

# Advanced Process Mining

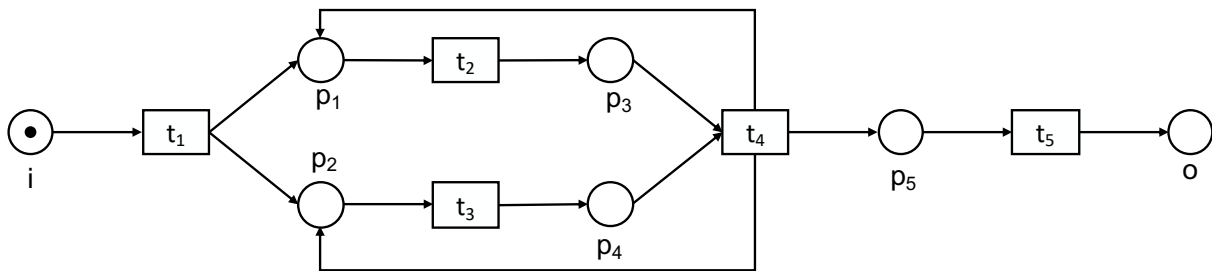
Summer term 2020

## Exercise sheet 1

Recap

### Exercise 1: Soundness

Is the following Petri net sound?



### Solution

To evaluate soundness the following four criteria have to be true:

- i) *Safe*: Places cannot hold multiple tokens at the same time.
- ii) *Proper completion*: If sink place is marked, all other places should be empty.
- iii) *Option to complete*: It is always possible to mark the sink place.
- iv) *No dead transitions*: All parts of the model are potentially reachable

For the Petri net above the evaluation is as follows:

- i) *Safe*: The place  $p_5$  can accumulate more than one token, therefore the net is not safe. ✗
- ii) *Proper completion*: Because of the AND-split in  $t_4$  at least two tokens will always remain in the net when the sink place is marked. ✗
- iii) *Option to complete*: No matter the sequence or combination transitions are enabled, it is always possible to mark the sink place. ✓
- iv) *No dead transitions*: Given the initial marking of the Petri net, all parts of the model are potentially reachable. ✓

It therefore follows that the Petri net is not sound.

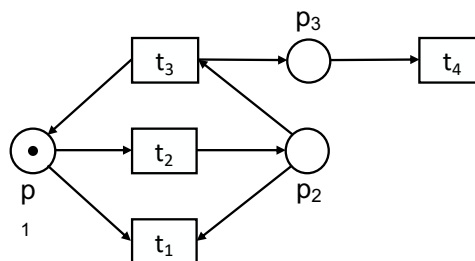
## Exercise 2: Liveness and Reversibility

Create a Petri net with the following properties:

- The Petri net is not live
- The Petri net is reversible
- The initial marking has exactly one token
- Deadlock free
- Petri net is not bounded

### Solution

A possible solution is the following Petri net:



### Exercise 3: Workflow Nets

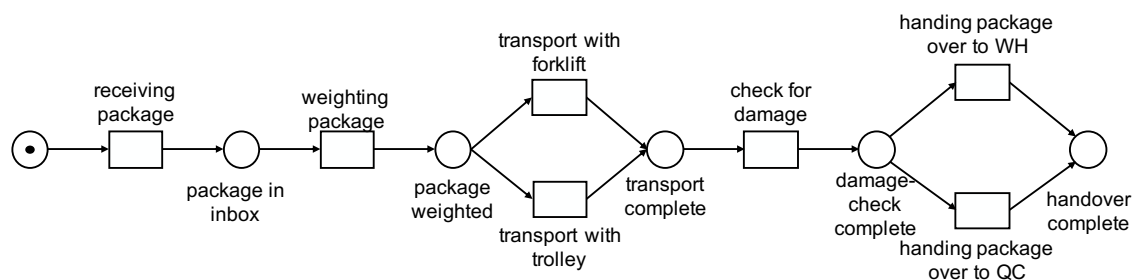
- a) Model the following scenario as a workflow net:

When a package is received there are two different ways the package is processed depending on its weight. Heavy packages will be transported with a forklift to the warehouse. For light packages a trolley is used. If the package is damaged during transport the package has to be redirected to the quality control department. Otherwise it will be handed over to the warehouse. The process ends once the package has been handed over.

- b) Is it possible for a workflow net to have more than one source place or sink place? Justify your answer.
- c) How can a mix-up between tokens from different cases be prevented in a model?

### Solution

- a)



- b) Sink and source places have to be unique. A net with an additional sink or source place would no longer be *strongly connected*.
- c) The process model is instantiated. That means for every new instance a copy of the same workflow net is created and tokens of different cases are not mixed.

**Exercise 4: Case vs. Trace**

A	B	C	D	E
1	1336	2020-03-02 04:12:30	receive specimen	Kenzo
	1349	2020-03-03 04:07:56	quality control	Helena
	1365	2020-03-04 20:14:51	pack specimen	Hans
	1376	2020-03-05 03:24:29	forward to shipping	Hans
2	1374	2020-03-05 19:03:53	receive specimen	Paloma
	1378	2020-03-06 00:52:49	quality control	Helena
	1394	2020-03-06 13:55:45	discard specimen	Helena
3	1370	2020-03-09 23:19:48	receive specimen	Kenzo
	1352	2020-03-10 18:53:53	quality control	Helena
	1367	2020-03-11 06:18:06	pack specimen	Hans
	1368	2020-03-12 01:34:52	forward to shipping	Hans
4	1337	2020-03-12 11:47:09	recieve specimen	Kenzo
	1334	2020-03-12 23:35:01	quality control	Helena
	1353	2020-03-16 00:56:33	pack specimen	Sabrina
	1355	2020-03-16 17:49:18	forward to shipping	Sabrina

- Inspect the table above and replace the placeholder names of the columns with the correct terms. Give a short description of the term.
- Explain the relationship between the following terms and give an example for each: *case*, *trace*, *event* and *activity*.

## Solution

### a) A: Case ID

The first column contains a unique case ID. A case is composed of a sequence of activities, also referred to as trace.

### B: Event ID

An event ID is a unique identifier and helps differentiate between activities with the same name.

### C: Timestamp

The timestamps provide information about the timeline in which activities were executed. Depending on the scenario, the timestamp can be indicated to the second one or less exact. Some event logs additionally provide the information when the activity was started and when completed. If only one timestamp is given for an activity and it is not indicated otherwise, it should be assumed that this is the time the activity was completed.

### D: Activity

This column denotes the activity that has been executed at the indicated time.

### E: Resource

A resource does not necessarily need to be included in an event log. It links a person, machine etc. to the executed activity. Information about the resource can provide very useful insights for process mining, but especially when it involves individual persons, concerns about privacy should be raised.

- b) A *trace* is an execution of a process and is a sequence of *events* (and not *activities*). It represents a pattern of recorded *activities*. Exactly one of these *traces* are found in a *case*. Multiple *cases* can contain the same *trace* i.e. the same sequence of recorded *activities*. An event corresponds to an *activity*, a unique identifier, a timestamp and additional information about the *event*.

Examples:

**Case** Case 1:  $\langle a, c, d, e, f \rangle$ , Case 2:  $\langle a, c, d, e, f \rangle$

**Trace** Trace x:  $\langle a, c, d, e, f \rangle$ , Trace y:  $\langle a, c, d, c, d, e, f \rangle$

**Event** [caseID: 1, eventID: 48736, Timestamp: 2020-04-10 07:32:34, Activity: a]  
[caseID: 2, eventID: 48738, Timestamp: 2020-04-10 07:39:34, Activity: c]

**Activity** *a, c* etc. are used as placeholders for proper activity names like *receive goods, contact customer* etc.