service providers. All relevant information is provided by the external service provider's market. The necessary technologies (e. g. service repositories and well described services based on standardized description languages) for a quick and mostly automated evaluation and integration of service providers are established.

As one of the most important information the availability of capacity on the external service provider's market has to be determined: The time frame a until an order can be executed is exogenous but can be determined from the information provided by the market. The absence of a service level concerning the availability of capacity on the external service provider's market therefore carries risk. With $a > 0$ orders cannot be executed immediately and this exogenous waiting time might be too long to support the in-house unit in executing orders within the service level agreed. This risk has to be considered within the processing costs for the external service provider's market to make an appropriate order routing decision.

3.5. Execution costs

To determine the execution costs for the order routing decision the following cost functions and parameters are necessary: The execution time t_i of the in-house unit for one order depends on its individual characteristics. Based on historical data the statistical distribution of t_i is stated. The total number of orders executed in-house is denoted with o_i. There are fixed costs c_f per unit capacity. The execution itself might cause additional variable costs c_v per order.

The fixed costs considered for one unit of capacity have to be calculated by cost accounting. They contain recurring costs of capacity e. g. wages of employees, running costs for the workplace and other equipment, overhead costs as well as all non-recurring costs building up this capacity.

For the external service providers the following characteristics have to be considered: The execution time t_e for one order depends on its individual characteristics. Based on historical data the statistical distribution of t_i is stated. There are no fixed costs but variable costs c_e which come up with the external execution of an order. These include not only the price for order execution to be paid to an external service provider but also the costs related with the evaluation and integration of the service provider. As prices may differ between different external service providers or even within one external service provider depending on its utilization, the respective price has to be provided along with the information about availability. The total number of externally routed orders is denoted with o_e.

3.6. Detailed objective function

Now the trade-offs considered within the optimization problem are described: First, providing too much capacity causes excessive costs of (idle) capacity. Providing too little causes excessive follow-up costs regarding the service level guaranteed. Second, the additional trade-off between the waiting cost resulting from the queue in front of the in-house unit and the time frame until an appropriate service provider can be identified and integrated is determined. With all characteristics stated above, the detailed objective function minimizing the total operating costs now can be stated as follows:

$$\min_y c = c_f\, y + c_v\, o_i + c_e\, o_e + c_g(\lambda, y, o_i, o_e, s, t_i, t_e, a)$$

3.7. Model analysis

To solve the optimization problem it is not sufficient to evaluate the two queuing systems representing the in-house unit and the external service provider's market separately. The service system has to be evaluated as a whole as the two queuing systems interact when executing orders arriving randomly.

Although queuing theory provides a strong mathematical foundation, this cannot be done analytically since the two queuing systems have different characteristics, especially concerning their distribution of processing times. They cannot be integrated to a service system or service network for which a mathematical model offers an analytical analysis or solution respectively. However, to derive interpretable results a discrete-event simulation as used by Hlupic and Robinson [11] is suitable. Thereby a simulation based optimum (referred to as "optimal capacity" hereinafter) can be determined. The effect of on demand integration capability then can be examined if the capacity optimization problem is solved with and without the external service provider's market.

To perform the discrete-event simulation an order routing algorithm has to be added to the service system deciding whether an incoming order is routed to the in-house unit or to the external service provider's market. With this routing algorithm the two interacting queuing systems are linked. Regarding to the cost-based optimization problem, the routing of orders has to be made on processing costs. These costs subsume all characteristics which have to be taken into account, e. g. the current processing time with regard to the service level agreed to our customer, fixed and variable costs, quantity discounts or minimum purchasing quantity as described in the previous sections.

The routing algorithm determines the processing costs and works as follows: With each arrival of an order the processing costs of both queuing systems are evaluated and the one with lower processing costs ("preferable execution") is chosen. Therefore the algorithm first determines the processing time for each queuing system. For the in-house unit it is easily determinable as the state of the system is known: It depends on the capacity available, the arrival rate of orders and the execution time. For the external service provider's market, the time frame a until free capacity is available, has to be retrieved to determine the processing time. Second, along with variable execution costs and the costs possibly incurring from the service level agreed, the processing costs can be calculated.

For further analysis we now introduce a case study based on a real world example to perform the discrete-event simulation. It uses the model setting described above and implements the necessary routing algorithm.

4. Simulation based on real world example of securities trading and settlement process

We consider the securities trading and settlement process with all necessary activities to be executed when securities are sold or bought e. g. via the stock exchange. This process is a typical case addressed with our model. It is a business process mostly sourced from a specialized business partner called "transaction bank" (see figure 1 for the case study's three-stage supply chain). A large number of orders have to be processed in time to meet regulatory standards and to avoid penalties or losses of interest when payments are not executed in time. Therefore detailed service levels are agreed. With few exceptions this process is fully digitalized and standardized through regulations and cross-company agreements. Nevertheless some manual interventions are necessary, especially after an order is placed and the corresponding transaction is closed. Within the settlement process for example, digitalized documents have to be checked, files and reports have to be completed or fees must be calculated.

Staffing the in-house unit charged with these manual interventions is usually an important optimization problem for the transaction bank. The margins for this business process are small and therefore the capacity of the in-house unit should be kept as small as possible to reduce the corresponding costs to a minimum. However the limited time for execution has to be taken into account. Along with the volatile arrival rates of incoming orders there is a trade-off between idle times or delayed execution respectively. Therefore these manual interventions performed on orders are the starting point for capacity optimization.

4.1. Input data

The necessary input data is identified as follows: The transaction bank accepts orders every bank working day between 8 a.m. and 6 p.m. Analyzing historical data reveals different peaks concerning the arrival rate of orders requiring manual intervention depending on exogenous factors. Dividing the ten hours of order acceptance in six time frames, the arrival rate within each time frame is approximated by an exponential distribution as summarized in table 1.

The manual interventions performed on an order take 16 minutes in average within the in-house unit as well as at the external service provider's market. In this special case, idle capacity cannot be used to accelerate the execution of orders as only one employee can work on one order. Cost accounting reveals that one unit of capacity y causes fixed costs amounting to EUR 330 a working day. There are no additional variable costs.

In the financial services industry it is necessary to execute orders in time, e. g. to meet external deadlines and to avoid losses of interest. Furthermore especially for legal and reputational reasons it is necessary that no order is left unexecuted. The service level agreement between the broker and the transaction bank therefore consists of two deadlines: First, each order has to be processed and sent back to the broker within a fixed time frame. With regard to all other activities within the whole business process the execution time of all manual interventions on one order must not exceed 30 minutes. For each minute an order exceeds this time frame, a compensation payment is due. This payment increases with the length of exceedance and is calculated following EUR $0.02 * (min. exceeded)^3$. Second, there is a final processing deadline at 8:00 p.m. for each working day. For each order which is not processed within this deadline the compensation rises to a penalty of EUR 100.

With a set of external service providers qualified for manual rework a fixed price of EUR 8.90 for an order was agreed. Based on historical data provided from these external service providers the waiting time in the queue in front of the external service provider's market is approximated (being a single customer to a set of service providers we assume no substantial effects on this waiting times routing orders externally). During a working day three time frames with different utilization of the external service provider's capacities are identified. Each time frame shows different waiting times for free capacity which can be approximated by a normal distribution as outlined in table 2 (to avoid negative values we used a truncated normal distribution within the simulation).

Orders sent to the external service provider's market have to wait for free capacity according to these time frames. The routing decision is made immediately after the orders arrived depending on the processing costs. Due to operational reasons subsequent changes to this decision are not possible.

4.2. Discrete event simulation set up

To determine the optimal in-house capacity the following procedure is applied: We perform multiple simulation experiments with increasing integer values for the capacity of the in-house unit. Each experiment consists of 1,000 simulation runs. For each run the total operating costs are determined. Starting the experiments with one unit of in-house capacity, we increase the value by one unit before the next experiment is started, which ensures the solution to be an integer value. This is done until the results of an experiment show that no waiting costs occur in front of the in-house unit for all runs. From this it follows a further increase of capacity does not have any positive effect concerning the total operating costs. Finally, comparing the average total operating costs for each experiment and choosing the one with the lowest costs then leads to the optimal in-house capacity.

With regard to the simulation time it is convenient that all working days are independently of each other (e. g. no unexecuted orders left due to the processing deadline) and the relevant events which determine the optimal in-house capacity are recurrent each bank working day. Thereby it is sufficient to simulate a single working day to determine the optimal capacity.

For each simulation run incoming orders are generated randomly following their statistical distributions. Whenever a new time frame is reached, the arrival rate is adapted. Concerning the external service provider's availability a random value is generated from the corresponding statistical distribution each time an order arrives. This random value represents the time frame the respective order has to wait until it can be executed externally. It is used by the routing algorithm to determine the processing costs of external execution.

Table 1. Arrival rates within a working day (mean number of orders per minute)

8:00 a.m. – 9:30 a.m.	9:30 a.m. – 11:30 a.m.	11:30 a.m. – 12:00 noon	12:00 noon – 1:30 p.m.	1:30 p.m. – 4:00 p.m.	4:00 p.m. – 6:00 p.m.
60	3	80	7	15	4

Table 2. Distribution parameters of the exogenous waiting time until an order can be executed at the external service provider's market (mean and standard deviation in minutes)

8:00 a.m. – 12:00 noon	12:00 noon – 2:00 p.m.	after 2:00 p.m.
μ = 25:00; σ = 8:00	μ = 12:00; σ = 2:10	μ = 10:00; σ = 7:00

The routing algorithm determines the current processing costs of both paths each time an order arrives. Then it routes the order to the path with lower costs. Thereby the processing costs of the in-house execution result from the service level agreement with the financial service provider only. There are no variable costs and all fixed costs are sunk costs which must not be taken into account. From the service level agreement costs can occur in two different ways: If an order cannot be processed ahead of the final processing deadline, the penalty has to be considered within the processing costs. Otherwise, if the agreed processing time per order is exceeded costs per minute are charged. For the external execution the processing costs consist of the variable cost per order and the costs resulting from the service level agreement determined analogous.

4.3. Simulation results

Identifying the optimal capacity level leads to the optimum outlined in table 3. The total operating costs are reduced by 14,730.39 EUR a working day if the external service provider's market is available on demand. The sharp drop in capacity shows that without the on demand integration of external service providers, a vast amount of overcapacity has to be kept to handle peak loads appropriate. Even if the risk connected to the on demand integration of external service providers as well as the variable cost for order execution are considered the total operating costs can be reduced by 12.4 %.

Table 3. Optimal capacity

on demand integration:	without ~	with ~
optimal y [units]:	261	61
corresponding c [EUR]:	118,600.85	103,870.46

Figure 3 shows the total operating costs gained from the simulation depending on the capacity of the in-house unit without and with on demand integration. Analyzing the cost functions which sum up to the total operating costs in-depth reveals the following findings, especially the influence of the cost associated with the service level agreement on the total operating costs:

For the scenario without on demand integration very small in-house capacity results in a high amount of unexecuted orders and the total operating costs are very high due to the corresponding penalties. With increasing capacity, more orders are executed during a bank working day ahead of the final processing deadline and the total operating costs decrease accordingly. Increasing capacity furthermore implies an interesting effect on the service level costs for orders exceeding the agreed processing time of 30 minutes: First, the increasing amount of orders which can be executed during a bank working day is accompanied by long waiting times. Reaching a certain point, additional capacity not only executes more orders but also reduces the waiting times for these executed orders in the queue in front of the in-house unit. These costs along with the linear fixed costs of capacity shape the total operating costs to a convex graph with a global minimum.

For the scenario with on demand integration the graph is shaped different. The most striking change is the reduced range of variation depending on the in-house capacity. The on demand integration capability therefore reduces the risk of allocating an inappropriate amount of capacity to the in-house unit.

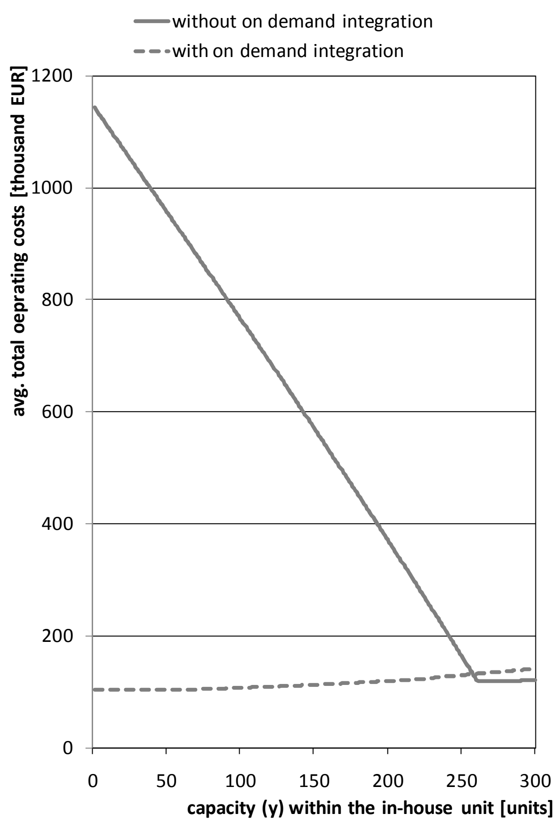


Figure 3. Total operating costs with and without on demand integration

This is due to the fact the external execution allows the execution of orders which are left unexecuted in the other scenario. Without the on demand integration capability, assigning too little capacity to the in-house unit inevitable leads to unexecuted orders. The external execution in contrast enables the execution whereby an external execution is associated with variable execution costs and risky waiting times. The high penalty ensures the routing algorithm decision chooses the external execution. Therefore the peak of the total operating costs occurring in the case without the on demand integration capability with a capacity lower than 261 units is cut off.

But there is another reason for the different shape of the total operating costs: With on demand integration capability not only more orders are executed ahead of the final processing deadline. Moreover it reduces the waiting times in the queue in front of the in-house unit and for this reason the waiting costs. Orders arriving in peak times (e. g. the early morning) can be routed to the external service provider's market and do not have to queue up in front of the in-house unit where the waiting times increase. This main advantage reveals figure 4, too. It shows the waiting time in front of the in-house unit during a working day for the scenario without and with on demand integration capability when the optimal capacity is assigned. By using external service providers on demand this waiting time can be reduced significantly in absolute terms as well as its deviation.

Furthermore it should be mentioned that the waiting costs in front of the in-house unit are shaped similar to the other scenario. Again, with increasing capacity more orders are executed in-house and the waiting costs rise until the point additional capacity ensures reduced waiting times in front of the in-house unit.

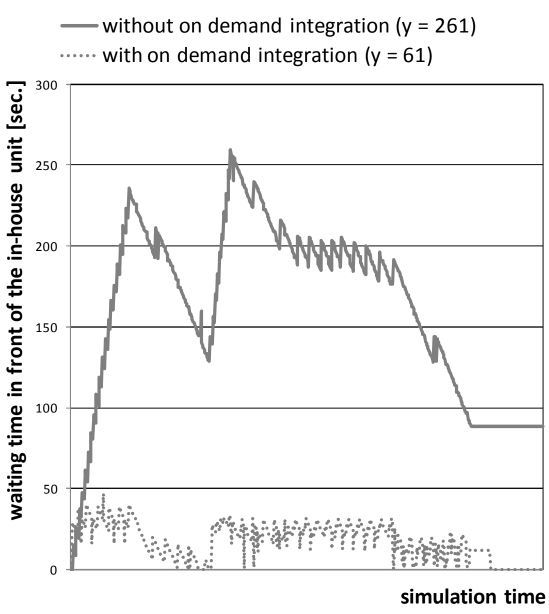


Figure 4. Waiting time in queue in front of the in-house unit during a working day

5. Summary and further research

This paper examines the effect of on demand integration capability on a capacity optimization problem for business processes when overcapacities at an external service provider's market can be used to handle peak loads. After modeling the three stage supply chain in general, a case study was introduced to quantify the effects.

The case study shows a sharp drop of optimal capacity when the on demand integration of external service providers is possible. The vast amount of overcapacity necessary without on demand integration capability to handle peak loads can be reduced significantly, even if the risk connected to the on demand integration of external service providers as well as the variable cost for order execution are considered appropriately. Furthermore, the risk of allocating an inappropriate amount of capacity to the in-house unit can be reduced.

Regarding the applicability of our model we would like to highlight the following: First, only a small part of possible settings which can be evaluated using the model are considered within the case study. With the simulation approach various different settings, e. g. dependencies between single working days or different execution times depending on the characteristics of an order can be taken into account, too. Second, the model also can be applied to determine the optimal capacity of a service level agreement the provider is going to conclude with a dedicated external service provider instead of have its own in-house unit. Here again, booked capacity can be optimized and costs can be reduced by routing peaks on demand to the market. Altogether the approach of an on demand routing helps to fully use the whole capacity available on the market through a flexible allocation to free resources. This is an important issue especially in terms of a sustainable usage of resources.

The main challenges for the applicability of the model are: Various historical data is needed as input for the simulation. As internal data about incoming orders or execution times are traceable or can be derived from contractual agreements, gathering data from external service providers could be difficult. As the proposed on demand routing of orders to the external service provider's market supports these companies to ensure their capacity utilization this might be an incentive to provide the relevant information. Furthermore there has to be a market providing all necessary tasks which have to be outsourced. Following recent developments discussed under the label of "cloud computing", show that a wide range of very different services is supplied already. Meanwhile providers offer business processes as a service, too, which might be an advertising message only but shows the efforts of service providers to perform more specialized tasks for their clients.

Additionally we would like to highlight, that the model presented leaves room for improvement as it currently applies only to scenarios where one activity or consecutive activities of a business process are executed at the external service provider's market. Modeling more complex business processes would require a queuing network considering e. g. different arrival times, processing times and a more complex layout of activities suitable for external execution.

As a subsequent need of further research the extension of the model in the following way seems useful: In addition to an external service provider's market with on demand integration, contracted agreements with different pricing options (flat fee, pay-per-use with agreed service level etc.) should be modeled. Comparing these contractual agreements as different opportunities to handle peak loads and their effects on capacity optimization may reveal further insights optimizing the supply chain considered. Finally we would like to mention, that legal reasons (as a different application domain) have to be discussed when using sub-contractors instead of an in-house unit

6. References

[1] B. Adenso-Diaz, P. Gonzalez-Torre, and V. Garcia, "A capacity management model in service industries", International Journal of Service Management, Vol. 13 (#3) 2002, pp. 286-302.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Soica, and M. Zaharia, "A View of Cloud Computing", Communications of the ACM, Vol. 53 (#4) 2010, pp. 50-58.

[3] A. Bassamboo, S. R. Ramandeep, and J. A. van Mieghem, "Optimal Flexibility Configurations in Newsvendor Networks: Going Beyond Chaining and Pairing", Management Science, Vol. 56 (#8) 2010, pp. 1285-1303.

[4] A. Bassamboo, R. S. Randhawa, and A. Zeevi, "Capacity Sizing Under Parameter Uncertainty: Safety Staffing Principles Revisited", Management Science, Vol. 56 (#10) 2010, pp. 1668-1686.

[5] K. Braunwarth, and C. Ullrich, "Valuating Business Process Flexibility achieved through an alternative Execution Path", Proceedings of the 18th European Conference on Information Systems, ECIS, Pretoria, South Africa, 2010, pp. 1-13.

[6] H. U. Buhl, M. Röglinger, S. Stöckl, and K. Braunwarth, "Value orientation in process management – Research gap and contribution to economically well-founded decisions in process management", Business & Information Systems Engineering, Vol. 3 (#3) 2011, pp. 163-172.

[7] L. Chen, and M. Nunez, "Business Process Integration of Multiple Customer Order Review Systems", IEEE Transactions on engineering Management, Vol. 57 (#3) 2010, pp. 502-512.

[8] J. P. Degabriele, and D. Pym, "Economic Aspects of a Utility Computing Service", HP Technical Reports, 2007, pp. 1-20.

[9] L. Dong, and E. Durbin, "Markets for Surplus Components with a Strategic Supplier", Naval Research Logistics, Vol. 52, 2005, pp. 734-753.

[10] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, Fundamentals of Queuing Theory, Wiley, Hoboken, NJ, 2008.

[11] V. Hlupic, and S. Robinson, "Business process modeling and analysis using discrete-event simulation,", Proceedings of the 1998 Winter Simulation Conference, Washington DC, USA, 1998, pp. 1363-1369.

[12] H. Lee, and S. Whang, "The Impact of a Secondary Market on the Supply Chain", Management Science, Vol. 48 (#6) 2002, pp. 719-731.

[13] I. Lee, "A Simple Model for Determining the Optimal Capacity Investment", in D. Neumann, M. Baker, J. Altmann, and O. F. Rana, Economic models and algorithms for distributed systems, Birkhäuser, Basel, 2010, pp. 209-220.

[14] T. Liu, "Revenue Management model for on-demand IT services", European Journal of Operations Research, Vol. 207, 2010, pp. 401-408.

[15] S. Netessine, G. Dobson, and R. Shumsky, "Flexible Service Capacity: "Optimal Investment and the Impact of Demand Correlation", Operations Research, Vol. 50 (#2) 2002, pp. 375-388.

[16] A. Rappaport, Creating shareholder value: the new standard for business performance, Free Press, New York, 1986.

[17] B. Tomlin, "On the Value of Mitigation and Contingency Strategies for Managing Supply Chain Disruption Risks", Management Science, Vol. 52 (#5) 2006, pp. 639-657.

[18] B. Tomlin, and Y. Wang, "On the Value of Mix Flexibility and Dual Sourcing in Unreliable Newsvendor Networks", Manufacturing & Service Operations Management, Vol. 7 (#1) 2005, pp. 37-57.

[19] J. van Mieghem, and N. Rudi, "Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities", Manufacturing & Service Operations Management, Vol. 4 (#4) 2002, pp. 313-335.

[20] M. Vehlow, and C. Golkowsy, Cloud Computing: Navigating the Cloud, Survey by PricewaterhouseCoopers AG, Frankfurt am Main, Germany 2010.