Math-heuristic to solve the airline recovery problem considering aircraft and passenger networks

Mat-heurística para recuperação de malhas aéreas considerando aeronaves e passageiros

José Carlos Fontoura Guimarães¹, Nicolau Dionísio Fares Gualda²

¹University of São Paulo, São Paulo – Brazil, jcf.guimaraes@terra.com.br
²University of São Paulo, São Paulo – Brazil, ngualda@usp.br

ABSTRACT

The airline recovery problem involves aircraft, crew, and passenger networks impacted by disruptions. Models to solve the problem consider one or more of these networks, on an integrated way or not. It belongs to the NP-hard class, even considering only one network. This work presents a math-heuristic to solve the problem integrating the aircraft and the passenger networks. A model to restore the aircraft network with minimum cost was also developed. These two models compose a framework which permits the airline to obtain the cost impact of including the passenger network in the recovery problem. Both models were tested with real-world ROADEF instances using an Intel i7 microcomputer (16Gb of RAM) and a high-performance cluster node (HPC) with 512 GB of RAM. The microcomputer solved instances with up to 85 aircraft and 276 impacted flights in less than 30 minutes (imposed limit). The faster high-performance server reached solutions with minimum gap of 0 to 0.7% for the instances with higher number of flights. Total costs considering aircraft and passenger networks were very close to the aircraft network recovery results, showing a cost compensation which highlights the importance of solving the recovery problem integrating aircraft and passenger networks.

1. INTRODUCTION

1.1 Airline Operational Plan

The airline operational plan starts long before a flight takes off. Its objective is to provide an effective and efficient service considering customers’ demands and costs perspectives.
It encompasses (as shown in Figure 1):

- The preparation of a Flight Schedule to meet the market demand
- An aircraft schedule (also called aircraft rotation) assigning an aircraft (tail assignment) to each flight in order to attend the expected demand (passenger schedule), respecting flight operational constraints.
- A corresponding schedule of qualified cabin crew and pilots compliant to the regulatory and labor standards

Each of these three schedules can be represented by a resource network. They must be synchronized so that the flights operate as planned and as efficiently as possible.

![Figure 1. Process to build the Airline Operational Plan – Source: Adapted from (Castro and Oliveira, 2011)](image)

### 1.2. Airline disruption

An airline disruption occurs when at least one of the required resources to operate a flight is not available and ready at the flight’s planned departure time. It may affect the departure of one or more flights. Even minor disruptions can create ripple effects across the network and quickly affect the rest of the flight network.

Disruptions can be caused for several reasons. The main ones are classified as follows (Artigues et al., 2012):

- Flight delays: may be caused by delayed boarding of passengers, ground time longer than planned, ground staff strike, or delays in passenger or crew connections
- Flight cancellations
- Unavailability of an aircraft for a certain period of time, preventing it to be assigned to flight operations during that period
- Reduction of airport capacity over a certain period of time, due, for example, to weather conditions or to strike of airport personnel, with consequent reduction on the number of landings and takeoffs in the period.

The estimated annual cost of disruptions is in the order of millions of U.S. dollars, encompassing (Castro, 2014):

- Aircraft operating costs
- Crew overtime and use of reserve crew costs
- Loss of passenger revenue and direct costs related to affected passengers, e.g. hotel, transportation and feed costs.
1.3. The airline recovery problem

The solution of the airline recovery problem consists in returning the airline back to the planned schedule after an operational disruption. It is carried out by adjusting flight plans, aircraft assignments, crew assignments and passenger itineraries within a period of time called recovery period. It is addressed to minimize operational costs and service penalties.

Usually, the recovery period is of one day (Zhang et al., 2015), but it really depends on the disruption event (e.g. a snowstorm may impact an airport operation for several days).

The Operational Control Center (OCC) is responsible to restore the airline flight schedule back to the original planning condition after the occurrence of a disruption, within the established recovery span. This task is called "Airline Disruption Management" or "Airline Recovery".

The basic steps of the airline recovery process are shown in Figure 2. In the Operation Monitoring step, the OCC checks whether flights, cargo and baggage loading on aircraft and check-in of passengers and crews are operating as expected. If not, an analysis is performed immediately to decide on corrective actions. The required processing time the OCC usually has to solve the recovery problem is of as much as 20 or 30 minutes (Morais et al., 2018).

Figure 2. Airline recovery process - Source: Adapted from (Kohl et al., 2007)

Figure 3 shows the sequential procedure usually used to solve the airline recovery problem:

Figure 3. Sequential procedure for solving the airline disruption problem – Source: Adapted from Castro and Oliveira (2011).

This means that, after a disruption event, the OCC first builds the flight cancellation and delay plans, then adjusts the aircraft network, synchronizes the crew network with the aircraft network and, finally, adjust the passenger route network.

The priority of this approach is the return to operation of the most expensive assets, aircraft and crew. It does not look to the best result for customers. As a consequence, it may imply high costs related to passenger delays. Therefore, it does not provide optimal solutions to the problem (Artigues et al., 2012).
2. LITERATURE REVIEW

2.1. Airline recovery categories and approaches

Airline recovery overall approach can be classified according to the following categories (Castro et al., 2014):

- Non-integrated recovery - only one of the three networks - or dimensions - is considered.
- Non-simultaneous integrated recovery - airline recovery considers the three dimensions - aircraft, crews and passengers - but separately, not simultaneously.
- Simultaneous integrated recovery - the three dimensions - aircraft, crews and passengers are treated simultaneously, without imposing degrees of importance to them.
- Partially integrated recovery - airline recovery considers two dimensions, simultaneously or not.

Among the 60 articles on airline recovery analyzed by Castro et al (2014), 75% belonged to the non-integrated recovery class and only 25% were related to integrated or partially integrated recovery (of which, non-simultaneous integrated recovery: 26%, simultaneous integrated recovery: 13%, and partially integrated recovery: 61%)

Operations research tools dominate work on recovery of airline networks. 73% out of 60 papers analyzed by Castro et al (2014), published between 1998 and 2012, used OR techniques. 70% considered minimizing flight delays, flight cancellations and operating costs, and only 7% considered impacts to passengers.

Regarding the solution methods, 31% used mathematical programming, 29% used network flows. 22% used heuristics and 18% used a column generation method.

Below is a summary of important work related to Non-simultaneous integrated recovery, Simultaneous integrated recovery and Partially integrated recovery.

2.1.1. Non-simultaneous integrated recovery

Kohl et al. (2007) present a sequential modeling with a distinct approach for each of the problem dimensions, aiming to minimize operating costs and passenger costs. For the aircraft dimension, a local search heuristic was used; for the crew dimension, a whole programming approach was used with resolution per column generation; and for the passenger dimension, multi-commodity network programming was used. The model considers the cost of delay at the final destination for the relocated passenger, direct costs for rescheduled passengers (hotel, food, etc.), impact on customer perception and costs related to class upgrade and downgrade (in this case strongly impacting the customers experience and their willingness to fly again with the company).

Maher (2015, 2016) present a sequential approach using programming and resolution with a column generation method; passenger recovery occurs after the resolution for aircraft and crew dimensions. The passenger cost depends on the number of passengers relocated due to flight cancellation and also on the delay costs for passengers relocated to other flights.

Jafari and Zegordi (2010) use mixed mathematical programming modeling, applying benders’ decomposition method. Passenger costs depend on the number of passengers impacted by cancellations and on delays to passengers related to their final destination.

Jozefowiez et al. (2012) present a large search in the vicinity heuristic, considering, for passengers, the costs of delay, flight cancellation and cabin downgrade.
2.1.2. Simultaneous integrated recovery


2.1.3. Partially integrated recovery

Bratu and Barnhart (2006) present a multi-commodity network flow model for solving the partially integrated recovery problem considering aircraft and passenger dimensions. It incorporates operational cost and cost reduction modeling for passenger itineraries delays and cancellations.

Abdelghany et al. (2008) propose an integrated model for recovery of aircraft and crew dimensions by using a nonlinear mixed mathematical programming approach to minimize operational costs.

Gao (2007) proposes an approach to partially integrated recovery considering aircraft and crew dimensions. It uses a multi-commodity network flow modeling with Benders decomposition to minimize operational costs cancellations and delays.

Zhang (2008) incorporates ground transportation in a nonlinear programming model for partially integrated recovery, considering aircraft and passenger dimensions to minimize operating costs and passenger costs.

Eggenberg and Salani (2010) present a solution for partially integrated recovery considering aircraft and crew dimensions. It applies the column generation method for minimizing operating costs.

Zegordi and Jafari (2010) propose a partially integrated recovery solution considering aircraft and passenger dimensions. A heuristic based on ant colony, with the objective of minimizing operational costs, flight cancellations, delays and costs for passengers is used.

Bisaillon et al. (2011) present a solution for the partially integrated recovery considering aircraft and passenger dimensions. A large search in the vicinity heuristic to minimize operational costs and impacts on passengers is used.

Jafari and Zegordi (2011) present a mixed integer mathematical programming model for partially integrated recovery considering aircraft and crew dimensions, for minimizing operating costs.

Gomes (2014) presents an integer mathematical programming composition with heuristics, to solve the partially integrated recovery with aircraft and crew dimensions.

Sinclair et al. (2014) present a large search in the vicinity heuristic for the partially integrated recovery with aircraft and passenger dimensions, based on Bisaillon et al. (2011).

Hu et al. (2016) present a GRASP heuristic for the partially integrated recovery considering aircraft and passenger dimensions for minimizing operating costs and number of passengers affected.

Arikan et al. (2016) present a model based on mixed integer mathematical programming considering aircraft and passenger dimensions. The inclusion of the aircraft cruise speed control is considered as an alternative course of action.

Zhang et al. (2016) present a recovery solution for the aircraft and passenger dimensions using a three-step math-heuristic. It is addressed to treat cancellation and delay of flights for minimizing costs of passengers’ accommodation.
Marla et al. (2017) present an approach to the Partially integrated recovery considering aircraft and passenger dimensions. It includes the aircraft cruise speed control as an alternative course of action using a network flows model.

2.1.4. Non-integrated recovery

Morais et al. (2018) present a two-part math-heuristic to solve the aircraft recovery: a mixed-integer programming network flow model to obtain a new schedule with minimum flight cancellations and delays; and an integer linear programming model to minimize aircraft swaps. It includes disruptions due to aircraft requiring unexpected maintenance.

2.2. Time-space networks

The airline planning can be described as a time-space network (Morais et al., 2018), as shown in Figure 4. For each airport there is a set of nodes and each node represents a pair: [event at the airport - departure or arrival - , time of the event]. Nodes associated with the same airport are positioned in the same column, ordered from top to bottom, depending on the time of the event. The offer node for each airport represents the number of aircraft available at the airport at the beginning of the planning cycle. At the end of the planning cycle, the number of aircraft available at the airport characterizes a termination node. Nodes are connected by directed arcs, where the unit flow represents an aircraft. All flows start on offer nodes and end on termination nodes. At all nodes flow conservation is maintained.

Time-space networks are directed and acyclic graphs (Abdelghany et al., 2004). Each aircraft has a preparation time between one flight and another (min_turn_time). This time can be modeled as part of the previous flight time (allows simplification of modeling), i.e. a flight node ends at the earliest possible time of departure of the next flight, rather than at the actual landing time. Therefore, a departure node corresponds to the departure of a flight at a location l(j) at time t(j) and an arrival node corresponds to the arrival of a flight in l(j) at time t(j) - min_turn_time, because t(j) corresponds to the actual arrival time + min_turn_time = (readiness time for the next flight).

![Figure 4. Time-space network - Source: (Morais et al., 2018)]
2.3. Modeling flight alternatives and disruptions

Flight alternatives can be modeled as inserted arcs (Zhang et al., 2016). For each of the planned flights in the original Flight Schedule, a set of flights related to a delay of a fixed period of time is inserted. A suffix “-1”, “-2”, etc. is added to the original flight identification, according to the considered delay (e.g. fifteen minutes delay alternatives: original Flight234 departing 15:00, alternative Flight234-1 departing 15:15, alternative Flight234-2 departing 15:30).

Disruptions such as flight cancellations are modeled by cancelling the original flight and all respective inserted arcs. Other disruptions are modeled by introducing resources unavailability constraints. For example, aircraft unavailability are modeled by removing the aircraft from the solution. Airport operational restrictions are modeled as a departure and a landing capacity reduction (Zhang et al., 2016 and Morais et al., 2018).

Finding shorter paths in flight networks corresponds to solving major airline optimization problems (Maher, 2016): maximizing profit on aircraft routes, minimizing crew allocation cost, and minimizing travel cost for passengers.

The resolution of these problems should consider only viable routes, i.e. aircraft routes should consider the aircraft readiness schedule for the next flight – not the arrival time – and, in the case of crews, the subset of possible paths given regulatory restrictions.

As already mentioned, time-space networks are directed and acyclic; therefore, there are efficient algorithms to find shorter paths in these conditions - algorithms of complexity proportional to the number of arcs in the network and that require topological ordering of nodes (Abdelghany et al., 2004).

New optimized flight schedule with a minimum delay cost can be obtained by means of a multi-commodity network flow model. The model looks for the lowest cost flows in the disrupted time-space network with multiple flows and capacity limitation (Zhang et al., 2016, Morais et al., 2018).

2.4. Approaches considering networks interaction

Industry data indicate that recovery approaches considering the three networks on an integrated and simultaneous way permit to recuperate about 95% of the aircraft, crew and passenger networks in the same day of the disruption (Castro, 2014).

Meanwhile, sequential recovery approaches allow to recover around 80% to 90% of the aircraft and crew networks, but only about 60% of passengers are re-scheduled in the same period of time. So, the solution obtained includes significant costs not only for the airline (image losses and direct expenses related to the affected passengers), but also for the passengers - dissatisfaction, personal and financial impacts (Castro, 2014).

Recovery approaches that allow the interaction of at least two of the networks of the problem generate better solutions than the sequential approach. (Petersen et al., 2012).

Airline disruption problems are of the NP-hard class; that is, there are no polynomial algorithms for their resolution, regardless of how it is considered – integrated, partially integrated or not integrated (Yu and Qi, 2004). Hence, the need to develop other types of models, such as heuristics, meta-heuristics and math-heuristics, to solve the recovery problem.

Heuristic-related modeling approaches do not necessarily provide the optimal solution to the problem. However, they can derive near-optimal results for NP-hard problems in very short

The decision to develop a math-heuristic for solving the recovery problem in this research took in consideration the opportunity to explore the potential of this relatively novel approach to treat NP-hard problems. According to Caserta et al (2010), math-heuristic models combine heuristic techniques with operations research techniques in an iterative or interactive way, offering more efficient solutions.

3. DEVELOPED MODELS

3.1. Models purpose

This research aims at providing the airline with tools to decide on the recovery with priority to the aircraft schedule or to both the aircraft and the passenger schedules. To do so, two models were developed: a model for the aircraft network recovery with minimum flight schedule cost (Flight schedule cost model) and a Math-heuristic for solving the airline recovery problem considering both the aircraft and the passenger networks.

![Figure 5. Description of the models developed in this research](image)

The Flight schedule cost model aims to restore the aircraft schedule with minimum cost for the airline using a sequential approach. The first step recovers the flight schedule, the second step recovers the aircraft rotation and the last step identifies the disrupted itineraries due either to flight cancellations in Step 1, to flights cancelled for lack of aircraft assignment in Step 2, or to lack of feasible connection time in the resultant flight schedule after Step 2.
Table 1 - Models variables definitions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Output from</th>
<th>Input to</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Set of all planned flights in the original flight schedule</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>FC₀</td>
<td>Set of planned flights and the inserted arcs for each flight f ∈ F. The original flight is an inserted flight with delay cost equal to 0.</td>
<td>Step 1 modeling based on disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>F-cancel</td>
<td>Set of original flights that, as part of the disruption event, are cancelled as an input of the problem; thus, cannot be part of the flight schedule adjustment.</td>
<td>Step 1 modeling based on disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>P</td>
<td>Set of airports indexed by p</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>M</td>
<td>Set of types of movement in the airport e [arrival, departure] indexed by m</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>H</td>
<td>Set of all time slots in the time-space network indexed by h representing each time slot inside the recovery window for each hour interval from the interval [0min-59min]</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>FC_INₙ</td>
<td>Set of all inbound flights to node n - original and inserted arcs</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>FC_OUTₙ</td>
<td>Set of all outbound flights from node n - original and inserted arcs</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>N_INₙ</td>
<td>Set of all aircraft entrance nodes</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>N_OUTₙ</td>
<td>Set of all aircraft exit nodes</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>A</td>
<td>Set of aircraft indexed by a</td>
<td>Disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>N_INₐₙ</td>
<td>Number of aircraft entering node n, in the time-space network, at the start of the recovery window</td>
<td>Step 1 modeling based on disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>N_OUTₐₙ</td>
<td>Number of aircraft exiting node n, in the time-space network, at the end of the recovery window</td>
<td>Step 1 modeling based on disruption problem input</td>
<td>Step 1</td>
</tr>
<tr>
<td>Capₚₙₙₙₙ</td>
<td>Capacity of airport p ∈ P for departure and arrival activities in each time slot h ∈ H</td>
<td>Disruption problem input</td>
<td>Step 1 and Step 3 (math-heuristic)</td>
</tr>
<tr>
<td>cost_delayₙₚ</td>
<td>Cost due to delaying a flight f – choose an inserted arc</td>
<td>Step 1 modeling</td>
<td>Step 1</td>
</tr>
<tr>
<td>cost_cancelₚ</td>
<td>Cost due to cancelling flight f</td>
<td>Step 1 modeling</td>
<td>Step 1</td>
</tr>
<tr>
<td>FCₓᵣ</td>
<td>Set of planned flights and the inserted arcs for each flight f ∈ Rₙ. The original flight is an inserted arc with delay cost equal to 0.</td>
<td>Step 1</td>
<td>Step 1</td>
</tr>
<tr>
<td>C_changeₙₕ</td>
<td>Cost to change aircraft a from flight f</td>
<td>Step 2 modeling based on disruption problem input planned aircraft rotation</td>
<td>Step 2</td>
</tr>
<tr>
<td>C_not_assign</td>
<td>Cost of having an aircraft not assigned to flights in Xₙ</td>
<td>Step 2 modeling input</td>
<td>Step 2</td>
</tr>
<tr>
<td>STDₜₕ</td>
<td>Departure time of flight f with aircraft a</td>
<td>Step 2</td>
<td>Step 2</td>
</tr>
<tr>
<td>ETAₜᵣ</td>
<td>Arrival time of flight f with aircraft a</td>
<td>Step 2</td>
<td>Step 2</td>
</tr>
<tr>
<td>ETAₜₙ</td>
<td>Arrival time of inserted arc fc</td>
<td>Step 2</td>
<td>Step 3</td>
</tr>
<tr>
<td>STDₜₙ</td>
<td>Departure time of inserted arc fc</td>
<td>Step 2</td>
<td>Step 3</td>
</tr>
<tr>
<td>Rₜ</td>
<td>Flight solution from Step 2. For every rₜ ∈ Rₜ, rₜ = 1 if fc is an active flight or 0 otherwise (flight cancelled in Step 1 or cancelled in Step 2 due to not having an aircraft assigned to it). fc is an inserted flight or the original flight (inserted arc with delay = 0)</td>
<td>Step 2</td>
<td>Step 3</td>
</tr>
<tr>
<td>T</td>
<td>Set of itineraries indexed by τ. Each itinerary τ has a sequence of flights that must be operated. An itinerary is affected if there is a cancelled flight or if there is not enough time for passengers to connect between two consecutive flights in the itinerary</td>
<td>Disruption problem input</td>
<td>Step 3</td>
</tr>
<tr>
<td>Cₜ</td>
<td>Cost to disrupt an itinerary τ ∈ T</td>
<td>Disruption problem input</td>
<td>Step 3</td>
</tr>
<tr>
<td>Xₜ</td>
<td>Flight schedule input to Step 3 of math-heuristic, obtained after Step 1 and Step 2. For every xₜ ∈ Xₜ, xₜ = 1 if flight f is confirmed or 0, if f is cancelled in Step 1 - f is an inserted flight or the original flight (inserted arc with delay = 0)</td>
<td>Step 2</td>
<td>Step 3</td>
</tr>
<tr>
<td>connτ</td>
<td>Minimum passenger connection time between two consecutive flights in a passenger itinerary τ ∈ T</td>
<td>Disruption problem input</td>
<td>Step 3</td>
</tr>
<tr>
<td>yₜ</td>
<td>Binary decision variable, yₜ = 1 if original flight f is cancelled, 0 otherwise</td>
<td>Step 1</td>
<td>-</td>
</tr>
<tr>
<td>xₜ₀</td>
<td>Binary decision variable, xₜ₀ = 1 if the inserted arc fc is chosen, 0 otherwise</td>
<td>Step 1</td>
<td>-</td>
</tr>
<tr>
<td>rₜₕ</td>
<td>Binary decision variable, rₜₕ = 1 if aircraft a is assigned to flight f ∈ Xₜ, 0 otherwise</td>
<td>Step 2</td>
<td>-</td>
</tr>
<tr>
<td>wᵣₕₙₕ</td>
<td>Binary decision variable, wᵣₕₙₕ = 1 if aircraft a is not assigned to flights in Xₜ, 0 otherwise</td>
<td>Step 2</td>
<td>-</td>
</tr>
<tr>
<td>zₜ</td>
<td>Binary decision variable, zₜ = 1 if itinerary τ ∈ T is disrupted (not possible to fly it from start to finish due to a cancelled flight in the itinerary or lack of enough time to connect between any two consecutive flights in the itinerary), 0 otherwise</td>
<td>Step 3 (Flight schedule cost and math-heuristic)</td>
<td>-</td>
</tr>
<tr>
<td>x'ₜₚ</td>
<td>Binary decision variable, x'ₜₚ = 1 if inserted arc x'ₜₚ ∈ FCₚ is selected, 0 otherwise</td>
<td>Step 3 (math-heuristic)</td>
<td>-</td>
</tr>
</tbody>
</table>
To restore the flight schedule with minimum cost and to identify affected passengers, the directions proposed by Zhang et al. (2016) to obtain the new flight schedule and to identify disrupted itineraries were adopted. Step 2 of the Flight schedule cost model aims to recover flight rotation by minimizing the cost to change an aircraft’s original rotation and the cost of no assigned aircraft. The modeling approach is inspired in Morais et al. (2018) and adapted to a single time-space network.

The math-heuristic objective is to recover passenger itineraries that were disrupted due to lack of enough connection time and to identify the overall cost impact considering the delays inserted in the final flight schedule. This objective is accomplished in Step 3 of the math-heuristic by simultaneously considering, in the same mathematical model, two dimensions – aircraft network and passengers network. An optimal solution is searched for both the flight schedule delay cost and the disrupted passenger itineraries cost, as shown in Figure 3. It was implemented based on Zhang et al. (2016) modeling approach. Two simplifications are inherent to the developed models: all aircraft must have the same configuration and no aircraft maintenance is considered. Therefore, an aircraft failure is taken as a disturbance that removes the aircraft from the solution. These simplifications allow modeling with a single time-space network (Zhang et al., 2016).

Figure 5 shows the basic steps of the Flight schedule cost model and of the Math-heuristic. Their variables are defined in Table 1.

### 3.2. Flight schedule cost model

This model, composed of three steps, aims to restore the aircraft schedule with minimum cost for the airline and to identify disrupted passenger itineraries (due to any cancelled flight in the itinerary or to lack of enough time to connect between any two consecutive flights in the itinerary).

#### 3.2.1. Step 1 – Flight schedule adjustment

This step attempts to recover the original flight schedule by delaying or cancelling flights. The problem is solved through a multi-commodity network flow model. The mathematical modeling is based on finding the lowest cost flows in the disrupted time-space network with multiple flows and capacity limitation.

The objective function aims to obtain a new flight schedule minimizing flight cancellations and delays:

\[
FO = \min \left( \sum_{f \in F} \text{cost}_{\text{cancel}}(f) \cdot y_f + \sum_{f \in FC \in F} \text{cost}_{\text{delay}}(f) \cdot x_f \right)
\]  

(1)

Model restrictions:

Every \( f \in F_{\text{cancel}} \) must be cancelled and all its inserted arcs are also cancelled in Step 1:

\[
y_f = 1, \forall f \in F_{\text{cancel}}
\]

(2)

\[
\sum_{f \in FC \in F} x_f = 0, \forall f \in F_{\text{cancel}}
\]

(3)

For each flight only one option can be chosen: one inserted arc – original flight is an inserted arc with delay equal to 0 - or cancel:

\[
\left( \sum_{f \in FC \in F} x_f \right) + y_f = 1, \forall f \in F
\]

(4)

Each of the nodes in the time-space network must maintain flow balance.

Balancing is modeled as follows for each aircraft entry node:
Balancing is modeled as follows for each aircraft output node:

$$\sum_{f \in FC\text{-}OUTn} x_{fc} = N \cdot I \cdot N_n, \forall n \in N \cdot I$$

(5)

Airport capacity restrictions are:

$$\sum_{f \in FC\text{-}JNn} x_{fc} \leq N \cdot OUT_n, \forall n \in N \cdot OUT$$

(6)

3.2.2. Step 2 – Aircraft rotation

This step aims to obtain the new aircraft rotation. It receives as input the solution obtained in Step 1.

The objective function aims to minimize the cost of the new aircraft rotation, given by the cost due to changing the original aircraft rotation and the cost of the aircraft with no assigned flight:

$$FO = \min \sum_{f \in FC} \sum_{a \in A} (C_{\text{change}_{a,f}} \cdot r_{a,f} + C_{\text{not assign}} \cdot w_{a,f})$$

(8)

Model restrictions to obtain the new aircraft rotation:

Each flight has an assigned aircraft or not. It is important to note that a flight with no aircraft assignment will be cancelled in Step 2:

$$\left(\sum_{a \in A} r_{a,f}\right) + w_{a,f} = 1, \forall f \in X_f$$

(9)

Flights cancelled in Stage 1 (whether original or inserted arcs) cannot have assigned aircraft:

$$\sum_{a \in A} r_{a,f} = 0, \forall f \in F_{\text{cancel}}$$

(10)

The departure time of an aircraft (considering its readiness time) must be later than the aircraft's arrival time after the previous flight:

$$STD_{f_2} \geq (r_{a,f_1} \cdot ETA_{a,f_1}), \forall f_2 \text{ consecutive to } f_1, \forall a \in A, \forall f_1, f_2 \in X_f$$

(11)

3.2.3. Step 3 – Identifying disrupted passenger itineraries

The objective of this step is to identify the affected itineraries, either by cancellation of flights in the previous steps or by lack of enough time for passengers to accomplish their flight connections in the itinerary.

The objective function aims to minimize the amount of disrupted itineraries:

$$FO = \min \sum_{\tau \in T} c_{\tau} \cdot z_{\tau}$$

(12)

Model restrictions are as follows.

An itinerary is affected if it contains a cancelled flight:

$$z_{\tau} = 1, \forall \tau \text{ with any cancelled flight}$$

(13)

For each pair of flights on an itinerary, if the departure of a f2 flight, consecutive to f1, is later than the arrival of f1 plus the passenger's connection time, then the connection is lost. Otherwise, the itinerary is lost:

$$(\sum_{f \in FC\text{-}f} r_{fc} \cdot STD_{f}) - (\sum_{f \in FC\text{-}f_1} r_{fc} \cdot ETA_{f}) - \text{conn}_{\tau} \geq -L \cdot z_{\tau}, \forall (f_1, f_2) \in T, \exists \text{Big } L$$

(14)

3.3. Math-heuristic

The Math-heuristic also has three stages. Stages 1 and 2 are identical to those modeled in item 3.2, but Step 3 is addressed to reduce the loss of itineraries.

3.3.1. Step 3 – Adjustment of flight schedule to reduce disrupted itineraries

The objective of this step is to minimize the number of passengers with affected connections.
It looks for a new adjustment of the light network, by introducing flight delays. The aircraft rotation utilized is the one obtained in Step 2.

Objective function: to minimize flight schedule delays and number of disrupted itineraries:

\[
FO = \min \left( \sum_{f \in F} \sum_{c \in FC} \text{cost}_{fc} \cdot x_{fc} + \sum_{t \in T} c_t \cdot z_t \right)
\]  

Model restrictions:

Each flight can only have associated with it an inserted arc:

\[
\sum_{c \in FC} x_{fc} = 1, \forall f \in X_f
\]  

The departure time of a flight must be later than the landing time of the previous flight:

\[
x_{f_2} \cdot STD_{f_2} \geq x_{f_1} \cdot ETA_{f_1}, \forall f_2 \text{ consecutive to } f_1, f_1, f_2 \in FC_f
\]  

An itinerary is affected if it contains a cancelled flight:

\[
z_t = 1, \forall t \text{ with at least one cancelled flight}
\]  

For each pair of flights on an itinerary, if the departure time of \( f_2 \) flight, consecutive to \( f_1 \), is later than the time of arrival of \( f_1 \) plus the connection time of the passenger, then the connection is not lost. Otherwise the itinerary is affected:

\[
(\sum_{f \in FC} x_{fc} \cdot STD_{fc}) - (\sum_{f \in FC'} x_{fc} \cdot ETA_{fc}) - \text{conn}_t \geq -L \cdot z_t, \forall (f_1, f_2) \in T, \text{Big } L
\]  

Airport capacity restrictions:

\[
\sum_{f \in FC} x_{fc} \leq \text{Cap}_{p,m,h}, \forall p \in P, \forall m \in M, \forall h \in H
\]  

4. TESTS AND RESULTS

The models were implemented in Python 3.7, C++ (Step 2 only), Gurobi 8.1.1. They were tested thanks to a set of real world instances created by ROADEF for an operations research challenge in 2009 (Palpant et al, 2009). The following information were considered for each instance: initial programming in terms of the number of flights per day of the recovery window; and numbers and details of the aircraft fleet, of the airports and of the affected passenger itineraries during the recovery window. The following disruptions are considered: flight delays in minutes, flight cancellations and reduction of airport capacity in specific time slots during the recovery window. Table 2 and Table 3 show the characteristics of the test instances.

The number of flights in Table 2 and Table 3 correspond to one day of operation and the number of flights in the modeling is proportional to the size of the recovery window, since the airline flight schedule problem is replicated for each day of the operating window. Table 4 presents results obtained by the Flight schedule cost model. Table 2 present results obtained by the Math-heuristic. The inserted arcs were limited to a maximum delay of two hours in relation to the original flight and the upper limit of the recovery window. Two simplifications are inherent to the developed models: all aircraft must have the same configuration and no aircraft maintenance is considered – although available in the test instances these information were not considered in the modeling.

Instances A1, A2, A3, A6, A7, and A8 – Table 2 – were run on an Intel i7 machine with 16Gb of RAM and Windows operating system. Instances B1, B2, B3, B4, B6, B7, B8, B9, A4, A9, A5 and A10 - with the characteristics presented in Table 2 and Table 3 - were processed using a high-performance cluster node - HPC - with Intel(R) Xeon(R) CPU E7- 2870 @ 2.40GHz and 512 GB of RAM. Constraint (11), in Step 2, generates a number of constraints in the order of \( |A| \cdot |X| \)², where \( |A| \) = number of aircraft and \( |X| \) the number of impacted flights, i.e., the solution
space grows exponentially with the number of aircraft and flights - one of the characteristics of an NP-hard problem (Gao et al., 2009).

In instances "B" the restriction (11) generated the approximate amount of 430,000,000 constraints in the optimizer software, occupying almost all the available memory (512 Gb) of the High Processor Computing equipment during the model loading step for solution execution. Instances B5 and B10 could not be run on the cluster node because they exceeded the available memory capacity during the execution of Step 2 of the models - both have a larger recovery window than the others (approximately half a day of operation more)

Table 2 - Test instances description - Source: Palpant et al., 2009

<table>
<thead>
<tr>
<th>Instances</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of flights to be analyzed per day of the recovery window</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
<td>608</td>
</tr>
<tr>
<td>Number of aircraft</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>Number of airports</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Number of itineraries</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
<td>1943</td>
</tr>
<tr>
<td>Number of airports affected with reduced capacity</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of delayed and cancelled flights</td>
<td>63</td>
<td>107</td>
<td>83</td>
<td>41</td>
<td>0</td>
<td>63</td>
<td>107</td>
<td>83</td>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>Recovery window period of time</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
<td>7/Jan 12:00</td>
</tr>
</tbody>
</table>

Table 3 - “B” test instances description - Source: Palpant et al., 2009

<table>
<thead>
<tr>
<th>Instances</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
<th>B8</th>
<th>B9</th>
<th>B10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of flights to be analyzed per day of the recovery window</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
<td>1422</td>
</tr>
<tr>
<td>Number of aircraft</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
</tr>
<tr>
<td>Number of airports</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Number of itineraries</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
<td>11214</td>
</tr>
<tr>
<td>Number of airports affected with reduced capacity</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Number of delayed and cancelled flights</td>
<td>229</td>
<td>254</td>
<td>228</td>
<td>229</td>
<td>0</td>
<td>229</td>
<td>254</td>
<td>228</td>
<td>229</td>
<td>0</td>
</tr>
<tr>
<td>Recovery window period of time</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
<td>7/Jan 04:00</td>
</tr>
</tbody>
</table>

To compare the Flight schedule cost model with the Math-heuristic, a Total Cost was defined as:

\[
\text{Total Cost} = \text{Flight Schedule Delay Total Cost} + \text{Total Cost of Affected Passenger Itineraries.}
\]

The Flight Schedule Delay Cost is calculated in terms of the amount of delay introduced in the adjusted flight schedule.

A value of the schedule delay per minute of €100 was used (Muligan, 2019). The cost of each passenger itinerary is provided as a parameter in Euros (€).

Cancelled Itineraries Delta Cost in Table 5 compares the Cancelled Passenger (Pax) Itineraries Total Cost obtained by the Math-heuristic and by the Flight schedule cost model:

\[
\text{Cancelled Itineraries Delta Cost} = \left( \frac{\text{Math-heuristic Cancelled Pax Itineraries Total Cost} - \text{Flight schedule cost model Cancelled Pax Itineraries Total Cost}}{\text{Flight schedule cost model Cancelled Pax Itineraries Total Cost}} \right)
\]
Table 4 - Results obtained by the Flight schedule cost model

<table>
<thead>
<tr>
<th>Flight schedule cost model results</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B9</td>
<td>B8</td>
<td>B7</td>
<td>B6</td>
</tr>
<tr>
<td>Sch Delay Step 1 (h)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Canc Flt Step 1</td>
<td>2.72</td>
<td>2.72</td>
<td>2.72</td>
<td>2.72</td>
</tr>
<tr>
<td>Canc Flt Step 2</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Flt. Step 2sch</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Flt. Step 3sch</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Pax In.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Canc. Flt Step 3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Canc Pax Delta</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Sch Delay Total Cost ($1000)</td>
<td>1.29</td>
<td>1.29</td>
<td>1.29</td>
<td>1.29</td>
</tr>
<tr>
<td>Delta Total Cost (h)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Delta Total Cost ($1000)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 5 - Results obtained by the Math-heuristic

<table>
<thead>
<tr>
<th>Math-heuristic results</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sch Delay Step 1 (h)</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Canc Flt Step 1</td>
<td>2.72</td>
<td>2.72</td>
<td>2.72</td>
<td>2.72</td>
</tr>
<tr>
<td>Canc Flt Step 2</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Flt. Step 2sch</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Flt. Step 3sch</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td>Canc Pax In.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Canc. Flt Step 3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Canc Pax Delta</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Sch Delay Total Cost ($1000)</td>
<td>1.29</td>
<td>1.29</td>
<td>1.29</td>
<td>1.29</td>
</tr>
<tr>
<td>Delta Total Cost (h)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Delta Total Cost ($1000)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Delta Total Cost column in Table 5 compares the Total Cost obtained by the math-heuristic and by the Flight schedule cost model:

Delta Total Cost = (Math-heuristic Total Cost – Flight schedule cost model Total Cost)/Flight schedule model Total Cost).

Step 2 was limited to a 12 hours execution and the Gurobi Gap % column in Table 4 and Table 5 indicate the gap obtained after the execution time indicated in column Δt (min).

The Math-heuristic presented lower costs of affected itineraries and Total Costs very close to the ones obtained by the Flight schedule cost model in the majority of the tests instances.
So, the cost of delays introduced to recover more passengers did not substantially affect the Total Cost for the airline. This is a very important result. It alerts that it is worthwhile to compare the aircraft network recovery solution with the corresponding aircraft and passenger one. This alert and the fact that the developed models permit the airline to compare these alternative recovery options represent contributions of this work.

The Math-heuristic was designed to reduce the number of disrupted passenger itineraries due to lack of enough connection time after the aircraft schedule adjustment. This is accomplished by inserting additional delays in the flight network beyond those caused by disruption. These solutions should have a lower cost of cancelled itineraries than the flight schedule cost model solution. However, the total cost may eventually be higher than that of the flight schedule cost solution, depending on the cost of the inserted delays.

Table 3 - Results obtained by the Math-heuristic

<table>
<thead>
<tr>
<th>Quantity of Affected Flights during the Recovery Window</th>
<th>Recovery Period (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>276</td>
</tr>
<tr>
<td>A2</td>
<td>153</td>
</tr>
<tr>
<td>A3</td>
<td>214</td>
</tr>
<tr>
<td>A4</td>
<td>337</td>
</tr>
<tr>
<td>A5</td>
<td>928</td>
</tr>
<tr>
<td>A6</td>
<td>276</td>
</tr>
<tr>
<td>A7</td>
<td>153</td>
</tr>
<tr>
<td>A8</td>
<td>214</td>
</tr>
<tr>
<td>A9</td>
<td>337</td>
</tr>
<tr>
<td>A10</td>
<td>928</td>
</tr>
<tr>
<td>B1</td>
<td>1838</td>
</tr>
<tr>
<td>B2</td>
<td>1821</td>
</tr>
<tr>
<td>B3</td>
<td>1838</td>
</tr>
<tr>
<td>B4</td>
<td>1838</td>
</tr>
<tr>
<td>B6</td>
<td>1838</td>
</tr>
<tr>
<td>B7</td>
<td>1821</td>
</tr>
<tr>
<td>B8</td>
<td>1838</td>
</tr>
<tr>
<td>B9</td>
<td>1838</td>
</tr>
</tbody>
</table>

The Math-heuristic actually led to reductions on the costs of cancelled passenger itineraries due to lack of enough connection time between flights for sixteen out of eighteen test instances (for two of them no improvements to passenger disruption cost were achieved). The achieved cost reductions represent an indicator of the effectiveness of the proposed Math-heuristic to solve the airline recovery problem considering aircraft and passenger networks.

Table 6 shows the number of affected flights during the recovery window for each instance. The solution time of the Flight schedule model and of the Math-heuristic for instances with 85 aircraft and 276 impacted, as presented in instances A1, A2, A3, A6, A7 and A8, is acceptable by the industry (up to 30 minutes).

However, usual enterprise computer (Intel i7 machine with 16Gb of RAM and Windows operating system) could not solve instances with a quantity of aircraft greater than 85 and an impacted flight greater than 276 within the industry acceptable timeframe (30 minutes).

Failure to achieve the optimal solution for the larger instances, even using a high-performance HPC-USP cluster node, confirms the NP-hard nature of the problem.

Meanwhile, the high-performance server – HPC allowed to extend the execution time of the model Step 2 to up to 12 hours in order to reach a solution with minimum gap for instances with a high number of flights. The gap ranged from 0% to 0.07%.
5. CONCLUSIONS

This work presented a math-heuristic to solve the airline recovery problem considering aircraft and passenger networks in partially integrated, partially interactive and non-simultaneous way. It seeks to reach the best solution for both the airline and the affected passengers. A model to restore the flight schedule with minimum cost for the airline was also developed.

These two models permit to compare results of the recovery with and without the passenger dimension. This type of comparison was pursued by this research. It allows the airline to decide on the pros and cons due to integrating or not the passenger dimension in the recovery process.

This airline recovery problem is of the NP-hard class, that is, there are no polynomial algorithms for its resolution, regardless of how it is considered – integrated, partially integrated or not integrated (Yu and Qi, 2004). Hence, the proposition of the Math-heuristic for solving it.

The math-heuristic modeling was chosen due to the fact that it represents a relatively novel modeling approach addressed to solve NP-hard problems.

The real world instances created by ROADEF for an operations research challenge in 2009 were used to test the models.

The solution times of the Flight schedule model and of the Math-heuristic for instances with up to 85 aircraft and, 276 impacted flights as in instances A1, A2, A3, A6, A7 and A8, are compatible with the industry requirement (up to 30 minutes).

Instances with a quantity of aircraft greater than 85 and with an impacted flight greater than 276 could not be solved in the industry acceptable timeframe (30 minutes) by an usual computer (Intel i7 machine with 16Gb of RAM and Windows operating system).

Failure to achieve the optimal solution for the larger instances, even using a high-performance HPC cluster node, confirms the NP-hard nature of the problem.

However, the high-performance server – HPC allowed to extend the execution time of the model Step 2 to up to 12 hours, in order to reach solutions with minimum gaps. The gap values ranged from 0% to 0.07%.

The applications of the Math-heuristic to the ROADEF instances presented, for 11 out of 18 instances, total costs of recovery with a difference of as much as 1.0 % to the ones obtained by the Flight schedule cost model. For 14 out of 18 instances the difference was of less than 3.7%. To say, the cost of the delays introduced to recover more passengers did not substantially affect the airline recovery cost for the tested instances.

This result permits to realize the importance of solving the recovery problem looking for the best solution for both the airline and the passengers. It justifies the purpose of this research to develop models to permit the airline to analyse the cost impacts related to including the passenger dimension in the recovery attempt.

ACKNOWLEDGEMENTS

The authors acknowledge CAPES for the support to the Post-graduate Program at EPUSP; CNPq for a Research Productive Fellowship (Process 307622/2015-0); and LPT/EPUSP for technical support.

REFERENCES


