Incorporating User Preferences within an Optimal Traffic Flow Management Framework

Joseph Rios∗   Kapil Sheth†

NASA Ames Research Center, Moffett Field, CA 94035

Sebastian Gutierrez-Nolasco‡

UC Santa Cruz, Moffett Field, CA 94035

The effectiveness of future decision support tools for Traffic Flow Management in the National Airspace System will depend on two major factors: computational burden and collaboration. Previous research has focused separately on these two aspects without consideration of their interaction. In this paper, their explicit combination is examined. It is shown that when user preferences are incorporated with an optimal approach to scheduling, runtime is not adversely affected. A benefit-cost ratio is used to measure the influence of user preferences on an optimal solution. This metric shows user preferences can be accommodated without inordinately, negatively affecting the overall system delay. Specifically, incorporating user preferences will increase delays proportionally to increased user satisfaction.

I. Introduction

Traffic management’s mission, as described by the Federal Aviation Administration (FAA), is to “balance air traffic demand with system capacity to ensure the maximum efficient utilization of the National Airspace System (NAS).” The service provider in the United States, the FAA endeavors to involve the airlines in the decisions affecting the airlines’ performance. Currently, that process of inclusion is called Collaborative Decision Making (CDM). In parallel to the fostering of collaboration between the service provider and the users of the airspace, investigation into the incorporation of optimal decision making to Traffic Flow Management has been on-going. These two Traffic Flow Management research areas (collaboration and optimization) are not generally considered simultaneously.

Several proposed approaches to Traffic Flow Management explicitly consider some form of equity through collaboration. Equity may be achieved through the use of auctions or credits by distributing delay

∗Research Aerospace Engineer, Systems Modeling and Optimization Branch, Mail Stop 210-15, Member AIAA.
†Research Aerospace Engineer, Systems Modeling and Optimization Branch, Mail Stop 210-15, Associate Fellow AIAA.
‡Senior Software Engineer, UC Santa Cruz, Mail Stop 210-8.
amongst airlines, or through enforcement of a first-come, first-served policy. An alternative, egalitarian option is to allow each flight to be given an equal cost to delay to obtain a global cost minimization. A more equitable form of that sentiment would be to allow airlines to rank the importance of their own flights by increasing the cost to delay “important” flights and then allow the service provider to find a global schedule with consideration of these rankings as demonstrated by Sheth and Gutierrez-Nolasco. In that study, there is a heuristic used to generate a schedule as opposed to an optimal approach as will be used here. Ranieri, Castelli and Martin proposed a collaboration-optimization concept wherein auctions are used to allocate departure slots to flights. The work presented here differs in terms of scale (thousands of flights and hundreds of resources) and allocation method.

In terms of computational feasibility, this work will leverage results and systems developed in previous research. Namely an optimal decomposition approach developed by Rios and Ross is used to solve the large-scale Traffic Flow Management model developed by Bertsimas and Stock Patterson. This model provides a delay-cost-optimal solution given a set of trajectories and capacities. It is capable of assigning both airborne and ground delay. In this paper, the credits proposed by Sheth and Gutierrez-Nolasco’s scheme will be used as weights in the objective function of the Bertsimas-Stock Patterson model, thus marrying the concepts of optimal Traffic Flow Management and Collaborative Decision Making.

The Bertsimas and Stock-Patterson (BSP) model will be described in Section II. A more complete description of the credits concept is then described in Section III. Next, the input data (Section IV) and the experiments (Section V) are then fully described. Concluding remarks and potential future directions are then provided in Section VI.

II. Optimization Model

The Bertsimas-Stock Patterson (BSP) model is used to perform scheduling that minimizes delay costs. After presenting the model in the next section, the architecture of the solver will be described in the following section.

II.A. The Bertsimas-Stock Patterson Model

\[
\begin{align*}
\text{Min} & \quad \sum_f \left[ (c_f^a - c_f^g) \sum_{t,k=P(f,1)}^{} t(w_{f,t}^k - w_{f,t-1}^k) \\
& \quad + c_f^a \sum_{t,k=P(f,\text{last})}^{} t(w_{f,t}^k - w_{f,t-1}^k) \\
& \quad + (c_f^a - c_f^g) d_f - c_f^a r_f \right] \\
\text{Subject to:} & \quad \sum_{f: P(f,1)=k}^{} (w_{f,t}^k - w_{f,t-1}^k) \leq D_k(t), \quad \forall k \in \text{Airports}, t \in \text{Time} \\
& \quad \sum_{f: P(f,\text{last})=k}^{} (w_{f,t}^k - w_{f,t-1}^k) \leq A_k(t), \quad \forall k \in \text{Airports}, t \in \text{Time} \\
& \quad \sum_{f: P(f,i)=j, P(f,i+1)=j'} (w_{f,t}^{j'} - w_{f,t}^j) \leq S_j(t), \quad \forall j \in \text{Sectors}, t \in \text{Time} \\
& \quad w_{f,t+\mu(f,j)} - w_{f,t}^j \leq 0 \quad \forall f \in \mathcal{F}, j = P(f,i), j' = P(f,i+1) \\
& \quad w_{f,t}^j - w_{f,t-1}^j \geq 0 \quad \forall f \in \mathcal{F}, j \in (f's \text{ flight path}) \\
& \quad w_{f,t}^j \in \{0, 1\} \quad \forall f \in \mathcal{F}, j \in (f's \text{ flight path}), t \in \text{Time}
\end{align*}
\]
The interpretation of the decision variables, \( w_{j,t}^f \), is that they are valued 1 if and only if a flight \( f \) has entered sector \( j \) by time \( t \) and 0 otherwise. For further illustration, the set of variables for a given flight-sector pair can be seen as a step function from 0 to 1 with the change occurring at the first time the flight enters the sector which is enforced by Constraints (8).

The objective function describes the sum of weighted air and ground holding costs over all flights. Expression (1) denotes the departure time since the difference between the two decision variables will always be zero (note the definition of \( w \) except when the flight actually departs, at which time the difference will be 1 and the product with \( t \) will simply be \( t \)). Likewise, expression (2) simplifies to the arrival time for flight \( f \). Expression (3) takes into account the scheduled arrival and departure times (\( r_f \) and \( d_f \), respectively). Thus the overall result for each flight is the weighted cost of holding that flight where \( c_g^f \) is the cost for holding a flight on the ground for one time step and \( c_a^f \) is the cost for holding a flight in the air for one time step. Typically the problem is modeled with \( c_g^f < c_a^f \) as the preference is to hold flights on the ground rather than in the air. A reasonable value for the ratio of \( c_g^f \) to \( c_a^f \) is 2 to 1. This ratio is roughly based on the extra cost of fuel to hold in the air for one minute versus one the ground for the same length of time, but it could vary based on many parameters (aircraft type, where ground holding occurs, etc.). In Section III the number of credits assigned to a given flight will translate into more appropriate weights from the airlines’ perspective.

For each flight, the model is able to decide if the flight needs to be held anywhere (and for how long) in order to satisfy the airport departure, airport arrival, and sector capacity constraints (constraints (4), (5), and (6), respectively). Airports are denoted by \( k \) and sectors by \( j \). Each of the capacities is a function of time, \( t \). Each flight in the set of flights, \( F \), is described as an ordered list of distinct sectors, \( j \), from a set of sectors, \( J \), with earliest and latest feasible entry times for each of those sectors. For modeling purposes, airports are considered a subset of sectors with associated arrival and departure capacities. A sector in a flight path is denoted by \( P(f,y) \), where \( f \) is the flight and \( y \) is the ordinal representing its place in the flight path. For ease of notation, \( P(f,\text{last}) \) is used to represent the last sector (usually an airport) in \( f \)’s path.

The problem is also constrained by the physical and temporal limitations of the flights. Specifically, each flight spends, at least, the specified minimum sector time, \( \mu(f,j) \), in each of its sectors, \( j \) as described by constraints (7). Coupled with the set of constraints enforcing the logic of the variables, i.e. all 0’s before all 1’s as discussed above (constraints (8)), physical and temporal logic is satisfied.

The original model, as published, contains a set of constraints to represent connecting flights. Those constraints have been removed for this study. The original paper also describes several alterations to the model to accommodate other aspects of TFM. For example, route choice for flights can be incorporated into the model by introducing a new subscript to the decision variables indicating a route.

II.B. The Solver

While the detail of the BSP model has many positive attributes, it comes at a steep computational cost. The difficulty of solving the BSP model has been well-documented. Attempts to mitigate the computational concerns of this model and related models have included partitioning the problem in space or time, decreasing the time resolution (i.e. using fifteen-minute time steps instead of one-minute time steps), and selectively paring the list of flights on which to optimize.

The most successful attempt at reducing the runtime of the optimization procedure has been the implementation and application of Dantzig-Wolfe Decomposition. This approach allows for solving several smaller subproblems in parallel over the course of several iterations. The solutions to the subproblems are organized by a master problem, which, in turn, adjusts the objective functions of the subproblems to guide the convergence of the entire procedure. Since the problem is solved in parallel, the wall-clock time of the optimization procedure can be reduced by more than 90%. The details of this implementation are fully described in a previous paper.
### III. Credits

The concept of credit points was proposed in earlier research. Credits are a form of artificial currency and are used for specifying users' flight preferences. The purpose is to incorporate users' flight priorities and multiple route options during the pre-departure route filing process.

Credits are assigned to flights for the duration of a session. The session could be for a day or for two hours (as in this research.) At the beginning of each session, the users are provided a fixed number of credits based on the size of their operations during the session. The total number of credits allocated for each user is five times their number of flights operating within the session. The credits expire at the end of each session. It is expected that the users will typically expend between zero and ten credits for each flight, based on its importance. However, there is no limit to the number of credits assigned to a flight, other than the users total credit balance. Flights would be assigned credits by an airline using its own utility function. Users assign more credits to more important flights based on their utility function. The utility function could depend on a combination of several parameters, including flight distance, number of passengers, type of aircraft, load factor, crew connection, etc.

For this research, four utility function proxies are used for credit assignment. The first is a Gaussian distribution with a mean of five credits and a variance of 2.13. The second is a distance-based function. This function assigns higher priority to longer flights. Flights that travel 2,000 nmi or greater are assigned ten credits. For example, a flight distance of 1,800 nmi is assigned nine credits, while a flight with 1,600 nmi is assigned eight credits. All flights with 500 nmi or less are assigned three credits. The third distribution is based on the hub-and-spoke business model. Most major airlines in the United States have operations with a hub-spoke system with some number of hubs and connections to other major cities. In this distribution, any flight traveling to or from a hub city is assigned ten credits. A flight arriving or departing from a (non-hub) major city is assigned eight credits. All other user flights are assigned two credits. The last is a flat distribution of credits is implemented wherein five credits are assigned to each flight. This distribution roughly models the value of the flights in the service provider’s perspective: all flights have equal value. The various modes of credit assignment are summarized in Table 1.

<table>
<thead>
<tr>
<th>Assignment Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>All flights assigned five credits</td>
</tr>
<tr>
<td>Gaussian</td>
<td>All participating airlines' flights assigned random number of credits with mean of five.</td>
</tr>
<tr>
<td>Distance</td>
<td>All participating airlines' flights assigned credits based on distance of flight.</td>
</tr>
<tr>
<td>Hubs</td>
<td>All participating airlines' flights assigned credits based on their hubs.</td>
</tr>
</tbody>
</table>

Once all the users assign credits for each of their flights, the system evaluates each of the constraints in the National Airspace System. For example, a sector can handle 18 aircraft in a 15-minute interval, but 20 aircraft plan to go through that sector. The system ranks all aircraft based on their credits and requires the last two aircraft to either use their filed optional routes or accept a delay on the ground. This process continues until all flights reach their destination. In this implementation, all of the airspace and airport constraints are satisfied.

### IV. Input Data

The flight data used in the experiments was taken from historical records. Specifically, air traffic data from Wednesday, August 24th, 2005 was obtained from the FAA’s Enhanced Traffic Management System (ETMS). This represents the fairly dense weekday traffic on a day with minimal weather impacts. The choice of a relatively clear weather day implies that the majority of flights were not delayed or rerouted by air traffic control due to weather.

Given that the optimization model was refined in scope to target the continental US, all international flights...
were removed from the air traffic data. However, domestic flights flying through Canadian airspace were retained, but the Canadian sector constraints were ignored. The exclusion of international flights simplifies the data processing phase. Similarly, domestic flights that re-entered the same sectors several times or were missing sectors in their flight plans were removed.

Once the air traffic data contained clean domestic flights, a two-hour baseline scenario was generated using the heavy traffic volume of 3 to 5 p.m. Eastern Daylight Time (EDT). The baseline scenario contained 9,552 domestic flights, representing 179 users. Similarly, a three-hour scenario comprised of the previous scenario plus one more hour (i.e. 3 p.m. to 6 p.m.) was also generated. The three-hour scenario contained 12,657 domestic flights and was tested simply to verify the runtime would still be reasonable.

As far as national airspace capacity constraints are concerned, any sector or airport that was used by any flight in the system was included in the data set. This resulted in 974 sectors being included along with 905 airports. Since determining the operational capacity of airports is outside the scope of this paper, we used the FAA’s Operations Network (OPSNET) to estimate the capacity of the top seventy airports. Airports outside of this set were assumed to have a default value of thirteen departures and arrivals per fifteen minutes. Similarly, the default sector capacities known as Monitor Alert Parameters (MAP) were from the Future ATM Concepts Evaluation Tool (FACET).23

It is worth mentioning that many of the constraints generated by these resources (see Constraints (4) and (5)) were not near their maximum capacity levels and were, thus, simplified out of the problem formulation. For example, if only a few flights are using Moffett Federal Airfield (even in consideration of potential delays), then the capacity constraints generated for the BSP model could never be binding and, thus, could be removed from the formulation. This filtering of constraints is a by-product of the optimization procedure (i.e. the constraints are trivially satisfied and therefore not needed to solve for the optimum). There was no distinction made between low, high or super-high en route sectors, but Terminal Radar Approach Control (TRACON) sectors were given special status by allowing them infinite capacity. The basis for this decision is that the capacity of these sectors is typically driven by the ability of the airports in the TRACON to accept or depart aircraft and those constraints (again, constraints Constraints (4) and (5) are already included in the formulation).

V. Experiments and Results

The experiments will seek to measure any potential tradeoffs in overall (system-level) delay costs in order to satisfy user preferences. To achieve this, a flight schedule and capacity data are generated as described in Section IV and fed into two different scheduling systems. Each of these systems schedules the traffic to eliminate over-capacity resources (sectors and airports). For each method, the four distinct sets of user preferences (as described in Section III) are applied. This results in eight distinct scenarios. The high-level flow of the experiments is described visually in Figure 1. Note that the initial schedule and capacity data remain constant over all two-hour experiments.

![Figure 1. The general flow of the experiments.](image-url)
The number of credits that are assigned to a flight will be the value of $c_f^g$ (ground-holding cost) for that flight in the BSP model. This has the desired effect of making it more costly for the BSP model to assign delay to a flight with more credits. Since the BSP model is capable of holding flights in the air, a cost for air holding ($c_f^a$) must be generated for each flight. As discussed in Section II, a ratio of 2-to-1 air-to-ground costs is desired.\cite{16,17} Recall that the average number of credits available per flight is 5. If a flight is assigned a ground holding cost of, say, 2, then when doubling that value for the air holding cost the result is a cost to hold that flight in the air which is less than the average cost to hold flights on the ground. This results in excess air holding in the system. Therefore to assign air-holding costs to a flight, 5 is added to the ground-holding cost. This results in a system where the average air-holding cost is twice the average ground-holding cost, and the cost to hold a flight in the air is still greater than the average cost to hold a flight on the ground (i.e. 5). These calculations are formalized in Equations 10 and 11. Note that since the average number of credits assigned to any flight in any credit assignment mode is 5, the ratio of air to ground holding costs is maintained at 2:1 when averaged over all flights.

$$c_f^g = \text{num\_credits\_assigned}$$

$$c_f^a = \text{num\_credits\_assigned} + 5$$

The primary results of the optimization runs are presented in Table 2. Note that the delay costs for a given flight are found by multiplying the number of minutes of air or ground delay for that flight by that flight’s air or ground holding costs, respectively. Total delay costs are found by summing this value over all flights. This overall value is most clearly seen in the first row of Table 2 where all flights had a ground holding cost of 5 and an air holding cost of 10 ($((29 \times 10) + (213 \times 5) = 1355)$).

<table>
<thead>
<tr>
<th>Credit Assignment</th>
<th>Delay Costs</th>
<th>Delay (min)</th>
<th>Number of Delayed Flights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>System</td>
<td>Users</td>
<td>Air</td>
</tr>
<tr>
<td>Flat</td>
<td>1355</td>
<td>1355</td>
<td>29</td>
</tr>
<tr>
<td>Gaussian</td>
<td>1775</td>
<td>1103</td>
<td>26</td>
</tr>
<tr>
<td>Distance</td>
<td>1470</td>
<td>1306</td>
<td>21</td>
</tr>
<tr>
<td>Hubs</td>
<td>1425</td>
<td>1310</td>
<td>22</td>
</tr>
</tbody>
</table>

To measure the impact of including user preferences within an optimization framework, the “Improvement Metric”, $\Delta^+$, is introduced. It is simply the ratio of the delay cost decrease from the users’ perspective to the system delay cost increase from the service provider’s perspective. This metric is an interpretation of the benefit cost ratio from the discipline of cost-benefit analysis. The major assumption for this metric is that the service provider will always rank the importance of each flight equally, thus the schedule produced using the user preferences will never be better in terms of overall delay than the “Flat” mode. In other words, the “Flat” mode will produce a global minimum delay cost. To provide an interpretation of the metric, consider $\Delta^+ = 1$. This implies that the increase in system costs is equal to the decrease in users’ perspective of the costs. If the value of the $\Delta^+$ goes below 1, then the increase in system delay costs are not entirely offset by the decrease in users’ perspective of the costs. While a value above 1 denotes a larger improvement in users’ perspective of the delay versus the increase in system delay costs. Any value at or above 1 would likely denote that the incorporation of the users’ preferences is likely worthwhile while values below 1 would indicate some reason for concern in implementing users’ preferences. This is assuming that there is equal emphasis placed on the value of system delay and users’ perspective of delay. Changes in this assumption would change the value of 1 as the “tipping point.” The metric is formalized in Equation 12.

$$\text{Improvement Metric} = \Delta^+ = \frac{\text{(% user cost decrease from “flat” credits)}}{\text{(% system cost increase from “flat” credits)}}$$

It is illustrative to consider a toy example. Say for some scenario the overall system delay cost is four, which is a global minimum for the BSP model using the flat credit assignment. The cost from the perspective of the users, however might be six since they have a different way of valuing flights than the service provider. Note that there is the same amount of delay minutes in both perspectives, just differing delay costs. If the
optimization is re-run using the users’ credits for valuing flights, the system delay would not decrease as we already have a global minimum of four. Say the new system delay costs are now five when considering the users’ preferences, but the users’ perspective of the delay costs is now three (this would be the global minimum from the users’ perspective). The change in system delay cost is 25% (from four to five) while the change in user delay cost is 50% (from six to three). The value of $\Delta^+$ would then be $50/25 = 2$.

Since $\Delta^+$ needs to be calculated comparing the “Flat” credit assignment to each other method for assigning delays, three tables are now presented. Tables 3, 4, and 5 provide the results for the Gaussian, Distance, and Hubs credit assignments, respectively.

<table>
<thead>
<tr>
<th>Credit Assignment</th>
<th>System Delay Costs</th>
<th>User Delay Costs</th>
<th>$\Delta^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>1355</td>
<td>1327</td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>1775</td>
<td>1103</td>
<td>0.54</td>
</tr>
<tr>
<td>Delta Costs (%)</td>
<td>31.0%</td>
<td>16.9%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Credit Assignment</th>
<th>System Delay Costs</th>
<th>User Delay Costs</th>
<th>$\Delta^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>1355</td>
<td>1416</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>1470</td>
<td>1306</td>
<td>0.92</td>
</tr>
<tr>
<td>Delta Costs (%)</td>
<td>8.5%</td>
<td>7.8%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Credit Assignment</th>
<th>System Delay Costs</th>
<th>User Delay Costs</th>
<th>$\Delta^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>1355</td>
<td>1395</td>
<td></td>
</tr>
<tr>
<td>Hub</td>
<td>1425</td>
<td>1310</td>
<td>1.18</td>
</tr>
<tr>
<td>Delta Costs (%)</td>
<td>5.2%</td>
<td>6.1%</td>
<td></td>
</tr>
</tbody>
</table>

These results indicate that user preference could be reasonably accommodated (as measured by $\Delta^+$) without jeopardizing overall system performance. In this small set of experiments, the “hubs” method for assigning credits appears to perform best, but the margin over the “distance” method is small (1.18 versus 0.92). The fact that they both perform better than the Gaussian method is encouraging as it is expected that the distance- and hub-based methods more closely model how airlines might value their flights. This is with the understanding that each airline would likely define its own method for assigning credits that would vary to some degree from every method presented here.

All delay values presented thus far are produced using the optimal scheduler described in Section II. The credit-based method of producing a feasible schedule as described in Section III offers an opportunity for comparing two models in terms of the delays they generate. The credit-based scheduler is a heuristic versus an optimal method. The optimal scheduler does not consider rerouting, but rerouting is a key feature of the credit-based method. The optimal scheduler can implement air-holding, while the credit-based method only uses ground controls. Even with all of these differences, it is still beneficial to compare the solutions that the methods provide. So, given the initial schedule and capacities and a set of reroutes (which the optimal scheduler does not consider), the overall delay in air minutes and ground minutes can be calculated. The change in flight time caused by a reroute in the credit-based method is considered “air delay” for the sake of comparison and is weighted as such. The results are provided in Figure 2.

The comparison reveals that while the credit-based scheduler was successful on its own, it may yet be strengthened by the incorporation of an optimization procedure to perform the scheduling. It is interesting to note that even without rerouting, the optimal scheduler is able to find a lower delay cost schedule in every scenario.

When using an optimization model in Traffic Flow Management, the runtime of the optimization procedure is often a chief concern. Even though the BSP model has been shown solvable to optimality in a reasonable amount of time for the planning horizons considered here,

\[ \text{American Institute of Aeronautics and Astronautics} \]
Figure 2. Comparing delay values between optimal method and credit-based scheduler.

To verify that the optimization procedure is valid for a longer planning horizon, one additional run was completed. A three-hour scenario using the same start time as the two-hour scenario revealed that the optimization model was capable of producing the same quality results as the two-hours experiments demonstrated. The runtime was 245 seconds and the value of $\Delta^+$ when compared with the Hubs credit assignment was 1.39.

VI. Concluding Remarks

Future TFM decision support tools will need to meet several criteria for operational acceptance. Incorporation of user preferences is one key criterion and efficient automated optimization will likely be another. In this paper, an existing concept used to weight the importance of each of an airline’s flights is wedded with an established optimal approach for determining delay assignments. Results indicate that the solution from the wedded system does not degrade in quality. In fact, the wedded solution can be considered better than either of the individual solutions. The delay costs discovered for an instance in which user preferences are not considered do not significantly increase when user preferences are incorporated. Likewise, user preferences are generally met just as well when global delay costs are considered as when they are not. In addition, the incorporation of user preferences do not adversely affect the runtime qualities of the optimization approach. These results indicate that future scheduling tools need not sacrifice either aspect of quality in order to satisfy the service provider’s need to minimize global delay costs simultaneously satisfying the users’ desires.
to maximize their personal cost functions.

Future work may include incorporating the optimization model into human-in-the-loop simulations. The Bertsimas-Stock Patterson model could also be extended to handle multiple routes per flight in order to more fully exercise the credits concept.

References

1. FAA, Order JO 7210.3V Facility Operation and Administration, February 2008.