Chapter 11. Scheduling to Prioritize Demand

Scheduling Timeline and Applications

- **Scheduling** – process of deciding what work to assign to which resources and when to assign the work. Matches the demand to the resources.
  - Project management
  - Manufacturing
  - Service scheduling
  - Transportation scheduling
  - Patient scheduling
  - Workforce scheduling
  - Tournament scheduling

Table 11.1 Different Time Horizons for Scheduling

Time Period	Examples
Long (1 year or more)	- Scheduling the delivery of new aircraft over the next decade - MLB schedule for the next year
Medium (1 month ~ 1 year)	- Maintenance schedule for nuclear power plant - Hiring schedule of temporary workers for the next quarter
Short (1 minute ~ 1 month)	- Deciding the patient to treat next in the ER - Caller assignment in a call center

Scheduling Timeline and Applications

- **Materials requirement planning (MRP)** – a system that plans the delivery of components required for a manufacturing process so that components are available when needed but not so early as to create excess inventory.

- **Bill of materials** – list of components that are needed for the assembly of an item.

- **Revenue management** – practice of maximizing the revenue generated from a set of fixed assets (like cars in a rental fleet, rooms in a hotel, or tables in a restaurant).

Resource Scheduling

- **Job** – a flow unit that requires processing from one or more resources.

- **Sequencing (or priority) rules** – Decision rules to allocate the relative priority of jobs at a work center
  - First-come-first-served (FCFS)
  - Shortest processing time (SPT)
  - Earliest due date (EDD)

Performance Measures

Three key process performance metrics

- Inventory
- Flow Rate
- Flow time (vs. makespan)

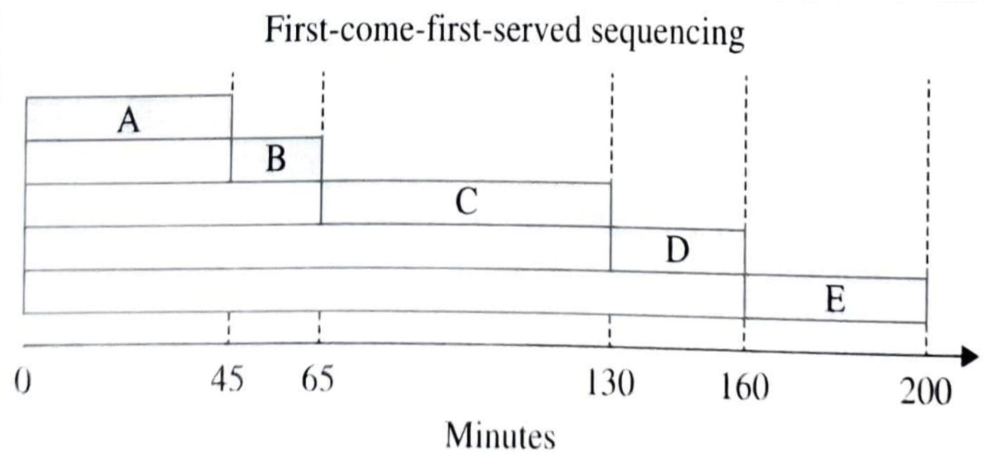
All three are relevant in a scheduling application, but according to Little's Law (I= RT, see Chapter 2), we only need to track two of them—once you know two of the metrics, you can evaluate the third.

Resource Scheduling – First-Come-First-Served

- **First-come-first-served (FCFS)** – a rule that sequences jobs to be processed on a resource in the order in which they arrived.

- **Example (p. 223)**

Job	**A**	**B**	**C**	**D**	**E**
Processing time (minutes)	45	20	65	30	40

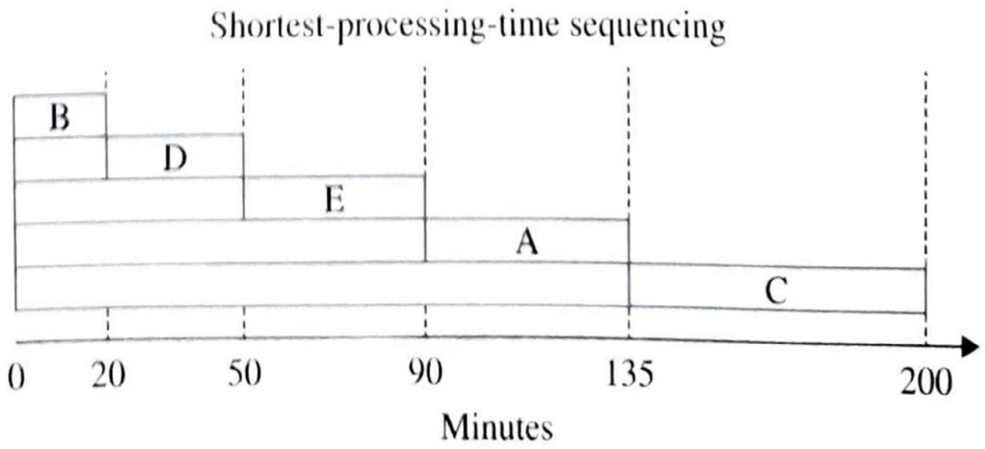


Resource Scheduling – Shortest Processing Time

- **Shortest processing time (SPT)** – a rule that sequences jobs to be processed on a resource in ascending order of their processing times.

- **Example (p. 223)**

Job	**A**	**B**	**C**	**D**	**E**
Processing time (minutes)	45	20	65	30	40



First-Come-First-Served vs. Shortest Processing Time

- **Calculation the average flow time:**

  The average flow time = (sum of job flow times)/ # of jobs

- **Example (p. 223)**
  - The average flow time of FCFS:

    (45+65+130+160+200) / 5 = 120 (min.)
  - The average flow time of SPT:

    (20+50+90+135+200) / 5 = 99 (min.)

First-Come-First-Served vs. Shortest Processing Time

- **Flow rate:**

  makespan = 200minutes

  5 jobs are processed

  flow rate = 5 jobs/200 minutes = 0.025 job/min

- The average inventory with FCFS:
  I = RT= 0.025 job/min $\times$ 120 min = 3 jobs
- The average inventory with SPT:
  I = RT= 0.025 job/min $\times$ 99 min = 2.5 jobs

First-Come-First-Served vs. Shortest Processing Time

SPT always gives a lower flow time than FCFS and the gap between the two is largest when there are many jobs waiting for service.

It has been proven to be **the very best sequencing rule**: if you want to minimize the average flow time (and in turn the average inventory of jobs in the system), you cannot do any better than SPT!

Limitations of Shortest Processing Time

- **Delay to determine processing times** – It takes too long to determine the processing time of the jobs; it is also possible that the only way to know the processing time for a job is to actually do the job.

- **Biased processing time estimates**. – Estimates of the processing times may be biased; it creates an incentive to distort the resource's perception of processing times because shorter jobs are processed earlier.

- **Fairness** – SPT raises concerns of fairness—it does not treat all jobs equally. This is the largest challenge for SPT.

Resource Scheduling with Priorities – Weighted Shortest Processing Time

- Jobs can vary in terms of their processing times and importance, or weight

- **Weighted shortest processing time (WSPT)** – a rule that sequences jobs to be processed on a resource in descending order of the ratio of their weight to their processing time. Jobs with high weights and low processing times tend to be sequenced early.

(Remark) Similarity with knapsack problem

Resource Scheduling with Priorities – Weighted Shortest Processing Time

- If all jobs have equal weights, then WSPT yields the same results as SPT.

- If there are different types of jobs and their weights are vastly different, then WSPT becomes a simple priority rule in which the resource works on the highest-priority job before switching to lower-priority jobs.

Resource Scheduling with Priorities – Weighted Shortest Processing Time

- **Example (p. 230)**

Job	A	B	C	D	E
Processing time (hours)	10	6	24	15	18
Weight/priority (cost per unit of time)	2	1	6	5	2

- The average cost of WSPT:
  $$(5 \times 15 + 6 \times 39 + 2 \times 49 + 1 \times 55 + 2 \times 73) / 5 = 121.6$$
- The average cost of SPT: 145.8
- The average cost of FCFS: 139.4

Resource Scheduling with Priorities – Weighted Shortest Processing Time

- **Emergency Severity Index (ESI)** – a scoring rule used by emergency rooms to rank the severity of a patient’s injuries and then to prioritize their care.

  1: resuscitation, 2: emergent ~ 5: nonurgent

  - Delay to determine priority or processing time
  - Biased priority or processing time estimates
  - Fairness

Resource Scheduling with Due Dates – Earliest Due Date

- **Percent on time** – the fraction of jobs that are completed on or before their due date.

- **Lateness** – the difference between the completion time of a job and its due date. For example, if a job is due on day 10 and it is completed on day 12, then it is late $12 - 10 = 2$ days.
  - Lateness can also be negative. For example, if the job is due on day 10, but it is completed on day 7, then it is –3 days "late," which means it is three days early. It might seem odd to have a negative lateness, which means the job is early, but this is a relatively common definition of lateness.

Resource Scheduling with Due Dates – Earliest Due Date

- **Tardiness** - if a job is completed after its due date, then the tardiness of a job is the difference between its completion time and its due date. = Max {Lateness, 0}
  - If the job is completed before its due date, then tardiness is 0. In some sense, the tardiness measure makes more sense than lateness because a job never has negative tardiness.

- **Tardy** – whether a job is completed after its due date.
  - If the job is completed after its due date, then tardy is 1. Otherwise 0. In some sense, someone may concern the number of tardy jobs.

Resource Scheduling with Due Dates – Earliest Due Date

- **Earliest due date (EDD)** – a rule that sequences jobs to be processed on a resource in ascending order of their due dates.

- **Example (p. 233)**

Job	**A**	**B**	**C**	**D**	**E**
Processing time (days)	2	4	6	8	10
Due date (days from today)	20	18	16	14	12

  - Let’s evaluate SPT and EDD in terms of flow time, lateness, and tardiness.

Resource Scheduling with Due Dates – Earliest Due Date

- **SPT**

Job	Processing Time (days)	Due Date (days)	Flow Time	Lateness	Tardiness	Tardy
A	2	20	2	-18	0	0
B	4	18	6	-12	0	0
C	6	16	12	-4	0	0
D	8	14	20	6	6	1
E	10	12	30	18	18	1
Average			14	-2	4.8	0.4
Maximum				18	18	1

- **EDD**

Job	Processing Time (days)	Due Date (days)	Flow Time	Lateness	Tardiness	Tardy
E	10	12	10	-2	0	0
D	8	14	18	4	4	1
C	6	16	24	8	8	1
B	4	18	28	10	10	1
A	2	20	30	10	10	1
Average			22	6	6.4	0.8
Maximum				10	10	1

Resource Scheduling with Due Dates – Moore & Hodgson's Algorithm

- **Moore & Hodgson's Algorithm** – an algorithm to minimize the number of tardy jobs
  1. Sort the jobs in EDD order
  2. Let S denote EDD schedule
  3. Find the first tardy job in S and suppose this is job j.
  4. Remove from S the job with the largest processing time from jobs $1, \ldots, j$
  5. Continue with step 3 for this new schedule S until all jobs are on time.

Resource Scheduling with Due Dates – Moore & Hodgson's Algorithm

- **Example**

$S = (\text{A}, \text{B}, \text{C}, \text{D}, \text{E}, \text{F})$

Job	**A**	**B**	**C**	**D**	**E**	**F**
Processing time (days)	3	4	10	10	8	6
Due date (days from today)	6	9	15	20	23	30
Flow Time	3	7	17	27	35	41
Tardy			1	1	1	1

The first tardy job is job C.

The largest processing time among (A, B, C) is job C.

Remove job C from *S* and Continue with step 3.

Resource Scheduling with Due Dates – Moore & Hodgson's Algorithm

- **Example**

$S = (\mathrm{A}, \mathrm{B}, \mathrm{D}, \mathrm{E}, \mathrm{F})$

Job	**A**	**B**	**D**	**E**	**F**
Processing time (days)	3	4	10	8	6
Due date (days from today)	6	9	20	23	30
Flow Time	3	7	17	25	31
Tardy				1	1

The first tardy job is <u>job E</u>

The largest processing time among (A, B, D, E) is <u>job D.</u>

Remove job D from S and Continue with step 3.

Resource Scheduling with Due Dates – Moore & Hodgson's Algorithm

- **Example**

$S = (\mathrm{A}, \mathrm{B}, \mathrm{E}, \mathrm{F})$

Job	**A**	**B**	**E**	**F**
Processing time (days)	3	4	8	6
Due date (days from today)	6	9	23	30
Flow Time	3	7	15	21
Tardy				

There is no tardy jobs in S.

The optimal schedule minimizing the number of tardy jobs is (A, B, E, F, C, D) (or (A, B, E, F, D, C)).

Multi-resource Scheduling – Largest-Processing-Time-first Heuristic

- **Multi-resource scheduling** – In general, scheduling deals with the problems with more than one resource. The types of resources can be the same or different.

- **Allocation and sequencing** – In multi-resource scheduling, it is important to determine which jobs assigns to a resource and how to order them.

- **Makespan** – the makespan is the total length of the schedule. This is equal to the maximum flow time of job.

Multi-resource Scheduling – Largest-Processing-Time-first Heuristic

- **Longest processing time (LPT)** – a rule that sequences jobs to be processed on a resource in descending order of their processing times.

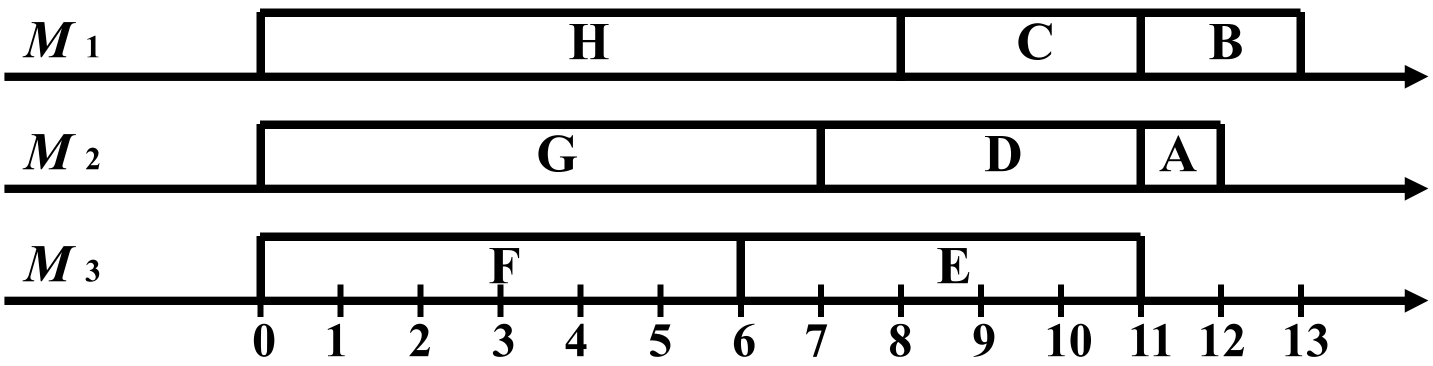
- **Largest-processing-time-first (LPT) Heuristic** – no exact algorithm has been developed for constructing optimal schedule of identical parallel machines. LPT heuristic provides a good schedule for minimizing the makepan of the identical parallel machines.
  1. Construct an LPT ordering of the jobs.
  2. Schedule the jobs in order, each time assigning a job to the machine with the least start time.

Multi-resource Scheduling – Largest-Processing-Time-first Heuristic

- **Example**

Job	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Processing time (days)	1	2	3	4	5	6	7	8

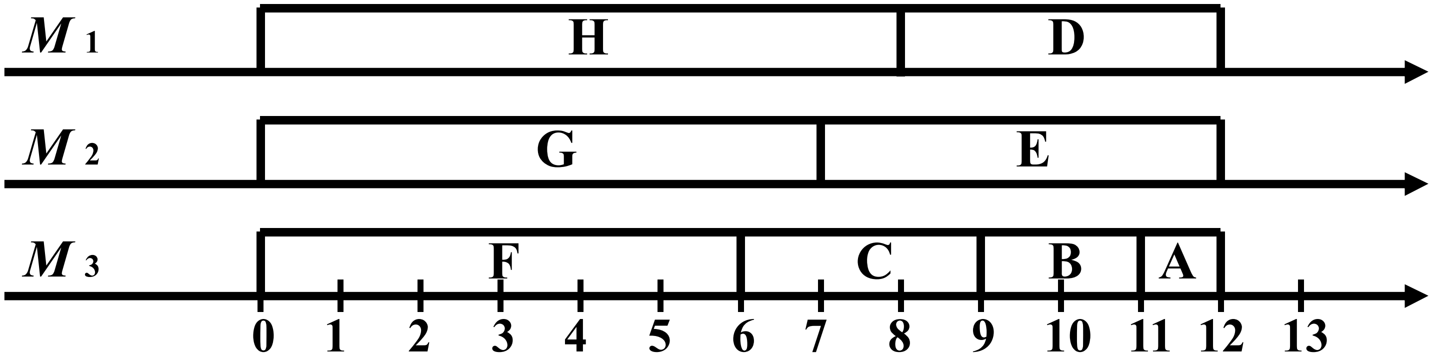
- **The result of LPT heuristic**



Note that it is not the optimal solution

Multi-resource Scheduling – Largest-Processing-Time-first Heuristic

- **Example**
  - **The optimal solution**



- **Worst case bound of LPT heuristic**

$$\frac{C_{\max}(\text{LPT})}{C_{\max}(\text{OPT})} \leq \frac{4}{3} - \frac{1}{3m}$$

where $C_{\max}(\text{LPT})$ be the makespan under the LPT rule,
$C_{\max}(\text{OPT})$ be the optimal makespan,
m is the number of machines.

Theory of Constraints

- **Theory of constraints** – an operation guideline that recommends managerial attention be focused on the bottleneck of a process.
  - The flow rate of a process is dictated by the flow rate through the bottleneck resource.
  - Identify the bottleneck.
  - Ensure that the flow rate through the bottleneck is maximized.
  - Schedule work through the process so that the bottleneck is never blocked or starved.

Theory of Constraints

TIME

The 25 Most Influential Business Management Books

There's never a shortage of new books about how to be more effective in business. Most of them are forgettable, but here are 25 that changed the way we think about management — from the iconic "How to Win Friends and Influence People" to groundbreaking tomes like "Guerilla Marketing" and quick reads like the "The One Minute Manager".

Like 250 Tweet Share

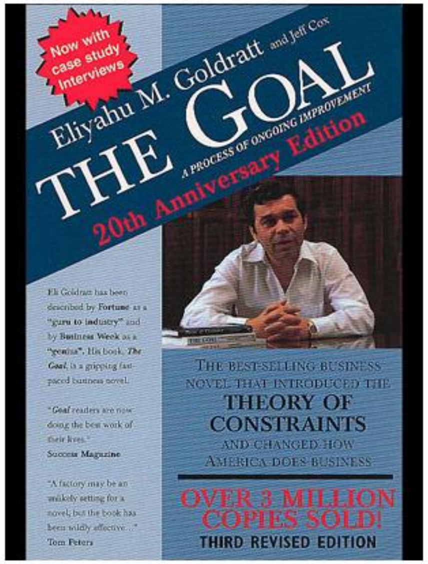
BUSINESS GUIDES

The Goal (1984), by Eliyahu Goldratt

10 *of* 25 VIEW ALL

By Nate Rawlings | Tuesday, Aug. 09, 2011

Eliyahu Goldratt's *The Goal* is unusual among business management books for at least two reasons. First, Goldratt wasn't a titan of industry, a b-school professor, or even a consultant, but rather a physicist. Second, *The Goal* is a novel. Centered on a production manager named Alex Rogo who has three months to turn around a deficient, unprofitable manufacturing plant, *The Goal* explains the "Theory of Constraints," which among other points incorporates the idiom, "A chain is only as strong as its weakest link;" and focuses on bottlenecks, the great hindrances to productivity. Rogo uses the Socratic method to help fix his marriage, then applies it to his plant crew, coming up with steps to solve the plant's problems. *The Goal* has been in print since 1984, and a revised third edition was released on the book's 20th anniversary. So does Rogo achieve his goal? You'll have to read it to find out.



Reservations and Appointments

An appointment system deals with matching supply with demand when there is sufficient capacity.

- Processing times are not perfectly predictable, so it is not clear when the next person should arrive for their appointment.
- People do not always arrive for their appointment—that is, demand, even scheduled demand, is not perfectly predictable.

Scheduling Appointments with Uncertain Processing Times

- **Makespan** – the total time to process a set of jobs. The makespan can never be smaller than the sum of the processing times, but it can be larger if the resource has some idle time.
- **Overbooking** – the practice of intentionally assigning more work to a resource than the resource can process so as to maintain a high utilization for the resource. Overbooking is meant to mitigate the consequences of no-show demand.

Table 11.1 Performance Measure Results with Appointments Spaced out at Different Intervals

		Interval between Appointments (min)		
Calculation	Measure	15	20	25
a	Average flow time (min)	51.1	23.80	19.25
b	Average processing time (min)	18.35	18.35	18.35
c	Number of patients	20	20	20
d	Makespan (hr)	6.13	6.62	8.15
$e = 60/b$	Capacity (patients/hr)	3.27	3.27	3.27
$f = c/d$	Flow rate (patients/hr)	3.26	3.02	2.45
$g = f/e$	Utilization	1.00	0.92	0.75
$h = a - b$	Average waiting time (min)	32.75	5.45	0.90

No-Shows

- **No-Show** – a customer who does not arrive for his or her appointment or reservation.
  - No-shows can be as high as 20 to 30 percent in many physician offices.
  - This is not good for the efficient use of the physician’s time.
  - Utilization decreases when no-shows are possible.
  - No-shows seem to help reduce average waiting times.

Overbooking

- Often happens for unforeseen circumstances
- Helps to off set no-shows
  - No-shows means unused resources – unused resources are costly to the company.
- One way to do a better job of matching supply with demand (irate customers)
  - Attempt to balance the cost of overbooking too much with the cost of overbooking too little (idle resources) ⇒ Chapter 18 (Revenue management)

Providing Solutions to Uncertainty

- Reduce the variability that causes the problem.
- Tackle the variability in processing times.
- Tackle no-shows.
  - Reminders
  - Charges for missed appointments 

- **Open-access appointment system** – an appointment system in which appointments are only available one day in advance and are filled on a first-come-first-served basis. (eg. 국내병원 당일 예약 시스템)

MRP Record with L4L Order

TABLE 7.8 **MRP Record with L4L order**

Item:	B100	Lot size:	L4L
Description:	Bicycle wheel	Lead time:	2 weeks
Beg. inv.:	40	Safety stock:	15

Week	Feb. 1	Feb. 8	Feb. 15	Feb. 22	Mar. 1	Mar. 8	Mar. 15	Mar. 22	Mar. 29
Gross requirements	0	124	0	176	100	70	0	70	100
Scheduled receipts	0	200	0	0	0	0	0	0	0
Planned OH inventory	40	116	116	15	15	15	15	15	15
Planned receipts				75	100	70		70	100
Planned order release		75	100	70		70	100		

Inputs for Material Requirements Plan

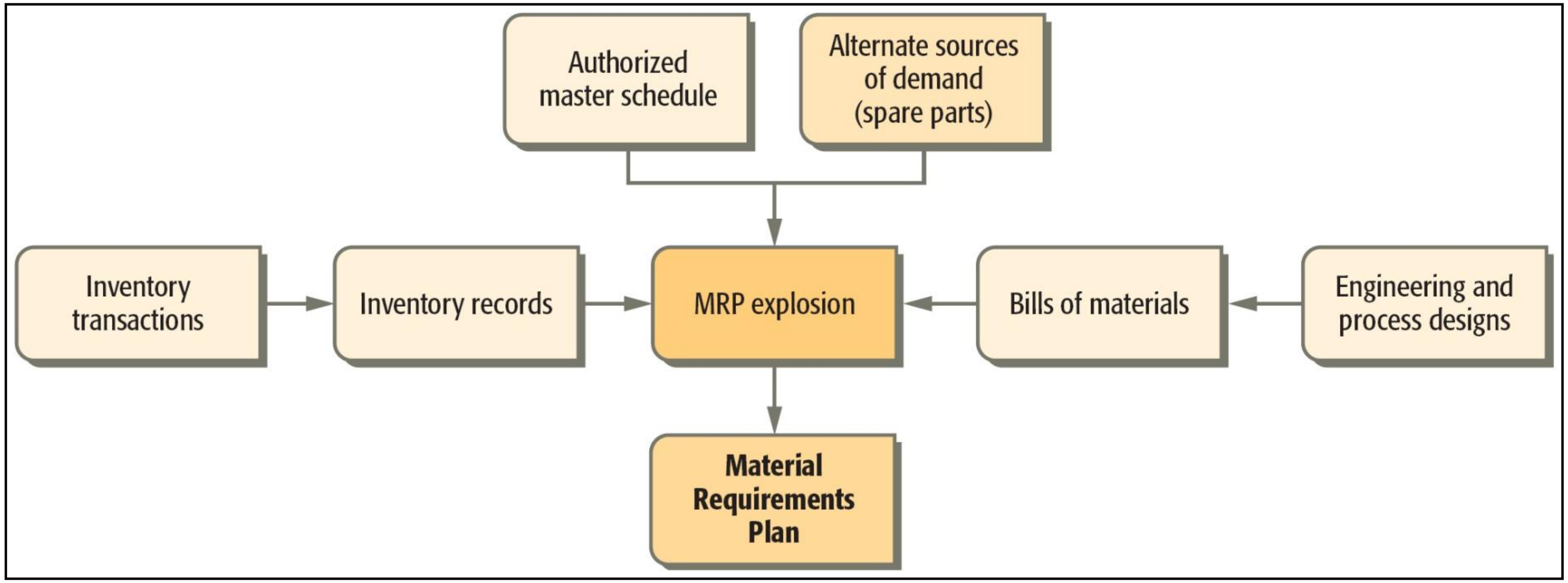


Illustration of a Rolling MRP Schedule

Master Production Schedule – Speedy Road Bicycle

45	0	62	0	88	50	35	50	35	100

Item: A130 — Lot size: L4L
Description: Frame assembly — Lead time: 1 week
Beginning inventory: 20 — Safety stock: 20

1 frame assembly for each bicycle

Week	1	2	3	4	5	6	7	8	9	10
Gross requirements	45	0	62	0	88	50	35	50	35	100
Scheduled receipts	45									
Planned OH inventory	20	20	20	20	20	20	20	20	20	20
Planned receipts			62		88	50	35	50	35	100
Planned order release		62		88	50	35	50	35	100	

Item: B100 — Lot size: 200
Description: Bicycle wheel — Lead time: 2 weeks
Beginning inventory: 40 — Safety stock: 15

2 wheels for each frame assembly = 2 x 62

Week	1	2	3	4	5	6	7	8	9	10
Gross requirements	0	124	0	176	100	70	100	70	200	
Scheduled receipts	0	200	0	0	0	0	0	0	0	
Planned OH inventory	40	116	116	140	40	170	70	200	200	
Planned receipts				200		200		200	200	
Planned order release		200		200		200	200			

Item: z125 — Lot size: 2500
Description: Spokes — Lead time: 3 weeks
Beginning inventory: 200 — Safety stock: 100

12 spokes for each wheel = 12 x 200

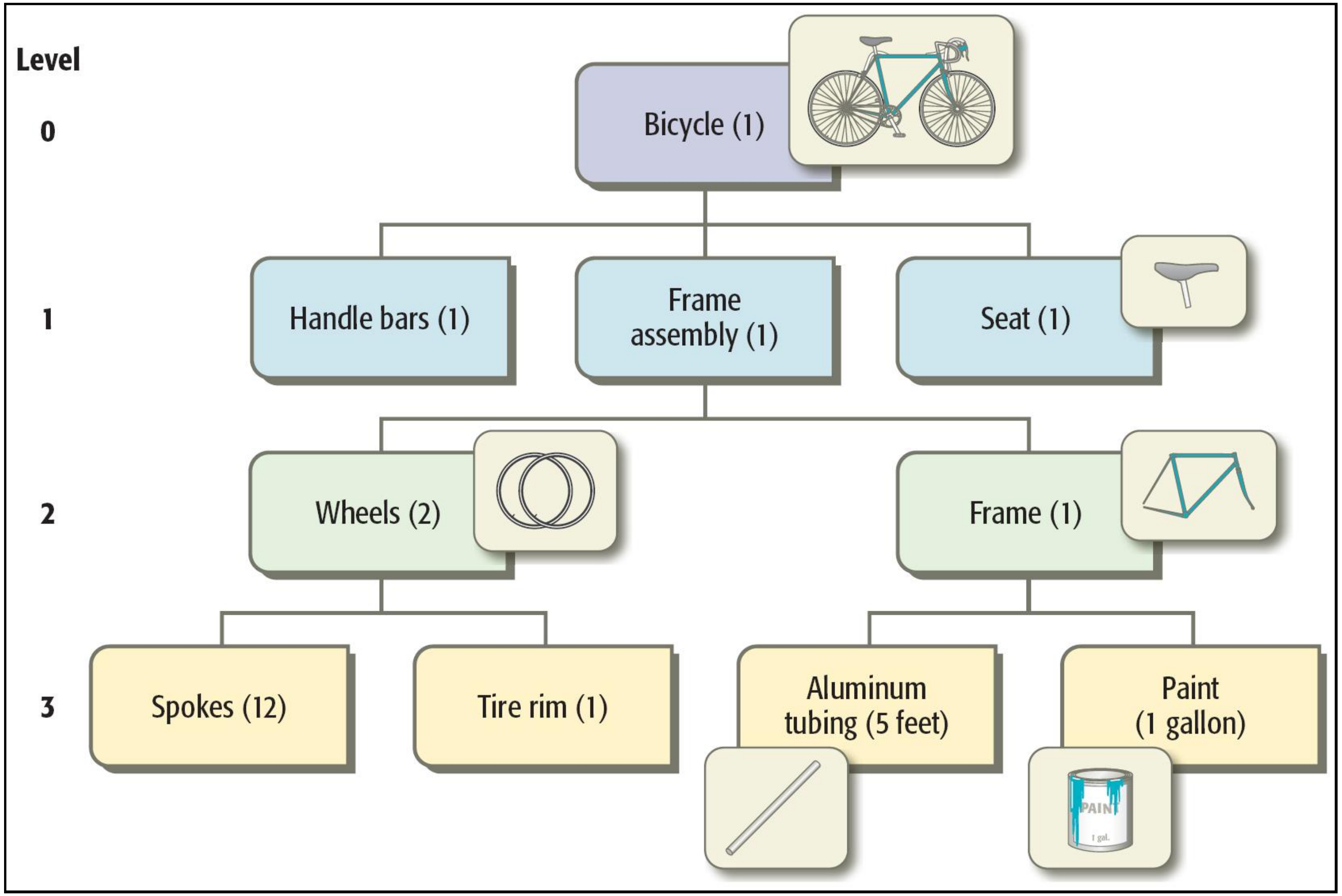
Week	1	2	3	4	5	6	7	8	9	10
Gross requirements		2400		2400		2400	2400			
Scheduled receipts		2500		2500						
Planned OH inventory	200	300	300	400	400	500	600	600	600	600
Planned receipts						2500	2500			
Planned order release	2500		2500	2500						

Item: D200 — Lot size: P = 3
Description: Tire rim — Lead time: 2 weeks
Beginning inventory: 220 — Safety stock: 10

1 tire rim for each wheel

Week	1	2	3	4	5	6	7	8	9	10
Gross requirements		200		200		200	200			
Scheduled receipts				390						
Planned OH inventory	220	20	20	210	210	10	10	10	10	
Planned receipts							200			
Planned order release	390				200					

Bill of Materials for a Bicycle



• First come, first served (FCFS): Jobs are processed in the order in which they arrive at the workstation.
• Last come, first served (LCFS): Jobs are processed in the order of last to first in which they arrive at the workstation.
• Shortest processing time (SPT): Jobs are processed in order of the processing time required at the workstation, with the job requiring the least processing time at the workstation scheduled first.
• Longest processing time (LPT): Jobs are processed in order of the processing time required at the workstation, with the job requiring the longest processing time at the workstation scheduled first.
• Earliest due date (EDD): Jobs are processed in the order in which they are due for delivery to the customer.
• Critical ratio (CR): Jobs are processed in order of increasing critical ratio (the ratio of time required by work left to be done to time left to do the work)
• SLACK - Jobs are processed in order of increasing slack time (time until due date minus remaining time to process)

"열차출발 전후 승차권 취소 매년 증가…4년간 위약금 980억원

송고시간 | 2020-09-27 07:00

김기훈 기자

고속버스도 승차권 반환·위약금 늘어…정동만 의원 "노쇼 최소화 대책 마련해야"

승차권 발매기 이용하는 시민들
[연합뉴스 자료사진]

(서울=연합뉴스) 김기훈 기자 = 열차 승차권을 예매했다가 출발을 전후해 취소·반환하는 '예약 부도'(노쇼·No Show) 건수가 해마다 늘고 있는 것으로 나타났다.