3. Starting with an mrp Model

Rather than creating a model from scratch, we begin with a venerable model called *materials requirements planning*. The model is often referred to as mrp with lower case letters (or sometimes "little-mrp") to make clear the distinction between mrp and MRP II. We will look at MRP II later.

The mrp model comes from a production planning perspective rather than an optimization perspective, but after we understand how it works, we can create an optimization model that corresponds to it. This model is a useful starting point for further modeling. So we will first understand the mrp model as it was originally given. It is a very practical model, so we will introduce an example early on. One does not need an optimization model for mrp, but we will use our model to further understand the limitations of mrp and we will use it as a basis for more sophisticated models.

The mrp model uses a lot of data about items and components. The term *Stock Keeping Unit* (SKU) is used to refer to items that are sold as well as components and components of components, etc. For each SKU we need to know

- the *lead time*, which is an estimate of the time between the release of an order to the shop floor or to a supplier and the receipt of the items;
- if there is a minimum production quantity (referred to as a *minimum lot size* for items that are manufactured in-house) or if there is a minimum order quantity for purchased items;
- the current inventory levels (for simplicity we include items scheduled for receipt during runs of mrp in earlier periods);
- components needed, which is often referred to as a bill of materials (BOM).

This list of data seems short, but can be very hard to obtain and maintain. The fact that mrp requires production personnel to provide and maintain these data is one of the reasons that mrp is often very popular with accountants.

3.1 An Example

We use a very small example to illustrate the notation. Suppose that there is a single end item with SKU AJ8172 that has a bill of materials as shown in Fig-

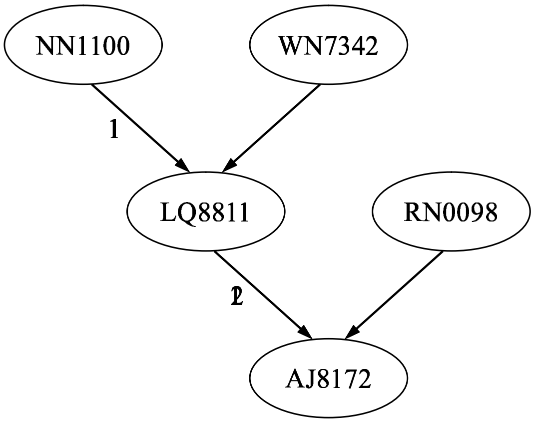

Fig. 3.1. BOM for a Simple Example

ure 3.1 and the properties given in Figure 3.2. In this figure, the components that are ordered from outside suppliers (e.g., RN0098) do not have a list of their components given since the mrp model typically stops at organizational boundaries.

While Figure 3.1 has a single end item, Figure 3.3 shows a slightly larger example that emphasizes that the product structure for mrp can be fairly general. In this example, there are two end items: AJ8172 and TR1777. They both use the component RN0098. The assembly of AJ8172 requires 1 of them and TR1777 requires three items of RN0098. We will generally make use of the simpler example, but the reader should bear in mind that mrp is intended for large numbers of SKUs with potentially many shared components or subassemblies.

3.2 mrp Mechanics

Materials Requirements Planning was defined operationally. Although it was invented before the term Just-in-Time (JIT) was popularized, in some sense mrp is a just-in-time system, even if it is not generally considered to be a "JIT" system. Production is planned to be done as late as possible but no later.

We make use of the so-called "low level" coding provided by mrp packages. This is an ordering of the parts such that the list begins with end-items and no item appears in the list before an item that contains it as a component. We assume that the parts are sorted in low level code order. We then proceed through the parts and for each one, we anticipate the need for lots to be ordered as inventories are depleted. Once we know when lots will be needed, we can subtract the lead time to determine when the order must be

1. AJ8172:
   Production Lead Time: 2 days
   Minimum Lot Size: 100
   Components: 2 LQ8811, 1 RN0098
   Initial Inventory: 90
2. LQ8811:
   Production Lead Time: 3 days
   Minimum Lot Size: 400
   Components: 1 NN1100, 1 WN7342
   Initial Inventory: 300
3. RN0098:
   Order Lead Time: 4 days
   Minimum Order Quantity: 100
   Components: N/A
   Initial Inventory: 100
4. NN1100:
   Order Lead Time: 1 day
   Minimum Order Quantity: 1
   Components: N/A
   Initial Inventory: 0
5. WN7342:
   Order Lead Time: 12 days
   Minimum Order Quantity: 1000
   Components: N/A
   Initial Inventory: 900

Fig. 3.2. Data for a Simple Example

released to the floor or the vendor. These release dates cause additions to the demand for component parts. When we eventually process the component SKUs, plans will be made to meet the accumulated demands for them, thus creating demands for their components and so on. Time must be broken into *buckets* such as days or weeks in an mrp model.

We can think of mrp logic as a plan for sequential toppling of a line of dominoes. If you know when you want the last domino to fall and you know the time for each of the dominoes to knock down the next one, you can calculate when the first domino should fall in order to achieve the desired result. The same logic underpins mrp.

Continuing with the example in Figure 3.2, suppose that the demand for AJ8172 in the next eight periods is 20, 30, 10, 20, 30, 20, 30, and 40. This results in the mrp production/inventory plan for AJ8172 given in Figure 3.4. We plan for receipts in the period when the inventory would be depleted without them. We then subtract the lead time in order to determine the last possible release date. Since we begin with an inventory of 90, the inventory would be depleted in period 5, so we plan to receive enough in period 5 to

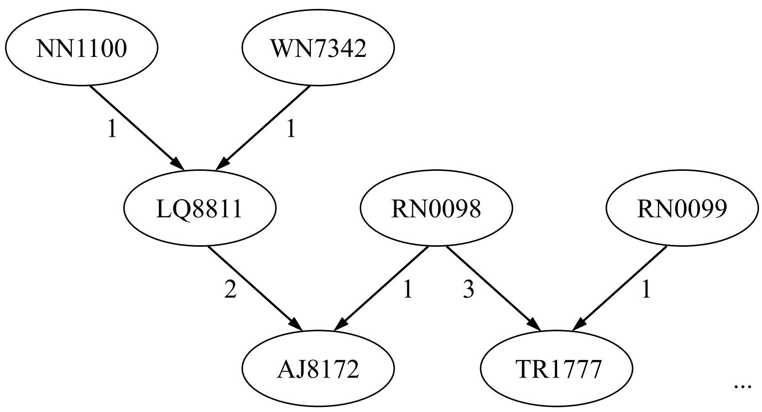

Fig. 3.3. BOM for a Slightly Bigger Example

avoid a shortage. In order to receive items in period 5, the order must be placed in period 3 because the lead time is 2. Since the minimum lot size is 100, the order is placed for that quantity.

	Day							
AJ8172	1	2	3	4	5	6	7	8
Demand	20	30	10	20	30	20	30	40
Inventory Plan (90)	70	40	30	10	80	60	30	90
Planned Receipts					100			100
Planned Releases			100			100		

Fig. 3.4. mrp Plan for AJ8172 Using Data from Figure 3.2

The planned releases of orders for AJ8172 creates demand for its components. There will be a need for 200 LQ8811s and 100 RN0098s in day 3 and day 6. This is shown in Figure 3.5. The initial inventory of WN7342 is adequate for the eight day planning horizon.

3.3 mrp Data

Central to mrp is the BOM, which gives the components that are combined to make other components and, ultimately, each end item. There are many ways to describe and display a BOM. Since we are interested in abstract models, we use the data $R(i, j)$ to give the number of SKU i's directly needed to make

LQ8811	Day 1	2	3	4	5	6	7	8
Demand			200			200		
Inventory Plan (300)	300	300	100	100	100	300	300	300
Planned Receipts						400		
Planned Releases			400					

RN0098	Day 1	2	3	4	5	6	7	8
Demand			100			100		
Inventory Plan (100)	100	100	0	0	0	0	0	0
Planned Receipts						100		
Planned Releases		100						

Fig. 3.5. mrp Plans for LQ8811 and RN0098 as a Result of Figure 3.4 Plan

one SKU j. This notation presupposes that the SKUs are in a numbered list such as Figure 3.2 so that the words "SKU i" have meaning. Any numbering of the parts can be used in the optimization models, but we use a low-level-coding to be somewhat consistent with mrp software. We let P represent the number of SKUs in the list. We use the numbering in Figure 3.2, which assigns a 1 to AJ8172, a 2 to LQ8811, and so on. Hence, the value of $R(2,1)$ is 2. The values of $R(3,1)$, $R(4,2)$, and $R(5,2)$ are 1. All others (e.g., $R(4,1)$, etc.) are zero.

Remember that time must be broken into "buckets" in an mrp model. The buckets are typically weeks or days. In our simple example we use days and we use T to represent the number of days in our planning model. In a typical mrp application T might be on the order of a few months or even a few years. Since T gives the last time bucket that we consider in the planning process, it is sometimes referred to as the *planning horizon*.

We use the notation $LT(i)$ to indicate the number of time buckets that one can expect between issuing an order for production or shipment of SKU i and receiving it. At this point, we are intentionally a little bit vague about the exact meaning. For the example data, $LT(1)$=2, $LT(2)$=3, etc.

Lot sizes are a standard part of mrp. Later, we will advocate elimination of most lot sizes as externally supplied data. But for now we want to remain consistent with common mrp practice, so we let $LS(i)$ be the minimum lot size for SKU i. In the example data, $LS(2) = 400$. For SKUs such as NN1100 that can be ordered in any quantity, LS should be given as 1. Other lot sizing conventions are possible, such as requiring that production be in multiples of a lot size. However, our models use a minimum production lot, which is satisfactory for avoiding unduly small production quantities for an SKU.

Our final pieces of data are the external demands for item i in period t, which is given as $D(i,t)$. Demands are clearly needed for end items. Of course, in many situations there are external demands for components as well because they are distributed to the maintenance organization, sold as replacement parts, or perhaps sold to competitors. This set of external demands is often called the *master production schedule*.

The data needed for an optimized mrp model are summarized in Table 3.1. A large number, referred to as M, is needed to force the computer to make some-or-none decisions that are needed to enforce the minimum lot sizes. This can be any number, provided it is larger than any possible production quantity. To avoid excessive roundoff error, one should try to use a number for M that is not more than, say, a hundred or a thousands times larger.

P	Number of SKUs
T	Number of time buckets (i.e., the planning horizon)
$LT(i)$	Lead time for SKU i
$R(i,j)$	Number of i's needed to make one j
$D(i,t)$	External demand for i in period t
$I(i,0)$	Beginning inventory of SKU i
$LS(i)$	Minimum lot size for SKU i
M	A large number

Table 3.1. Data for the mrp Formulation

3.4 mrp Optimization Formulation

An optimization model is not needed to use mrp, but we can create one and then extend it. In other words, our goal is to create an optimization problem that matches mrp not for its own sake but to get started with models that match classic planning systems. Using this model as a starting point, it is easy to go on to more sophisticated models.

We really have only one decision variable, $x_{i,t}$, which is the quantity of SKU i to start or order in period t. In order to enforce lot sizing rules, we need something that indicates production of an SKU in period t. What is needed is an *indicator variable*. This is a somewhat advanced topic, but we need to tackle it now in order to model classic mrp.

Note there are only two things in optimization models: data and variables. Something that indicates that there will be production in period t is clearly not data. We create a variable, $\delta_{i,t}$, that will be one if any of SKU i will be started in period t. We have to put in constraints to force the computer to make this variable behave this way. At first glance, this variable is redundant with $x_{i,t}$, but not equivalent. We will see that it serves a different role.

The following constraints must hold for all $i = 1, \ldots, P$ and $t = 1, \ldots, T$.

- Demand and materials requirement:

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t} \left(D(i,\tau) + \sum_{j=1}^{P} R(i,j) x_{j,\tau} \right) \geq 0$$

- Lot size requirement:

$$x_{i,t} \geq \delta_{i,t} LS(i)$$

- Modeling constraint for production indicator:

$$\delta_{i,t} \geq \frac{x_{i,t}}{M}$$

- Integer constraint for production indicator:

$$\delta_{i,t} \in \{0,1\}$$

- Non-negative production:

$$x_{i,t} \geq 0$$

Fig. 3.6. Constraints for mrp

Figure 3.6 gives the constraints needed for a model of mrp. These constraints are fairly complicated, but they provide a nearly complete description of mrp. The first constraint requires that the sum of initial inventory and production up to each period has to be at least equal to the total of external demand and demand for assemblies that uses the SKU. The summation is to $t - LT(i)$ for each period (there will be one constraint for each value of t) because of work that must be started LT periods before it can be used to satisfy demand. The product $R(i,j)x_{j,\tau}$ anticipates the demand for SKU i that results when it is a component of SKU j. This product will turn out to be zero for a lot of i, j combinations, but that does not present any special difficulty for a computer.

The demand and materials constraint in Figure 3.6 could have been written

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) \geq \sum_{\tau=1}^{t} \left(D(i,\tau) + \sum_{j=1}^{P} R(i,j) x_{j,\tau} \right).$$

The use of algebra to rearrange the terms of a constraint is entirely a matter of taste. There is no effect on the solution or the computational effort. We chose to put a zero on the RHS to emphasize that mrp requirements are all firm. The main point is that the term

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau}$$

captures the production that will be completed up to time t, while the term

$$\sum_{\tau=1}^{t}\left(D(i,\tau)+\sum_{j=1}^{P}R(i,j)x_{j,\tau}\right)$$

is the total demand that will have occured up to the same time period.

The lot size constraint specifies that if there is any production of a SKU during a period it must be at least as much as the minimum lot size. The modeling constraint for the production indicator forces it to take a value greater than zero if there is production for the SKU in the period, which the integer constraint forces it to be either zero or one. Many modeling languages eliminate the need for these two constraints by allowing for *semi-continuous* variables, which must be either zero or above a threshold; see §8.3.1 for more information. The final constraint forces the computer to pick only production values that are not negative.

All we need in order to use an optimization package to accomplish mrp is an objective function. The objective in mrp is to make things as late as possible but no later. So one possible objective is to minimize

$$\sum_{i=1}^{P}\sum_{t=1}^{T}(T-t)x_{i,t}$$

that will result in things being made "as late as possible" and we count on the constraints to enforce the "but no later" part of the mrp model. There are better objectives than this, some of them will be described later. However, for now our goal is to mimic mrp. Stated in classic optimization form, the mrp problem is given in Figure 3.7.

In order to illustrate that these models can be implemented almost directly using a modeling language and make the notion of a modeling language concrete, Chapter 7 is provided. In that chapter, we demonstrate how some popular modeling languages can be used to implement this simple model **mrp**. We also show how the example data can be entered and an optimal solution obtained.

3.5 Discussion of mrp

Notice that we do not require integer valued production quantities. This particular model results in integer valued production quantities, provided that the demands and minimum lot sizes are integers. After we extend the model, we could worry that it might be possible for the computer to report that the optimal plan involves producing 123.3 of some product in a period and this might not make sense for that particular product. However, this model is intended to be a plan and not a schedule. Furthermore, unless the

Minimize:

$$\sum_{i=1}^{P} \sum_{t=1}^{T} (T-t) x_{i,t} \qquad \textbf{(mrp)}$$

subject to:

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t} \left(D(i,\tau) + \sum_{j=1}^{P} R(i,j) x_{j,\tau} \right) \geq 0 \qquad i=1,\ldots,P,\ t=1,\ldots,T$$

$$x_{i,t} - \delta_{i,t} LS(i) \geq 0 \qquad i=1,\ldots,P,\ t=1,\ldots,T$$

$$\delta_{i,t} - \frac{x_{i,t}}{M} \geq 0 \qquad i=1,\ldots,P,\ t=1,\ldots,T$$

$$\delta_{i,t} \in \{0,1\} \qquad i=1,\ldots,P,\ t=1,\ldots,T$$

$$x_{i,t} \geq 0 \qquad i=1,\ldots,P,\ t=1,\ldots,T$$

Fig. 3.7. mrp Model

production quantities are very small, the production quantity can typically be rounded to the nearest integer producing errors that are much smaller than the estimation errors involved in the data for the problem.

Classic versions of mrp, as well as our optimization formulation, are intended only for certain types of bills of material. Bills of material where multiple SKUs are combined to make a new SKU work well. This is the case with many things such as computers and cars. Products where one item is used to produce multiple items are referred to as *divergent* BOMs. For divergent portions of the BOM, the entries in $R(i,j)$ will be fractional. For example, suppose SKU TB4-16 is a sixteen foot board and TB4-8 is an eight foot board. In this example R(TB4-16, TB4-8) will be one half. More complicated situations such as cycles in the BOM, require modification to the basic scheme. Such modifications are beyond our scope.

We close our discussion of mrp by considering some of its troubles along with some of its virtues. Computerized planning systems based on mrp have been in use for decades and mrp logic remains at the heart of the production planning module of many modern *Enterprise Resources Planning* (ERP) systems. Any system in broad use is sure to have some troubles and virtues.

3.5.1 Troubles

Of course, no production planning model can be perfect, but there are a number of well-known and very severe problems with mrp as we have described it. Perhaps the three most serious problems are

- the actual time to complete an order is usually a function of congestion rather than of the SKU,

- lot sizing can cause *nervousness*,
- there are no capacity constraints.

It turns out that these problems are related. The lack of capacity constraints results in a need for lot sizing and exacerbates the variable lead time problem considerably.

The underlying model for mrp is based on common thinking among data processing professionals in the early days of the application of computers to business problems. Data are collected into a database and then processed. The trouble with lead times is that they are given as static data. However, the time from issuance of an order to completion depends mainly on what work has to be done before the order "gets to the front of the line" of orders that await processing. Lead times are often weeks when the actual production time is hours.

One reason that lead times have to be much longer than production times is to account for machine failures. But even if the capacity is not overutilized and if the production resources are reliable, lead times can be variable due to waiting lines that form in front of bottleneck resources. This issue can be important but it is difficult to deal with, so it must be deferred to a later research oriented chapter. Apart from the (considerable) mitigation due to using capacity constraints, discussion the problem of variable lead times is deferred to Chapter 9.1.

A major part of the reason that lead times must be so long is to guard against periods when the resources are overbooked. This can be mitigated considerably by the use of capacity constraints, which are not included in mrp. As we shall see, MRP II was developed as a partial solution to this problem. We will develop optimization models that address capacity constraints much more effectively.

Nervousness is a phenomenon where small changes in demand result in large changes in production plans due to lot sizing rules. Consider the example given in §3.2. If the company received an order for ten more AJ8172's in period four, it would cause production to be shifted earlier and production for the entire lot of 400 LQ8811's would be shifted one period earlier as well. The dynamics of nervousness makes it demoralizing for production workers and as a result they often ignore the production plans produced by mrp systems. Who can blame them? First the production plan given in §3.2 is released to the floor, then a modest size order causes large changes in the schedule; meanwhile order cancellations can have a similar effect.

One response to this problem is to produce a "frozen zone" for end items that forbids changes in the schedule for some number of near-term time buckets. This seldom works. The reason is that customers do not care about mrp induced nervousness. They demand flexibility. As a result, orders get changed whether the master production schedule reflects it or not. In the worst case (a common case, unfortunately) expeditors manipulate inventories and production schedules to respond to customer needs. Once this practice begins, it is hard to stop. The mrp system assumptions are no longer valid at all, since components originally produced for one end item are used in another. Eventually, day-to-day scheduling is essentially done by expeditors and the mrp system is reduced to a raw materials procurement aid.

A major reason to use large lot sizes, or lot sizing rules at all, is to ensure that not too much productive capacity is used to changeover from one SKU to another. Lot sizes are at best a blunt instrument for accomplishing this. In reality, one cannot know how big the lot sizes need to be until the production schedule is complete. For resources that do not happen to be capacity constrained in a time period, the addition of more changeovers will not adversely affect throughput, so smaller, more flexible lots can be used. Conversely, for resources that are capacity constrained, a delicate tradeoff is needed between the use of small lots to provide flexibility in meeting customer needs and the use of large lots to maximize throughput. Setting lot sizing rules *a priori* is hardly delicate.

The solution is to simultaneously create production schedules and determine lot sizes that respect capacity limitations. This solution shows the final problem with mrp: there is nothing in mrp to guarantee production schedules that can actually be executed. That is to say, there is an excellent chance the production plans for many SKUs far exceed the capacity of the resources used to create them.

3.5.2 Virtues

Having said all that, we can say that an mrp model can still be very useful. For one thing, it is usually much better than no planning model at all. This is particularly true in industries with changing demand patterns where standard orders cannot be used. An mrp model can provide a good starting point for planning and for the ordering of raw materials.

The materials requirements estimates provided by an mrp system can be useful to the purchasing department because they provide an "earliest-case" estimate of requirements. To the extent that the plan exceeds available capacity, the actual production will take place later than the mrp plan. The idea of buying materials based on an "earliest-case" estimate is not consistent with modern notions of "just-in-time," but it is better than being late. Furthermore, for commonly used low cost raw materials, the mrp plan can provide essential information for purchasing.

Another reason to consider the use of an mrp model is the same reason that causes us to begin with it in this book. It is a simple model that can be the starting point for more sophisticated models. If you plan to use an mrp model in production and there are minimum lot size requirement, you should see §8.3.1 for a method of streamlining the model.

4. Extending to an MRP II Model

MRP II was inspired by shortcomings in mrp, and as a result the data processing orientation is preserved in MRP II. As was the case with mrp, we first explain the concepts behind MRP II, then we develop an optimization model to mimic and improve its behavior. After we have this model in hand, we extend it to produce a model that can give us production plans that trade off alternative capacity uses, holding inventory and tardiness in an optimized way. The letters MRP in MRP II stand for *Manufacturing Resources Planning* to make it clear that resources are considered in addition to materials as in mrp. The word "resource" is used to emphasize that any type of productive capability can be considered, not just machines. The Roman number II is intended to make it clear that it is an extension to materials requirements planning (mrp).

4.1 MRP II Mechanics

There are a number of well-known deficiencies in the model that underlies mrp. Potentially the most severe is the fact that it ignores capacity. To discuss this issue it is useful to remember that we are making a distinction between planning and scheduling as we described in §1.3. Although we have introduced it as a planning tool, mrp is also often used as a scheduling tool as well. A severe problem is that there is no guarantee that there will be enough capacity to actually carry out the plan produced by mrp. In fact, for capacity constrained production systems, it is seldom possible to implement an mrp plan as a schedule. This is debilitating when mrp is used as a scheduling tool, but is also bad for mrp as a planning tool because the absence of capacity considerations can make the plans so unrealistic that they are not useful.

The data processing professionals who were developing and selling mrp software in its early years recognized this deficiency and MRP II was developed in response to it. The database for mrp is extended to include routing and capacity information. Each production resource is entered into the database along with its maximum production during a time bucket. We will refer to the maximum production by a resource during a time bucket as its *capacity.* The list of resources used to produce a particular SKU is known as the *routing* for the SKU.

With this information the data processing specified by MRP II can be carried out. The processing begins by executing mrp to determine a production plan. Then, for each time bucket each SKU is "followed along its routing" and the utilization of each resource is updated. In the end, those resources whose capacity would be exceeded by the mrp plan are identified. The user can be given long and potentially confusing reports that dice and slice the following data for infeasibilities:

- resources,
- time buckets, and
- SKUs.

Reports are also generated for end-items that would use these "offending" SKUs and the time buckets in which the end-items would be produced.

The information concerning capacity infeasibilities can be used by the planner or some software to attempt to change the input data so that a feasible plan results. The most common method is to "repair" the master production schedule (i.e., change the timing of the external demand input data.) Such repairs are difficult, so one of our goals will be to use optimization models to produce feasible plans in the first place and eventually improve them later on. But before we do that, we provide the details of MRP II for the example given earlier. The example is too small to provide a realistic view of the issues associated with creating feasible plans in a real production setting, but it will be good enough for us to illustrate the mechanics of MRP II.

Before we can get started with a practical example, we need to deal with the issue of units of measurement. In a specific production facility capacity may be measured in hours, or tons, or pieces, or something else. Since we want to create abstract models, we will designate the capacity of every resource during every time bucket to be one. This will allow us to state resource utilizations as fractions in our models. We will represent the fraction of the capacity of resource k in one time bucket used by production of one SKU i as $U(i,k)$.

Suppose that we have only two resources for in-house production: HR-101 and MT-402. Further suppose that HR-101 has 80 hours of capacity per day and that MT-402 can produce 300 items per day. A capacity of 80 hours in one day would typically be achieved by a crew of 10 people (or "human resources"). There are only two items in Figure 3.2 that are produced in-house. Suppose that AJ8172 requires 10 minutes of HR-101 and that LQ8811 requires 5 minutes of HR-101 and also makes use of MT-402. Using common slang, we would say that LQ8811 is routed through both HR-101 and MT-402, while AJ8172 is routed only through HR-101. Note that even though the word "through" is used, it might be the case that HR-101 is a person or group of people who move from job to job. This routing is shown in an abstract way in Figure 4.1.

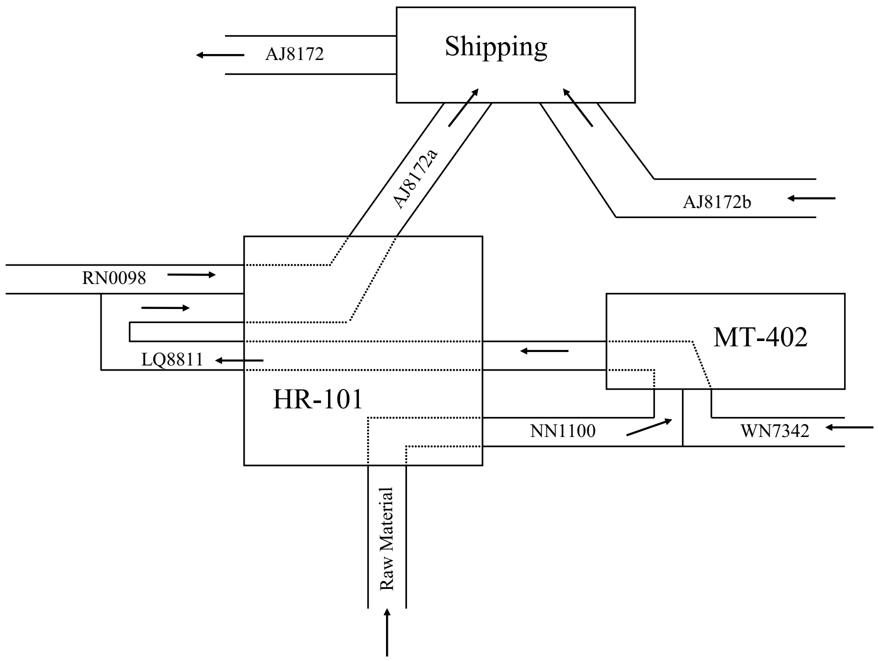

Fig. 4.1. Simple Routing Diagram

To use our notation, label HR-101 as resource 1 and MT-402 as resource 2. We then calculate that the utilization fraction of AJ8172 on HR-101 is

$$\frac{(10/60)\text{ hours}}{80\text{ hours}} = \frac{1}{480}.$$

Later, we will refer to this as $U(1,1)$. In similar fashion, we can compute the utilization fraction for one unit of LQ8811 of HR-101 and MT-402 as $\frac{(5/60)}{80} = \frac{1}{960}$ and $\frac{1}{300}$, respectively. We have summarized these utilization fractions in Table 4.1.

	Fraction Used By	
Resource	AJ8172	LQ8811
HR-101	1/480	1/960
MT-402		1/300

Table 4.1. Fraction of Capacity Utilized to Make Each SKU at Each Resource

Period	Resource	Utilization
3	HR-101	$\frac{100}{480} + \frac{400}{960} = \frac{5}{8}$
	MT-402	$\frac{400}{300} = \frac{4}{3}$
6	HR-101	$\frac{100}{480} = \frac{5}{24}$

Table 4.2. Anticipated Capacity Utilization for MRP II Example

This information allows us to compute the utilizations that would result from the mrp plan we developed in §3.2, which is given in Figures 3.4 and 3.5. Production of AJ8172 will utilize

$$100 \times \frac{1}{480}$$

of the capacity for HR-101 in periods three and six. In period three, production of LQ8811 will utilize $\frac{400}{960}$ of the capacity for HR-101 and $\frac{400}{300}$ of the capacity for MT-402. We have summarized the anticipated capacity utilization for the mrp production plan in Table 4.2.

We can see that the plan would not be possible because there is not enough capacity for MT-402 in period three. One solution is to move some of the production of LQ8811 to period two. Other possibilities include authorizing overtime for MT-402 or subcontracting some of the production of LQ8811.

It is not too hard to fix things with a small example, but when there are hundreds or thousands of SKUs interacting on many resources, it can be very difficult. In the next sections we describe ways to use optimization models and software to find good, and perhaps optimal, solutions based on the MRP II model.

4.2 MRP II Data and Constraints

As was the case with mrp, MRP II did not begin as an optimization model. However, we can create an optimization model that will mimic its behavior and do more. To be specific, we can mimic its intended behavior. That is, we can schedule things as late as possible without violating capacity constraints. The objective function for mrp is retained, but additional data are needed for constraints. The data needed to mimic MRP II are given in Table 4.3.

The data requirements are nearly the same as for mrp except that we have dropped the lot sizing information (for now) and added information about utilization. We use the same variables as for mrp, namely $x_{i,t}$, which is the quantity of SKU i to start or order in period t. We will not need $\delta_{i,t}$ unless we need to add information about changeovers. The major change from the

P	Number of SKUs
T	Number of time buckets
K	Number of resources
$I(i,0)$	Beginning inventory of SKU i
$LT(i)$	Lead time for SKU i
$R(i,j)$	Amount of SKU i needed to make one j
$D(i,t)$	External demand for SKU i in period t
$U(i,k)$	Fraction of resource k needed to make one unit of SKU i

Table 4.3. Data for a Simple MRP II Formulation

mrp model is the addition of a capacity constraint. The MRP II constraints are as follows:

- Demand and materials requirement for all times t and all SKUs i:

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t}\left[D(i,\tau) + \sum_{j=1}^{P} R(i,j)x_{j,\tau}\right] \geq 0$$

- Constrain capacity for some (or all) resources k and times t:

$$\sum_{i=1}^{P} U(i,k)x_{i,t} \leq 1$$

- Non-negative production for all SKUs i and times t:

$$x_{i,t} \geq 0$$

Stated in classic optimization form, the MRP II problem is given in Figure 4.2.

For the purpose of discussion it is useful to isolate the two most important constraints and refer to them by name. We will refer to

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t}\left[D(i,\tau) + \sum_{j=1}^{P} R(i,j)x_{j,\tau}\right] \geq 0$$

as the *materials requirements constraint*, and we will call

$$\sum_{i=1}^{P} U(i,k)x_{i,t} \leq 1$$

the *capacity constraint.*

If there are minimum lot sizes for some SKUs, then we must add the following constraints for those SKUs, i:

Minimize:

$$\sum_{i=1}^{P} \sum_{t=1}^{T} (T-t) x_{i,t} \qquad \textbf{(MRPII)}$$

subject to:

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t} \left(D(i,\tau) + \sum_{j=1}^{P} R(i,j) x_{j,\tau} \right) \geq 0 \qquad i=1,\ldots,P, \ \ t=1,\ldots,T$$

$$\sum_{i=1}^{P} U(i,k) x_{i,t} \leq 1 \qquad t=1,\ldots,T, \ \ k=1,\ldots,K$$

$$x_{i,t} \geq 0 \qquad i=1,\ldots,P, \ \ t=1,\ldots,T$$

Fig. 4.2. MRPII Model

$$\begin{aligned} x_{i,t} - \delta_{i,t} LS(i) &\geq 0 && t=1,\ldots,T \\ \delta_{i,t} - \frac{x_{i,t}}{M} &\geq 0 && t=1,\ldots,T \\ \delta_{i,t} &\in \{0,1\} && t=1,\ldots,T \end{aligned}$$

where $LS(i)$ is the minimum lot size for SKU i and M is a large number as in the mrp model.

If capacities are expected to change for a resource, k, then the capacity constraints for these resources must be written as

$$\sum_{i=1}^{P} U(i,k,t) x_{i,t} \leq 1.$$

4.3 Discussion of MRP II

Using classic MRP II software, problem **MRPII** would not be solved directly. Instead, problem **mrp** would be solved and then the capacity constraint for the **MRPII** model would be checked. To be more specific, suppose the solution to the problem **mrp** was given as $X(i,t)$ for the production of SKU i to plan to start in time t. In other words, the result of solving problem **mrp** provides values for the decision variables. Once these values are known, they become data for subsequent processing so we use an upper case X to indicate the values that are given.

Given the **mrp** solution, those SKUs for which

$$\sum_{i=1}^{P} U(i,k) X(i,t) > 1$$

can be identified as those that would violate the capacity constraints. They would be the subject of reports and the production planner would change the data in an attempt to find a solution to mrp that was also feasible for **MRPII**. By "change the data" we mean that due dates, lot sizes, and capacity utilizations would be changed. Due dates for end items are typically adjusted to be at a later date. These data would then be given to the software that uses the new data to compute a new solution to **mrp** and then checks the constraints for **MRPII**.

This process is very hard work for the planners. Even though MRP II software provides numerous reports, it is often still not possible for the planners to produce a capacity feasible schedule given that they often have only a few hours to produce it. With a few dozen time buckets and a few thousand SKUs, the problem is just too big to be solved easily by people, even if assisted by software that can check the constraints and solve problem **mrp**.

An important thing to note is that the classic iterative MRP II solution process that we have described results in an implicit, rather than an explicit objective function. The actual objective function that is implied by the MRP II solution process is not easily determined because the solution is obtained by changing the data rather than finding the best solution given the best estimate of the data. As the planners change the data during the struggle to find a good, feasible solution the objective function implicit in the solution meanders without clearly articulated direction. The fact that some modern ERP software makes it easy to change the data (for example to increase lead times) emphasizes the importance of this point.

In spite of the severe difficulties, at the time of this writing MRP II logic is central to modern ERP systems. Many academics are of the opinion that MRP II should be done away with. If they mean that we should dump the solution methods, then we agree, but if they mean dump it completely, then they are hopelessly misguided. Regardless of how one articulates the methods, one should include the requirements constraint and the capacity constraint. They represent physical reality. The processing for MRP II is not normally explained by giving the constraints as we have shown them, but the constraints are there. They have to be. They have to be there in any other system as well. If the software does not include them explicitly or implicitly, then the planners will have to enforce them or else produce a plan that is very unrealistic and therefore not very valuable. Ultimately, the constraints are present in the production process.

Direct solution of the optimization model is a much better idea and this is the basis of much of the newest planning software that is sold as modules or add-ins for ERP systems. In practice, the problem is bigger and harder to solve than the simple **MRPII** model that we have presented. However, **MRPII** provides us with a good jumping off point for more sophisticated models because it mimics a widely used planning tool.

We can and will embed these constraints in a model that captures costs and constraints that are important to the manufacturing organization or the supply chain. By solving the optimization problem directly, we can include in the objective function guidance for solutions. We want to find solutions that correspond to the goals of the organization over and above merely satisfying the two constraints.

4.4 Changeover Modeling Considerations

The simple **MRPII** model will be the basis for many additional features. However, we might also want to remove features. For example, not all resources need to be modeled. Often, it is easy to see that some resources do not pose a problem. Such resources should simply be omitted from the model.

One feature that has been dropped from the **mrp** model in creating the **MRPII** model is lot sizes. Usually, the valid reason for minimum lot sizes is that a significant effort or cost is required to changeover for production, hence small lots might not be cost effective. Setting a fixed lot size is a crude response to this problem, so we will try to build models that take changeovers explicitly into account.

However, there are cases where minimum lot sizes are needed. For example, some chemical processes must be done in batches with a minimum size. If needed, the δ variables can be put in for the SKUs that require lot sizes along with the lot sizing constraints for those SKUs.

In many production environments, the proper modeling of capacity requires modeling of changeovers. By the word “changeover” we mean the effort required to switch a resource from the production of one SKU to the production of another. In fact, it is changeover avoidance that results in the need for lots that are larger than what would be needed to satisfy immediate customer demands. Changeover modeling can be quite involved.

The first thing that we will need is $\delta_{i,t}$, which will be equal to one if any of SKU i will be started in period t. This, in turn, requires that we include the constraints introduced for **mrp** to enforce the meaning of the δ variables.

- Modeling constraint for production indicator for all SKUs i and times t:

$$\delta_{i,t} \geq \frac{x_{i,t}}{M}$$

- Integer constraint for production indicator for all SKUs i and times t:

$$\delta_{i,t} \in \{0, 1\}$$

4.4.1 A Straightforward Modification

In a simple model of changeovers, we use the data given in Table 4.4. As usual we do not insist that these data be given for every resource or every

SKU. In particular, we would expect that W values will essentially be zero for most or all SKUs. If they are all zero, then of course there is no need to add them to the model. The idea is that when one changes from one SKU to another, some material can be destroyed (i.e., wasted.) A very common example is that when production for SKU j is begun, a few items of that SKU have to be destroyed for quality control testing or a few defective items are produced while the machine is adjusted. For example, suppose for SKU 32, 100 items are created during a changeover that have to be discarded. In this case $W(32, 32)$ would be 100. In more complex situations, only some of the components of SKU j are wasted during the changeover so we allow for a more general data element, $W(i, j)$, to facilitate this.

$S(i,k)$	Fraction of resource k used to changeover to SKU i
$W(i,j)$	Waste of SKU i to changeover to SKU j

Table 4.4. Data for Changeover Constraints

The following replacements for the requirements and capacity constraints are needed:

- Demand and materials requirement for all times t and all SKUs i:

$$\sum_{\tau=1}^{t-LT(i)} x_{i,\tau} + I(i,0) - \sum_{\tau=1}^{t} \left[D(i,\tau) + \sum_{j=1}^{P} (R(i,j)x_{j,\tau} + W(i,j)\delta_{j,\tau}) \right] \geq 0$$

- Constrain capacity for all resources k and times t:

$$\sum_{i=1}^{P} (U(i,k)x_{i,t} + S(i,k)\delta_{i,t}) \leq 1$$

4.4.2 Production that Spans Time Buckets

In other environments, the fact that an SKU is produced in a time period does not necessarily mean that there will be a changeover in the time period because the production run might span multiple time periods. If this is the case, the expression $W(i,j)\delta_{j,\tau}$ will overstate the requirements and the expression $S(i,k)\delta_{i,t}$ will overstate the capacity utilization. The proper correction depends somewhat on the situation. One solution is to introduce variables $\gamma_{i,k,t}$ that are usually zero, but take the value one if SKU i will be the last product produced on resource k in time bucket $t-1$ and the first produced in time bucket t. We can change the capacity constraint for some (or all) resources k and times t to be

$$\sum_{i=1}^{P} [U(i,k)x_{i,t} + S(i,k)(\delta_{i,t} - \gamma_{i,k,t})] \leq 1.$$

In order to force the γ variables to have the appropriate meaning we must require that they be either zero or one and we must add the following constraints for all t and for those k and i of interest:

$$\delta_{i,t-1} + \delta_{i,t} \geq 2\gamma_{i,k,t} \tag{4.1}$$

$$\gamma_{i,k,t}/M \leq U(i,k) \tag{4.2}$$

$$\sum_{i=1}^{P} \gamma_{i,k,t} \leq 1. \tag{4.3}$$

The constraints labeled (4.1) allow γ to be one for SKU i on resource k only if there is production of SKU i in both periods. Constraints (4.2) ensure that we only set γ to one for SKUs i that are to be routed to resource k, which is done mainly to avoid spurious values of γ that can be confusing when reading the solution. Constraints (4.3) ensure that at most one product can span the time boundary on a specific resource k. In order to have constraints (4.1) make sense, we must define $\delta_{i,0}$ as data, which is a departure from our usual naming conventions that use parenthesis for data indexes. In other words, $\delta_{i,0}$ is set to one for the SKUs that are scheduled for production in the last period before the first period for which the model applies. In many cases, the planning model begins with the “next” period so the $\delta_{i,0}$ values would reflect the schedule for the “current” time bucket.

The drawback to this scheme is that increases in the number of integer variables can dramatically increase the time needed for solving the problem. This can be mitigated somewhat by modifying the constraints and providing the variables only for resources k that are bottlenecks with significant changeover requirements. In other words, we do not create variables $\gamma_{i,k,t}$ for those k that do not have significant changeovers or that do have plenty of capacity in every period.

4.4.3 Parallel Machines

When there are a number of identical machines that are grouped together, it is possible for an SKU to be run on more than one machine at a time. This means that the amount of capacity consumed by changeovers depends on the number of machines used.

This presents yet another modeling dilemma. To model parallel machines properly, a large number of integer variables must be introduced. If there are many such resource groupings, and if they are potential bottlenecks and many of the parts use the bottlenecks, then special purpose sequencing and assignment algorithms must be used. This sort of situation is beyond the scope of this book. However, if there are only one or two resource groups with parallel machines that are important then it can be reasonable to add their descriptions to a linear model.

One way to model them exactly is to treat each machine as a separate resource and then every SKU that can use the set of parallel machines will have alternative routings. Modeling of alternative routings is discussed in §6.1. Unless the number of machines is fairly small, this may add too much complexity.

A computationally simpler solution is to use an approximation. We can model the changeover time with an approximation based on typical uses of the machines. For example, we might let the changeover time be a multiple of the single machine changeover time for some SKUs to anticipate that more than one machine will be changed over for those SKUs. A somewhat better approximation can be obtained by modeling the effect of quantity on the utilization rather than modeling the changeovers explicitly.

One can use historical or engineering data to fit a function that predicts the utilization (including changeovers) based on the quantity produced. Consider a situation where a group of machines is typically used to process one family of SKUs and there is only a minor changeover required between members of this family. However, the machine group might also process a few other SKUs requiring a significant changeover. Consider one such SKU, i, for a group of parallel machines that is resource k. If just one part is produced the utilization will be quite high because a changeover will be required in addition to the processing time. If the same machine can be used to produce up to 100 including a changeover, then the per part utilization (given as $U(i,k)$) will be a decreasing function of x up to 100 parts, then increasing sharply, then decreasing again. Modeling this sort of non-linear constraint is deferred to §8.3.

4.4.4 Sequence Dependent Changeovers

In some production environments, the changeover effort depends not only on the SKU that is to be produced, but on the previous part as well. The data for this can be modeled by changing the S data element to have an additional subscript. We can use $S(i,j,k)$ to indicate the capacity used to switch production on resource k from SKU i to SKU j. It is then necessary to model the production sequence.

In order to make use of linear models, we need to have indicator variables for the sequence. For example, we could add a subscript to the δ variables so that the production sequence is indicated. Under this scheme, the variable $\delta_{i,j,t}$ is equal to one if product i is produced immediately before product j in period t. This results in a set of rather messy constraints. There are a number of choices for writing the constraints that force creation of a sequence. We give one method here that relies on creation of a fictional SKU number 0 that serves as the first and last SKU in the sequence for each period. Presumably $S(0, j, k)$ and $S(i, 0, k)$ can have an average changeover utilization value for the affected resources. Clearly, if there are only a few SKUs produced in each time bucket, then this model will not be very accurate and an even more complicated model may be needed.

This adds many new constraints and there will be a dramatic increase in the number of variables. If there are P products and T time periods, then there would be more than TP^2 integer variables. For even a small company with 12 periods in the planning horizon and 100 SKUs affected by sequence dependent changeovers, this is a large number of integers. If there are 50 periods and 1000 SKUs that must be modeled as having sequence dependent changeovers, then this model will be way too large for standard optimization software to handle in a straightforward way. The dramatic increase in model size might not be worth the increase in accuracy. But in some situations, bottleneck resources have significant changeover time requirements and different sequences result in dramatically different capacity utilizations. The models are even more complex if it is necessary to also capture the ability to continue production of an SKU across a time bucket boundary. In such situations, it may be necessary to use heuristic search methods to address the sequencing problems (see §8.4).

4.4.5 A Few Remarks About Changeovers

Attempts at proper modeling of changeovers illustrate that modeling is largely an art form. At some point, we must make simplifying assumptions. There are more complications than the ones described here. The ones we have described should help readers to formulate constraints that capture the essence of their particular situations.

The process of developing the models can provide spinoff benefits. For example, we can gain a small but useful insight by considering the difficulties imposed by changeovers. It has always been obvious that an increase in the amount of time available for production of parts would result from engineering changes to production processes that reduce the time required for each changeover. Furthermore, the nuisance of changeovers is felt by production management as well as shop floor workers. Over the past few decades there has been a significant engineering effort worldwide associated with reducing the time required to perform changeovers. Many of the techniques are under the rubric SMED, which is an acronym for single minute exchange of die. As the name implies, many of the ideas in SMED involve redesign of jigs, fixtures, dies, software and other components of production equipment associated with changeovers.

There has been a significant decrease in the time required for changeovers on many machines due to SMED related efforts. This has resulted in the liberation of time previously spent on changeovers that can now be spent producing parts. This is important and valuable, but our models suggest that there are additional benefits. By eliminating or reducing changeover times using such techniques we not only gain greater capacity, we also create a production facility that is easier to manage.

This is good news not just for lazy managers. It is good news for ambitious managers as well. The elimination of changeover time as a significant factor in production planning makes the problem much easier for optimization software. In this context, "easier" means that optimization can be performed much, much faster. This, in turn, means that the optimization can be performed more often as new information becomes available and this results in plans that are more up-to-date and more responsive to current customer needs and shop-floor realities.

Our models clearly demonstrate the improved management characteristics that result from setup reductions. As the models become simpler, they are easier to solve. But these simplicity benefits are present whether formal mathematical models and computers are used to find good production plans or if the plans are hashed out on a chalkboard. Good production plans are simply easier to obtain when changeover times are reduced.

The changeover modeling process can be a bit complicated, but yields important benefits. We can summarize as follows:

- Changeover time reduction is good not only because of increased capacity but also improved manageability;
- meanwhile modeling enables leveraging computing power to obtain solutions as well as the ability to gain insights via the models and the modeling process.

We return to issues related to creating and using production planning models in a supply chain. In order to make the right changeover decisions we need some cost information and perhaps alternative routing information. This is discussed in the next chapters.