# X International Business Process Intelligence Challenge

Vladimir Sudarev, Elena Popova, Igor Kashirin, Dmitry Usachev,
Makar Shadiyan, Anastasia Knishenko, Daniil Surnyaev

*Internal Audit Department, Rostov-on-Don, Russia*

## Analytical report of reimbursement process at TU/e

## Abstract

Any process can be executed in a different sequence of events. In some cases, the process is executed according to certain regulations, in other cases it is executed under the influence of various factors. These factors change the sequence and execution time of events in the process. Thus, to analyze the process, first of all, it is necessary to determine the reference (most popular) sequence of events in the process. To find the most popular sequence of the analyzed process, data will be clustered. The purpose of this report is to search and analyze the reference sequence of events and also to search for cases with non-standard behavior.

## Related Work

Figures and tables which illustrate behavior of particular group or cluster will be presented during the analysis. The time indicated in the tables and figures is median, taking into account all the data or a specific cluster.

First of all, it is necessary to carry out preprocessing of the data. To do this, we group the data by case_id and timestamp, then find the sequence of events, as well as the time that was spent on the transition between events. As a result, we received the table, shown below.

Table 1 – Event sequences for international trips

| | case_name | transact | time_diff |
|---|---|---|---|
| 0 | declaration 100000 | Start Log-->Declaration SUBMITTED by EMPLOYEE | 0.0 |
| 1 | declaration 100000 | Declaration SUBMITTED by EMPLOYEE-->Declaratio... | 693519.0 |
| 2 | declaration 100000 | Declaration APPROVED by ADMINISTRATION-->Decla... | 90019.0 |
| 3 | declaration 100000 | Declaration FINAL_APPROVED by SUPERVISOR-->Req... | 92624.0 |
| 4 | declaration 100000 | Request Payment-->Payment Handled | 276511.0 |
| 5 | declaration 100000 | Payment Handled-->Log End | 0.0 |
| 6 | declaration 100005 | Start Log-->Declaration SUBMITTED by EMPLOYEE | 0.0 |
| 7 | declaration 100005 | Declaration SUBMITTED by EMPLOYEE-->Declaratio... | 3.0 |
| 8 | declaration 100005 | Declaration APPROVED by ADMINISTRATION-->Decla... | 1513.0 |
| 9 | declaration 100005 | Declaration FINAL_APPROVED by SUPERVISOR-->Req... | 96068.0 |
| 10 | declaration 100005 | Request Payment-->Payment Handled | 103559.0 |
| 11 | declaration 100005 | Payment Handled-->Log End | 0.0 |

The sequence of events was obtained as a result of shifting the «concept:name» column up one position and filling in the empty values as «Log End». We initially added a new «Start Log» event for each case_id, with a timestamp that is at least one second less than the earliest timestamp in the log file. To calculate the execution time of the event, we shift the «time:timestamp» column.

Table 2 – Pivot table for column case_name and transact

| transact | Declaration APPROVED by ADMINISTRATION->Declaration APPROVED by BUDGET OWNER | Declaration APPROVED by ADMINISTRATION->Declaration FINAL_APPROVED by SUPERVISOR | Declaration APPROVED by ADMINISTRATION->Declaration REJECTED by BUDGET OWNER | Declaration APPROVED by ADMINISTRATION->Declaration REJECTED by SUPERVISOR | Declaration APPROVED by BUDGET OWNER->Declaration FINAL_APPROVED by SUPERVISOR |
|---|---|---|---|---|---|
| case_name | | | | | |
| declaration 100000 | 0 | 1 | 0 | 0 | 0 |
| declaration 100005 | 0 | 1 | 0 | 0 | 0 |
| declaration 100010 | 0 | 1 | 0 | 0 | 0 |
| declaration 100015 | 1 | 0 | 0 | 0 | 1 |
| declaration 100021 | 1 | 0 | 0 | 0 | 1 |
| declaration 100027 | 0 | 1 | 0 | 0 | 0 |
| declaration 100032 | 0 | 1 | 0 | 0 | 0 |
| declaration 100037 | 0 | 1 | 0 | 0 | 0 |

To carry out clustering, we need quantitative features. For this purpose, we built a pivot table, as shown in table 2. Indexes are identifiers with which we analyzed the data. Columns are transitions between events that we received by shifting the «concept:name»

column. The numbers indicate how many times a given transition between events occurs for a specific case_id. Depending on the number of unique events in the process, there can be many such transitions. Thus, it is necessary to use principal component analysis to reduce the feature space.

As a clustering algorithm, DBSCAN with Euclidean metric was used. As a result of clustering, we get the most popular sequence for two types of trips, domestic and international.

For domestic trips it's looks like this:

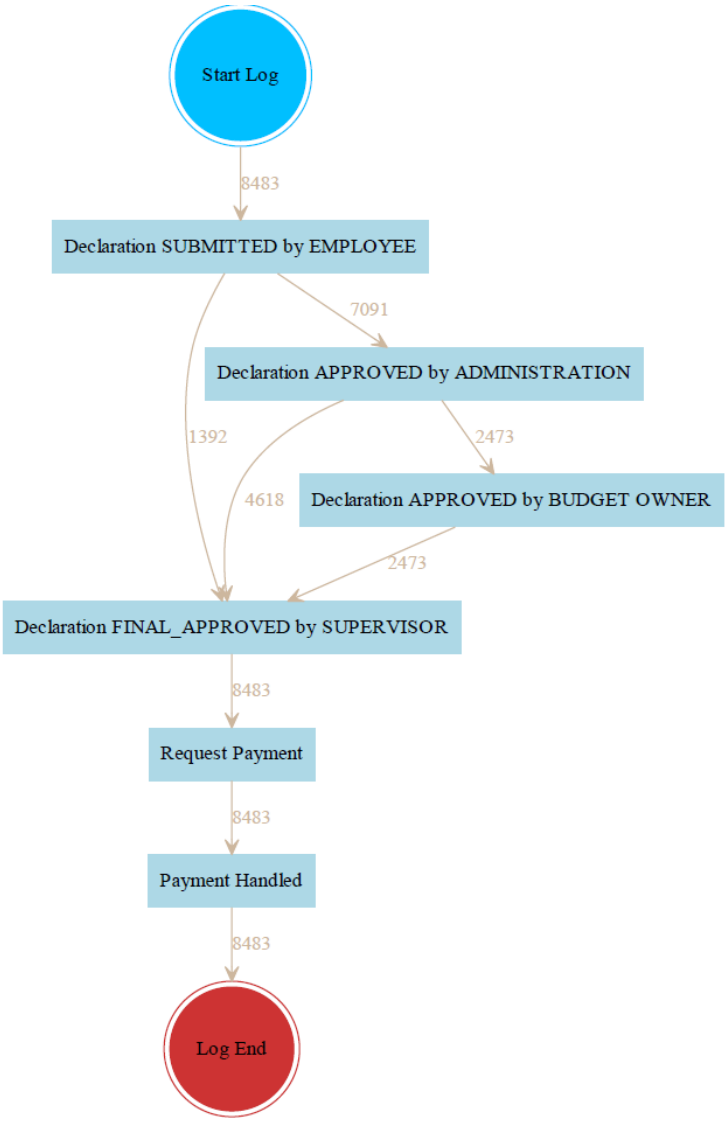


Figure 1 – Frequency graph for domestic trips (most popular sequence of process)

As seen from figure 1, 8483 cases (80 % of total case_id, 10500) have similar sequences of events, divided into three main groups. Approval by administration takes place in the largest group. In the group with 2473 cases budget owner approval is required in addition to administration approval. In the group with 1392 cases the approval is carried out directly by the supervisor, bypassing administration and budget owner. For analysis, we selected a group with 4618 cases. Figure 2 shows the execution time of each event in days:hours:minutes:seconds format.

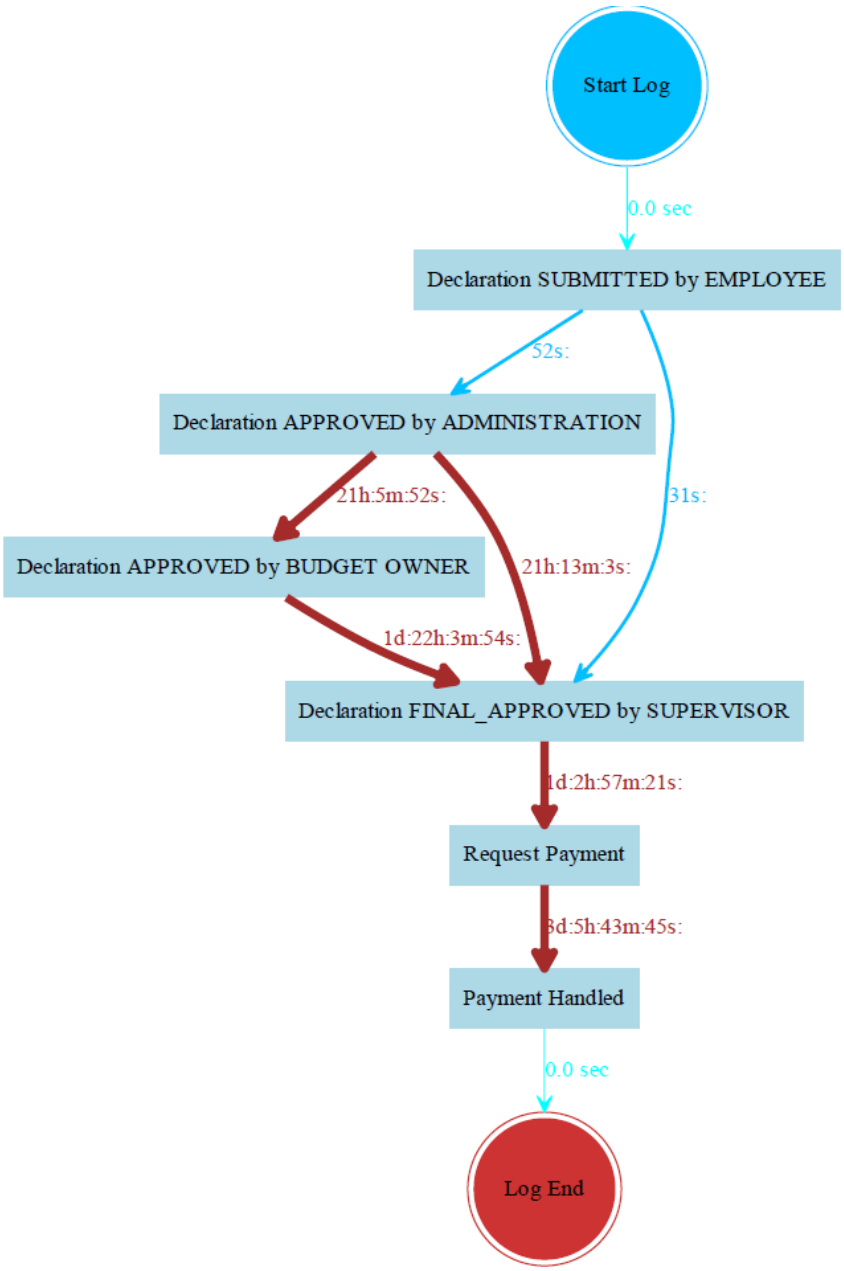


Figure 2 – Performance graph for domestic trips (most popular sequence of process)

Table 3 provides information about the execution time of each event for most popular sequence. Total duration of process is just over 5 days. As you can see, the longest stage is «Request Payment».

Table 3 – Event duration for most popular group of domestic trips

| Process From | Process To | Time spent |
|---|---|---|
| Declaration SUBMITTED by EMPLOYEE | Declaration APPROVED by ADMINISTRATION | 52 sec |
| Declaration APPROVED by ADMINISTRATION | Declaration FINAL_APPROVED by SUPERVISOR | 21 hours 13 minutes 3 sec |
| Declaration FINAL_APPROVED by SUPERVISOR | Request Payment | 1 day 2 hours 57 minutes 21 sec |
| Request Payment | Payment Handled | 3 days 5 hours 43 minutes 45 sec |
| | **Total:** | **5 days 5 hours 55 minutes 1 sec** |

Next, we analyzed the data for international trips. On figure 3 you can see most popular event sequences in international trips for cluster of 3107 case_id (48 % of total case_id, 6449). As in the case of domestic trips, branching occurs at the approval stage. Moreover, a declaration rejection stage with re-submission by employee is added.

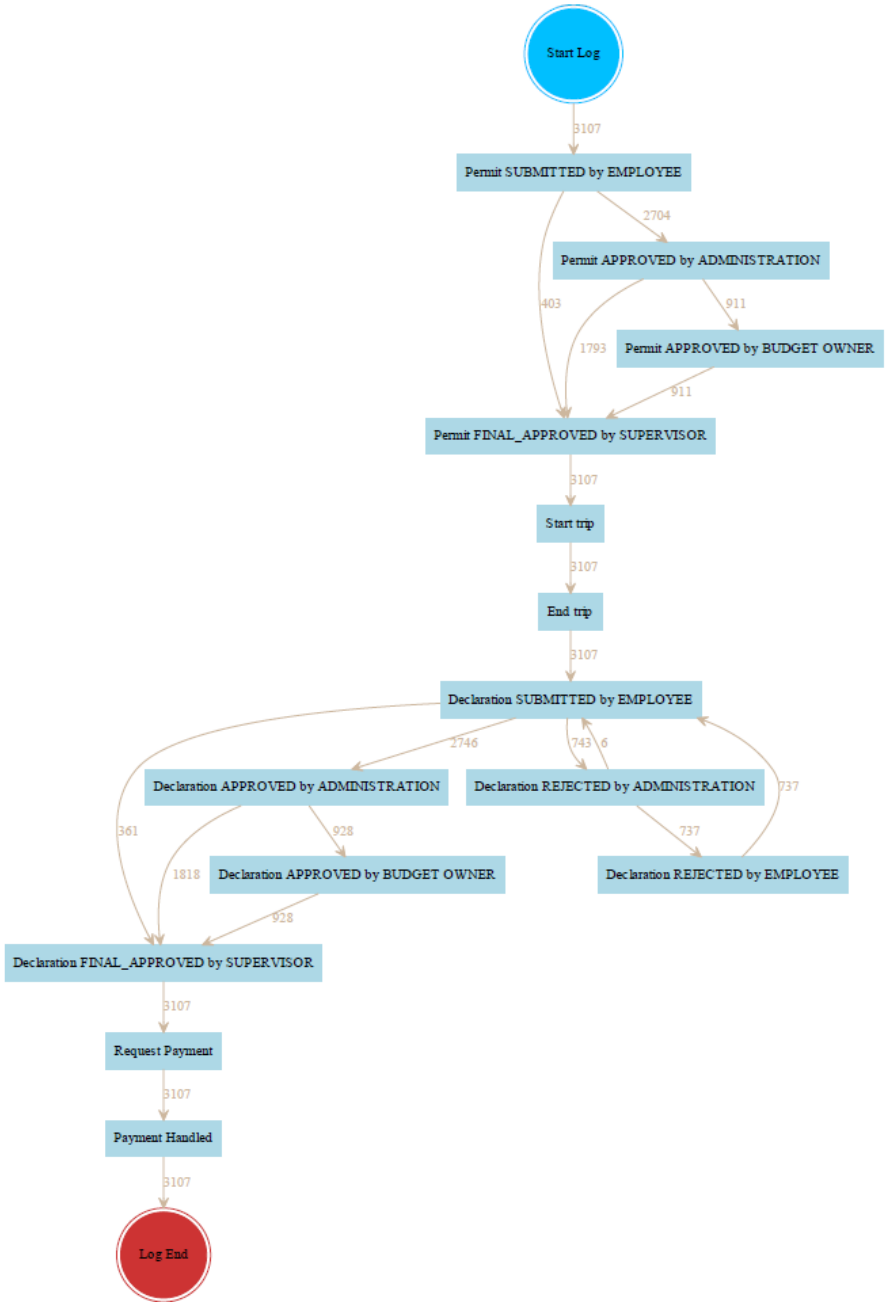


Figure 3 – Frequency graph for international trips (most popular sequence of process)

Table 4 shows duration of stages for most popular event sequences in international trips.

Table 4 – Event duration for most popular sequence of international trips

| Process From | Process To | Time spent |
|---|---|---|
| Permit SUBMITTED by EMPLOYEE | Permit APPROVED by ADMINISTRATION | 5 sec |
| Permit APPROVED by ADMINISTRATION | Permit FINAL_APPROVED by SUPERVISOR | 22 hours 3 minutes 28 sec |
| Permit FINAL_APPROVED by SUPERVISOR | Start trip | 26 days 15 hours 2 minutes 24 sec |
| Start trip | End trip | 4 days |
| End trip | Declaration SUBMITTED by EMPLOYEE | 5 days 11 hours 6 minutes 23 sec |
| Declaration SUBMITTED by EMPLOYEE | Declaration APPROVED by ADMINISTRATION | 2 minutes 58 sec |
| Declaration APPROVED by ADMINISTRATION | Declaration FINAL_APPROVED by SUPERVISOR | 1 day 4 hours 18 minutes 44 sec |
| Declaration FINAL_APPROVED by SUPERVISOR | Request Payment | 1 day 3 hours 52 minutes 43 sec |
| Request Payment | Payment Handled | 3 days 5 hours 10 minutes 22 sec |
| | **Total:** | **42 days 19 hours 52 minutes 23 sec** |
| | | 5 days 13 hours 24 minutes 47 sec |

The total duration of the process is almost 43 days. The stage of approval of the declaration takes more than 5 days, for which the «Request Payment» is the longest (more than 3 days).

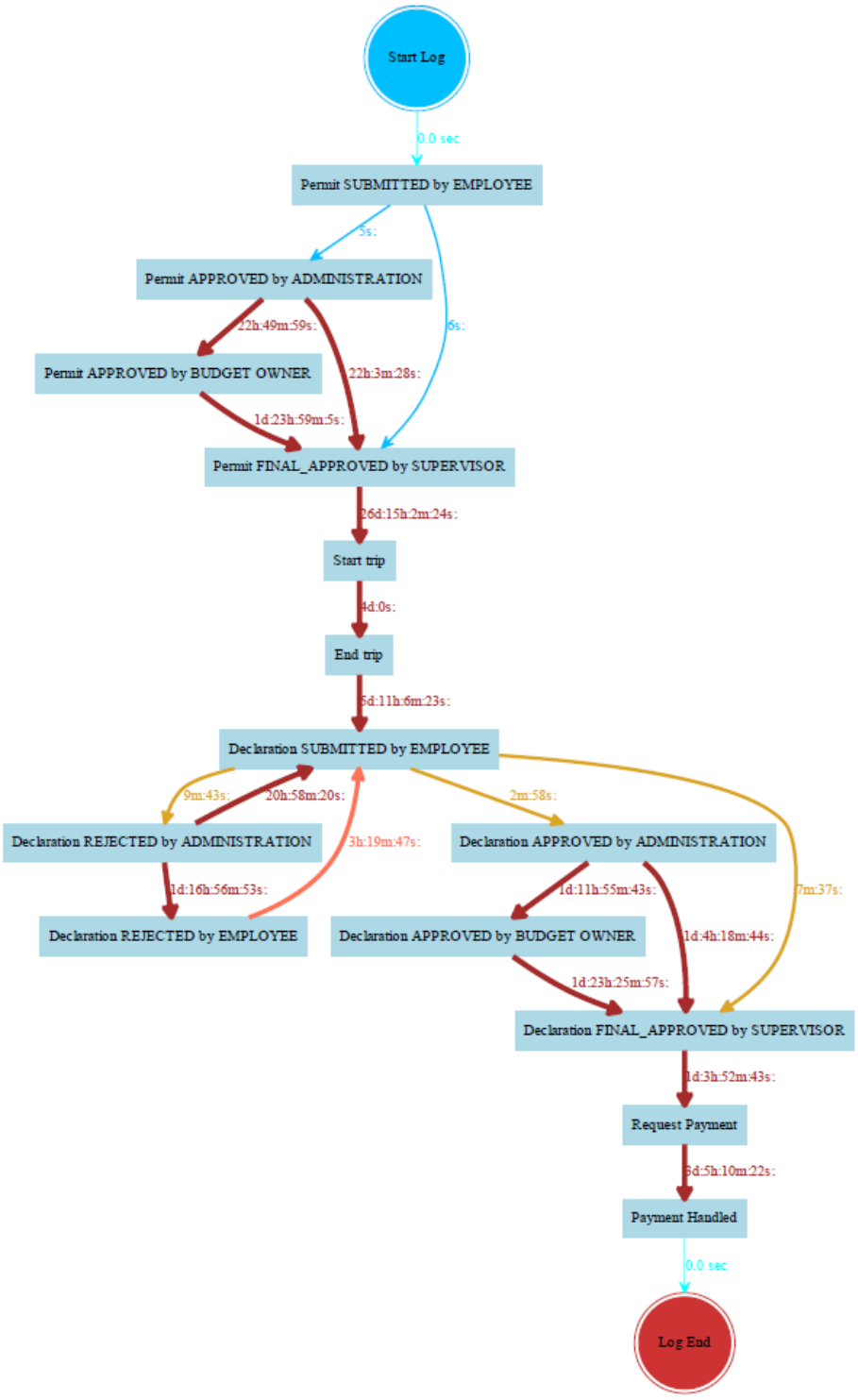


Figure 4 – Performance graph for international trips (most popular sequence of process)

If you compare domestic and international trips, you can see, that declaration approval and payment stages is almost identical in both types of trips. (Table 3 and Table 4). On figure 4 you can see, that on Permit Approved and Declaration Approved stages there is an event with the participation of the Budget Owner, which increases the approval time by more than 2 times.

Table 5 shows information about number of rejected declarations, and also number of declarations that was never approved.

Table 5 – Number of rejected and non-approved declarations

| | Get rejected | Never approved |
|---|---|---|
| Domestic trips | **1301 (12.4%)** | **456 (4.3%)** |
| International trips | **1576 (24.4%)** | **262 (4.1%)** |

«Get rejected» is used for events which have been rejected at any stage in the process. «Never approved» are the events that do not have «Payment Handled» in the sequence of process events. Percentages reflect proportion of the total number of case_ids.

Next, we looked at the end options of the log. For international trips, information is presented in table 6:

Table 6 – End of event sequences for all data in international trips

| | events | counts |
|---|---|---|
| **0** | Payment Handled-->Log End | 5646 |
| **1** | End trip-->Log End | 593 |
| **2** | Declaration REJECTED by EMPLOYEE-->Log End | 130 |
| **3** | Declaration SAVED by EMPLOYEE-->Log End | 54 |
| **4** | Declaration REJECTED by MISSING-->Log End | 11 |
| **5** | Permit REJECTED by MISSING-->Log End | 8 |
| **6** | Request Payment-->Log End | 3 |
| **7** | Send Reminder-->Log End | 2 |
| **8** | Declaration REJECTED by SUPERVISOR-->Log End | 1 |
| **9** | Declaration FINAL_APPROVED by SUPERVISOR-->Log... | 1 |

Information for domestic trips is presented in table 7:

Table 7 – End of event sequences for all data in domestic trips

| | events | counts |
|---|---|---|
| **0** | Payment Handled-->Log End | 10043 |
| **1** | Declaration REJECTED by EMPLOYEE-->Log End | 284 |
| **2** | Declaration SAVED by EMPLOYEE-->Log End | 134 |
| **3** | Declaration REJECTED by MISSING-->Log End | 30 |
| **4** | Declaration REJECTED by ADMINISTRATION-->Log End | 5 |
| **5** | Declaration REJECTED by SUPERVISOR-->Log End | 4 |

For domestic trips the situation is pretty common, however in international trips there are some cases, in which the last event of process is «Request Payment» without payment. On Figure 5 event sequences is presented, which is identical for three cases (145798, 146418, 16408).

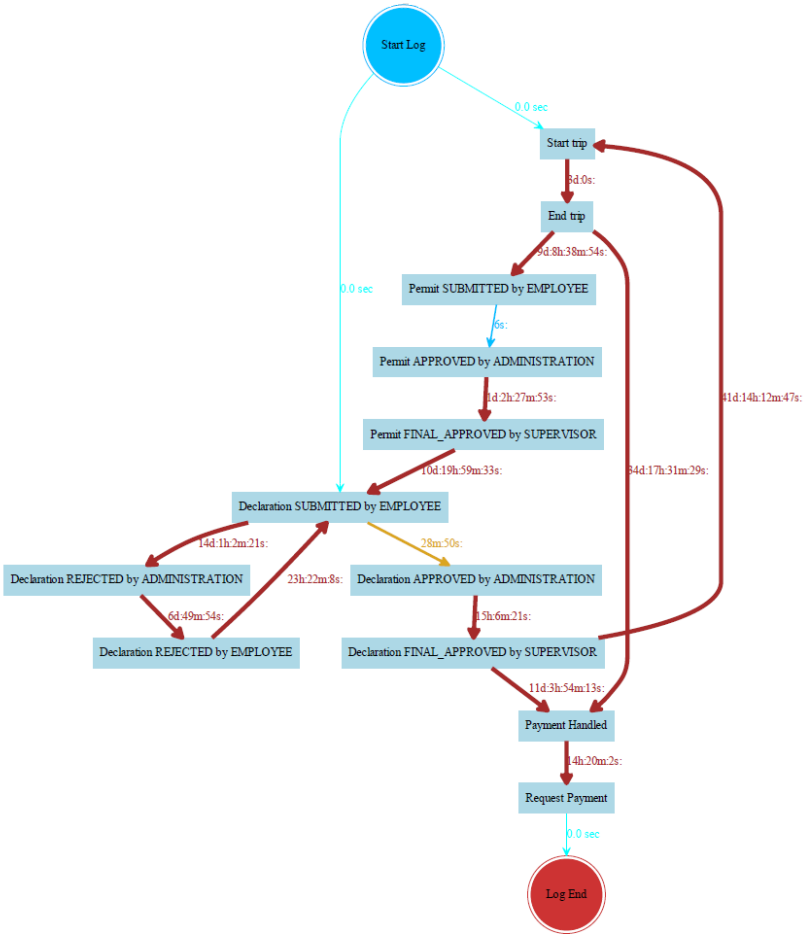


Figure 5 – Event sequences for 3 case_id of international trips

For some reason a second Request Payment was sent, despite the fact that «Payment Handled» already was committed. In table 8 the transitions between events, that have event «Payment Handled» are presented.

Table 8 – Payment event sequence for international trips

| | Payment event sequence | Counts |
|---|---|---|
| 0 | Request Payment-->Payment Handled | 6149 |
| 1 | Payment Handled-->Log End | 5646 |
| 2 | Payment Handled-->Start trip | 415 |
| 3 | Payment Handled-->End trip | 121 |
| 4 | End trip-->Payment Handled | 15 |
| 5 | Permit REJECTED by MISSING-->Payment Handled | 10 |
| 6 | Start trip-->Payment Handled | 8 |
| 7 | Declaration FINAL_APPROVED by SUPERVISOR-->Pay... | 5 |
| 8 | Payment Handled-->Request Payment | 3 |
| 9 | Payment Handled-->Send Reminder | 2 |

Here we see another interesting case for international trips. Payment was made after the trip was rejected, however the trip has already been made. This case is identical for the following case_ids (79145, 81231, 83795, 86449, 79948, 81233, 83798, 81229, 81235, 83801, 86449). An example is shown on figure 6:

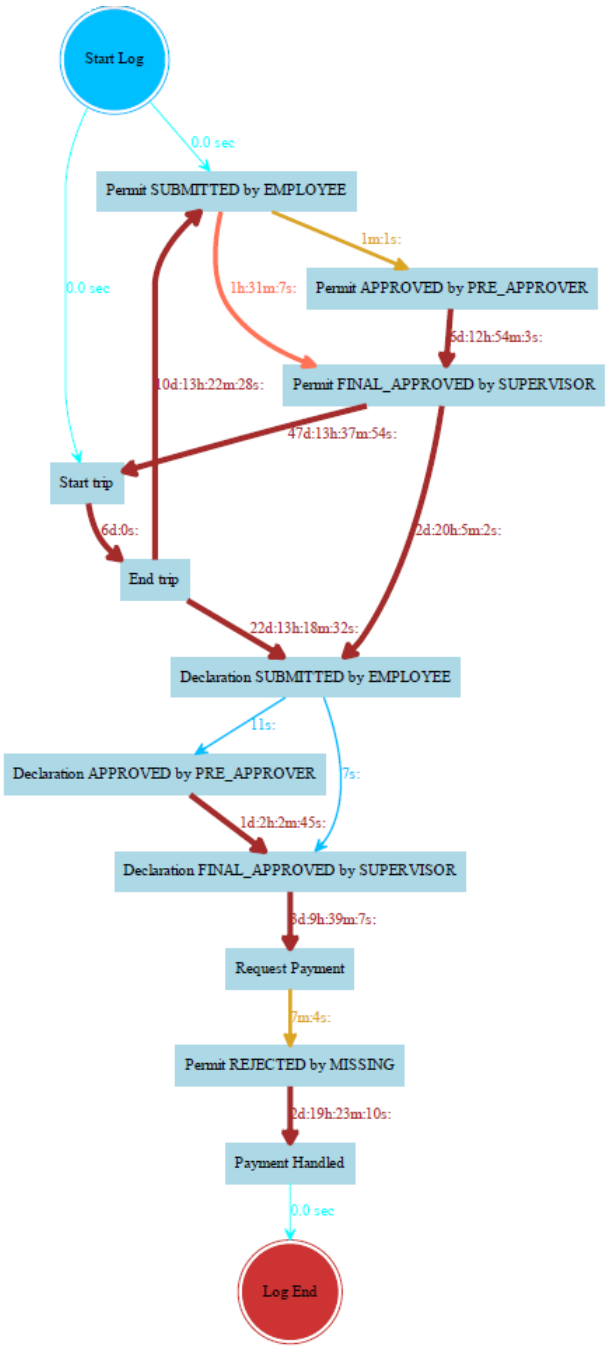


Figure 6 – Event sequences for 10 cases of international trip

## Conclusion

We present an overview of the reference sequence of events for two types of process (Domestic trips and International trips). The median execution time of events has been determined as well as bottlenecks in the process. We considered the cases of non-standard behavior in the process which can be both a logging errors and actions, that violate regulations for implementation of the process.