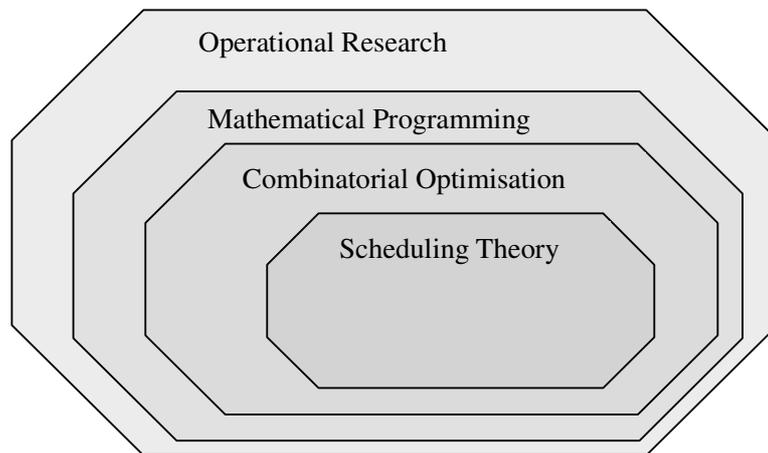


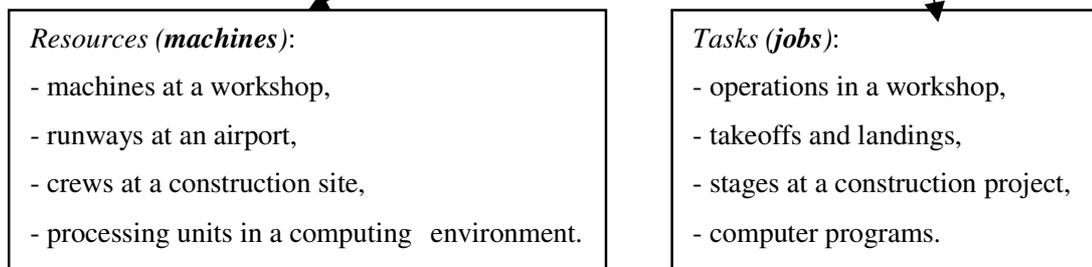
1. Introduction

As an independent branch of Operational Research, Scheduling Theory appeared in the beginning of the 50s. In addition to computer systems and manufacturing, scheduling theory can be applied to many areas including agriculture, health care and transport.



Scheduling deals with the problems of optimal arrangement, sequencing and timetabling.

Scheduling is a decision-making process of allocating **limited resources** to **activities** over time.



A *schedule* is a job sequence determined for every machine of the processing system.

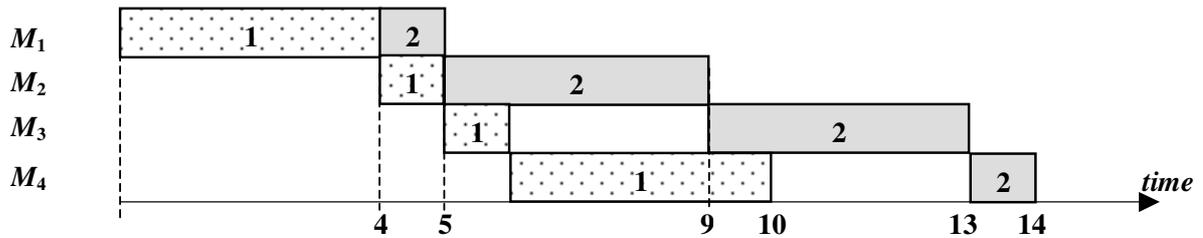
Standard scheduling requirements say:

- a job cannot be processed by two or more machines at a time,
- a machine cannot process two or more jobs at the same time.

Depending on the type of scheduling system, specific constraints should be satisfied (jobs may be released at different times, there may be allowed preemption of jobs by other jobs, etc.).

Gantt chart is a horizontal bar chart that graphically displays the time relationships between the different tasks in a project:

- the X-axis represents the time,
- the Y-axis represents machines,
- a colour and/or pattern code may be used to indicate operations of the same job.



Scheduling theory covers more than 10,000 different models. The models are specified according to three-field classification $\alpha\beta\gamma$ where

- α specifies the **machine environment**,
- β specifies the **job characteristics**, and
- γ determines the **optimality criterion**.

Machine environment

Single stage systems

If there is a single machine ($m=1$), each job should be processed by that machine exactly once.

If there are several parallel machines $\{M_1, M_2, \dots, M_m\}$, each job can be processed by any machine.

$$\alpha = \begin{cases} 1 - \text{single (dedicated) machine} \\ \quad p_j - \text{processing time of job } j \\ P - \text{identical parallel machines,} \\ \quad p_{ij} = p_j - \text{processing time of job } j \\ \quad \text{on machine } i \end{cases}$$

Multistage systems

Each job should be processed on each machine from the set $\{M_1, M_2, \dots, M_m\}$.

All machines are different.

$$\alpha = \begin{cases} F - \text{flow shop,} \\ \quad \text{job } j \text{ is processed first on machine 1,} \\ \quad \text{then on machine 2, ..., and finally on} \\ \quad \text{machine } m \\ J - \text{job shop,} \\ \quad \text{each job has its own route to follow} \\ O - \text{open shop,} \\ \quad \text{each job can be processed by} \\ \quad \text{the machines in an arbitrary order} \end{cases}$$

Job environment

There are n jobs $N=\{1,\dots,n\}$.

Processing time of job j on machine i is p_{ij} . If there is a single machine, then processing time of job j does not depend on machine number and it is denoted by p_j .

For job j there may be given also

r_j – release time (the time the job arrives at the system),

d_j – due date (the time the job is promised to the customer),

w_j – weight (the importance of job j).

Preemption (*pmtn*) implies the processing of any job can be interrupted and resumed later.

Optimality criterion

The schedule can be characterised by starting or completion times of all operations of the jobs. The objective is to construct a schedule, that minimizes a given objective function F . Usually function F depends on job completion times $C_j, j=1,\dots,n$, where C_j is the completion time of the last operation of job j .

The most common objective functions are

Makespan	$C_{\max} = \max \{C_j j=1,\dots,n\}$
Total completion time	$\sum C_j = \sum_{j=1}^n C_j$
Total weighted completion time	$\sum w_j C_j = \sum_{j=1}^n w_j C_j$

Other objective functions depend on due dates d_j . We define for each job j

$L_j = C_j - d_j$ - lateness of job j ,

$E_j = \max\{0, d_j - C_j\}$ - earliness,

$T_j = \max\{0, C_j - d_j\}$ - tardiness,

$U_j = \begin{cases} 0 & \text{if } C_j \leq d_j \\ 1 & \text{otherwise} \end{cases}$ - unit penalty.

The corresponding objective functions can be defined as follows:

Maximum lateness	$L_{\max} = \max \{L_j j=1,\dots,n\}$
Total (weighted) tardiness	$\sum_{j=1}^n T_j \quad (\sum_{j=1}^n w_j T_j)$
Total (weighted) number of late jobs	$\sum_{j=1}^n U_j \quad (\sum_{j=1}^n w_j U_j)$

Examples:

- 1) $1|r_j, pmtn|L_{\max}$ is the problem of finding a preemptive schedule on one machine for a set of jobs with given release times such that maximum lateness is minimised.
- 2) $P|p_j=1|C_{\max}$ is the problem of scheduling jobs with unit processing times on m identical parallel machines such that the makespan is minimised.
- 3) $J3|p_{ij}=1|C_{\max}$ is the problem of minimising maximum completion time in a three-machine job shop with unit processing times.

In-class exercise 1: Consider a scheduling problem with n readers and two books.
Classify the following scheduling models:

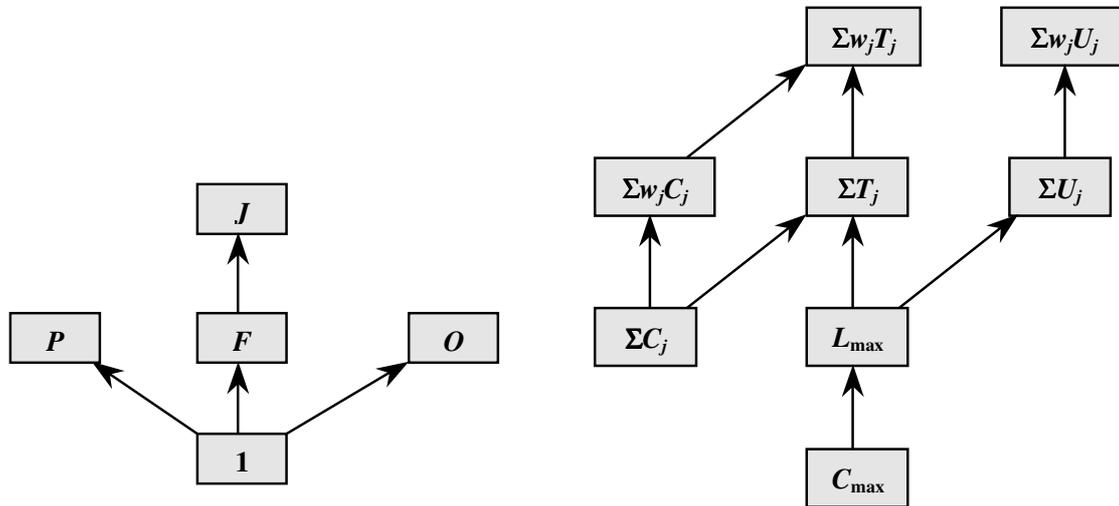
Machines	Jobs	Objective	$\alpha \beta \gamma$
Two volumes of one book	Readers	Finish reading as soon as possible	$F2 C_{\max}$
Two volumes of one book	Readers	Minimise the cost of late book return	
Two different (independent) books	Readers	Finish reading as soon as possible	
Two different (independent) books	Readers (each reader has its own "reading sequence")	Finish reading as soon as possible	

In-class exercise 2: Classify the examples of scheduling problems:

<i>Publishing industry:</i> typesetting, actual printing, binding, packaging. Different items have different processing times depending on the book size, the number of copies, etc. The objective is to produce all items as soon as possible.	$F4 C_{\max}$
<i>Clothing industry:</i> cutting, sewing, pressing, packing. The objective is to produce all items as soon as possible.	?
<i>Steel mills:</i> different rods or girders pass through the set of rollers in their own orders with their own temperatures and pressure settings. The objective is to produce all items as soon as possible.	
<i>Repair of cars in a garage:</i> replace tires, repair gear box, check brakes, repair headlights, etc. The objective is to repair all cars in a garage as soon as possible.	
<i>Completing several pieces of CW</i> so that the maximum lateness is minimised. CW j is released at time r_j , requires p_j days for completion and has a due date d_j .	
<i>Revision schedule:</i> starting on 13/12/2004, revise the material of n modules by their exam dates. Revision time for module j is p_j .	
<i>Literature review for FYP</i> should be based on n library books. Book j can be read in p_j days and it should be returned by its due date d_j . The library charges 30p per day on each overdue book. The objective is to minimise the total fine.	

Complexity Hierarchy

There is a certain complexity hierarchy among scheduling systems and objective functions as shown in the diagrams below. The arrows go to harder problems.



C_{\max} reduces to L_{\max} by setting $d_j=0$ for all j .
 ΣC_j reduces to $\Sigma w_j C_j$ by setting $w_j=1$ for all j .
 ΣC_j reduces to ΣT_j by setting $d_j=0$ for all j .
 $\Sigma w_j C_j$ reduces to $\Sigma w_j T_j$ by setting $d_j=0$ for all j .

In this course we demonstrate most scheduling methods and techniques. Most algorithms will be illustrated using scheduling software systems LEKIN and LiSA.

LiSA - Library of Scheduling Algorithms, Otto-von-Guericke Universität, Magdeburg, Germany,
 Project Leader: Professor Heidemarie Bräsel,
<http://fma2.math.uni-magdeburg.de/~lisa/>

LEKIN - Educational Scheduling System, Stern School of Business, New York University,
 Project Leader: Professor Michael Pinedo,
<http://www.stern.nyu.edu/om/pinedo/lekin/>

The following books are considered to be standard texts on the topic:

1. J.Blazewicz, K.H. Ecker, E.Pesch, G.Schmidt, J. Weglarz "Scheduling Computer and Manufacturing Processes", Springer, 1996, 2001.
2. P. Brucker "Scheduling Algorithms", Springer, 1995, 1998, 2001.
3. S. French "Sequencing and scheduling", Ellis Horwood, 1982.
4. M. Pinedo "Scheduling: Theory, Algorithms, and Systems", Prentice Hall, 1995, 2001.
5. M. Pinedo "Operations Scheduling with Applications in Manufacturing and Services", Irwin/McGraw Hill, 1999.