How to do conformance checking in ProM

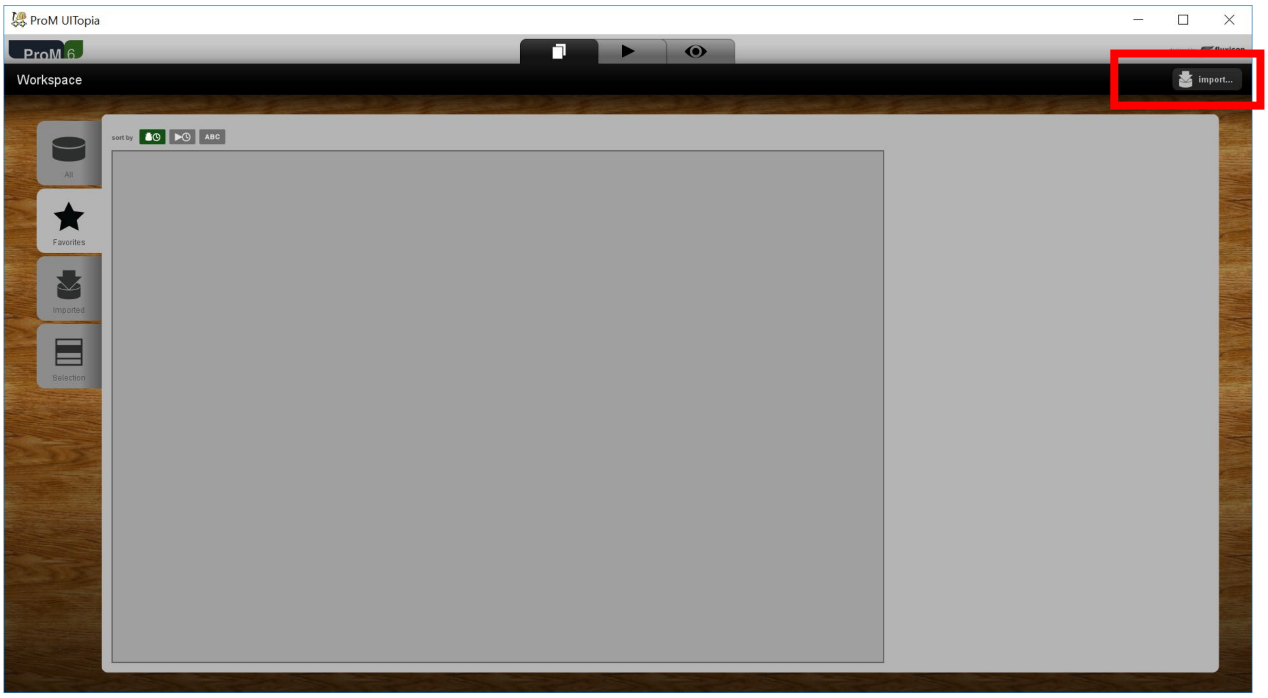
Josep Carmona
Boudewijn van Dongen
Matthias Weidlich
Andreas Solti

This tutorial provides a step-by-step manual for conformance checking in ProM, using data from the BPI challenge 2012 as well as some models for that data. The theoretical background is provided by the book "Conformance Checking", available through Springer.

The ProM toolkit can be downloaded from www.promtools.org. This tutorial uses ProM Lite, but all versions of ProM can be used for conformance checking as long as the package "Alignment" is installed in the package manager.

Starting ProM

After launching ProM, the main screen is shown.

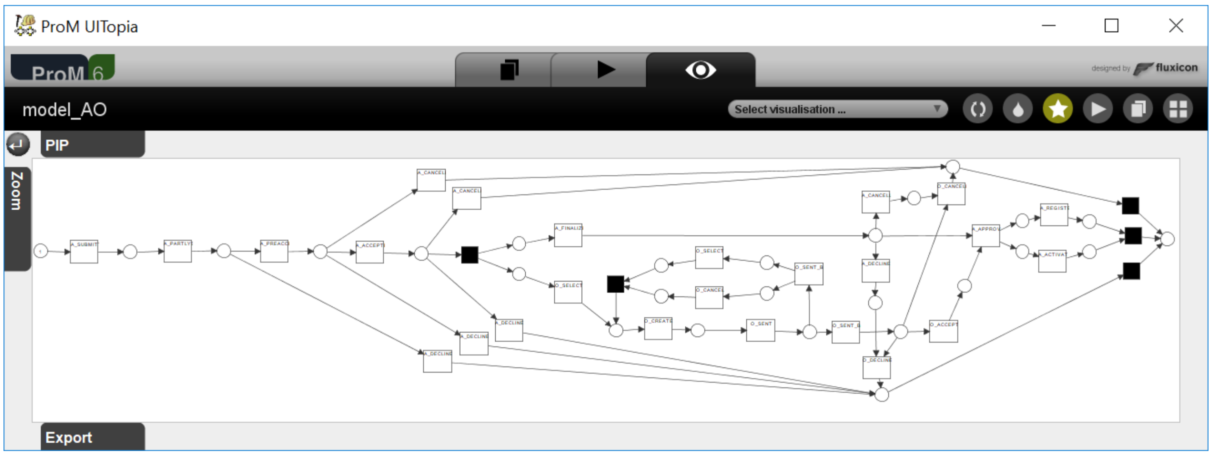


Opening files

Files can be loaded by clicked on the *Import* button, which is highlighted through a red box in the figure above. In the laboratory session, we use data from the BPI Challenge 2012 (https://data.4tu.nl/repository/uuid:3926db30-f712-4394-aebc-75976070e91f) and models provided through Canvas

There are three models for this dataset, namely a model for "A_" activities, a model for the "O_" activities and a model for the "AO_" activities. Therefore, the first step in conformance checking is to filter the event log to keep only those activities that are relevant for the model under consideration.

In the tutorial, we focus on the "AO_" model, which looks like this:

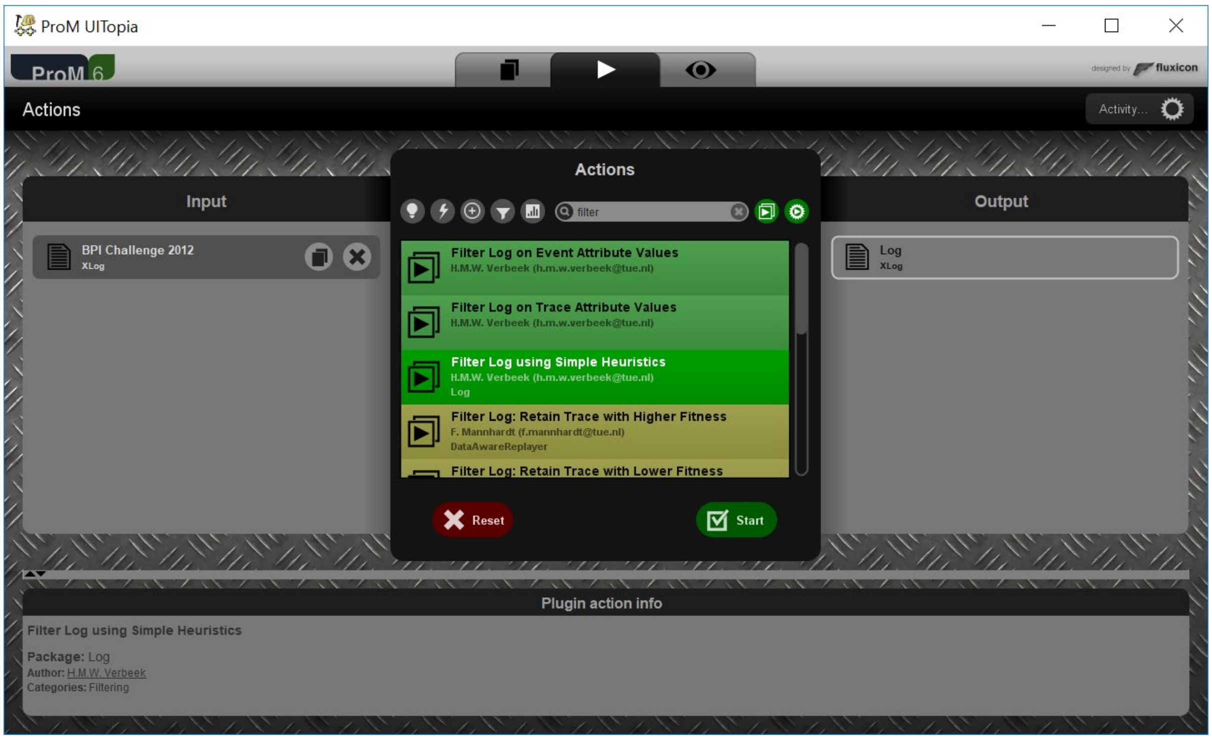


All three models are man-made, i.e. they are not the result of a mining algorithm The models have properties like duplicate transitions (transitions with identical labels) and routing transitions (transitions that are black and cannot be observed).

A correct execution of the model starts with a token in the initially marked place and ends with one token in the final place. The question for conformance checking is: What is the relation between the behaviour in the log and the behaviour described in the model.

Filtering the log

For filtering, we use the simple heuristics filter which is contained in a plugin called "Filter Log using Simple Heuristics".



BPI Challenge 2012
XLog

Log
XLog

Filter Log on Event Attribute Values
H.M.W. Verbeek (h.m.w.verbeek@tue.nl)

Filter Log on Trace Attribute Values
H.M.W. Verbeek (h.m.w.verbeek@tue.nl)

Filter Log using Simple Heuristics
H.M.W. Verbeek (h.m.w.verbeek@tue.nl)
Log

Filter Log: Retain Trace with Higher Fitness
F. Mannhardt (f.mannhardt@tue.nl)
DataAwareReplayer

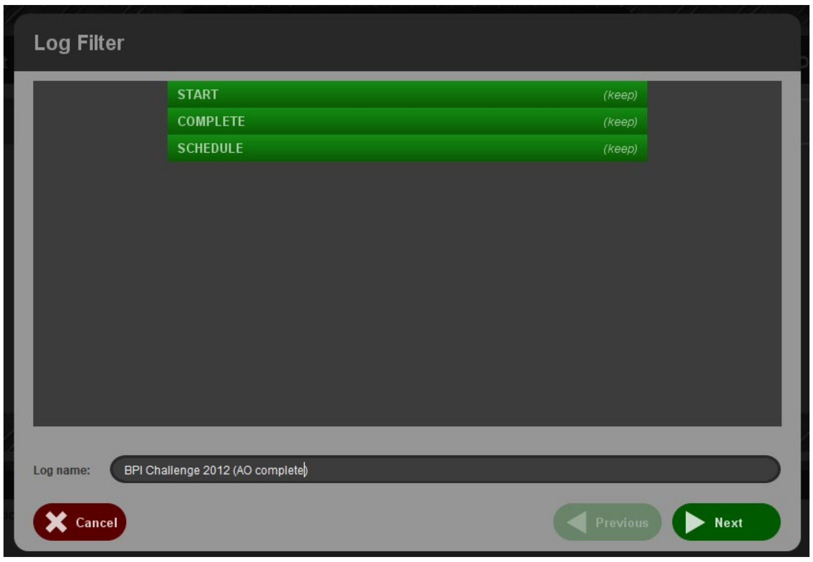
Filter Log: Retain Trace with Lower Fitness

Filter Log using Simple Heuristics

Package: Log
Author: H.M.W. Verbeek
Categories: Filtering

After starting this plugin, you get a wizard that guides you through the filtering steps. To only keep the activities that start with "A_" or "O_" and only complete cases, we need to perform filtering as follows:

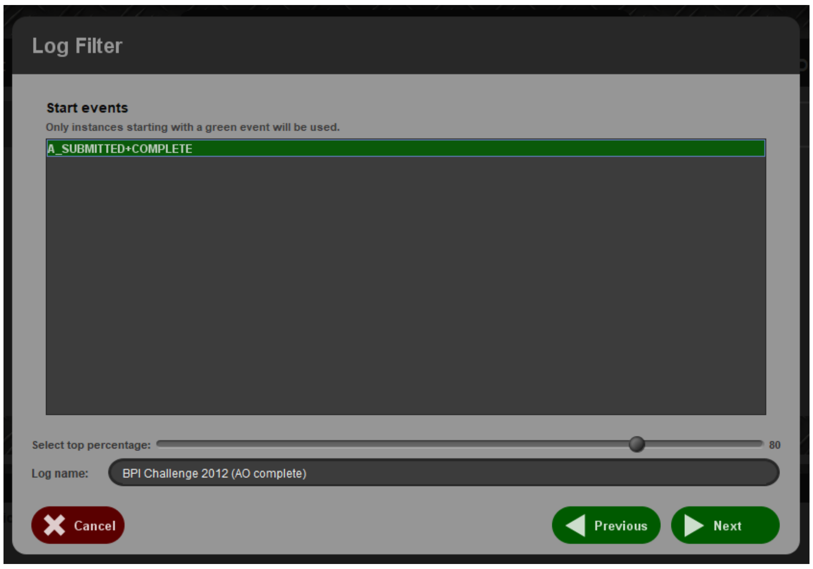
1) Keep all lifecycle types and change the names accordingly.



START	(keep)
COMPLETE	(keep)
SCHEDULE	(keep)

Log name: BPI Challenge 2012 (AO complete)

2) Keep all start events


Log Filter

Start events

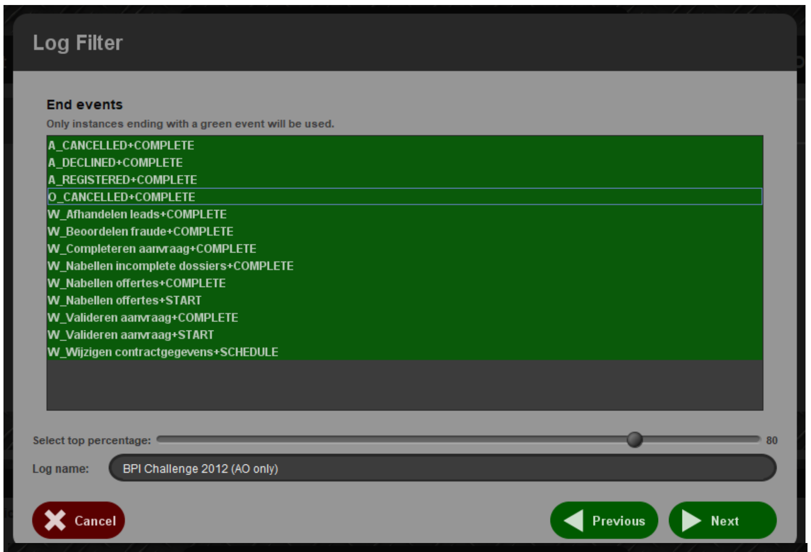
Only instances starting with a green event will be used.

A_SUBMITTED+COMPLETE

Select top percentage: 80

Log name: BPI Challenge 2012 (AO complete)

3) For now, keep all end events


Log Filter

End events

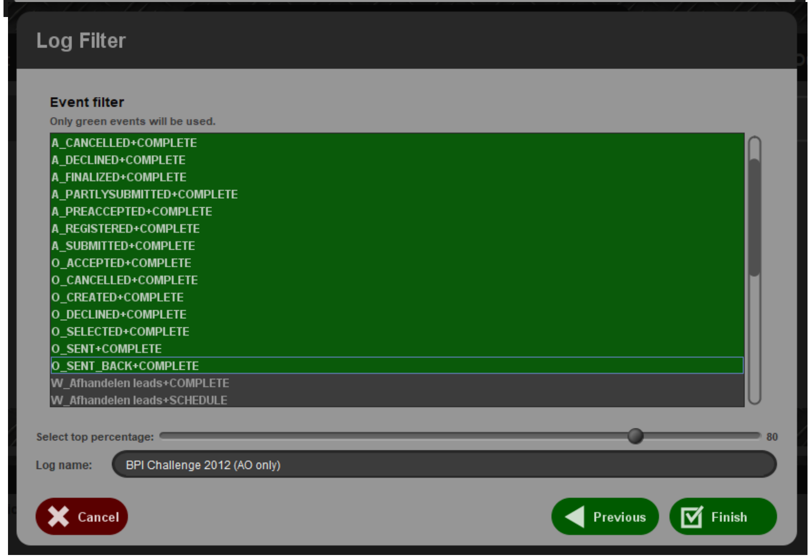
Only instances ending with a green event will be used.

A_CANCELLED+COMPLETE
A_DECLINED+COMPLETE
A_REGISTERED+COMPLETE
O_CANCELLED+COMPLETE
W_Afhandelen leads+COMPLETE
W_Beoordelen fraude+COMPLETE
W_Completeren aanvraag+COMPLETE
W_Nabellen incomplete dossiers+COMPLETE
W_Nabellen offertes+COMPLETE
W_Nabellen offertes+START
W_Valideren aanvraag+COMPLETE
W_Valideren aanvraag+START
W_Wijzigen contractgegevens+SCHEDULE

Select top percentage: 80

Log name: BPI Challenge 2012 (AO only)

4) Finally, keep all events that start with "A_" or with "O_"


Log Filter

Event filter

Only green events will be used.

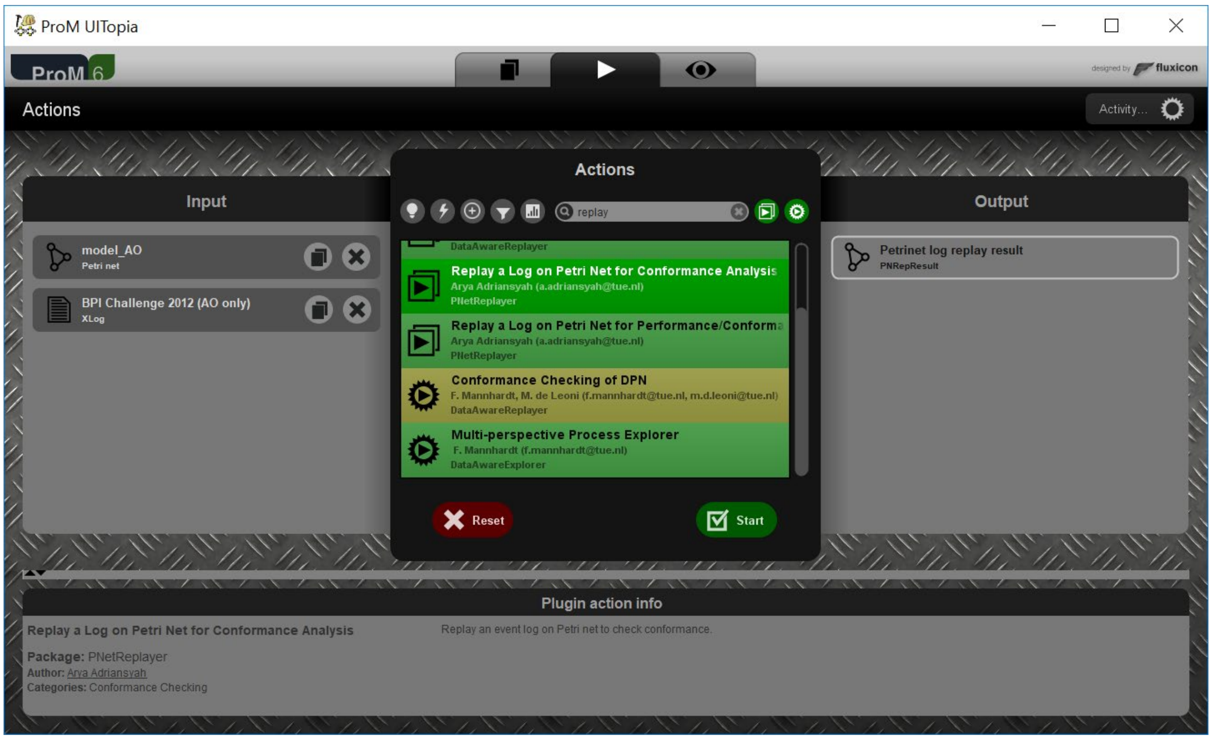
A_CANCELLED+COMPLETE
A_DECLINED+COMPLETE
A_FINALIZED+COMPLETE
A_PARTLYSUBMITTED+COMPLETE
A_PREACCEPTED+COMPLETE
A_REGISTERED+COMPLETE
A_SUBMITTED+COMPLETE
O_ACCEPTED+COMPLETE
O_CANCELLED+COMPLETE
O_CREATED+COMPLETE
O_DECLINED+COMPLETE
O_SELECTED+COMPLETE
O_SENT+COMPLETE
O_SENT_BACK+COMPLETE
W_Afhandelen leads+COMPLETE
W_Afhandelen leads+SCHEDULE

Select top percentage: 80

Log name: BPI Challenge 2012 (AO only)

Conformance Checking

After selecting both a log and a model, select the “Replay a log on Petri net for Conformance Analysis” plugin.



model_AO
Petri net

BPI Challenge 2012 (AO only)
XLog

DataAwareReplayer

Replay a Log on Petri Net for Conformance Analysis
Arya Adriansyah (a.adriansyah@tue.nl)
PNetReplayer

Replay a Log on Petri Net for Performance/Conforma
Arya Adriansyah (a.adriansyah@tue.nl)
PNetReplayer

Conformance Checking of DPN
F. Mannhardt, M. de Leoni (f.mannhardt@tue.nl, m.d.leoni@tue.nl)
DataAwareReplayer

Multi-perspective Process Explorer
F. Mannhardt (f.mannhardt@tue.nl)
DataAwareExplorer

Petrinet log replay result
PNRepResult

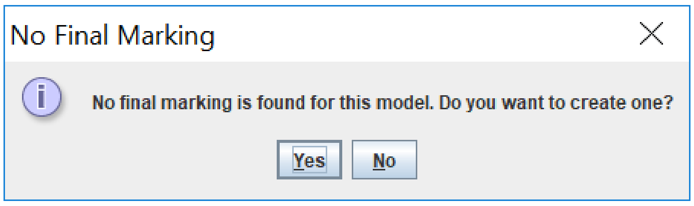
Replay a Log on Petri Net for Conformance Analysis

Replay an event log on Petri net to check conformance.

Package: PNetReplayer
Author: Arya Adriansyah
Categories: Conformance Checking

Conformance setup

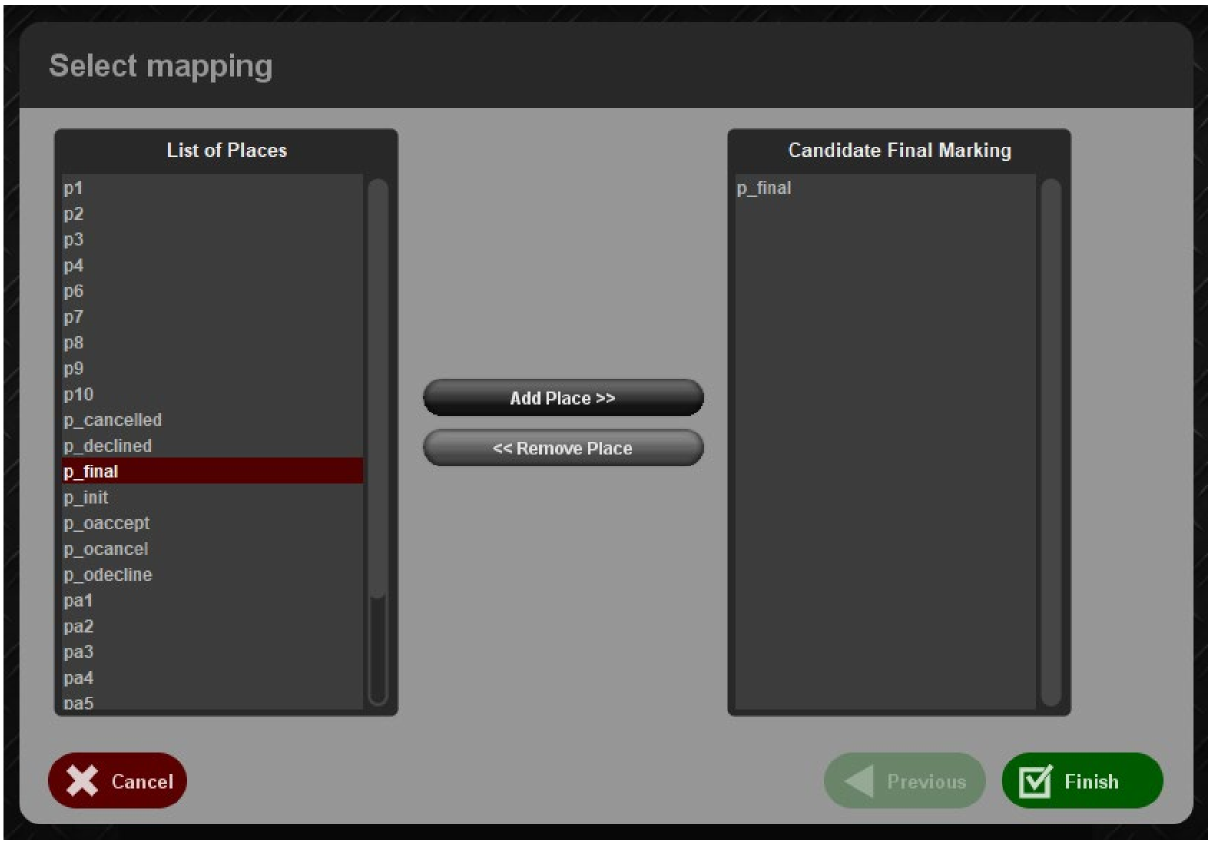
If no final marking is known for the model yet, you will be asked to provide one.



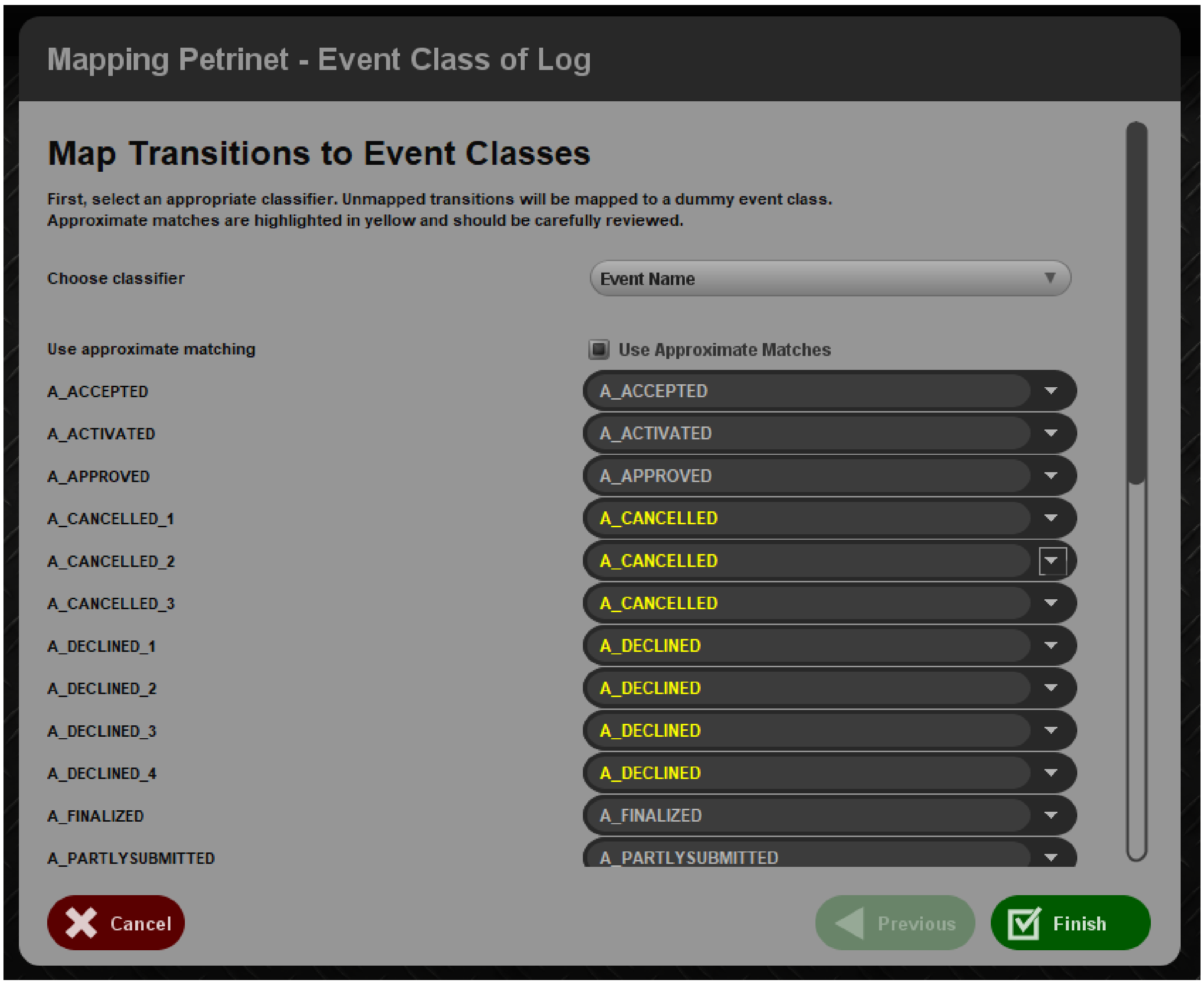
No Final Marking

No final marking is found for this model. Do you want to create one?

For our model, we need the model to terminate in the marking [p_final] and that is what we specifiy in the following dialog:



In the next step, we need to link labels of events in the log to labels of transitions in the model. By default, ProM will suggest to use the “MXML Legacy Classifier” which for this log is not the right choice. Instead, we select the “Event Name” classifier.


Mapping Petrinet - Event Class of Log

Map Transitions to Event Classes

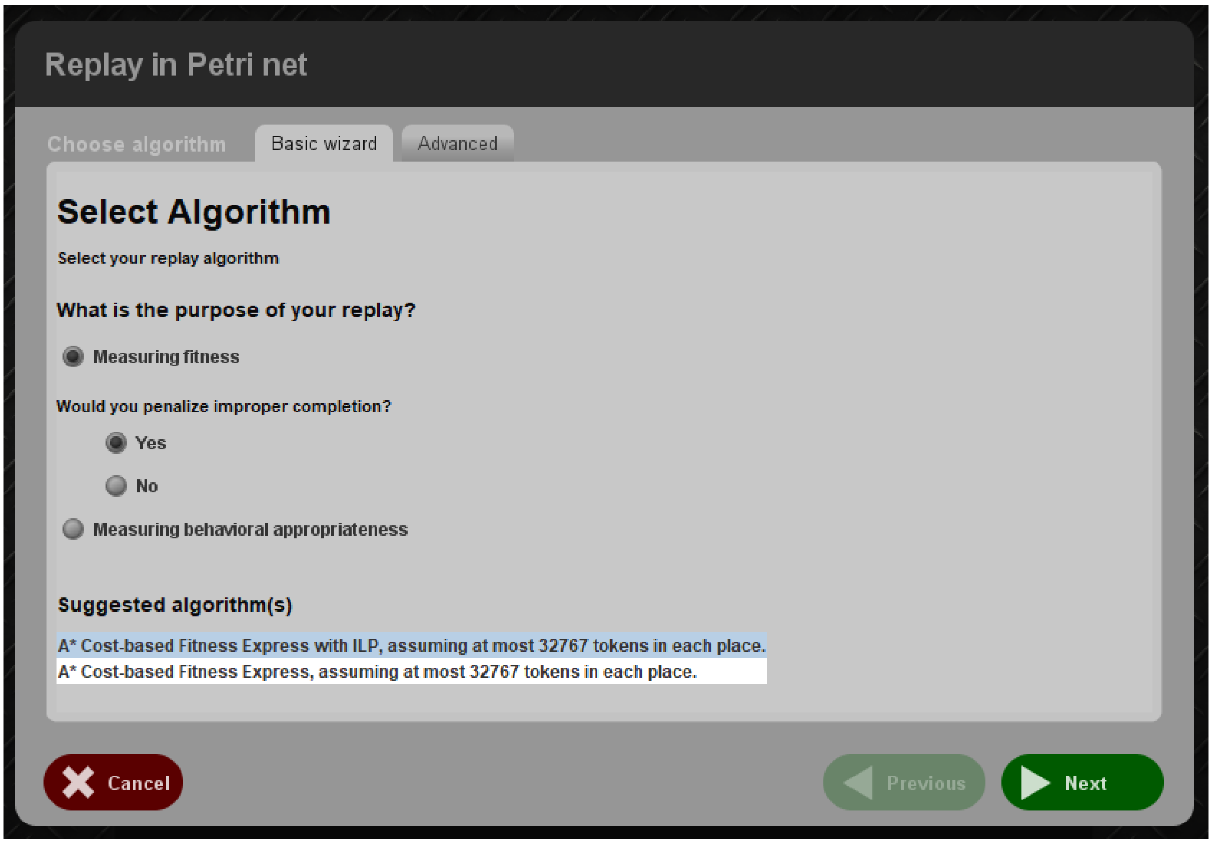
First, select an appropriate classifier. Unmapped transitions will be mapped to a dummy event class.
Approximate matches are highlighted in yellow and should be carefully reviewed.

Choose classifier	Event Name
Use approximate matching	Use Approximate Matches
A_ACCEPTED	A_ACCEPTED
A_ACTIVATED	A_ACTIVATED
A_APPROVED	A_APPROVED
A_CANCELLED_1	A_CANCELLED
A_CANCELLED_2	A_CANCELLED
A_CANCELLED_3	A_CANCELLED
A_DECLINED_1	A_DECLINED
A_DECLINED_2	A_DECLINED
A_DECLINED_3	A_DECLINED
A_DECLINED_4	A_DECLINED
A_FINALIZED	A_FINALIZED
A_PARTLYSUBMITTED	A_PARTLYSUBMITTED

The tool then proposes a mapping based on string similarity. Labels that match fully are shown in grey. Labels that do not match fully are shown in yellow and these should manually be checked. In our case, the automatic mapping is correct and we can click “Finish”

Conformance Checking

After setting up the model with an initial and final marking as well as the mapping between the log and the model, we continue with conformance checking.


Replay in Petri net

Choose algorithm | Basic wizard | Advanced

Select Algorithm

Select your replay algorithm

What is the purpose of your replay?

[✓ Measuring fitness]

Would you penalize improper completion?

[✓ Yes] No

Measuring behavioral appropriateness

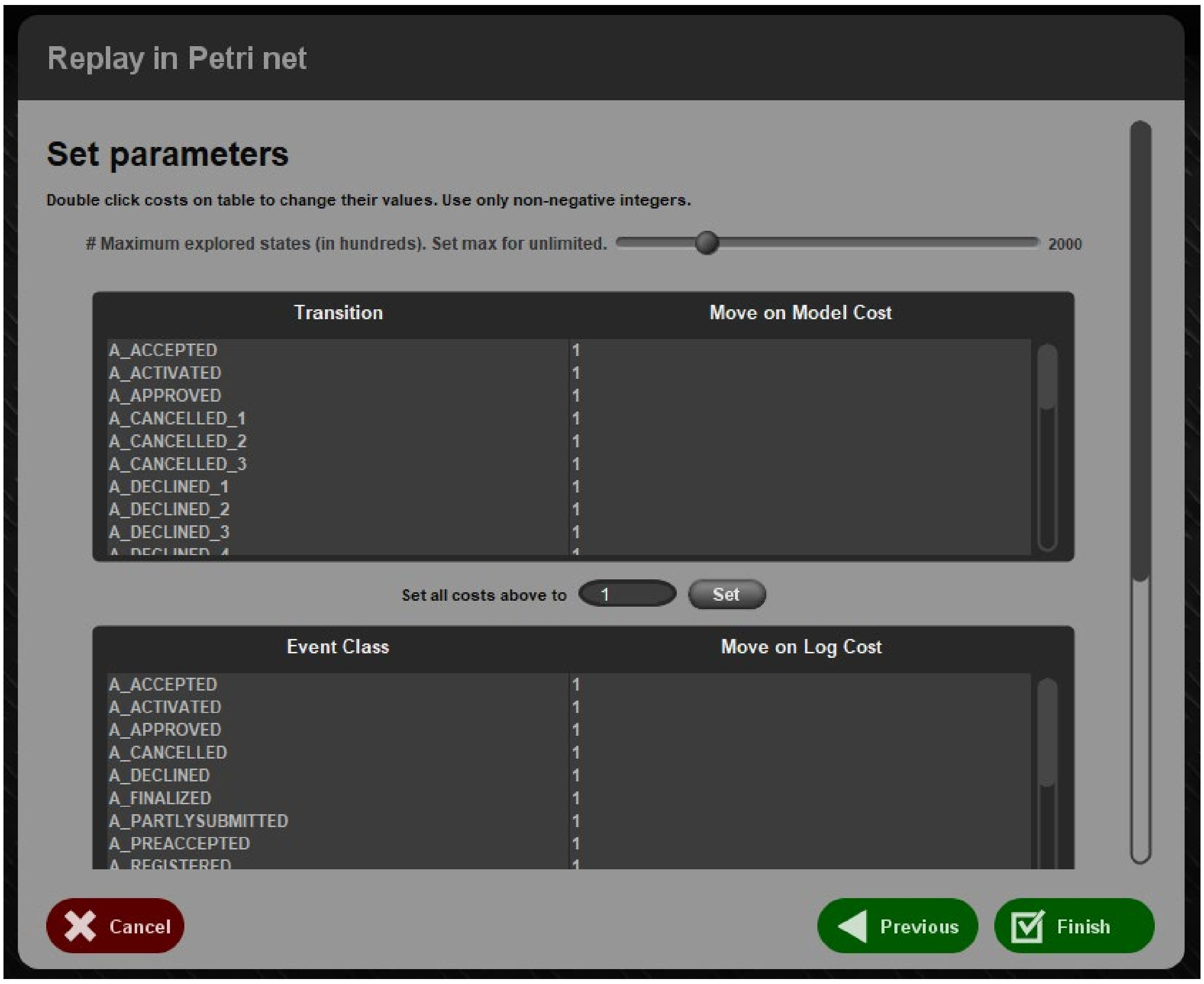
Suggested algorithm(s)

A* Cost-based Fitness Express with ILP, assuming at most 32767 tokens in each place.

A* Cost-based Fitness Express, assuming at most 32767 tokens in each place.

ProM contains many techniques, but the default suffices for this tutorial:

In the next screen, we can set the costs of deviations.


Replay in Petri net

Set parameters

Double click costs on table to change their values. Use only non-negative integers.

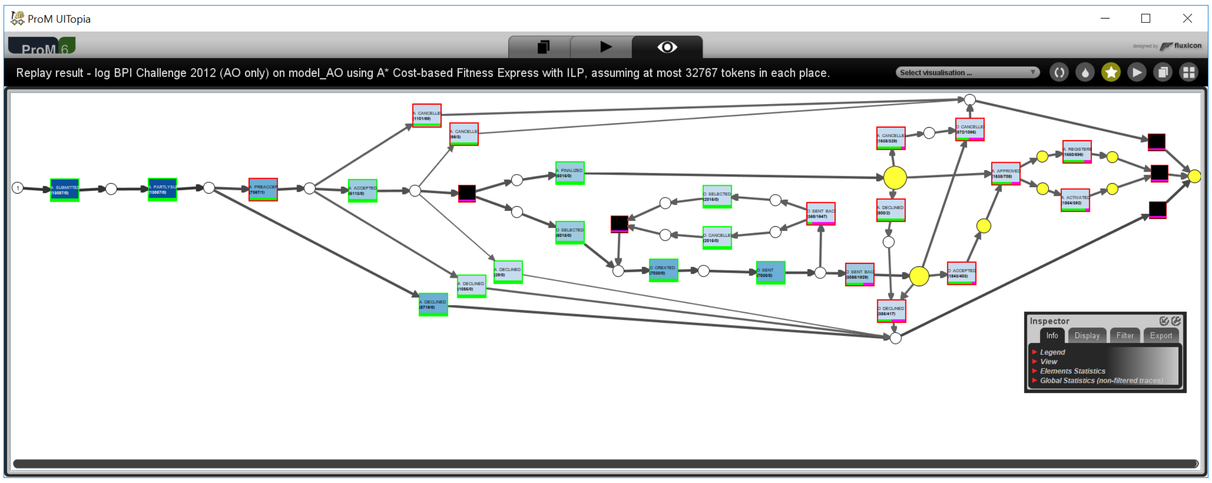
Maximum explored states (in hundreds). Set max for unlimited. 2000

Transition	Move on Model Cost
A_ACCEPTED	1
A_ACTIVATED	1
A_APPROVED	1
A_CANCELLED_1	1
A_CANCELLED_2	1
A_CANCELLED_3	1
A_DECLINED_1	1
A_DECLINED_2	1
A_DECLINED_3	1
A_DECLINED_4	1

Set all costs above to 1 Set

Event Class	Move on Log Cost
A_ACCEPTED	1
A_ACTIVATED	1
A_APPROVED	1
A_CANCELLED	1
A_DECLINED	1
A_FINALIZED	1
A_PARTLYSUBMITTED	1
A_PREACCEPTED	1
A_REGISTERED	1

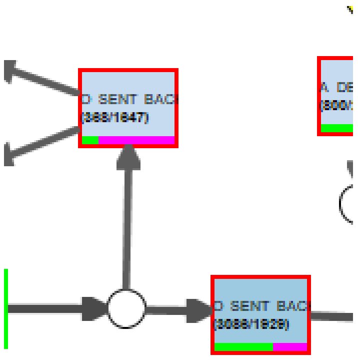
Again, the default cost function suffices and after clicking the analysis starts. The resulting screen is a model with colors indicating the deviations:


ProM UITopia

Replay result - log BPI Challenge 2012 (AO only) on model_AO using A* Cost-based Fitness Express with ILP, assuming at most 32767 tokens in each place.

The darker blue the fill is, the more frequent an activity in the log. The bottom of each transition shows green/purple bar indicating the fraction of correct/incorrect executions. Places that are coloured yellow indicate that there was an event in the log that could not be explained when there was a token in that place.

What is striking in this picture is the large amount of purple bars in the "O_SENT_BACK" transition. Apparently, these transitions are often necessary to explain a trace in the log, while they did not appear in the log, i.e. they appear to have been skipped during execution of the process.

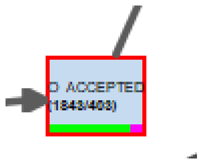


At this stage, it's up to you to contact the process owner and the check if this is the case and if so, why?

The explanation here is that the computer system used only allows for one offer to be created for each application, while customers prefer to receive multiple offers to choose from. Call agents therefore create an offer, print it and cancel it without the offer ever leaving the bank, nor being sent back. After that the call agent create a new offer, print it and send two offers in one envelope to the customer. If then a signed offer is received back and it happens to be the first offer created, then this loop is repeated again to re-create the first offer.

A clear example of human call agents working around the intended behaviour.

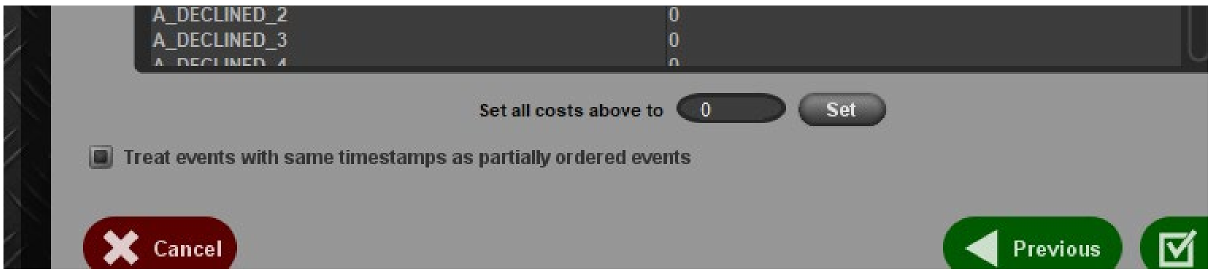
A more interesting problem is the fact that the transition "O_ACCEPTED" appears to be skipped frequently. It is listed as being skipped 403 times out of 2,246 cases. This would imply that for 2,246 cases an offer was not accepted, but a loan was approved and activated (money payed to the customer).



Dealing with time

Unfortunately, the event log has a property that many logs have. Multiple events have the same time stamp, i.e. according to the log, they occurred at the exact same time and hence, in the log, the ordering of these events is arbitrary. As the ordering in the model is enforced by the places, this leads to false deviations.

To resolve this, do the replay again, but this time in the dialog where you can set the costs of deviation, select the option to treat events with the same timestamps as partially ordered events.

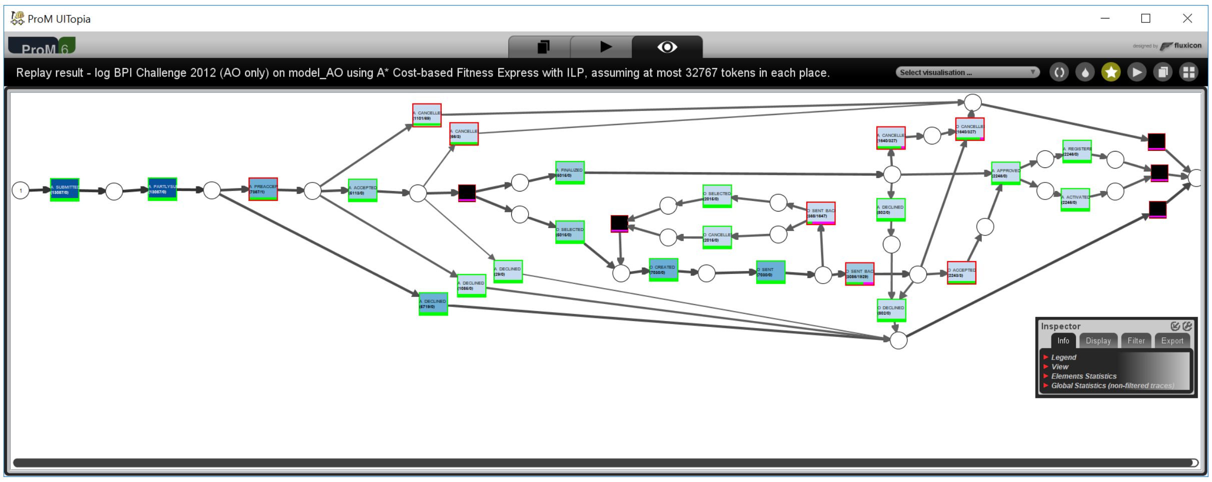


A_DECLINED_2	0
A_DECLINED_3	0
A_DECLINED_4	0

Set all costs above to 0

✓ Treat events with same timestamps as partially ordered events

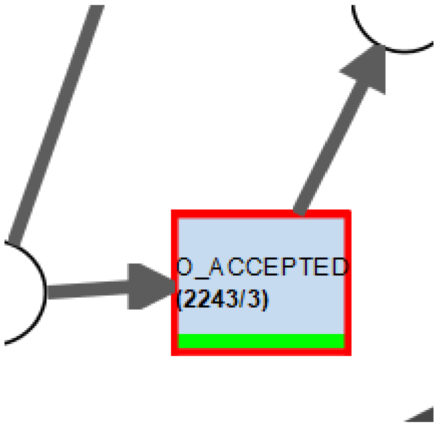
The result is a model with fewer deviations:



ProM UITopia

Replay result - log BPI Challenge 2012 (AO only) on model_AO using A* Cost-based Fitness Express with ILP, assuming at most 32767 tokens in each place.

The deviations around "O_SENT_BACK" are still there, as expected. However, now, there appear to be only three cases where "O_ACCEPTED" was skipped out of 2,246.

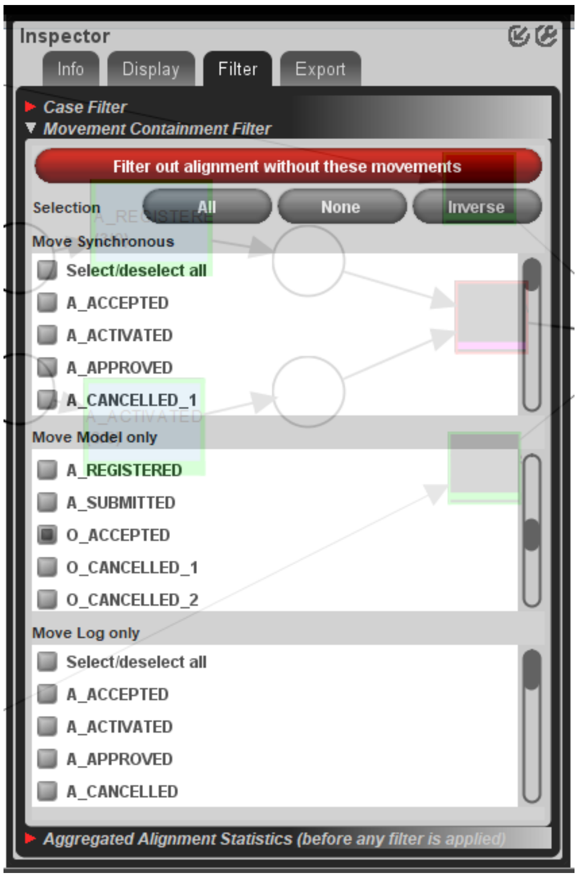


Filter based on deviations

Using the filter dialog, we filter the log to keep only those cases in which this activity "O_ACCEPTED" was skipped.

For this, we use the "Movement Containment Filter" in the inspector panel and we filter out all cases without the move on model "O_ACCEPTED". (to do this, uncheck "O_ACCEPTED" in the middle list, then click inverse and then "Filter…")

The result should now show:



Inspector

Info | Display | Filter | Export

- Case Filter
- Movement Containment Filter

Filter out alignment without these movements

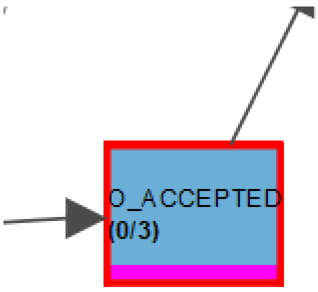
Selection: All | None | Inverse

Move Synchronous
- Select/deselect all
- A_ACCEPTED
- A_ACTIVATED
- A_APPROVED
- A_CANCELLED_1

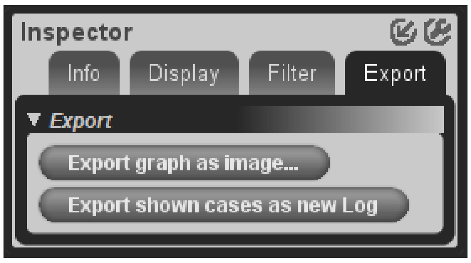
Move Model only
- A_REGISTERED
- A_SUBMITTED
- O_ACCEPTED
- O_CANCELLED_1
- O_CANCELLED_2

Move Log only
- Select/deselect all
- A_ACCEPTED
- A_ACTIVATED
- A_APPROVED
- A_CANCELLED

- Aggregated Alignment Statistics (before any filter is applied)



After filtering, we can export the log using the export tab in the inspector panel.



After exporting and giving the log an appropriate name, you obtain a log with 3 cases and 45 events.

Using the log inspector, you can manually verify that indeed, for cases 177083, 180310 and 198310, no offer was accepted, while the loans were activated, i.e. money was transferred to the client. A total of 18,000+20,000+25,000 = 63,000 euro.

This is a clear compliance violation.

Final Notes

ProM is an academic prototype. The authors have made huge efforts in making this tool accessible and functional for a large number of use cases. If you find bugs, please report them during the lectures or on the ProM Forum, but don't expect immediate fixes.

About the files

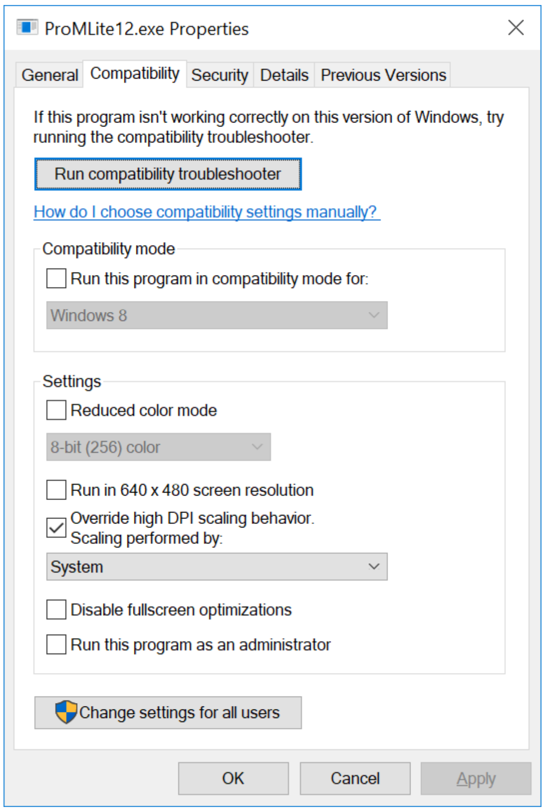
The log used was provided by a company for the BPI Challenge 2012. This challenge is organized yearly by the process mining research group in Eindhoven and challenges participants from all over the world to show their process mining abilities.

Currently known bugs/issues:

High resolution screens: The chart area cannot scale beyond a certain resolution. Beyond 2560x1440, the fonts get stretched and the location of the tooltip doesn't match the actual events. This a Java issue and can, at this point, not be circumvented.

Workaround 1: On high resolution screens, keep the ProM window smaller than 2560x1440 physical pixels.

Workaround 2: select the ProMLite12.exe file, go to properties and under compatibility, select "Override high DPI scaling behavior. Scaling performed by System". Note that this only works if the scaling is such that the scaled screen has a resolution less than 2560x1400. For a 4K screen, this is 150% scaling.


ProMLite12.exe Properties

If this program isn't working correctly on this version of Windows, try running the compatibility troubleshooter.

How do I choose compatibility settings manually?

Compatibility mode

Run this program in compatibility mode for: Windows 8

Settings

Reduced color mode: 8-bit (256) color

Run in 640 x 480 screen resolution

[✓ Override high DPI scaling behavior. Scaling performed by: System]

Disable fullscreen optimizations

Run this program as an administrator