



Process Mining Project

Giada Gabriele 235799, Federica Trovato 242498 @UniCal

House Building

Indice

1. Introduzione	2
2. Modello BPMN	3
3. Workflow Net (PN-A)	4
4. Simulazione BIMP (MXML-A)	5
5. Model mining	7
5.1 VIM-A	7
5.2 VIM-B	8
6. Conformance Checking	9
6.1 PN-A con F-XES-A	10
6.2 PN-A con XES-B	11
6.3 PN-B con F-XES-A	14
6.4 PN-B con XES-B	15
7. Conclusioni	17

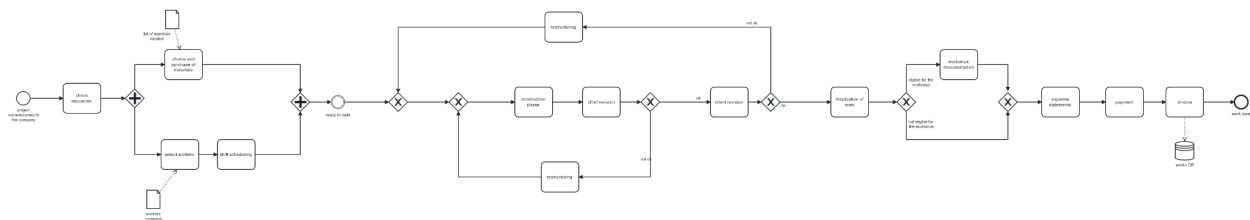
1. Introduzione

In questo report, analizzeremo il processo di costruzione di una casa, utilizzando le tecniche di process mining. Questa analisi prende in esame il processo sin da quando il progetto della casa viene commissionato all'impresa edile, e termina nel momento in cui la casa è pronta. Vengono analizzati gli aspetti riguardanti l'approvvigionamento dei materiali necessari, il lavoro manuale degli operai e le varie pratiche fiscali, che comprendono, ove previsto, anche l'utilizzo dell'ecobonus in fase di pagamento.

Inizialmente abbiamo modellato il processo utilizzando BPMN, un linguaggio molto diffuso per la modellazione di processi aziendali, poi ne abbiamo simulato un event log, utilizzando il simulatore BIMP; questo ci ha permesso di simulare il processo in uno scenario reale e identificare eventuali colli di bottiglia o potenziali inefficienze. Successivamente utilizzando WoPED (Workflow Petri Net Designer), un software utilizzato per simulare ed analizzare i processi aziendali, abbiamo creato un Petri Net. Ed infine, abbiamo analizzato l'event log ed eseguito il controllo di conformità, confrontando il processo reale con quello modellato, utilizzando ProM.

2. Modello BPMN

Il nostro modello BPMN riportato di seguito è stato creato con la piattaforma bpmn.io:

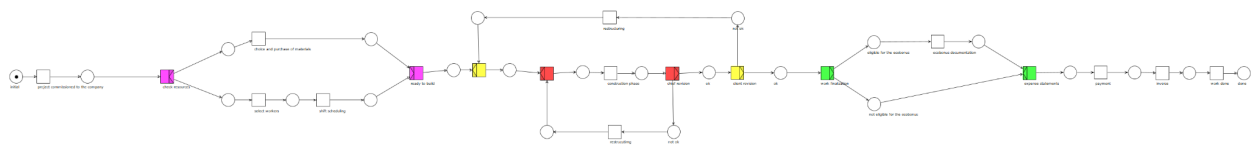


1. Modello BPMN

Il processo inizia con il commissionamento del progetto da parte del cliente alla ditta incaricata per la costruzione, seguono poi le fasi di approvvigionamento dei materiali da costruzione, la selezione degli operai da parte della ditta con relativi turni di lavori; solo dopo queste fasi preliminari, i lavori possono iniziare. Inizia così la fase di costruzione, che è poi seguita dalla revisione dei lavori da parte del capo cantiere, e questa, se ha esito negativo comporta una fase di ristrutturazione/rifacimento dei lavori non soddisfacenti, altrimenti se ha esito positivo si può procedere alla fase di revisione da parte del cliente. Quest'ultima in caso negativo avrà la stessa sorte della fase precedente, e quindi si passerà alla fase di modifiche dei lavori effettuati ma non soddisfacenti, in caso di approvazione da parte del cliente si passerà invece alla finalizzazione dei lavori. A lavori ultimati si procede con le pratiche fiscali e burocratiche se il cliente utilizzerà l'ecobonus statale, altrimenti si salterà questa fase di documentazione dell'ecobonus, e si procederà con le varie fatture e pagamenti; solo adesso potrà terminare il nostro processo.

3. Workflow Net (PN-A)

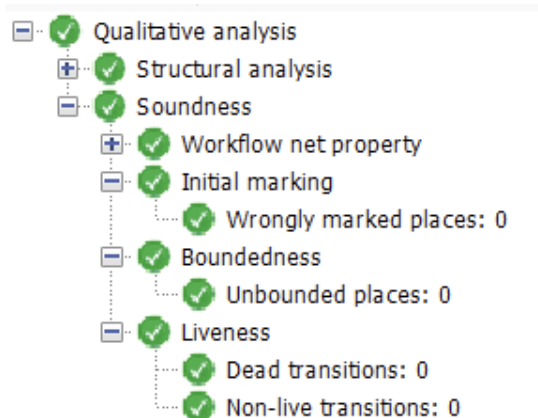
Il Workflow Net è un tipo di Petri Net, con caratteristiche specifiche. È stato ottenuto utilizzando WoPeD, un software utile per modellare processi aziendali.



2. PN-A

Analizzando tramite “operator coloring” si può notare come le transizioni siano utilizzate nella maniera corretta, inoltre, effettuando il token game abbiamo verificato l’effettiva funzionalità del Workflow Net.

L’analisi semantica riportata di seguito ci dimostra che il workflow net rispetta la struttura e il soundness:



3. Analisi semantica

4. Simulazione BIMP (MXML-A)

Abbiamo generato il log MXML-A utilizzando il simulatore [BIMP](#), che ci permette di eseguire una simulazione del nostro modello di processo assegnando i possibili tempi di completamento per ogni attività.

I criteri della simulazione sono i seguenti:

- contiene 50 istanze di processo totali
- l'orario di arrivo è specificato con una media di 60 minuti e con una deviazione standard di 30
- due tipi di schedulazione dei lavori, rispettivamente:
 - Default: Monday/Friday, from 00:00 to 23:59.
 - 24/7: Monday/Friday, from 09:00 to 17:00.


Scenario Specification

Inter arrival time	Mean *	Std deviation *	Time unit
Normal ▼	60 ▾	30 ▾	Minutes ▼

Total number of process instances *

50 ▾ % to exclude from sta ▾ ☒ Use the same % for tail

Scenario start date and time

December 5th 09:00 

Currency

EUR ▼

4. Specifiche dello scenario

Timetables / Work schedules +

Name	Begin day	End day	Begin time	End time	Remove
Default	Monday ▾	Friday ▾	09:00 	17:00 	
24/7	Monday ▾	Sunday ▾	00:00 	23:59 	



5. Schedulazione lavori

Di seguito le probabilità con cui sono stati definiti i vari gateway (XOR):

Gateways



N/A

Type: Exclusive (XOR)

client revision	Probability *	75 
restructuring	Probability *	25 



N/A

Type: Exclusive (XOR)

ecobonus documentation	Probability *	35 
not eligible for the ecobonus	Probability *	65 

N/A

Type: Exclusive (XOR)

restructuring	Probability *	25 
finalization of work	Probability *	75 

6. Probabilità per gli XOR

5. Model Mining

In questo progetto abbiamo ricavato due modelli dai due log a nostra disposizione, ovvero un event log filtrato (F-XES-A dalla simulazione BIMP) ed uno 'noise' (XES-B). Utilizzando ProM e il plug-in **Mine with Inductive Visual Miner**, siamo stati in grado di ottenere questi due modelli. In particolare, abbiamo ottenuto un modello normativo, che chiameremo VIM-A, dalla simulazione BIMP e un modello descrittivo, che chiameremo VIM-B, dal registro eventi alterato XES-B.

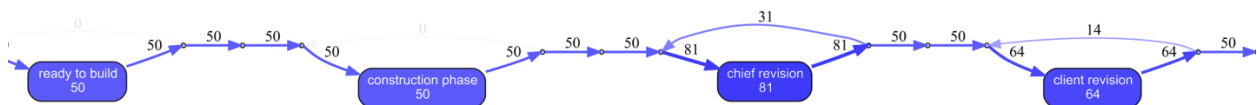
5.1 VIM-A

Per derivarlo abbiamo personalizzato le soglie per **Activities and Paths**, che sono rispettivamente 1 e 0.9:



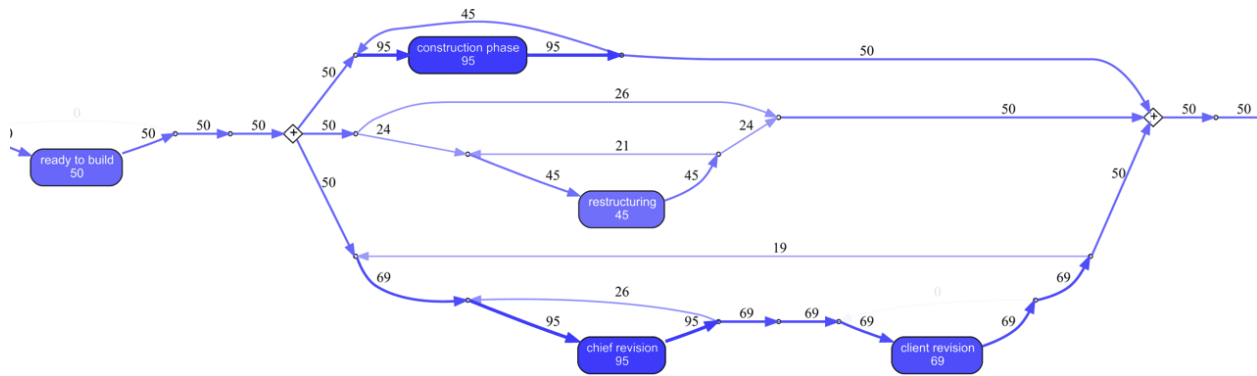
7. VIM-A

Mantenendo le soglie standard, nella parte centrale dello schema avremmo perso tutta la parte relativa ai cicli, manca la fase **restructuring**, mentre **chief revision** e **client revision** non vengono rappresentate correttamente:



8. VIM-A con soglie standard

Alzando invece il valore della soglia dei path, notiamo come l'intero XOR split-join (con cicli) viene mostrato correttamente:



9. VIM-A con soglie personalizzate

5.2 VIM-B

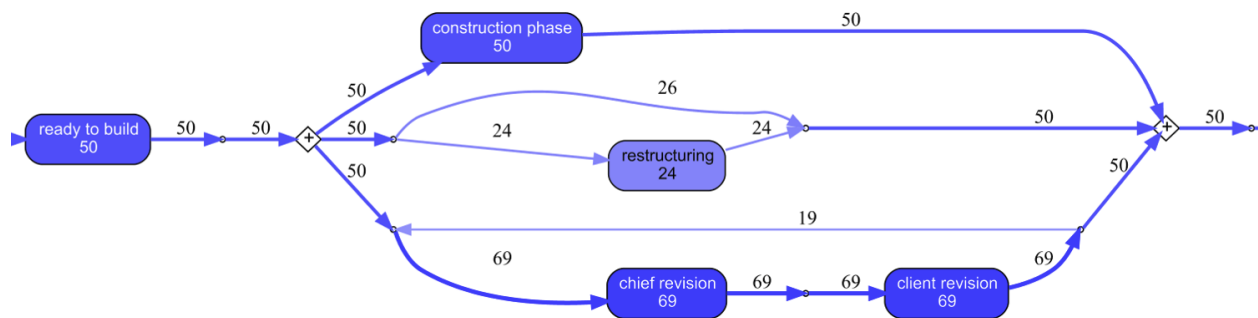
Utilizzando XES-B, e lo stesso plugin citato precedentemente per ricavare VIM-A, abbiamo derivato un altro modello: VIM-B. Anche qui abbiamo personalizzato le soglie che sono rispettivamente 1 e ≈ 0.93 :



10. VIM-B

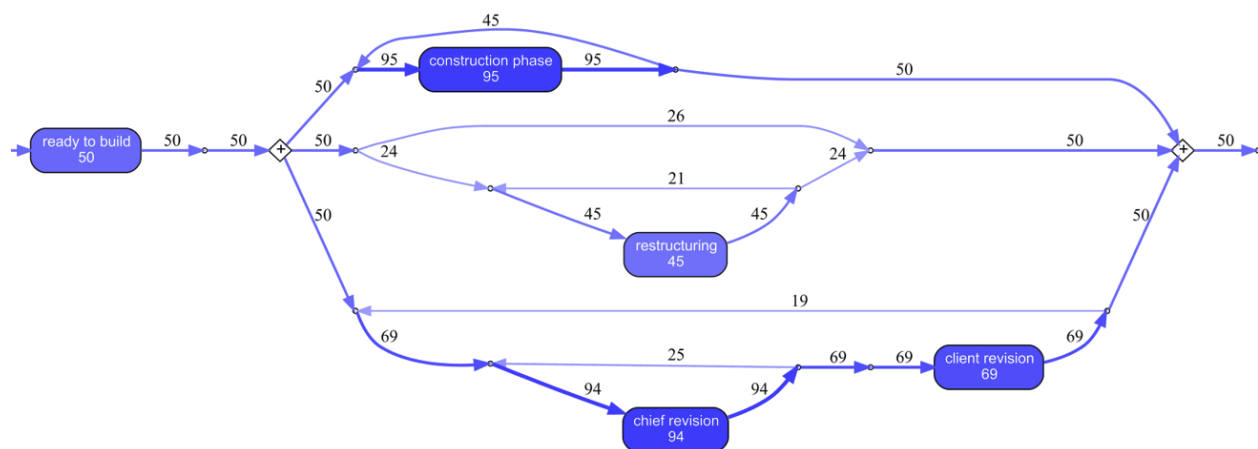
Mantenendo le soglie standard nella parte relativa ai cicli avremmo perso (problema principale) il ciclo di ritorno su **chief revision**, in questo modo sarebbe possibile tornare indietro solo dopo il fallimento

di **client revision**:



11. VIM-B con soglie standard

Invece, personalizzando il valore dei path notiamo come, anche in questo caso, lo schema viene rappresentato correttamente:



12. VIM-B con soglie personalizzate

6. Conformance Checking

Il controllo di conformità consiste nel confrontare un modello di processo con un registro di eventi dello stesso processo, per verificare se quest'ultimo è conforme al modello e viceversa, e per identificare le eventuali aree in cui i processi non funzionano come previsto. Rende possibile trovare deviazioni tra il processo effettivo e il processo previsto del modello, al fine di trovare discrepanze e correggerle.

Utilizzeremo:

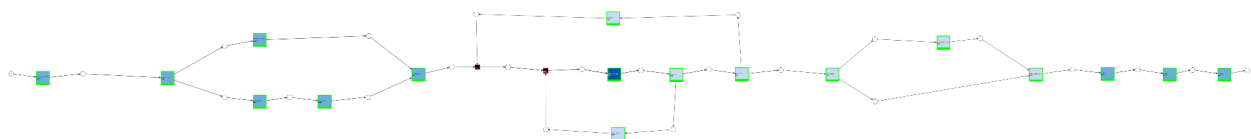
- PN-A, (modello normativo) è il workflow net modellato manualmente
- PN-B (modello descrittivo) è il workflow net ottenuto dal 'noisy' event log
- XES-B, è il 'noisy' event log
- F-XES-A, è l'event log simulato, filtrato con il plug-in "**Filter Events**"

Noi, utilizzando il plug-in **Replay a Log on Petri Net for Conformance Analysis**, abbiamo effettuato quattro controlli di conformità, rispettivamente:

- PN-A con F-XES-A
- PN-A con XES-B
- PN-B con F-XES-A
- PN-B con XES-B

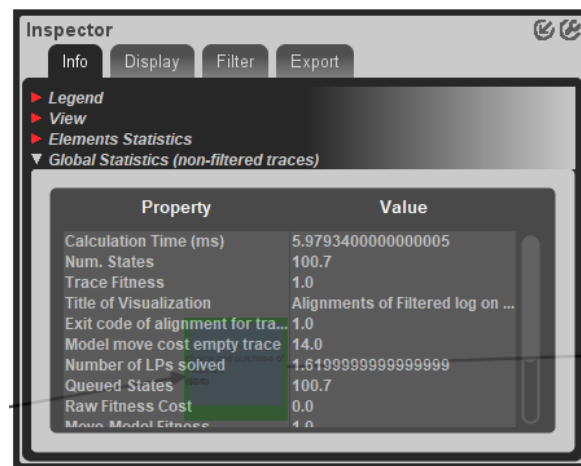
Per verificare la correttezza di un processo, ovvero quanto un processo effettivo sia conforme a quello previsto, verrà considerato il valore *Trace Fitness*, che va da 0 a 1 in base alla presenza di eventi mancanti, ripetuti, o collocati in modo errato rispetto al flusso previsto.

6.1 PN-A con F-XES-A



13. Replay con PN-A e F-XES-A

Come previsto per questo controllo di conformità, visto che sia il log F-XES-A che il modello PN-A sono stati ottenuti dallo modello BPMN, il valore di *Trace Fitness* sarà pari a 1, ovvero il modello effettivo si adatta al 100% a quello previsto, quindi nessuna istanza diverge dal comportamento atteso.

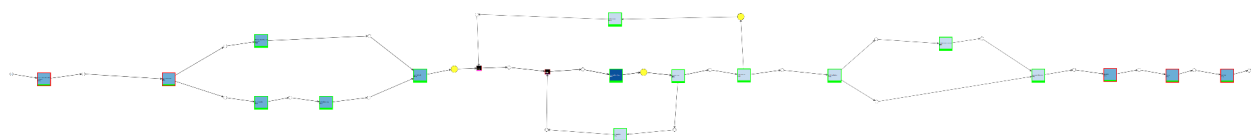


Property	Value
Calculation Time (ms)	5.979340000000005
Num. States	100.7
Trace Fitness	1.0
Title of Visualization	Alignments of Filtered log on ...
Exit code of alignment for tra...	1.0
Model move cost empty trace	14.0
Number of LPs solved	1.6199999999999999
Queued States	100.7
Raw Fitness Cost	0.0
Mean Model Fitness	1.0

14. Statistiche del replay con PN-A e F-XES-A

6.2 PN-A con XES-B

Qui confrontiamo il modello normativo con l'event log 'noise' , perciò, data la presenza di 'rumori' in XES-B, non ci aspettiamo un valore di *Trace Fitness* pari a 1 ma inferiore.

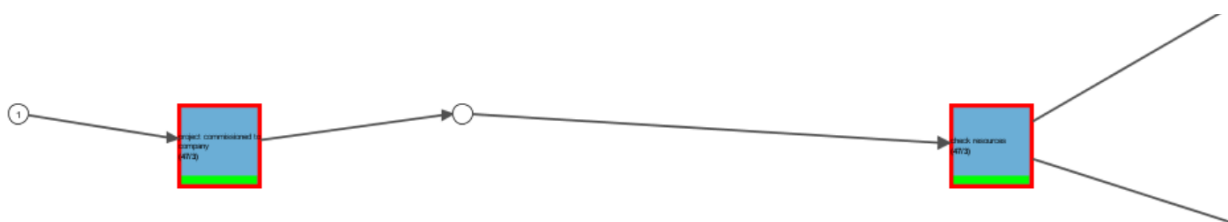


15. Replay con PN-A e XES-B

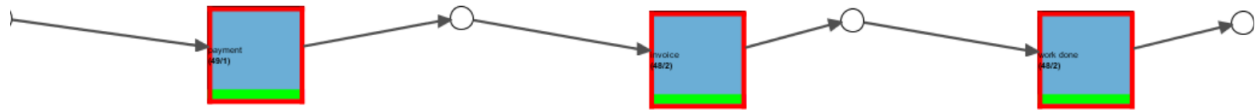
Property	Value
Calculation Time (ms)	3.737420000000006
Num. States	92.1
Trace Fitness	0.9904419844861021
Title of Visualization	Alignments of Filtered log on ...
Exit code of alignment for tra...	1.0
Model move cost empty trace	14.0
Number of LPs solved	1.5000000000000002
Queued States	92.1
Raw Fitness Cost	0.28
Mean Model Fitness	0.9904419844861021

16. Statistiche del replay con PN-A e XES-B

Alcune tracce sono anomale, nel senso che contengono sequenze di attivazione che normalmente non sono previste nel modello che abbiamo costruito e che non sarebbero possibili nel processo aziendale che abbiamo analizzato. Questo è il motivo per cui questi passaggi non possono essere replicati in modo sincrono e quindi il valore di *Trace Fitness* è inferiore a 1. Nelle due immagini di seguito possiamo vedere quali transizioni hanno il problema dell'esecuzione non sincrona, ci sono problemi di allineamento causati da un movimento nel modello della sequenza di partenza e finale che non è presente nella traccia.

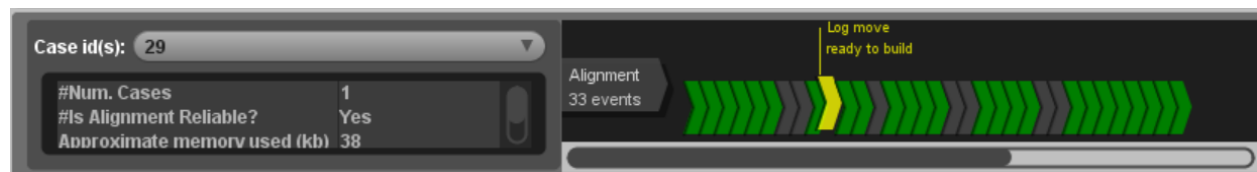


17. Initial transitions problem



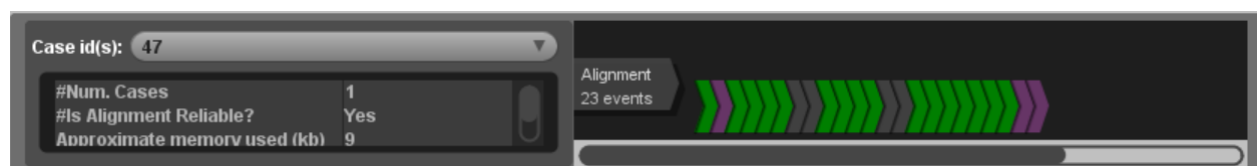
18. Final transitions problem

Se controlliamo l'allineamento è evidente che ci sono dei problemi, non è possibile 'accendere' alcune transizioni perché quelle non sono nella traccia, e questo porta ad eseguire uno spostamento nel modello. Il problema si ripete per più di 1 traccia e la maggior parte delle volte è sempre lo stesso. Sono evidenti i problemi di allineamento causati dalla presenza di sequenze di accensione errate, di alcune transizioni che vengono eseguite nel momento sbagliato o che non vengono eseguite proprio.



19. Vista Log-Model dell'istanza 29 di XES-B

Qui nel caso 29, c'è un problema di allineamento, *'ready to build'* non è nel posto in cui dovrebbe essere, e ciò ovviamente non può coincidere con un *Trace Fitness* uguale a 1.



20. Vista Log-Model dell'istanza 47 di XES-B

Qui nel caso 47 il problema è che mancano tre transizioni corrispondenti ai tre blocchi viola, e ciò comporta una riproduzione errata del modello.

6.3 PN-B con F-XES-A



21. Replay con PN-B e F-XES-A

Comparando PN-B con F-XES-A, avremo un valore di *Trace Fitness* approssimativamente pari a 0,95.

Property	Value
Calculation Time (ms)	3.865999999999997
Num. States	67.98
Trace Fitness	0.9511984128802522
Title of Visualization	Alignments of Filtered log on ...
Exit code of alignment for tra...	1.0
Model move cost empty trace	14.0
Number of LPs solved	1.0999999999999999
Queued States	68.0
Raw Fitness Cost	1.8399999999999994
Mean Model Fitness	4.0

22. Statistiche del replay con PN-B e F-XES-A

I problemi rilevati in questo controllo di conformità, sono problemi di allineamento, di movimenti asincroni sui log, come mostrato nell'immagine successiva spesso le transizioni si trovano in posti errati, e ciò chiaramente non garantisce l'accensione di una successione corretta. Inoltre è evidente come manchino cicli essenziali per la corretta riproduzione del modello come il ciclo che permette di tornare indietro dopo aver eseguito la transizione '*chief revision*'.



23. Transizioni PN-B e F-XES-A

6.4 PN-B con XES-B



24. Replay con PN-B e XES-B

Inspector

Info Display Filter Export

Legend
View
Elements Statistics

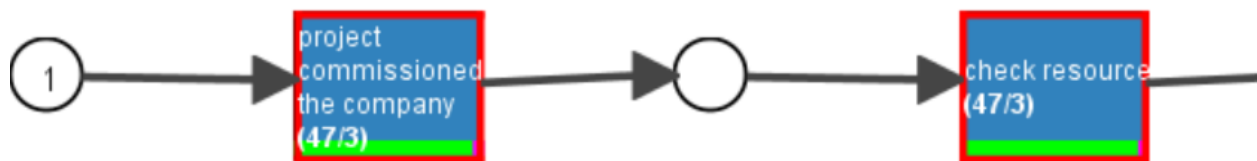
Global Statistics (non-filtered traces)

Property	Value
Calculation Time (ms)	2.2674199999999995
Num. States	67.55999999999999
Trace Fitness	0.9433352539803287
Title of Visualization	Alignments of Filtered log on ...
Exit code of alignment for tra...	1.0
Model move cost empty trace	14.0
Number of LPs solved	1.0799999999999996
Queued States	67.6
Raw Fitness Cost	2.0599999999999996
Mean Model Fitness	0.0957750403644457

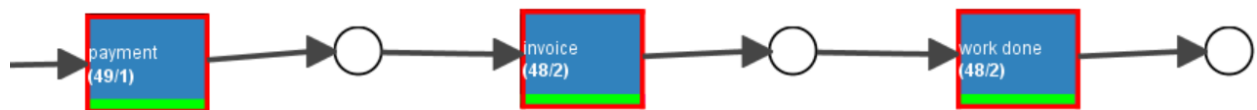
25. Statistiche del replay con PN-B e XES-B

In questo caso ci sono:

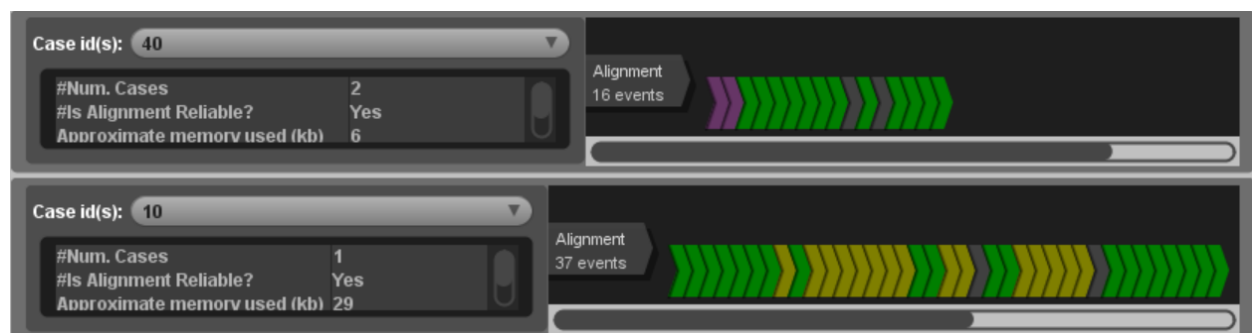
- da PN-B — F-XES-A, la mancanza di alcuni cicli, che porta quindi all'impossibilità di rieseguire alcune transizioni essenziali, infatti nelle immagini seguenti vedremo dei salti
- PN-A — XES-B, la presenza di molti problemi di transizione, rendendo così impossibile avere un risultato di riproduzione corretto. Qui di seguito gli stessi problemi riportati dalla precedente analisi di XES-B.



26. Initial transition problems



27. Final transition problems



28. Transizioni PN-B e XES-B

7. Conclusioni

Petri Net e Event Log	Trace Fitness
PN-A e F-XES-A	1
PN-A e XES-B	0.99
PN-B e F-XES-A	0.95
PN-B e XES-B	0.94

Notiamo da questo schema finale come ovviamente la conformance checking migliore è quella tra PN-A e F-XES-A, essendo prodotti dal BPMN con l'event log che ha solo gli eventi *completi* al suo interno, così come il caso peggiore è, come previsto, PN-B e XES-B derivando da schemi alterati e con 'rumore' all'interno. In generale la trace fitness ottenuta è comunque positiva avendo tutti valori superiori a 0.90.