Chapter 2
Process Modeling and Analysis

prof.dr.ir. Wil van der Aalst
www.processmining.org
Productivity improvements

- **Adam Smith** (1723-1790) showed the advantages of the division of labor.
- **Frederick Taylor** (1856-1915) introduced the initial principles of scientific management.
- **Henry Ford** (1863-1947) introduced the production line for the mass production of “black T-Fords”.
- Since 1950, computers and digital communication infrastructures are the most dominant factor influencing business processes and their management.
Role of models

- Operations management, and in particular operation research, is a branch of management science heavily relying on modeling.
- Models are used to reason about processes (redesign) and to make decisions inside processes (planning and control).
- A process model may be used to discuss responsibilities, analyze compliance, predict performance using simulation, and configure a WFM system.
Problems of models

• The model describes an idealized version of reality.
• Inability to adequately capture human behavior.
• The model is at the wrong abstraction level.
• Therefore, we advocate the use of event data:
  - Process mining allows for the extraction of models based on facts.
  - Moreover, process mining does not aim at creating a single model of the process.
  - Instead, it provides various views on the same reality at different abstraction levels.
  - For example, users can decide to look at the most frequent behavior to get a simple model (“80% model”).
  - However, they can also inspect the full behavior by deriving the “100% model” covering all cases observed.
Transition systems

[Start]
- Register request

[c1,c2]
- Examine casually

[c2,c3]
- Check ticket

[c3,c4]
- Examine thoroughly

[c1,c4]
- Examine casually

[c5]
- Reinitiate request

- Pay compensation

- Reject request

[End]
Petri nets

- **Start register**
- **Request**
- **Examine thoroughly**
- **Examine casually**
- **Check ticket**
- **Decide**
- **Pay compensation**
- **Reject request**
- **Reinitiate request**

**Places:**
- Start
- Register request
- Examine (casually)
- Examine (thoroughly)
- Check ticket
- Decide
- Pay compensation
- Reject request

**Transitions:**
- AND-split
- XOR-split
- XOR-join

**Tokens:**
- Token
Reachability graph
How many states?

(a) (c)

(b)

(c)
WF-nets and soundness

start

\[\text{Cloud}\]

end
YAWL

start

register request

examine thoroughly

OR-split

OR-join

pay compensation

end

examine casually

decide

check ticket

reject request

new information

c1

c2

c3

multiple instance task

composite task

cancellation region

condition (like a place in a Petri net)
task (i.e., an atomic activity)
AND-split

AND-join

XOR-split

XOR-join

OR-split

OR-join

start

end
deferred choice pattern using the event-based XOR gateway
Event-Driven Process Chains (EPCs)

start
register request

OR
e1
examine thoroughly

OR
e2
examine casually

AND
e3
check ticket

OR
e4
decide

AND

e5
pay compensation
end

XOR

e6
reject request
end

function

AND-split connector
AND-join connector

XOR-split connector
XOR-join connector

OR-split connector
OR-join connector

start event
end event

intermediate event
Vicious cycle paradox
Causal nets (C-nets)

- **a** register request
- **b** examine thoroughly
- **c** examine casually
- **d** check ticket
- **e** decide
- **f** reinitiate request
- **g** pay compensation
- **h** reject request
- **z** end

**Splitting operations**:
- XOR-split
- AND-split
- OR-split

**Joining operations**:
- XOR-join
- AND-join
- OR-join
Why C-nets?

• Similar to heuristic nets and representation used by genetic miners.
• Fits well with mainstream languages (BPMN, EPCs, YAWL, BPEL, etc.).
• Model XOR, AND, and OR, but no silent steps or duplicate activities needed.
• Loose interpretation (focus on replay semantics rather than execution semantics).
Another C-net

Diagram showing:
- Start booking (a)
- Book flight (b)
- Book car (c)
- Book hotel (d)
- Complete booking (e)

Connections indicate the sequence of booking activities.
WF-net interpretation of C-nets
(only valid sequences!)

a
start booking

b
book flight

c
book car

d
book hotel

e
complete booking

WF-net:

- Start with task a
- Optionally book flight
- Optionally book car
- Optionally book hotel
- Complete booking with task e
Non-sound C-nets

(a) unsound because there are no valid sequences

(b) unsound although there exist valid sequences
Unbounded C-net
Example: YAWL
Performance analysis, e.g., simulation in BPM|one

[Image of a BPM tool interface and a data table]
Limitations of model-based analysis

• Verification and performance analysis heavily rely on the availability of high quality models.
• When the models and reality have little in common, model-based analysis does not make much sense.
• There is often a lack of alignment between hand-made models and reality.
• Process mining aims to address these problems by establishing a direct connection between the models and actual low-level event data about the process.
• Process discovery techniques allow for viewing the same reality from different angles and at different levels of abstraction.