Reading: Ch. 5. Flow Rate and Capacity Analysis (excerpt) *(Managing Business Process Flows: Principles of Operations Management)*

Author: Anupindi, Ravi; Chopra, Sunil; Deshmukh, Sudhakar D.

Editor: N/A

Publisher: Prentice Hall Publication Date: 2012 Pages: 102-109

Course: COMR_V 398 201 202 2024W2 Introduction to Business Processes and Operations
Course Code: 202 Term: 2024 Winter Term 2

Department: COMR_V

Chapter

5

Flow Rate and Capacity Analysis

INTRODUCTION

Anticipating an explosion in the demand for cars in China—estimated to increase at over 20 percent per year in 2010 and beyond—car manufacturers are investing billions of dollars in plant, equipment, and personnel in order to expand their production capacity. Toyota, with expected sales of over 800,000 cars in 2010, is counting on its new plant in Changchun, Jilin Province, to go online in 2012 with the production capacity of 100,000 cars per year. Similarly, Nissan, a major Toyota competitor, is increasing its own capacity from less than 600,000 to 900,000 cars per year. Not to be outdone, Volkswagen is increasing it's own capacity by more than 10 percent to 850,000.

Like Toyota, Nissan, and Volkswagen, every company is striving to match its capacity to expected demand by deploying the appropriate level of resources and by maximizing the effectiveness at which these resources are utilized. This is the topic of this chapter.

In Section 5.1, we examine how flow rate and capacity can be measured. In Section 5.2, we define the effective capacity of a resource pool, and show that the effective capacity of the process depends on that of its bottleneck resources. In Section 5.3, we examine how product mix decisions impact the effective capacity of a process and its profitability. In Section 5.4, we discuss capacity waste and introduce the notions of theoretical capacity and capacity utilization. Finally, in Section 5.5, we study some key ideas to improve the capacity of a process.

5.1 FLOW RATE MEASUREMENTS

As we have defined in earlier chapters, throughput, or average flow rate, of a stable process is the average number of flow units that flow through the process per unit of time. Capacity is the maximum sustainable throughput. Throughput and capacity, which indicate a "scale" of a process, are extremely important metrics of performance: Since the flow of units through the process represents the creation of economic value, it follows that the higher the throughput, the greater the value generated by the process. Capacity is also important from the perspective of managing process flow times since insufficient process capacity may lead to congestion and excessive waiting time. For these reasons, keeping track of the flow rate of a process is one of the most fundamental tasks of management in any organization.

Throughput is expressed in terms of number of flow units per unit of time, such as customers per day, tons per shift, cars per hour, dollars per month, and patients per year. Analogous to the estimation of average flow time outlined in Chapter 4, the average flow rate (throughput) of a stable process, R, can be measured by the following three-step procedure:

1. Observe the process over a given, extended period of time.
2. Measure the number of flow units that are processed by the process over the selected period of time.
3. Compute the average number of flow units per unit of time.

As mentioned earlier, the capacity is the maximum sustainable flow rate. It can be measured by observing the system in periods of heavy congestion in which the flow rate is limited by (and therefore equal to) capacity.

The lean operations literature, which began with the Toyota Production System (Chapter 10), often describes throughput in terms of **takt time**. Derived from the German word for rhythm or beat, takt time is the *reciprocal of throughput*. The concept is particularly useful in the context of synchronized assembly lines, where it represents the average activity time at each workstation (takt times for the assembly of mass-produced cars are on the order of 1 minute). Takt time is sometimes also called cycle time, but some authors use cycle time as a synonym for flow time. To avoid confusion, we do not use the term cycle time in this book.

5.2 RESOURCES AND EFFECTIVE CAPACITY

In general, the capacity of a system depends on the level of resources deployed by the system and on the effectiveness at which these resources are utilized. The capacity of any given process is typically quite difficult to analyze, mainly due to the subtle and complicated ways in which the various resources can interact. In this section we provide a simple and useful approximation, called the effective capacity.

5.2.1 Resources and Resource Pools

As we discussed in Chapter 1, activities are performed by capital and labor resources. Each activity may require one or more resources, and each resource may be allocated to one or more activities. For example, in the process of making bread, raw materials—flour, salt, butter, water, and so forth—are transformed into loaves of bread. The entire process requires performing such activities as mixing the ingredients, kneading the dough, forming the loaves, placing them in the oven(s), and baking them. In turn, these activities use such resources as mixers, bakers, and ovens. A given resource—for instance, a baker—may be used by several activities, such as mixing, kneading, and forming dough. Similarly, a given activity, such as loading the oven, may require multiple resources, such as a baker and an oven.

A **resource pool** is *a collection of interchangeable resources that can perform an identical set of activities. Each unit in a resource pool* is called a **resource unit**. For instance, in the case of bread making, three flexible bakers, each of whom can perform any baking activity, would be viewed collectively as a single resource pool containing three resource units. On the other hand, if each of the three bakers specialized in a separate activity (mixing, kneading, and forming dough, respectively), they would be regarded as three separate resource pools, each consisting of a single resource unit.

Combining separate resource pools into a single, more flexible, pool able to perform several activities is called **resource pooling**. It is a powerful operational concept that can significantly affect not only process flow rate and process capacity but also flow time (as we will see in Chapter 8).

5.2.2 Effective Capacity

The **unit load of a resource unit** is the *average amount of time required by the resource unit to process one flow unit, given the way the resource is utilized by the process*. The **effective capacity of a resource unit** is the *inverse of the unit load*. It represents the maximum sustainable flow rate through the resource unit, if it were to be observed in isolation. The **effective capacity of a resource pool** is the *sum of the effective capacities of all the resource units in that pool*. As an illustration, consider an insurance agent (a resource unit) whose job is to file residential insurance claims (flow units). Assume, for example, that on average, the agent spends 12 minutes per claim (unit load). Then, the effective capacity of the resource unit is 1/12 claims per minute (60/12 = 5 claims per hour). If two agents were available to process claims, then the effective capacity of the resource pool would be 2 × 5 = 10 claims per hour.

Formally, if we denote the unit load (at resource pool i) by T_i and the number of resource units by c_i

$$\text{Effective capacity (of resource pool } i\text{)} = c_i/T_i \qquad \textbf{(Equation 5.1)}$$

In practice, the agent is likely to handle various types of claims which require different amounts of time. The unit load in this case represents the average amount, over all types of claims. This issue is taken up in detail in Section 5.3. Also, if the various resource units are not identical in terms of their effective capacities, then the effective capacity of the resource pool will be the sum of the effective capacities of each resource unit in the pool. For the rest of this chapter, however, we will assume that all units in a resource pool are identical.

Effective capacities of different resource pools may vary. Since all resource pools are required to process each flow unit, no process can produce output any faster than its **bottleneck**—*the "slowest" resource pool of the process*. Thus, we define the **effective capacity of a process** as *the effective capacity of the bottleneck*.

The effective capacity of a process is a very useful concept for managing capacity, as Equation 5.1 provides a simple and practical way for connecting process capacity to overall resource levels, given the effectiveness at which resources are employed.

EXAMPLE 5.1

Health maintenance organizations (HMOs) provide their customers with all-inclusive medical service for a fixed monthly fee. To secure services, they contract with physicians and hospitals that provide their services on a fee-per-service basis. When members of an HMO receive medical service, the providing physician or hospital submits a claim to the HMO for reimbursement. NewLife Finance is a service provider to HMOs. For a small fee, it performs the entire claims processing operation on behalf of the HMO.

Table 5.1 Effective Capacity for Physician Claims, NewLife Finance

Resource Pool (i)	Unit Load (minutes per claim) (T_i)	Effective Capacity of a Resource Unit (claims per minute) ($1/T_i$)	Number of Units in the Resource Pool (c_i)	Effective Capacity of a Resource Pool (claims per minute) (c_i/T_i)
Mailroom clerk	1.00	1/1 = 1.00	1	1.00
Data-entry clerk	5.00	1/5 = 0.20	8	1.60
Claims processor	8.00	1/8 = 0.125	12	1.50
Claims supervisor	2.50	1/2.5 = 0.40	5	2.00

Processing a physician claim consists of the following operations:

1. Claims billed by physicians arrive by mail and are opened and date-stamped by the mailroom clerk. They are then placed into a data-entry bin.
2. Data-entry clerks enter date-stamped applications—first in, first out—into NewLife's claims-processing system. Data-entry clerks must check claims for proper formatting and completeness of data fields before they input claims into the system. If a claim is not legible, fully completed, or properly formatted, it must be sent back to the physician for resubmission. Once entered, claims are stored in a processing inventory called "suspended claims."
3. Claims are assigned to a claim processor for initial processing.
4. Processed claims are transferred by the system to a claim supervisor for inspection and possible alterations.
5. Claims are returned to their original claim processors who complete the transaction and issue instructions to accounts payable for settlement.

The process involves five steps, performed by four resource pools, namely mailroom clerks, data entry clerks, claim processors, and claim supervisors (Steps 3 and 5 are performed by the same resource pool, the claims processors). Table 5.1 lists, for each of the four resource pools, the unit load, its inverse, that is, the effective capacity of the resource unit, and the number of resource units. Based on this information, the last column of the table computes the effective capacity of each of the resource pools.

As can be seen, the bottleneck of the process is the pool of mail room clerks; and the effective capacity of the entire system equals 1.00 claim per minute, or 60 claims per hour.

How many professionals are required to achieve capacity of, say, 80 claims per hour? Using the effective capacity as our guideline, we note that an additional "third" (20/60) of a mailroom clerk is necessary. If part-time solutions are not available, one additional clerk must be hired. That will increase the capacity of the mailroom to 120 claims per hour, which is more than sufficient. Note, however, that the capacity of the *process* will not increase to 120: The bottleneck in this case shifts to the pool of claims processors, which has just enough capacity to handle 90 claims per hour.

5.2.3 Capacity Utilization

The throughput of a process, R, is the average number of flow units processed over a given period of time. Throughput of a process may not equal capacity because of external constraints such as low outflow rate (due to low demand rate, meaning an external bottleneck in the output market) or low inflow rate (due to low supply rate, meaning an external bottleneck in the input market).

For the i th resource pool, the **capacity utilization** of the resource pool, denoted by u_i, is defined by the relation:

u_i = Throughput/effective capacity of the i th resource pool **(Equation 5.2)**

Capacity utilization indicates the extent to which resources—which represent invested capital—are utilized to generate outputs (flow units and ultimately profits). It is defined for each resource pool independently. The capacity utilization of the process is defined by the bottleneck resource pool. We illustrate the concept in Example 5.2:

EXAMPLE 5.2

Assume that the average number of claims processed by NewLife Finance during a given month was measured to be 400 per day. The effective capacity of the various resources is as indicated in the second column of Table 5.2. The third column of the table lists the capacity utilization of the various resources:

Table 5.2 Capacity Utilization for NewLife Finance

Resource Pool (p)	Effective Capacity of a Resource Pool (claims per 8-hour day)	Capacity Utilization (u_i)
Mailroom clerk	1.00 × 480 = 480	400/480 = 83%
Data-entry clerk	1.6 × 480 = 768	400/768 = 52%
Claims processor	1.5 × 480 = 720	400/720 = 56%
Claims supervisor	2.0 × 480 = 960	480/960 ≃ 50%

Notice that, by definition, the bottleneck resource is the most highly utilized resource. If the throughput were to increase, that resource will be the first to hit full utilization. The capacity utilization of the entire process is 83 percent, given by the bottleneck.

5.2.4 Extensions: Other Factors Affecting Effective Capacity

The calculations of effective capacity discussed in the previous section ignore several important factors. First, they assume that resources handle units sequentially, or one unit at a time, rather than in load batches (imagine loaves of bread baked in an oven). Second, they assume that all resources are available for the same amount of time (imagine a factory with some units running a second shift). Finally, we have ignored the effects of setups or switching between products (imagine an operating room that needs to be reset between different types of surgery). Often these assumptions do not hold. We discuss how Equation 5.1 can be adjusted to accommodate these factors in Appendix 5.1.

5.3 EFFECT OF PRODUCT MIX ON EFFECTIVE CAPACITY AND PROFITABILITY OF A PROCESS

Firms often produce several products simultaneously. Since various products utilize resources at different rates, the effective capacity depends on the products produced and their proportions in the mix. This observation has an important business implication. In most organizations, sales/marketing departments make product mix decisions. Since such decisions affect the process capacity (a major driver of profitability), input from the operations group, which is responsible for production, is required. In Example 5.3,

Table 5.3 Unit Loads for Various Products, NewLife Finance

Resource Pool	Unit Load (Physician) (minutes per claim)	Unit Load (Hospital) (minutes per claim)
Mailroom clerk	1.00	1.50
Data-entry clerk	5.00	6.00
Claims processor	8.00	8.00
Claims supervisor	2.50	4.00

Table 5.4 Effective Capacity for Hospital Claims, NewLife Finance

Resource Pool (i)	Unit Load (minutes per claim) (T_i)	Effective Capacity of a Resource Unit (claims per minute) ($1/T_i$)	Number of Units in the Resource Pool (c_i)	Effective Capacity of a Resource Pool (claims per minute) (c_i/T_i)
Mailroom clerk	1.50	1/1.50 = 0.66	1	0.66
Data-entry clerk	6.00	1/6.00 = 0.17	8	1.33
Claims processor	8.00	1/8.00 = 0.125	12	1.50
Claims supervisor	4.00	1/4.00 = 0.25	5	1.25

we demonstrate the dependence of capacity on the product produced. In Sections 5.3.1 and 5.3.2, we examine the issue of product mix.

EXAMPLE 5.3

Assume that, in addition to processing physician claims, NewLife also handles claims submitted directly by hospitals. The process used to handle hospital claims is the same process used for physician claims. However, the unit loads required for the various operations are different. Table 5.3 contrasts the unit loads required for the two types of products.

In Table 5.4, we recompute the effective capacity of the process, using the unit loads that pertain to hospital claims (second column). The calculations of the other columns follow the logic of Example 5.1.

Thus, the capacity of the process is only 0.66 claims per minute (40 per hour) for hospital claims, as opposed to 60 claims per hour for physician claims.

5.3.1 Effective Capacity for Product Mix

For a process that produces several types of products simultaneously, we can represent the overall flow of the various products by constructing an (artificial) flow unit which represents the entire mix of the various products. We can calculate the unit load of the mix by averaging the unit loads of the individual products, using the weights of the mix, as illustrated in Example 5.4:

EXAMPLE 5.4

Currently, NewLife handles a product mix of 60% physician claims and 40% hospital claims. What is the effective capacity of the process?

Table 5.5 Unit Loads for Various Products, NewLife Finance

Resource Pool	Unit Load (Physician) (minutes per claim)	Unit Load (Hospital) (minutes per claim)	Unit Load (60%–40% mix) (minutes per claim)
Mailroom clerk	1.00	1.50	1.20
Data-entry clerk	5.00	6.00	5.40
Claims processor	8.00	8.00	8.00
Claims supervisor	2.50	4.00	3.10

Table 5.5 lists again the unit loads of the two types of claims. The fourth column, listing the unit loads of the 60%–40% mix, is computed by taking the weighted average of the previous two columns. For example, the unit load for the mailroom clerk equals $60\% \times 1.0 + 40\% \times 1.5 = 1.20$.

We can compute the effective capacity of the product mix in the same way as that of an individual product, using the averaged unit load instead of the individual values. This is illustrated in Table 5.6:

Table 5.6 Effective Capacity for 60%–40% Product Mix, NewLife Finance

Resource Pool (i)	Unit Load (minutes per claim) (T_i)	Effective Capacity of a Resource Unit (claims per minute) ($1/T_i$)	Number of Units in the Resource Pool (c_i)	Effective Capacity of a Resource Pool (claims per minute) (c_i/T_i)
Mailroom clerk	1.20	1/1.20 = 0.83	1	0.83
Data-entry clerk	5.40	1/5.40 = 0.185	8	1.48
Claims processor	8.00	1/8.00 = 0.125	12	1.50
Claims supervisor	3.10	1/3.10 = 0.32	5	1.61

As can be seen, the effective capacity in this case is 0.83 claims per minute, or 50 claims per hour. As expected, this falls between the effective capacities of the individual products (1 and 0.66 claims per minute respectively). However, the effective capacity of the mix is *not* equal to the 60%–40% weighted average of the respected effective capacities. Rather, it is the unit load, the inverse of the effective capacity, which is equal to the 60%–40% weighted average of the respected total unit loads.

In the case of NewLife, the pool of mailroom clerks is the bottleneck resource pool for every product mix. However, in general, a change in the product mix can affect not only the effective capacity but also the bottleneck.

5.3.2 Optimizing Profitability

Which of the two claims processed by NewLife Finance—physicians or hospitals—is more profitable? To answer this question we need to supplement the data on capacities for the two products with financial information concerning revenues and variable costs.

The **unit contribution margin** of each flow unit is *its revenue less all of its variable costs*. For instance, if the revenues per unit for physician and hospital claims are \$5.50 and \$6.75 per unit respectively, and that the variable costs are \$0.5 and \$0.75, then the unit contribution margins for the two products are $5.5-0.5 = \$5$ and $6.75-0.75 = \$6$ per unit, respectively.

On first sight it may seem that the product with the highest unit contribution margin is the most profitable, and thus one may conclude that hospital claims are more profitable than physician claims ($6 per unit is more than $5 per unit). However, as we see below, assessing the profitability of products solely on the basis of contribution margin per unit ignores the essential role that capacity plays in determining profitability. This is demonstrated in Example 5.5:

EXAMPLE 5.5

Table 5.7 summarizes the information about the two products:

Table 5.7 Effective Capacity and Contribution Margins, NewLife Finance

	Physician Claims	Hospital Claims
Effective Capacity (units per hour)	60	40
Contribution margin ($ per unit)	5.00	6.00

Consider first the case of processing only physician claims. Since the capacity is 60 per hour, and the margin per unit is $5, we can generate $60 \times 5 = \$300$ per hour. On the other hand, if we process only hospital claims, we can generate $40 \times 6 = \$240$ per hour. Thus, even though the contribution margin per unit for hospital claims is larger, it is the less profitable product. Clearly in this case, the fact that we can make a higher profit on each unit of hospital claims is more than offset by the fact that we can process fewer units per hour.

As the example demonstrates, the relevant criterion in determining the profitability of products is not the contribution *per unit* but the contribution *per unit of time*. In essence, this metric corresponds to viewing capacity and throughput in terms of financial flows rather than in terms of flow of physical units ($300 of contribution margin per hour rather than 60 claims per hour). The concept combines the relevant financial information, namely contribution margin per unit, with the operational concepts of capacity and throughput.

In practice, product mix problems are likely to be larger, more complicated, and involve other considerations, such as demands for the various products and other marketing constraints. In Appendix 5.2 we briefly present a general methodology, called Linear Programming, for handling such issues.

5.4 CAPACITY WASTE AND THEORETICAL CAPACITY

In most cases, the routine operation of a process involves a considerable amount of capacity waste due to factors such as resource breakdown, maintenance, quality rejects, rework and repetitions, setups between different products or batches, non-value-adding activities, and so forth. The unit load, which is used to determine effective capacity, is an aggregation, given the way the resources are currently being utilized, of the "productive" as well as the "wasted" time. However, if capacity waste is large, we may want to turn our attention to waste elimination; and thus, it is useful to "segregate" the wasted capacity. We discuss this issue in this section.

5.4.1 Theoretical Capacity

The **theoretical unit load of a resource unit** is *the minimal amount of time required to process a flow unit, if all waste were eliminated*. The **theoretical capacity of the resource unit** is *the reciprocal of the theoretical unit load*. It represents the maximum sustainable