ON THE INTERACTION BETWEEN RESOURCE FLEXIBILITY AND FLEXIBILITY STRUCTURES

Zeynep Akşin[†], **Fikri Karaesmen**[††], **and Lerzan Örmeci**[††]

[†] *Graduate School of Business, Koç University, Istanbul, Turkey, zaksin@ku.edu.tr*
[††] *Dept. of Ind. Eng., Koç University, Istanbul, Turkey, fkaraesmen@ku.edu.tr, lormeci@ku.edu.tr*

Abstract: Most service systems consist of multi-departmental structures corresponding to multiple types of service requests, with possibly multi-skill agents that can deal with several types of service requests. The design of flexibility in terms of agents' skill sets and assignments of requests is a critical issue for such systems. Managers also have a choice in determining the percentage of capacity that will have the designed flexibility structure. A similar problem exists in the manufacturing of multiple products in multiple plants. In this paper, we investigate two issues regarding flexibility design: what is the amount of resource flexibility needed in each department and how does this interact with the flexibility structure in place.

Keywords: Flexibility, Capacity, Random Demand, Cross-Training, Network Flows

1 Introduction

This paper considers service systems with multi-departmental structures having possibly multi-skill servers that treat several types of service requests. In any such system, it is possible to have a different mix of skill sets with a different number of servers belonging to each skill set. It is well known that more flexibility leads to better operational performance. However given that there are costs associated with creating and maintaining this flexibility, and difficulties managing the resulting more complex system, it is desirable to understand the value of this flexibility in more depth. This paper will focus on providing a better understanding of the relationship between different flexibility structures, flexible capacity, and value. The following questions are relevant in this setting: Given a particular skill set, what proportion of the servers should be cross-trained in these (as opposed to remaining dedicated to their special skill)? In other words, how much resource flexibility should one plan for? How does the skill set design or demand variability affect this decision? By building on earlier work in flexibility design, we establish a structural property, and then illustrate the interactions between flexibility structure and resource flexibility through some numerical examples.

A well known application of this flexibility design problem in a manufacturing setting is the problem studied by Jordan and Graves [6]. In this setting, the departments are different plants or production lines, while the customer types represent different products to be produced in these production facilities. Process flexibility constitutes the ability of producing a product in multiple plants or production lines. The model that we study in this paper is identical to the one in Jordan and Graves [6]. However, in our setting a proportion of the plant capacity is allowed to remain dedicated to a single product. We explore the choice of this proportion and how it interacts with the process flexibility structure in place.

The remaining parts of this paper are organized as follows. Related literature is reviewed in the next section. Section 3 introduces the model and presents the results on the diminishing returns property of flexibility. The numerical results on flexibility/capacity interactions are presented in Section 4. We end with concluding remarks in Section 5.

2 Literature Review

The importance of flexibility in service delivery is well known. A significant source of service delivery process flexibility comes from the use of cross-trained servers. While the practice itself is widespread, there is little formal evaluation of the value of this type of practice from an operations standpoint. Pinker and Shumsky [9] consider trade-offs between capacity and quality for cross-trained workers in service systems. Numerous studies in manufacturing have looked at the case of flexible workers and their impact on performance in terms of operational measures like throughput. Most of these studies analyze specific work-sharing schemes in queueing network models (Van Oyen et al. [11], and references therein). Karaesmen et al. [7] investigate flexibility in the context of field service design. These papers assess the value of certain workforce flexibility practices in given settings, however do not tackle the broader question of designing the type of flexibility in these systems.

More generally, the benefits and design of flexibility in operations have been studied extensively (DeGroote, [3]; Sethi and Sethi, [10]). An important stream consists of papers that address the capacity investment problem in the presence of flexible resources (Fine and Freund [4]; Van Mieghem [12]; Netessine et al. [8]). These papers assume a certain form of flexibility and then explore the question of the ideal level of this flexibility and how it relates to value under uncertain demand. A similar capacity investment (staffing) problem is addressed in Harrison and Zeevi [5] and Chevalier et al. [2], however these papers consider queueing and loss systems respectively as the underlying service system. The analysis in this case is further complicated by the routing and scheduling of calls to multiple queues. Like the current paper, Chevalier et al. [2] explore the optimal proportion of flexible and dedicated servers in a particular loss system with two types of servers. Servers can either be dedicated or fully flexible in their analysis. We consider the possibility of having servers with partial flexibility, by also allowing for flexibility designs where not all flexible servers are generalists. Thus our analysis is more general in terms of the flexibility structures that are considered. However we assume capacity to be deterministic as in Jordan and Graves [6], and only explore the value of flexible capacity, without considering its costs.

The objective of this paper is to represent flexibility structures through a network flow model as in Jordan and Graves [6] and Aksin and Karaesmen [1] and formulate the resource flexibility question in terms of this model. By doing this, we are able to relate the flexible capacity question to the earlier analyzed flexibility structure design problem. Our analysis demonstrates the close interaction between the amount of resource flexibility and flexibility structure, thereby pointing to an important direction for future research.

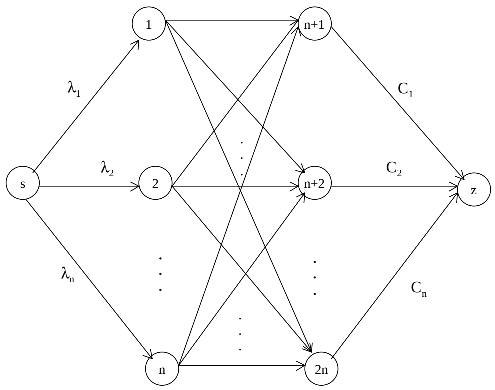

Figure 1: An n class service system with full flexibility

3 The Model and Analysis of Process Flexibility

Consider a service system with multiple customer types. Customer types differ in terms of their service requirements. Servers specialize by customer type, but can be flexible with overlapping skill sets, allowing them to treat customer requests from different types. The service system can be represented as a directed graph $G = (N, A)$ with a set of nodes N, of which one is a source node and one a sink node, and a set of arcs A whose elements are ordered pairs of distinct nodes. Some standard definitions are useful to formalize the description of this network. A directed arc (i, j) emanating from node i is said to have tail i, terminating in node j known as the head of the arc. For an arc $(i, j) \in A$, the node j is said to be adjacent to node i. The node adjacency list $A(i)$ is the set of adjacent nodes, $A(i) = \{j \in N : (i, j) \in A\}$. The indegree of a node is the number of incoming arcs of that node and its outdegree is the number of outgoing arcs.

An instance of a network that represents the service system is depicted in Figure 1. This graph illustrates a system with n customer types given by the set of nodes $I = \{1, 2, ..., n\}$, served by servers in n departments, given by the set of nodes $J = \{n + 1, n + 2, ..., 2n\}$. Note that since servers are assumed to be organized by their primary skills, the number of customer types is equal to the number of departments. The case where the number of customer types is larger than the number of departments can also be treated within this framework, where the additional classes can be served by dummy departments with no servers in them. The arcs emanating from the source node s and terminating in nodes $i \in I$ represent the service demand, and have capacity given by the demand vector $\lambda = (\lambda_1, ..., \lambda_n)$. This vector represents the realization of demand for a given period. The arcs emanating from nodes $j \in J$ and terminating in the sink node z represent the capacity of each department. These arcs have a capacity given by the vector $\mathbf{C} = (C_1, ..., C_n)$. The arcs (i, j) with $i \in I$ and $j \in J$ represent the flexibility of the system. Whenever a customer of type $i \in I$ can be served by a server of type $j \in J$, an arc (i, j) with infinite capacity is added to the network. The network in Figure 1 illustrates a case where all customers can be treated by all servers, i.e. where the system has full flexibility. In a system with n departments, full flexibility implies that each node $i \in I$ has outdegree equal to n. In general, the outdegree of node $i \in I$ represents the number of possible routings for customers of type i, and the indegree of a node $j \in J$ represents the number of skills a server of type j has. Assuming that each customer request of type $i \in I$ is worth r to the system, the problem of maximizing the value generated by a given configuration for a demand realization $\boldsymbol{\lambda}$ is equivalent to the maximum flow problem for this network. We refer to the maximal flow as the throughput and denote it by $T(\boldsymbol{\lambda}, \mathbf{C})$. Focusing on a different context i.e. process flexibility in manufacturing, Jordan and Graves [6] consider a random demand vector $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ where the performance measure of interest is the expected throughput $E[T(\boldsymbol{\lambda}, \mathbf{C})]$. In this case, a maximum flow problem is solved for each realization of the random demand vector $\boldsymbol{\lambda}$ and the expectation is taken over all realizations of $\boldsymbol{\lambda}$. Jordan and Graves also discuss the relevance of the expected throughput maximization objective and relate it to a number of other possible objectives. Our main focus here is also on this random demand model and the expected throughput criterion but unlike Jordan and Graves we allow partial resource flexibility.

Now let s and z denote the source and sink nodes respectively, where the source node is connected to nodes $1, 2, ..., n$ and the sink node is connected to nodes $n+1, n+2, ..., 2n$ and let x_{kl} denote the flow on the arc (k, l). For a given demand vector $\boldsymbol{\lambda}$, the maximum flow problem can be expressed as the following linear program:

$$\begin{array}{cccl}
T = \max & \sum_{i=1}^{n} x_{si} & & \\
x_{si} & \leq & \lambda_i & \text{for } i = 1, 2, ..., n \\
x_{jz} & \leq & C_{j-n} & \text{for } j = n+1, n+2, ..., 2n \\
x_{si} & = & \sum_{j|(i,j)\in A} x_{ij} & \text{for } i = 1, 2, ..., n \\
x_{jz} & = & \sum_{i|(i,j)\in A} x_{ij} & \text{for } j = n+1, n+2, ..., 2n \\
x_{si} & \geq & 0 & \text{for } i = 1, 2, ..., n \\
x_{jz} & \geq & 0 & \text{for } j = n+1, n+2, ..., 2n \\
x_{ij} & \geq & 0 & \text{for all } i, j|(i,j) \in A
\end{array}$$

Now assume that the proportion of cross-trained (flexible) capacity of resource j is α. In other words while the total capacity of resource j is still C units, the capacity that it can allocate to other demand types is only αC. This imposes the following additional constraint on the maximum flow problem:

$$\sum_{i|(i,j)\in A, i\neq j-n} x_{ij} \leq \alpha C_{j-n} \text{ for } j = n+1, n+2, ..., 2n \qquad (1)$$

The solution of the above maximum flow problem with the additional constraint set (1) is easy for a given demand vector. The expected throughput can then be obtained by simulation. General properties on the structure of the solution are more difficult to obtain and will be investigated through numerical examples in the next section. There is, however, one property on the effect of increasing α that can be shown.

Property: The expected throughput, $E[T(\boldsymbol{\lambda}, \mathbf{C})]$, is increasing and concave in α.

Proof: α appears on the right hand side of the maximum flow LP. It is well-known that the objective function of an LP is increasing and concave in the right-hand side. In addition, both monotonicity and concavity are preserved under the expected value which proves the desired result.

4 Preliminary Numerical Results

Our numerical example considers a service system with three departments and three customer types. Once again, our focus is on systems where each department has servers who have one main skill which

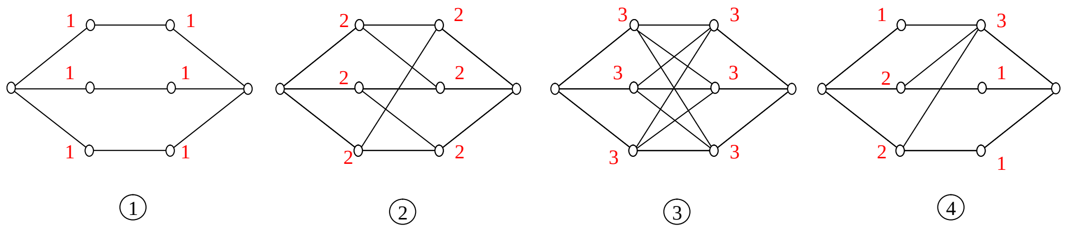

Figure 2: The Different Flexibility Structures Considered

coincides with customer types. Figure 2 presents the four different flexibility structures that will be considered. In the first structure, all servers are dedicated. This is a structure with no flexibility. The other three structures are so-called chain structures (Jordan and Graves [6]). The authors define a *chain* structure as a group of directly or indirectly connected products (customer types) and plants (departments). Structures 2 and 3 are symmetrical chains where the servers have respectively 2 and 3 different skills. Structure 4 is an extreme case where department 1 has servers that are cross-trained in all three skills but departments 2 and 3 are dedicated.

The service capacities are assumed to be 20 units per period in each of the departments. Demand for each skill is exponentially distributed with mean 20 units per period. The expected throughput is estimated by simulation. We report the normalized expected throughput which is the ratio of the expected throughput to the total available capacity.

The first set of results pertains to the monotonicity in α. The results are depicted in Figure 3, where Flex$j$ denotes the expected throughput of a system with flexibility structure j, as shown in Figure 2. As shown, the expected throughput increases in a concave manner as a function of α regardless of the flexibility structure. When structures 2 and 3 are compared, it is observed that the initial investment in flexible capacity (going from no flexible capacity to 20% flexible capacity) is more beneficial in terms of expected throughput in structure 3. However at 100% flexibility, the performance of the two structures are very close. Structure 4 manifests a different behavior. Its performance is significantly inferior to the symmetrical structures in 2 and 3. It also benefits little from additional flexible resources. This implies that flexibility structure and the amount of resource flexibility have joint effects. More flexible structures not only have better performance but also benefit from additional investments in flexible capacity.

The second set of results investigates the effects of the flexibility structure for a given resource flexibility level. In Figure 4, we compare the expected throughput of structures 1, 2, and 3. For 100% resource flexibility, it was proven in Aksin and Karaesmen [1] that the marginal increase in expected throughput in going from structure 1 to structure 2 is higher than the marginal increase in going from structure 2 to structure 3. The corresponding curve in the figure indicates that the marginal increase in going from structure 1 to 2 is indeed much more significant. Moreover, this property is observed to be true regardless of the resource flexibility level. Relatively small but well designed improvements in flexibility structure (structure 2 instead of structure 1) bring significant benefits in terms of performance.

The final issue we investigate is demand variability. In order to perform this, we compare the performance of the system with exponentially distributed demand with that of an identical system with a less variable demand distribution. In particular, the reference demand distribution is a Gamma

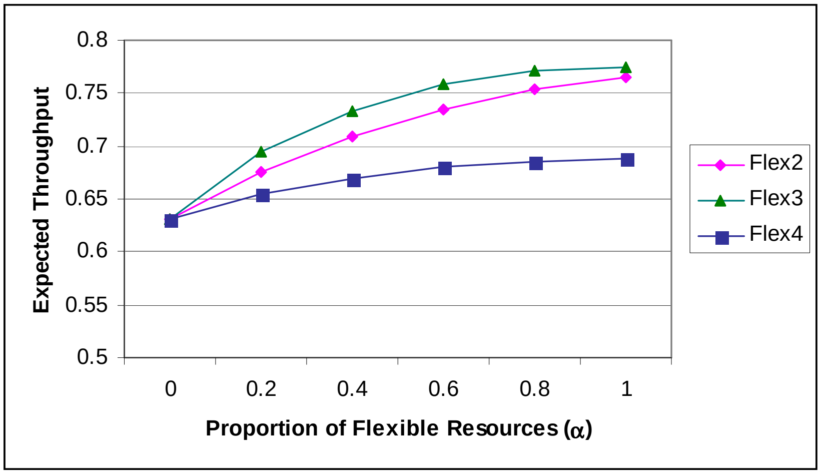

Figure 3: The Expected Throughput as a Function of the Proportion of Flexible Resources

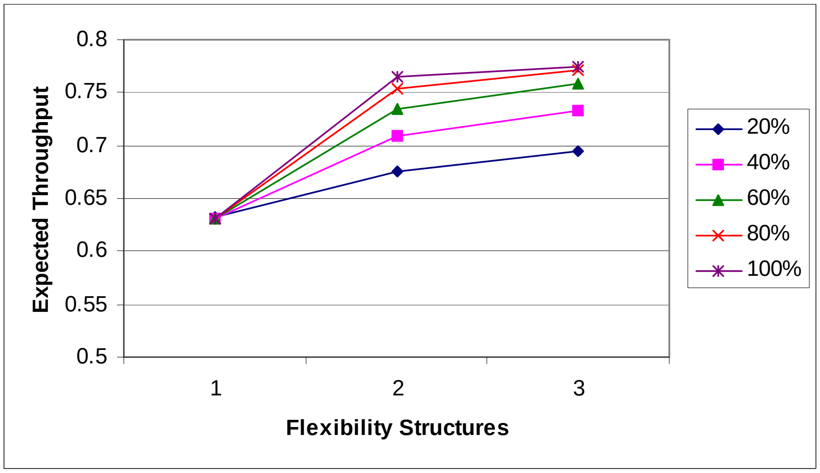

Figure 4: The Expected Throughput as a Function of Flexibility Structure

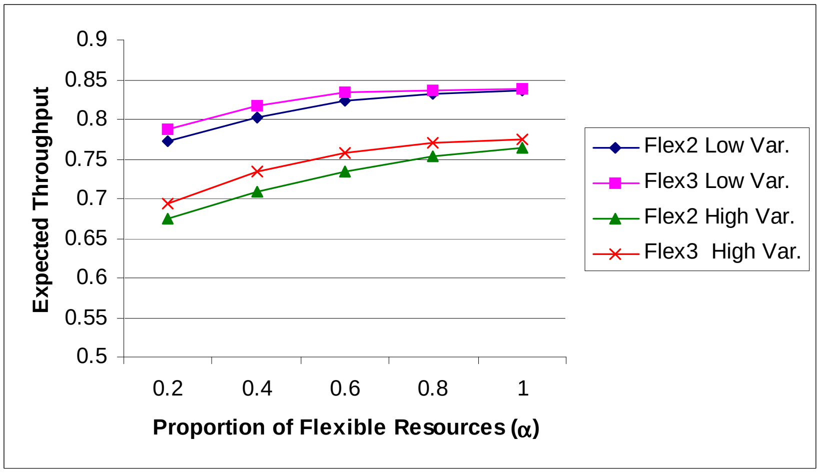

Figure 5: The Expected Throughput as a Function of Flexibility Structure and the Variance of the Demand Distribution

distribution with shape parameter 2 and scale parameter 10. Note that the two distributions have identical means but the Gamma distribution not only has a lower variance (200 compared to 400 for the exponential distribution) but is also less variable than the exponential distribution in the sense of a convex stochastic order. Figure 5 summarizes the results. The first immediate observation is that increased demand variability hurts performance in a significant manner as expected. Moreover, there is little difference in the performance of structures 2 and 3 when demand variability is low but the difference is more pronounced when demand variability is high. Increased resource flexibility is always beneficial for the high variability system. For the low variability system, increased resource flexibility beyond $\alpha > 0.6$ has little benefit. It seems like when the demand variability is high and the flexibility structure is limited, additional resource flexibility always pays off.

5 Conclusion and Perspectives

In this paper, we investigated the interaction between resource flexibility levels and flexibility structures, and their effects on system performance using a single-period network flow model with random demand. This investigation has revealed a number of interesting preliminary results that should be verified more thoroughly in future work.

The interaction effect between flexibility structure and the amount of resource flexibility points to a need for the joint optimization of these two features in such systems. In Chevalier et al. [2] an optimization problem is formulated to determine the optimal proportion of fully flexible and dedicated servers. A cost structure, wherein each additional skill has a cost in addition to the cost of a dedicated server is assumed. A rule where 20 percent of the servers are flexible is shown to be near-optimal in most cases. While the underlying system, a loss system with overflow from dedicated to fully flexible servers, is different from ours, a comparison of the results is instructive. Our concavity property supports their general result of a limited amount of resource flexibility to obtain most of the benefits while keeping costs to a minimum. However, we think that the consideration of limited flexibility structures may lead to a different proportion of flexible servers in their results, since the same benefits may be obtained at much lower cost in this case. The interaction effect indicates that managers should first design their flexibility structure, and then address the question of amount of resource flexibility. Our numerical examples also indicate that the variability in customer arrivals may further emphasize this interaction.

The single-period model here is a simplified version of the stochastic dynamic problem where demand is distributed to different departments depending on the workloads over time. In general, this latter problem takes the form of a queueing control problem. Future work will focus on verifying the properties observed here for that problem.

References

[1] Akşin, O.Z. and Karaesmen, F. "Designing Flexibility: Characterizing the Value of Cross-Training Practices". *INSEAD, Working Paper*, February 2002.

[2] Chevalier, P., Shumsky, R.A., and Tabordon, N "Routing and Staffing in Large Call Centers with Dedicated and Flexible Servers". Working Paper, March 2004.

[3] De Groote, X. "The flexibility of production processes: a general framework". *Management Science*, 40:7 933-945, 1994.

[4] Fine, C.H. and Freund, R.M. "Optimal investment in product-flexible manufacturing capacity". *Management Science*, 36:4 449-466, 1990.

[5] Harrison, J.M. and Zeevi, A. "A Method for Staffing Large Call Centers Based on Stochastic Fluid Models". *Manufacturing and Service Operations Management forthcoming*, 2004.

[6] Jordan, W.C. and Graves, S.C. "Principles on the benefits of manufacturing process flexibility". *Management Science*, 41:4 577-594, 1995.

[7] Karaesmen F., F. van der Duyn Schouten and L. van Wassenhove, "Dedication vs. Flexibility in Field Service Operations", *Working Paper, Center for Economic Research, Tilburg University, revised version*, 2003.

[8] Netessine, S., Dobson, G. and Shumsky, R. "Flexible service capacity: optimal investment and the impact of demand correlation". *Operations Research*, 50:2, 375-389, 2002.

[9] Pinker, E. and Shumsky, R. "Efficiency-quality tradeoff of crosstrained workers". *Manufacturing and Service Operations Management*, 2:1 , 2000.

[10] Sethi, A.K. and Sethi, S.P. "Flexibility in manufacturing: a survey". *International Journal of Flexible Manufacturing Systems*, 2:289-328, 1990.

[11] Van Oyen, M.P., Gel, E.G.S., and Hopp, W.J. "Performance opportunity for workforce agility in collaborative and noncollaborative work systems". *IIE Transactions*, 33:9, 761-777, 2001.

[12] Van Mieghem, J.A. "Investment strategies for flexible resources". *Management Science*, 44:1071-1078, 1998.