Distributed Mechanisms for Determining NAS-Wide Service Level Expectations: Year 1 Report

by

Michael Ball
Cynthia Barnhart
Antony Evans
Mark Hansen
Yi Liu
Prem Swaroop
Vikrant Vaze

May 19, 2011

1. Overview of Problem and Results

1.1 Background

One of the prime NextGen objectives is to move to a performance-based air traffic management system. Such a system should be capable of making intelligent trade offs among different performance criteria when deciding on specific traffic flow management actions. Viewed at a high level the objective of this research is to provide a means for the flight operator community to collectively set service level expectations and thereby define the criteria for making the required performance trade offs.

One can also view the problem to be solved based on current operational challenges. Within today's US National Airspace System (NAS), flight operators influence the decisions of FAA traffic managers in many ways and on a variety of levels. Every two hours, the FAA's air traffic control system command center holds a strategic planning telecon in which airlines express their opinions on options being considered. When ground delay programs (GDP's) are being planned, airlines frequently call the command center to collaborate on decisions involving GDP parameter settings and even the decision on whether or not to implement the GDP. At the local level airlines will frequently interact with Center or Tracon personnel to express opinions on flow management strategies. Such interactions are certainly quite legitimate and desirable in that they allow the NAS users who are impacted by various FAA decisions to help the FAA understand their priorities and the impact of FAA actions. At the same time, these interactions can be somewhat ad hoc and at times can lead to inequities. The flight operators who are most aggressive in expressing their views can sometimes have a disproportionate influence on decisions that affect a broad range of others who are less vocal. Of even greater import, the verbal input focuses on the mechanisms, e.g. GDP's, and not on the underlying performance objectives. The overall goal of our research proposed is to formalize these interactions under a mechanism that allows all flight operators to participate but at the same time provides those most impacted by an FAA decision to have the largest influence.

1.2 Description of Desired System Architecture

The result of this research should be a mechanism that allows the NAS flight operator community to jointly set the service expectations that will guide NAS decision-making on a given day or during given time period.

The 5 performance criteria of interest for this activity are:

1) capacity, 2) cost effectiveness, 3) efficiency, 4) flexibility, 5) predictability.

These are taken from the list of 11 globally endorsed performance areas. In today's world, on any given day-of-operations, the FAA will make tradeoffs among these criteria in designing an operational traffic flow management (TFM) strategy. For an example, an "aggressive" approach might yield very high expected capacity/throughput, but at the expense of delaying the times when final decisions on releasing flights are made, thus reducing predictability. A highly cost effective strategy might seek to reduce airborne delay but sacrifice capacity/throughput. The goal of this research is to allow the flight operator community to set service expectations and thus make tradeoff among these criteria directly, before actual TFM strategies and mechanisms are determined.

Figure 1 provides a high level view of the process to be investigated.

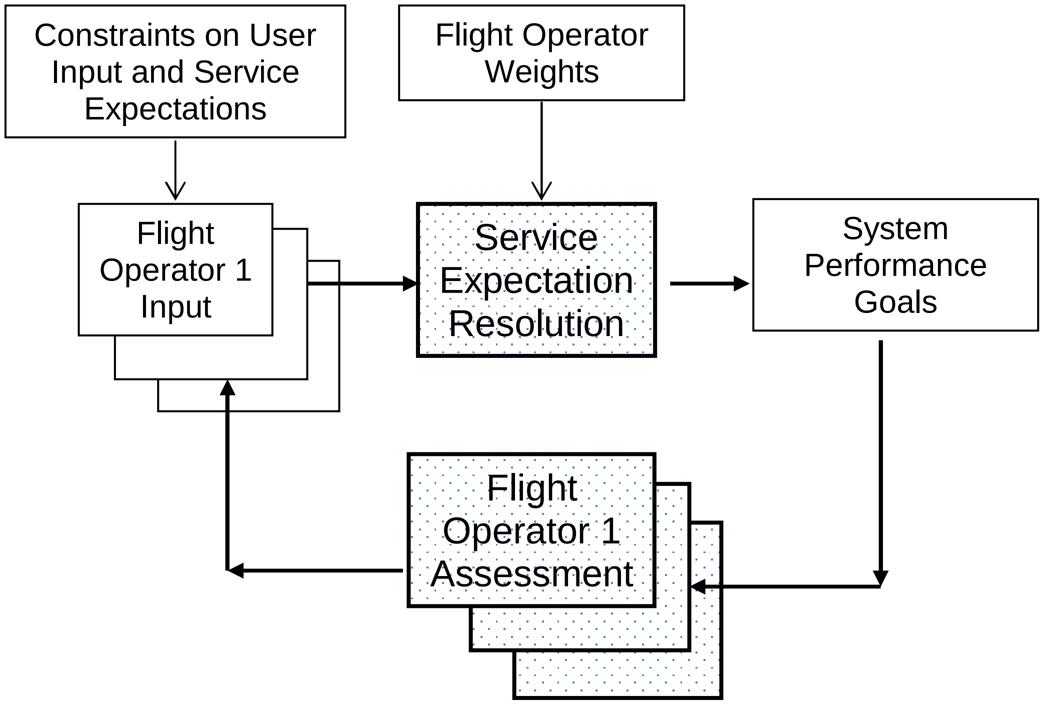


Figure 1: Architecture for Service Expectation Setting Process

The ultimate output of the process is a set of *system-wide performance goals* (upper right-hand box) that will be used by TFM specialists to set TFM strategies and to choose specific traffic management initiatives (TMI's). The process starts with *inputs from each flight operator* (set of boxes on upper left). These can be viewed as the service expectation desired by that flight operator; however, for reasons to be discussed later they may take on a different form. The *Service Expectation Resolution* process takes the inputs from all flight operators and produces the set of system-wide performance goals. Further, it provides individual *flight operator feedback*. This feedback should allow each flight operator to assess the impact of the system-wide performance goals on its operational performance. This is done by the *Flight Operator Assessment* process. This process will typically be proprietary to each flight operator. As part of the assessment the flight operator may determine appropriate adjustments to its input in order to influence the system-wide performance goals and, thereby, to improve its operational performance. It is anticipated that this feedback loop will be executed several times until a form of equilibrium is reached.

1.3 Possible Meta-Models for Structuring Process

The problem setting just outlined does not immediately fit in a standard analytic framework. Thus, we consider multiple paradigms for framing the problems and explore the strengths and weaknesses of each. Each of these suggests a different way for structuring and analyzing the problem. Some we explore in more detail than others.

Investment alternatives: A flight operator could view each performance criterion as an investment option. In order to achieve its overall performance goals the flight operator would spread its investment out among several options just as investment managers to balance expected return and risk. Flight operators would be allocated a type of artificial currency in proportion to their relative importance with respect to the ATFM challenge under consideration. At each iteration of the feedback loop illustrated in Figure 1, each flight operator would allocate its currency among the various performance criteria. After observing the overall performance of the current allocation, i.e. the system wide performance goals, the flight operators would potentially change their allocations to achieve a better result on the next iteration.

Voting: It is perhaps most natural to determine a single decision for a group of participants (the flight operators) using some type of voting mechanism. Under the setting we propose, a small set of candidate performance vectors would be proposed (a candidate performance vector is a "feasible" set of specific values for each performance criterion). Each flight operator would cast a vote on these performance vectors. Since there are many "candidates" in this case, we propose a mechanism (instant runoff voting) that provides for an effective way of determining a single winner from among many candidates. While such an approach does determine a result in a single iteration, it could be very valuable to allow the dynamic generation of additional candidate performance vectors (candidates) – this could be accomplished using multiple iteration of the feedback loop in Figure 1.

Game Theoretic Analysis: One can view the process illustrated in Figure 1 in game theoretic terms, most specifically as a multi-player non-cooperative game. In such a setting, each player submits a strategy (the flight operator inputs). The set of all strategies together with the game structure determines a payoff for each player (the cost or value each flight operator associates with the system wide performance goals determined by the ANSP). A fundamental question one asks about such a game is whether a Nash Equilibrium exists. That is, does there exist a set of strategies where no flight operator could unilaterally adjust its own strategy to produce a "better" outcome. A Nash Equilibrium can be viewed as a solution to the game in the sense that once all players arrive at their equilibrium strategy there is no incentive to deviate from them. Further analysis can be performed and other insights can be gained using game theoretic tools. These help to better understand the structure of the mechanism and aid in its design.

1.4 Summary of Research Results

Four general categories of research were carried out:

1) Overall problem modeling and structuring: all schools;
2) Game theoretic analysis: MIT

3) Approximate models for trading off performance metrics: UC Berkeley
4) Voting schemes: U of Maryland

Items 2) and 4) addressed the voting and game theoretic analysis topics discussed in the previous section. Item 1) included a preliminary analysis of the investment alternatives approach discussed in the previous section. Item 3) produced functional models of how performance metrics of interest change in response to changes in the parameters of traffic management initiatives. This work also produced models that quantified the manner in which one could trade off one performance metric with another. This work represents a key ingredient in the broader models needed to structure and evaluate the various approaches under consideration for the overall problem.

We now summarize our findings relative to the three approaches discussed in the previous section.

Investment alternatives: Under this approach one views a flight operator as seeking to maximize some value function (return on investment). To do this the flight operator seeks to invest in each of several options (the performance criteria). Here each flight operator would be allocated some kind of artificial currency to use to makes its investments. If a flight operator perceives that one criterion, say capacity, is too low then that flight operator would shift some of its investment away from some other performance criterion of lesser importance, say cost effectiveness, and reinvest that amount freed up in capacity. The process would iterate until no flight operator could find any shift in investment that would improves its value function.

Our preliminary analysis of this paradigm found several major challenges.

In a standard investment or bidding (combinatorial auction) model, there is an amount of each product available and an investor would seek through his or her bids to obtain a share of each of the available products. The overall value achieved by one investor would be determined by the amount of each product that that investor purchased. For the ATM problem we seek to model, the products (performance criteria) are not being split up. Rather, performance criteria that receive a high investment level would have a high individual value. The overall value achieved by one investor would be determined by the values of each performance criteria. That is, the outcome will be a single vector of performance criteria and each investor/flight operator will derive a different value based on their individual value functions. A major challenge in applying such a model is the so-called ***free rider problem***. If one investor was confident that another would invest heavily in say capacity, then that investor could focus its investments on other criteria knowing the capacity would be “taken care of” by others. This leads to “strategic behavior” on the part of the bidders/investors and generally unstable processes and poor outcomes. These problems can be viewed as arising from the fact that the investment model does not fit the problem we are trying to solve well.

In spite of the aforementioned challenges, it would be very valuable to obtain from flight operators some valuation of the relative importance of one criterion vs another. Thus, we hope to incorporate ideas of this type on top of a more stable approach.

Voting: We have developed a practical voting approach for addressing the problem of interest. The core of the process uses the instant runoff voting (IRV) method. IRV is appropriate in a context where there are several “candidates” running and a single winner is desired. In our case, flight operators vote on several alternative performance vectors and the process determines a single winning vector. While in theory the process could produce an answer in a single iteration we employ multiple iterations so that additional performance vectors can be added from one iteration to the next. These vectors could be provided by the ANSP or by any individual flight operator. While the overall process is well defined and can reliably produce a result, further research is required to refine the various steps through additional analysis and experimentation.

Game Theoretic Analysis:

We were able to model the problem of interest as a multiple-player non-cooperative game. Furthermore, in certain cases we were able to show that the game had a variety of desirable properties, in that players played the game truthfully and equilibrium solutions existed and could be obtained. On the other hand, the solutions obtained were not necessarily “desirable” in the sense that they tended to be “extreme” solutions in which one player got their most desirable option and another did not. Specifically, “compromise” solutions were not achieved. In one case, there was some advantage achieved by the use of randomized strategies and in which a flight operator chose among input strategies based on a probability distribution. We anticipate that such strategies will be a “tough sell” in practice. Our general conclusion from this analysis is that a practical approach should generate a large number of candidate performance vectors and then have a voting procedure choose among them. This is what has been proposed in the Section 3 of this report. It is also the case that the analytic framework we have developed should be useful in studying the voting mechanism.

2. Representation and Analysis as a Multi-player Non-Cooperative Game (authors: Antony Evans, Vikrant Vaze, Cynthia Barnhart)

2.1 Background

It is anticipated that the feedback loop in Figure 1 will be executed several times until a form of equilibrium is reached. While one view of the flight operator inputs is a set of desired service expectations, another view is a set of investments into each of the service areas. We can view the problem as one where each flight operator determines its investment level so that, relative to the investments by the other flight operators, it optimizes the system-wide service expectations. In fact, one could seek a type of Nash Equilibrium where no flight operator could unilaterally adjust its own investment levels to produce a “better” set of system performance expectations (here better is relative to the performance objectives of that flight operator). Such a game theoretical approach is illustrated in Figure 2.

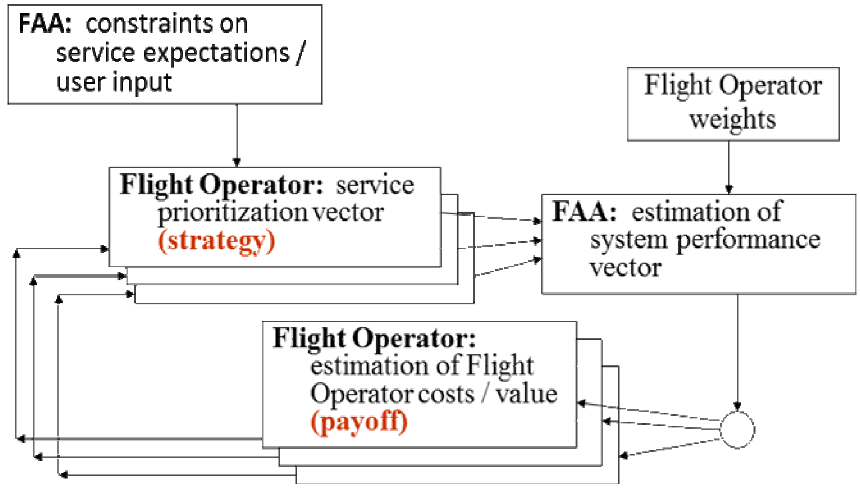


Figure 2: Viewing the service level setting process as a multi-player non-cooperative game.

Mathematically, this can be represented as follows:

- Flight Operators: $k=1,2,\ldots,K$
- System performance metrics, or Goals: $g=1,2,\ldots,5$
- FAA’s initial estimate of the system performance vector: $G'=(G_1,G_2,G_3,G_4,G_5)$
- Each flight operator would specify a modified performance vector:

  $I'_k = (I_{1k},I_{2k},I_{3k},I_{4k},I_{5k})$
- Characterize the valid inputs as changes to G': $\Delta'_k=(\Delta_1, \Delta_2,\Delta_3,\Delta_4,\Delta_5)$

  such that: $I'_k = G' + \Delta'_k$
- The set of feasible Δ'_k is defined by the feasible region $FEAS\_\Delta$:
  - Parameters: for each g: Δ^-_g , Δ^+_g , b_g
  - Constraints: $\Delta^-_g \le \Delta_{gk} \le \Delta^+_g$ for all g and k (1)

    $\Sigma_g\, b_g\, \Delta_{gk} = 0$ for all k (2)
- Performance goal vector of combined user preferences:

  $G^* = (G^*_1,G^*_2,G^*_3,G^*_4,G^*_5)$
  - Flight Operator weights: $(w_1, \ldots, w_K)$ with $\Sigma_k\, w_k = 1$ and $w_k \ge 0.$
  - Combine user preferences according to: $G^* = \Sigma_k\, w_k\, I'_k$
- Payoff to Flight Operator k: $P_k = \Sigma_g\, a_{gk}\, G^*_g$

In the initial model described above, a proposed overall architecture is as follows:

- Initiation: The FAA specifies an initial system performance goal vector (*G'*) for an initiative at a resource.

- Form of user input: Flight operators specify their preferred performance goal vectors (I'_k) by specifying the difference (Δ'_k) from FAA's initial estimates (*G'*). The validity of this input is then checked by the FAA to ensure that the specified differences (Δ'_k) belong are feasible (i.e., fall within the feasible set $FEAS\_\Delta$)

- Performance goal resolution process: The FAA determine the system-wide performance goal vector (G^*) by taking a weighted average of all user inputs. The weights may be proportional to the relative number of operations each flight operator has that are impacted by the initiative. The weighted average of valid user inputs (i.e., Δ'_k that fall within the feasible set) will always be valid (i.e., will also fall within the feasible set).

- Form of FAA feedback to flight operator: The FAA specifies the resultant performance goal vector (G^*) based on the previous round of inputs.

- Flight operator assessment and response: Based on the feedback, flight operators determine the best response that maximizes their own payoff, and submit a revised input.

- Convergence to final solution: The above process is repeated a set number of times or until convergence to a stable equilibrium.

To demonstrate this architecture, a number of simulations were run with randomly generated data. These simulations assume only **two** performance goals, and **two** flight operators. Truthful solutions are identified for each flight operator, which represent the performance goals that would maximize each flight operator's payoff independently. However, because of the nature of mechanism described above, it may be more profitable for flight operators to request performance goals that are not in fact their "truthful" solution! This is because the performance goal resolution process applies a linear combination of flight operator preferences: $G^* = \Sigma_k w_k I'_k$
Thus, if a flight operator knew what the other flight operator was requesting, it may shift its own request such that the linear combination of both requests coincided with its profit maximizing solution. This is called gaming. In the simulation results presented below, as "strategic" solution represents the result of such gaming by both flight operators.

1.4 Sample Simulation Results

A sample simulation result is presented below. In these results, linear flight operator payoff functions were randomly generated, and are shown in Figure 3. Separate payoff functions were generated for performance goal, for each flight operator.

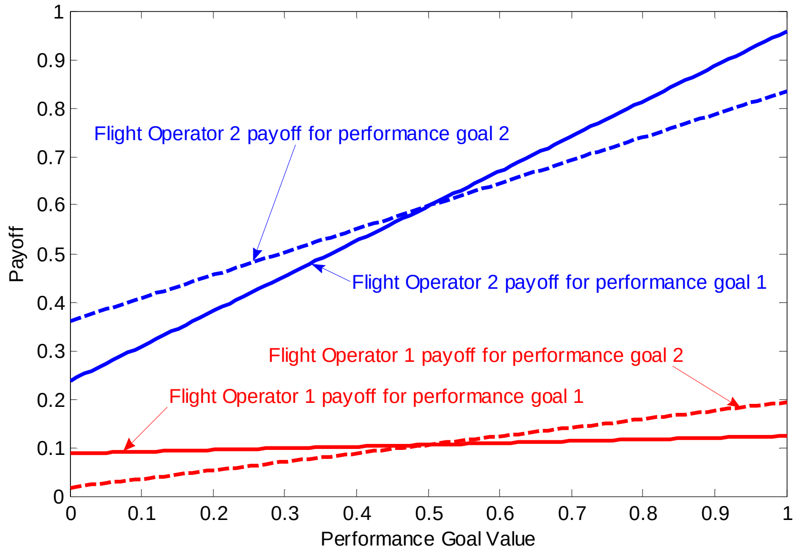


Figure 3. Sample linear payoff functions

A convex feasible set is also specified, the upper boundary of which is defined by the solid black line in Figure 4. Any combination of performance goals falling inside (to the left of) this boundary therefore represents a valid user input. By overlaying lines of constant payoff for each airline, we can identify their truthful solutions. If each line of constant payoff was shifted to the right, while remaining parallel to the lines shown, the flight operator's payoff would increase. Therefore, the truthful solution is the point at which the lines of constant payoff are tangent to the feasible region, as shown in Figure 4. These points represent the preferred combinations of the two performance goals for each flight operator, if they were assured that they would get what they asked for.

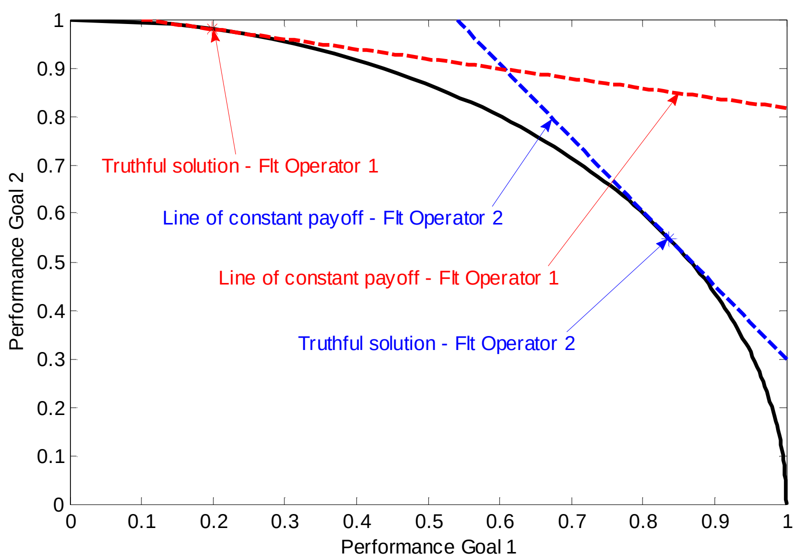


Figure 4. Truthful solutions

When we simulate the model framework described in Section 2.2 above, allowing the flight operators to strategically select their inputs according to what the other flight operator has requested in the previous round, the results are as shown in Figure 5.

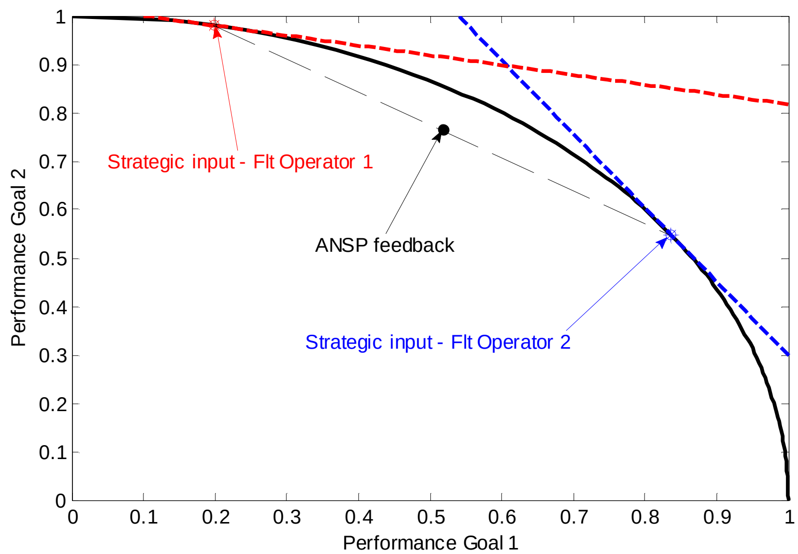


Figure 5. Solution of strategic game

As can clearly be seen by comparing the results of figures 4 and 5, the strategic solutions requested by the flight operators are the same as their truthful solutions. The result of the linear combination combining them is a new solution (labeled ANSP feedback in Figure 5), that falls inside the feasible region. This result is attractive because it suggests that there is no gaming, but is not system optimal, because it does not fall on the boundary of the feasible region. This latter point will be addressed in more detail later.

1.5 General Properties of Game

Using game theory, a number of statements can be made about a more general problem of a similar nature to that simulated above, as follows:
For the model described above, in which **flight operator payoffs are linear** and the **feasible region is convex**, it can be proved that, regardless of the problem data, number of flight operators, and performance goals:

1. There exists an equilibrium.

2. The equilibrium is devoid of any probabilistic decisions.

3. If each flight operator has a unique most preferred point at which they want to operate (i.e., the flight operators know what they want), then the equilibrium is unique.

4. The feedback process converges to this unique equilibrium point (in fact this will always happen in exactly one iteration).

5. The flight operators have **no** incentive to strategically submit different preferences than their real preferences. The equilibrium will therefore be truthful (i.e., there will be no gaming).

6. The system performance goal G^* will be an interior point, and therefore not pareto optimal.

1.6 Non-linear Payoff Functions

In reality, flight operator payoffs are unlikely to be linear for any of the performance criteria described in Section 2.1. Flight operator payoffs are often ‘concave increasing’ rather than linear. For example, beyond a certain threshold, decreases in capacity lead to faster than linear increases in passenger re-accommodation costs, reserve and delay crew costs, airline recovery costs, etc, as illustrated in Figure 6. This is also the case for other performance criteria.

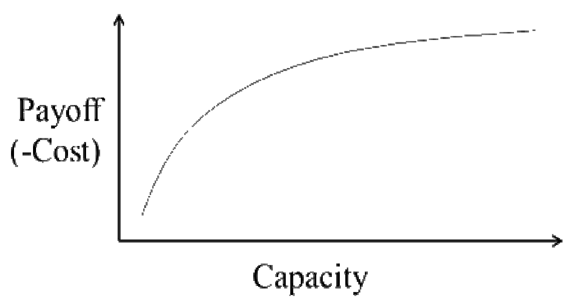


Figure 6. Concavity of Payoff Functions

This case can also be simulated: Instead of specifying linear flight operator payoff functions, concave increasing functions can be specified, as follows:

- Payoff to flight operator k: $P_k = \Sigma_g f_{gk}(G^*_g)$, where f_{gk} are concave increasing functions of G^*_g
  - For example, $f_{gk}(G^*_g) = a(G^*_g)^2 + b(G^*_g) + c$, with $a<0, b>0$

Everything else in the simulation is identical to that simulated above, including performance goal resolution process, which is still a linear combination of user performance inputs, i.e., $G^* = \Sigma_k w_k I'_k$

Again, the simulation is run with randomly generated data for two performance goals and two flight operators. Concave increasing flight operator payoff functions were randomly generated, and are shown in Figure 7.

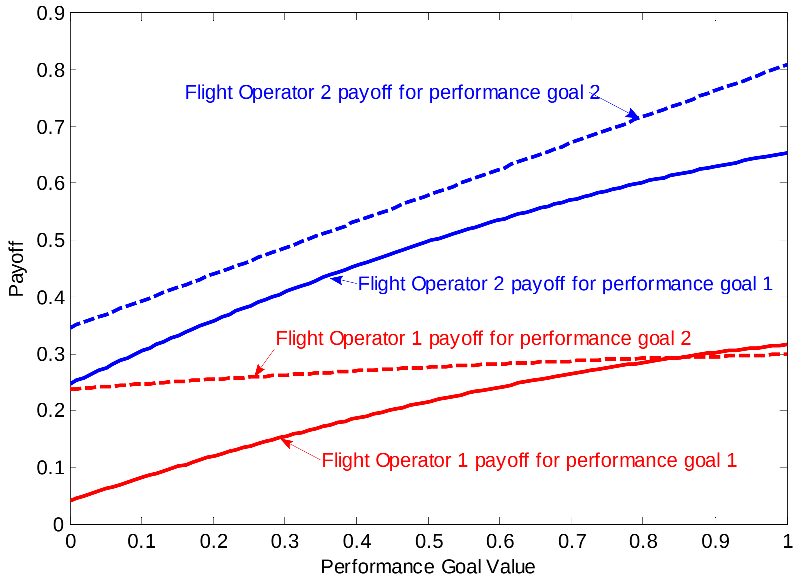


Figure 7. Sample concave increasing flight operator payoff functions

A convex feasible set, with lines of constant payoff for each airline (which are now convex, because of the concave payoff functions), and the corresponding truthful solution for each flight operators are shown in Figure 8. Again, these points represent the preferred combinations of the two performance goals for each flight operator, if they were assured that they would get what they asked for.

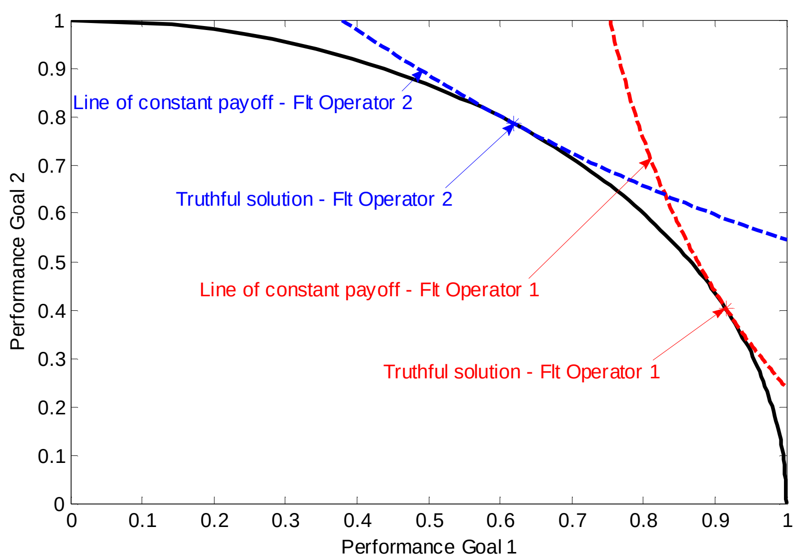


Figure 8. Truthful solutions, with concave payoff functions.

Again, we simulate the model framework described in Section 2.2 above, allowing the flight operators to strategically select their inputs according to what the other flight operator has requested in the previous round. The results are as shown in Figure 9.

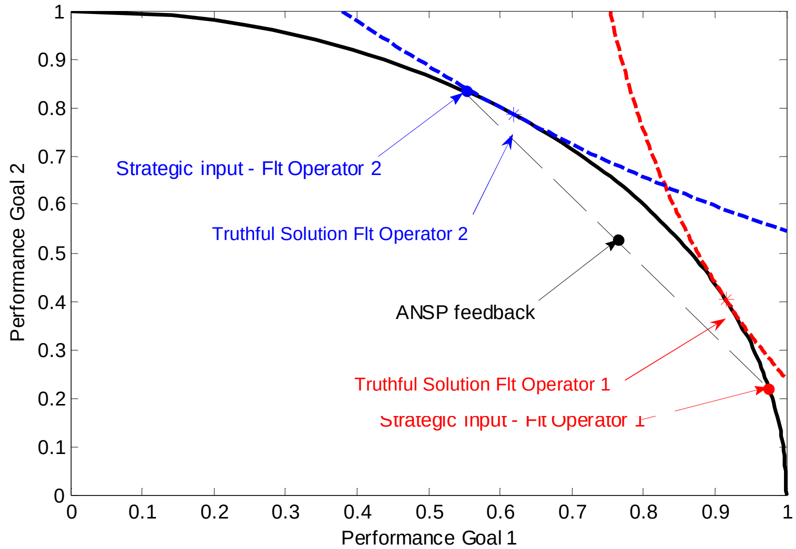


Figure 9. Solution of strategic game, with concave payoff functions

As can immediately be seen in Figure 9, the strategic solutions and truthful solutions of both flight operators no longer coincide. This is because of the linear combination used to generate the combined solution. Each flight operator can be seen to attempt to "pull" the solution away from its competition, in an attempt to shift the FAA's feedback closer to its truthful solution. Of course, both flight operators attempt to game the system in this way, so the result is that the final solution is still situated between the two flight operator preferences.

The result above suggests that, with more realistic concave payoff functions, flight operators are likely to attempt to game the system, and will not be truthful. In the simulation, the degree to which flight operators game is a function of concavity of their payoff functions, i.e., highly concave payoff functions lead to greater gaming.

1.7 General Properties of Game with Concave Payoff Functions

As in Section 2.3.2, we can use game theory to make a number of statements about the more general problem, as follows:

For the model described above, in which **flight operator payoffs are concave increasing** and the **feasible region is convex**, it can be proved that, regardless of the problem data, number of flight operators, and performance goals:

1. There exists an equilibrium.
2. The equilibrium is devoid of any probabilistic decisions.
3. Uniqueness **cannot** be proved.
4. Convergence **cannot** be proved.
5. An outcome need not be truthful (i.e., there may be gaming). It is dependent on the problem data.

6. The system performance goal G^* will also be an interior point, and therefore not pareto optimal.

Two of the properties in the list above are not preferred for the problem addressed. Firstly, a solution that is not pareto optimal is unlikely to be system optimal, in the sense that payoffs for all flight operators could still be increased (in a manner that maintained fairness). This suggests that the linear combination of user preferences may not be the best approach for combining the user preferences. Secondly, gaming can cause significant problems because the flight operators are not expressing their truthful preferences. In some cases, with minimal gaming, this may not be a problem. However, in other cases, where flight operators game to a larger extent, they may make extreme requests, which can confuse the situation, and make it appear that system solutions are significantly further from the flight operators real preferences than they actually are. It may also result in a system solution that is significantly more favorable for one flight operator than others, introducing issues to do with equity. An example of this is provided in the following section.

1.8 Alternatives to a Linear Combination of User Preferences

Two alternatives were considered to taking a linear combination of the flight operator preferences in order to address some of the problems described above. These are:

- To push the FAA feedback (G^*) out to the boundary of the feasible set.
- Randomly choosing one of the flight operator inputs (I'_k) for system performance goal vector (G^*). The probability of choosing each flight operator inputs (I'_k) may be based on percentage of operations affected by the initiative (w in the formulation described in Section 2.2).

Pushing the G out to the Boundary*

Using the same inputs as for the simulation described in Section 2.5, a simulation was run in which the FAA feedback (G^*) was taken as identical to the previous linear combination, but pushed out to the boundary of the feasible set. This ensures that the FAA feedback is always pareto optimal. The flight operator solutions (both truthful and strategic), in this case, are presented in Figure 10.

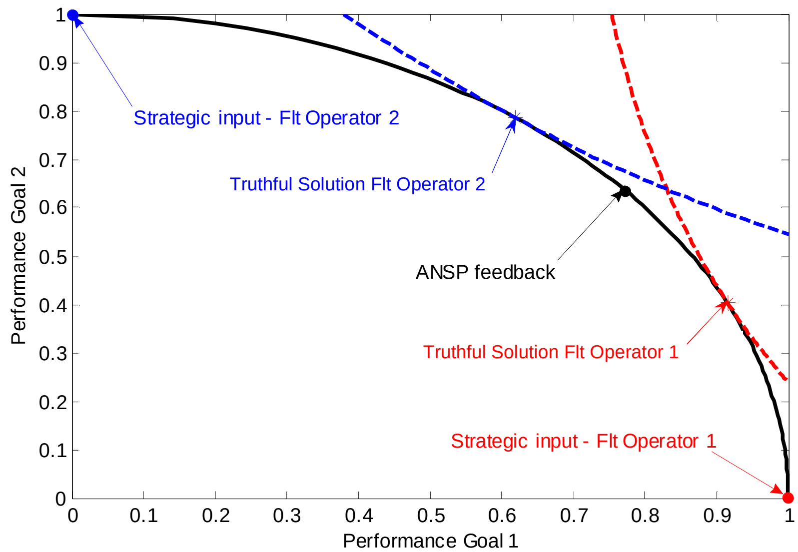


Figure 10. Truthful and Strategic Solution: Pushing *G** to Boundary

This result shows that, when the FAA feedback (G^*) is pushed out to the boundary of the feasible set, this feedback is pareto optimal (on the boundary of the feasible set), but there is extensive gaming by both flight operators. The game always converges with at least one flight operator requesting one of the corner points of the feasible set. The flight operator that reaches a corner point first is at a significant disadvantage because it is no longer able to "pull" the system solution towards its true preference, while the other flight operator is still able to do so. The result is that the final system solution is closer to the second flight operator's true preference. In the case where only one flight operator reaches a corner point, the final system solution would equal the other flight operator's true preference. This approach therefore has significant issues with equity.

*Random Choice of one of the Flight Operator Inputs for G**

Again, using the same inputs as for the simulation described in Section 2.5, a simulation was run in which one of the flight operator inputs (I'_k) was randomly chosen for the system performance goal vector (G^*). The probability of choosing each flight operator input (I'_k) was based on the percentage of operations affected by the initiative (i.e., w in the formulation described in Section 2.2). The flight operator solutions (both truthful and strategic), in this case, are presented in Figure 11.

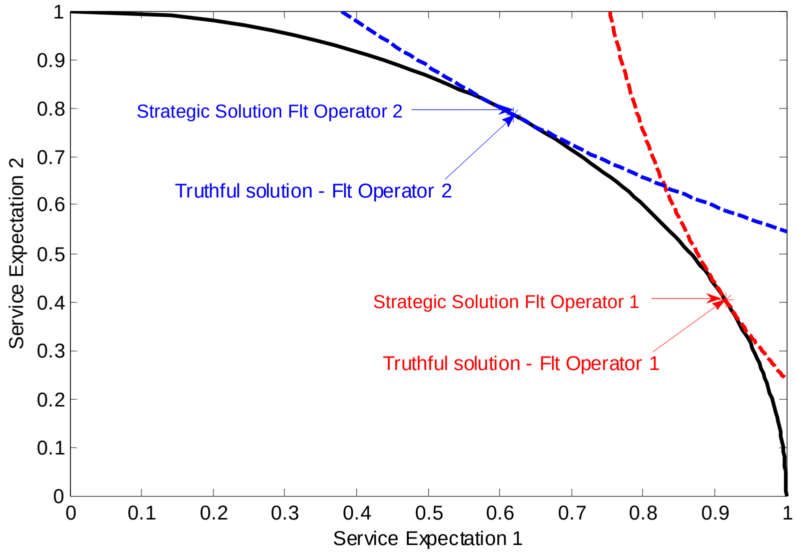


Figure 11. Truthful and Strategic Solution: Random choice of $G*$

This result shows that, when one of the flight operator inputs (I'_k) is randomly chosen for the system performance goal vector ($G*$), the strategic and truthful solutions coincide. The reason for this is that flight operators are not incentivized to game in any way because, if their solution is not chosen, their input does not affect the chosen solution in any way. Thus, they are incentivized to submit truthful solutions, and this is regardless of how the probabilities are defined to randomly choose one of the flight operator inputs. The system goal vector (G*) is also pareto optimal (on the boundary of the feasible set). The disadvantage of this solution is that it does not account for that fact that the payoff of any chosen solution could vary significantly across flight operators. The chosen solution may therefore have highly disproportionate impacts of each flight operator. A solution may exist that has lowest overall impact on all flight operators. This is dealt with in greater detail in the following section.

With the random choice of a flight operator input (I'_k) for the system performance goal vector ($G*$), all the properties described in Section 2.4 apply, as well as the system performance goal being pareto optimal.

1.9 Further Considerations

There are a number of further considerations that must be considered when applying the framework described in this document to the real world.

Firstly, for any given initiative, it may be preferable for the FAA to specify a set of discrete combinations of performance goals, or options, for which flight operators must indicate their preferences, instead of allowing a continuous spectrum of combinations of performance goals. This allows consideration of a framework that allows voting on individual options. Such a voting framework may also be designed in such a way as to induce flight operators to indicate their relative costs or payoffs across different options.

This would provide information about whether a flight operator was almost indifferent about two options, or if it valued one significantly over another. This information would be useful because it would allow the FAA to select the option that maximized the (weighted) payoff across all flight operators.

Secondly, flight operators serve networks that may be impacted by more than one initiative in any given day. Flight operators may therefore have specific priorities for which initiatives they would like to have tailored to their preferences, and which initiatives they are indifferent about. They may also have preferences across multiple initiatives, e.g., they may only want one option in one initiative, if they can get another option in another initiative (this may particularly be the case for simultaneous initiatives in different geographical areas), but not otherwise. Again, a framework that allows voting may allow flight operators to indicate which facilities are most important to them, or which combinations of options they prefer.

A framework that allows voting is introduced in the following section.

2. Voting-based Approach **(authors: Prem Swaroop, Michael O. Ball)**

2.1 Background

The Service Expectations Problem (SEP) aims to collect input from competing flight operators in the form of target NAS-wide performance metrics. A possible starting point would be for the ANSP to announce the likely deviations of five performance metrics from their normal levels. These deviations can be represented as a vector with five elements, each corresponding to a specific metric. Assuming the performance criteria of: 1) capacity, 2) cost effectiveness, 3) efficiency, 4) flexibility, 5) predictability, an example candidate vector could be {0.9, 0.8, 0.95, 0.76, 0.85}, representing a 10% reduction in capacity, 20% reduction in cost effectiveness, 5% reduction in efficiency, and so on.

The flight operators then return with their feedback and preferences, which are then used to update the ANSP's target vector. This process is expected to be iterative: the flight operators submit their preferences, the ANSP updates and announces the winning target vector; the airlines then fine-tune and re-submit their preferences, and so on, until an acceptable vector is determined.

The SEP has elements of multi-attribute, consensus-based decision making, in which a common burden is shared among multiple parties. Issues of fairness and equity would arise in such settings, more so because the ANSP is a public entity with an explicit goal to be equitable towards all the flight operators. For this procedure to be acceptable in the long-term, it is vitally important to ensure that all the players perceive the mechanism to be fair to all of them. In this section, we propose a voting mechanism to handle various aspects of the SEP, and present the research agenda.

2.2 Consensus-based Fair Multi-Voting Mechanism: Introduction

We propose the following conditions be met by the mechanism:

single winner determination. The mechanism should result in a single winning vector.
confidentiality. The private information requirements from the flight operators should be minimal.
practicality. The procedure should be easy to administer, and not involve time-consuming information gathering and / or processing steps.
consensus-building. The winning vector should have "maximum acceptability" among the airlines.
equitable. The mechanism should be perceived to be fair to all parties involved from the outset.
strategy-proof. As far as possible, the mechanism should discourage gaming, and encourage truth-telling behavior.

We propose a multi-round voting mechanism that attempts to meet the above objectives as we describe below. An instant run-off voting (IRV) is proposed to determine a single winner among a pool of candidate vectors. IRV is known to be a practical method that requires only preference rankings to be elicited from the voters, which is a relatively simple task. Instead of asking for detailed payoff functions that may infringe upon airlines' private information, it asks for preference ordering for the top few candidates alone, further reducing information load. Hence, we meet conditions (i)-(iii) by use of IRV.

Before proceeding, we present a simple example of IRV in the following table. 5 airlines express their preference ordering as given for the 4 candidate vectors. As the first iteration finds no absolute majority winner, the Vector 2 being least-preferred is eliminated in this iteration. Next, the airline preference ordering is updated for iteration 2. Note for example, if an airline had ranked Vector 1, as its first choice then, upon elimination of vector 2, the second choice of that airline would become its first choice, its third choice would become its second, etc. On the next iteration, Vector 3 is eliminated. In the final iteration, Vector 1 is found to have absolute majority, and is declared winner.

Iteration		Vector 1	Vector 2	Vector 3	Vector 4
1	Airline 1	1	3	4	2
1	Airline 2	2	4	3	1
1	Airline 3	1	3	2	4
1	Airline 4	3	2	1	4
1	Airline 5	4	2	3	1
1	*Most preferred by*	*2*	*0*	*1*	*2*
1	*%age of total*	*40%*	*0%*	*20%*	*40%*
1	*Decision*		*Eliminate*		
2	Airline 1	1	-	3	2
2	Airline 2	2	-	3	1
2	Airline 3	1	-	2	3
2	Airline 4	2	-	1	3
2	Airline 5	3	-	2	1
2	*Most preferred by*	*2*	-	*1*	*2*
2	*%age of total*	*40%*	-	*20%*	*40%*
2	*Decision*		-	*Eliminate*	
3	Airline 1	1	-	-	2
3	Airline 2	2	-	-	1
3	Airline 3	1	-	-	2
3	Airline 4	1	-	-	2
3	Airline 5	2	-	-	1
3	*Most preferred by*	*3*	-	-	*2*
3	*%age of total*	*60%*	-	-	*40%*
3	*Decision*	*Winner*	-	-	*Eliminate*

An IRV implementation requires the specification of further details, like tie-breaking rules, whether multiple least-preferred vectors are simultaneously eliminated, whether all candidates must be necessarily ranked by all voters, etc. The simple example shown glosses over these for brevity.

In order to address conditions (iv)-(vi), we need to buttress IRV. Instead of an absolute majority, we propose to build a consensus: a candidate vector would need a larger than absolute majority to be declared a winner. In addition, for reasons discussed in Section 1,we propose to use a weighting scheme to reflect the differing importance of various flight operators. Further, unlike IRV, we propose to have more than just one round. And, unlike traditional voting schemes, we propose to add more candidate vectors in each round, instead of restricting the choice set to only the initial set of proposed vectors. Finally, we wish to study the entire mechanism through simulation and game-theoretic models to verify its susceptibility to strategic voting and nomination.

We now give an example of a weighted IRV in the following table. The weights emphasize the heterogeneity among the airlines (and other stakeholders), and let them have a proportional representation in the decision-making process. A more detailed discussion of weights follows later.

An absolute majority is again not found in the first iteration, which eliminates the least-preferred Vector 2. The second iteration has same outcome as the first, as Vector 2 was not the most preferred candidate for any airline. Iteration 2 eliminates the Vector 3 as it has the minimum support; rankings are then updated for iteration 3. Iteration 3 now finds Vector 1 to have an absolute majority.

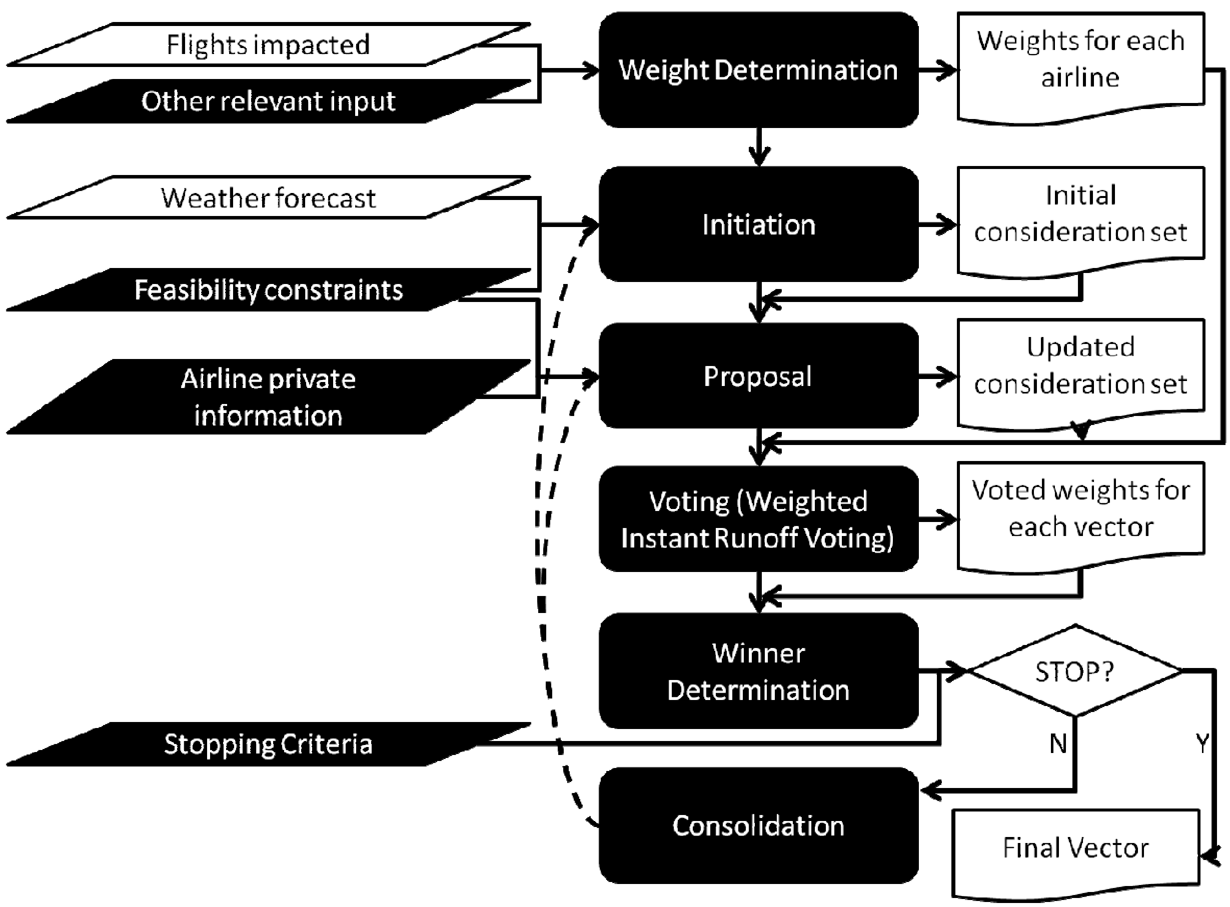
Iteration		Vector 1	Vector 2	Vector 3	Vector 4	Airline Weight
1	Airline 1	1	3	4	2	6
1	Airline 2	2	4	3	1	6
1	Airline 3	1	3	2	4	3
1	Airline 4	3	2	1	4	4
1	Airline 5	4	2	3	1	4
1	*Weighted top preference*	*9*	*0*	*4*	*10*	*23*
1	*%age of total weight*	*39%*	*0%*	*17%*	*43%*	
1	*Decision*		*Eliminate*			
2	Airline 1	1	-	3	2	6
2	Airline 2	2	-	3	1	6
2	Airline 3	1	-	2	3	3
2	Airline 4	2	-	1	3	4
2	Airline 5	3	-	2	1	4
2	*Weighted top preference*	*9*	-	*4*	*10*	*23*
2	*%age of total weight*	*39%*	-	*17%*	*43%*	
2	*Decision*		-	*Eliminate*		
3	Airline 1	1	-	-	2	6
3	Airline 2	2	-	-	1	6
3	Airline 3	1	-	-	2	3
3	Airline 4	1	-	-	2	4
3	Airline 5	2	-	-	1	4
3	*Weighted top preference*	*13*	-	-	*10*	*23*
3	*%age of total weight*	*57%*	-	-	*43%*	
3	*Decision*	*Winner*	-	-	*Eliminate*	

Like any voting procedure, IRV is known to suffer from shortcomings. Firstly, a ``Condorcet" winner is not guaranteed to be an IRV winner, that is, it is possible in an IRV that a candidate that would win against every other candidate in pairwise contests may not be declared an IRV winner. Furthermore, IRV does not meet independence of irrelevant alternatives (IIA) criterion, that is, an IRV winner may be different if a candidate who cannot be a winner is included as an option. Thus, IRV is prone to strategic nomination. Thirdly, according to Gibbard-Satterthwaite theorem, no voting system is entirely immune to strategic voting unless it is dictatorial or incorporates an element of chance. Fourthly, IRV is not a proportional voting system; proportionality being an established measure of fairness, larger airlines may have an advantage if IRV alone were to be used.

The above shortcomings serve to highlight our case of buttressing the simple IRV with the proposed enhancements in order to meet conditions (iv)-(vi). While adding more rounds seems counter-productive -- IRV after-all eliminates the need for multiple rounds -- it is necessitated due to our proposal to add more candidates. As the simple example above indicates, Vector 1 is deemed winner by a rather thin margin in the final iteration. If a consensus-building approach were to be used with a requirement of 2/3'rd majority for the winner, then this round alone would not have produced a satisfactory winner. Moreover, the initial rankings show that it was indeed less preferred than Vector 4, and would have lost to it in a pairwise contest.

We propose to (a) allow flight operators to propose their own candidate vectors, and (b) replace the running candidate vectors with new candidate vectors that may have a better chance of being a consensus winner. In order to make new proposals, we shall be using the data on previous rank-orderings, and our understanding of the feasible vectors. This obviously would require defining clear stopping criteria.

With the above introduction, we now present the mechanism details.



2.3 Consensus-based Fair Multi-Voting Mechanism: Details

Please refer to the accompanying flow-chart for a schematic view of our proposal. We explain the various processes in the following subsections. The "black-boxes" depict the proposed research agenda items.

3.3.1 Weight Determination.

The ANSP initializes the voting process by first allocating weights to the airlines. This assignment procedure will be based on agreed-upon rules, and must be perceived to be fair by the flight operators. Research would be required to determine equitable procedures.

One way is to assign the weights proportional to the number of flights impacted due to the weather conditions on the given day of operations. However, this may treat smaller airlines unfairly, as they could effectively become irrelevant. This drawback could be mitigated by using a transformation with a decreasing rate of increase, e.g. a square-root, or logarithm function (The ANSP would then take the ceiling of the fractional numbers). Yet another way could be to take a monotonically increasing step-functional form of the number of flights impacted: an airline is assigned weight of say w_1 if number of its flights impacts is within say $1 \ldots U_1$, it gets weight of w_2 ($> w_1$) if its flights impacted is within $(U_1 + 1) \ldots U_2$, and so on.

An altogether different procedure may incorporate randomization. Flight operators with more flights impacted than say U would get an allocation as above; the smaller flight operators would be grouped together, and one (or more) of them would be chosen at random and assigned the weight for all smaller airlines' flights impacted. Hence, as the procedure gets repeated in practice, all the smaller airlines would get a significant say in the decision every once in a while.

At the outset, it only seems fair that the number of flights impacted be taken as the sole criterion for determining weights for the decision-making. This metric has a direct relevance to the problem at hand, is based on publicly available data, and has an element of randomization as well (instances of bad weather are unlikely to have exactly the same composition of number of flights impacted on any two occasions). However, the weather conditions typically play out somewhat similarly in given geographies, and the overall fleet composition for airlines may be expected to be stable in the geography.

For example, consider Northeast region of the U.S. in the winter season. Flow Constrained Areas may be similar all over the winter whenever bad weather hits as the snow-storms move northward and eastward from the gulf region. Also, the broad mix of flights operated by the different airlines can be expected to be broadly stable for the season. Hence, while the exact weights may be different over multiple instances, the dominance or otherwise of particular airlines may remain more-or-less the same. It seems appropriate therefore to consider different approaches and allocation formulas that may incorporate more criteria into the weight determination process.

Research should be directed at the overall impact of weight allocation process on the decision-making. While the ex-ante process must be deemed fair in itself, it would further boost confidence in the proposal if it could also be seen as being fair ex-post.

3.3.2 Initiation.

The ANSP next introduces a set of initial performance vectors that are all individually feasible. In practical terms, these may correspond to various ATFM alternatives available for the given day of operations and weather. The initial set may be a null set, implying that the ANSP wishes to let the airlines propose the performance vectors. The vectors in the initial set are required to be feasible, and constitute the *consideration set*.

3.3.3 Proposal.

Each airline may propose new feasible vector(s) along with its preference ordering across the entire consideration set. Each new vector is screened: it must be feasible, and it must have a minimum rank given by the airline (to discourage cheap talk). Only the screened vectors are then added into the consideration set.
Again, the models described in Section 4 may be useful for feasibility verification. The research described in Section 2 may be useful in determining the individual airline proposals.

3.3.4 Voting.

A weighted IRV is now conducted, and the final results across all the candidate vectors are shared with all the flight operators, without divulging the individual airlines' rankings.

3.3.5 Winner Determination / Stopping Criterion.

A pre-specified stopping criterion would be used to adjudge whether the voting rounds should be closed and a winner vector be announced. An example could be simple majority (over 50% weight). A consensus-based stopping criterion may look for higher agreement, say 80%. Research will be required in establishing a good stopping criterion.
IRV assumes proportionality in preferences over rankings. An flight operator has to necessarily pronounce a vector to be of a lower rank despite it being indifferently favoring two (or more) vectors. A more nuanced preference elicitation may be effected by asking flight operators to allocate their assigned weights to each candidate vector. This would also allow a flight operator to give similar weights to similarly preferred vectors. As it requires additional effort from the flight operators, this procedure may be proposed for the top few winning candidates alone, instead of the entire consideration set.

Establishing a winner fairly may also require agreement by a minimum number of flight operators -- small or big, or a mix of the two. In other words, it may be perceived to be more equitable to have, say, at least 50% of the operators to agree on a vector before declaring it to be a winner. This would again come with additional complexity in winner-determination procedure. Such rule-based procedures are common in voting literature, and research should be done to establish a good winner-determination model.

A simple stopping criterion may be required that ensures that at least a minimum number of rounds are conducted before a final winning vector is announced. This would ensure that initial perceptions alone do not dominate, and a chance is given to elicit preferences and form a consensus. Another simple stopping criterion may specify a maximum number of rounds in the interest of time.

3.3.6 Consolidation.

The stopping criterion would specify whether a consolidation of performance vector is required to be done by the ANSP. The preferences elicited in the previous rounds could be exploited by the ANSP in this phase. The consolidation may involve either selecting a few vectors from subsets of “similar" candidates and eliminating the rest, and / or proposing new candidate vectors in place of a subset of ”similar" vectors that may have the maximum acceptability given the votes and the subset.

Broadly, the SEP feasible region would be discrete-continuous. ATFM TMIs themselves form a finite and discrete set, and each TMI has a possibly continuous range of feasible parameters. Each feasible vector then has a potential for votes (or weights) from each airline; and hence a likelihood of being accepted as a winning candidate. SEP thus seeks to establish a mechanism to efficiently search the feasible space for a vector that has maximum acceptability.

This assumes that the ANSP has no particular objective function that it wishes to optimizes, but is only interested in basic feasibility to ensure safety in operations. In a more general setting, the ANSP may wish to optimize on one (or more) objective function(s), and thus may influence the likelihood of acceptance directly.

This procedure would be conducted by the ANSP, and may involve a procedure akin to column-generation. Research will be required to establish fair and useful ways to exploit the preference data. Weight-based procedure as highlighted above may be used for this consolidation phase.

3.3.7 Repeat.

The consideration set would now be updated, and process repeated either from ”Proposal" phase or just “Voting" phase. Reverting to a Proposal round may be fair to the flight operators, especially if the ANSP makes a mistake in its assumptions about the acceptability of the consolidated vector(s). However, it may prolong the entire proceeding, and may be prone to filibustering, wherein an airline may continually propose newer vectors instead of acting rationally. Disincentives toward this may be implemented by allowing a fixed number of proposals to each airline; however this would require some judgment from the ANSP.

2.4 Conclusion

The history of voting has been fraught with paradoxes and impossibility results. Research will be required to establish an overall mechanism that would be fair, practical, and leads to a useful outcome in most of the likely scenarios that may arise. Research may incorporate elements of voting theory, game theory, information and decision analysis, statistical likelihood methods, and semi-continuous optimization, including mixed integer programming.

4. Performance Metrics and Trade-offs in Ground Delay Programs **(authors: Yi Liu, Mark Hansen)**

4.1 Introduction

Before we design a distributed mechanism whereby flight operator community can collectively express its performance goals, it should be clear that what the performance criteria should be. In this chapter, we will introduce our performance metrics for capacity, efficiency and predictability and show how these metrics may be traded off in the design of Ground Delay Programs under capacity uncertainty. All the metrics are constructed such that the values are between 0 and 1. The GDP has two decision variables: the duration and the scope. We focus on the effect of the duration of a GDP on the performance metrics. The impact of GDP modification in response to updated information is ignored, but has been considered in our ongoing work.
The remainder of this chapter is organized as follows. The GDP model for this analysis is set up in 4.2. Then we describe our performance metrics and discuss how they will change with the duration of the GDP in 4.3. A realization of the results has been presented in 4.4. And finally the analysis is concluded in 4.5

4.2 Model Specification

The situation is examined at a single airport. We assume the scheduled arrival demand rate is a constant, λ (flight/hour). For a normal day, the capacity is at a constant high level C_H. When the GDP is initiated on a bad day, the capacity will decrease to a constant low level C_L. The duration of the GDP is assumed to be T. Or say we assume the capacity of this airport will go up to the normal level after T. Due to errors in predication, in the real case the capacity may increase at time τ, which can be different from the assumed time T, as we can see from Figure 1.

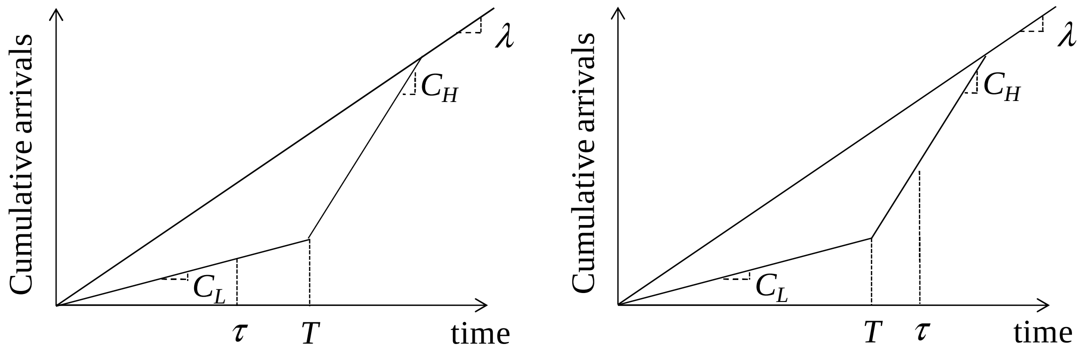


Figure 1: Queueing Diagrams of Possible Scenarios

When the GDP is initiated, T is set but τ is unknown. We assume τ has a uniform distribution between t_{min} and t_{max}. Conceptually, if T is set close to t_{min}, τ is very likely to be larger than T. In this case, capacity will be more fully utilized but there will be less predictability. On the contrary, if T is set close to t_{max}, then τ has a big chance to be smaller than T. As a result, it is very possible that the capacity will be underutilized and unnecessary delay will be resulted. However, delay is highly predictable.

4.3 Performance Metrics

Metrics have been constructed for three performance criteria: capacity, predictability and efficiency.

4.3.1 Capacity Utilization

Once the GDP is initiated, flights will be delayed and queue will build up until the capacity goes back to the normal level. After that, queue starts to vanish. N_R is denoted as the realized throughput from the beginning of the GDP until when there is no more delay. It should be noticed that the realized throughput will be smaller than the throughput that would be possible if the capacity actually goes up before the assumed time. Define N_I as the ideal throughput that would have been possible through this time if perfect information is available. The metric of capacity utilization, α_c, is then constructed as the ratio of N_R to N_I. When τ is less than T, the capacity will be underutilized for a period. The situation is illustrated with Figure 2. As a result, α_c is less than 1.

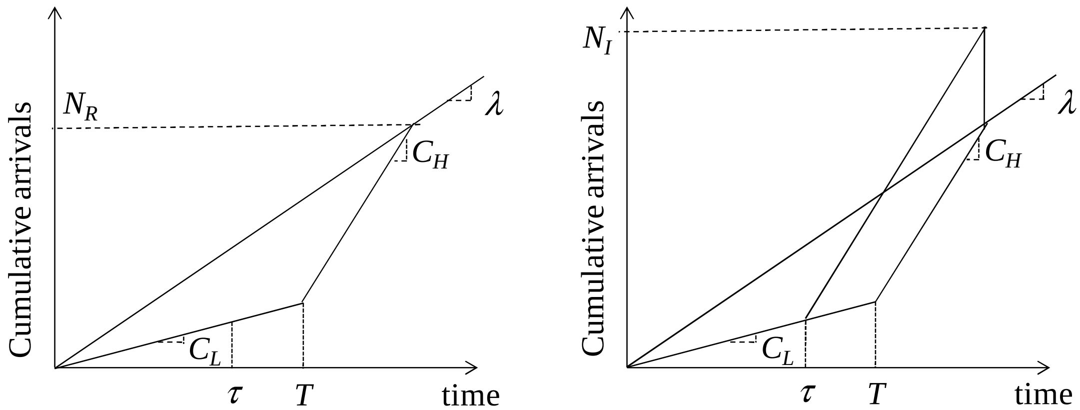


Figure 2: Values of the Realized Throughput and the Ideal Throughput when τ is less than T

Oppositely if the capacity increases after T in the real situation, then the capacity has been overestimated. We may need to further extend the GDP or airborne delay will occur. In either case, the realized throughput will be the same as the ideal throughput because all the available capacity is utilized and nothing can be improved even perfect information can be obtained at the beginning. So α_c is equal to 1.

With consideration to both situations, we can get the following equations for the capacity metric as a function of τ:

$$\alpha_c(\tau) = \begin{cases} \dfrac{\lambda T}{\lambda\tau + C_H(T-\tau)}, if\ \tau < T \\ 1, if\ \tau > T \end{cases}$$

To evaluate how the duration of the GDP will impact on the metric, we must take expectation of it over τ, which is showed below:

$$E[\alpha_c] = \frac{t_{max} - T}{t_{max} - t_{min}} - \frac{\lambda T}{(C_H - \lambda)(t_{max} - t_{min})} \cdot log\left(\frac{\lambda T}{C_H T - (C_H - \lambda)t_{min}}\right)$$

A realization in Section 2.3 will show how capacity will change with T based on this formula.

4.3.2 Predictability Ratio

Predictability is assumed to be equivalent to a measure of delay variability. When the duration of the GDP is determined, we may calculate the total flight delay. We call it D_P, which is the flight delay as planned at the beginning of the GDP. We further define the total realized delay as D_R and the predictability metric α_P as the ratio of D_P to D_R. When the capacity recovers before the assumed time T, since higher capacity is available than assumed and no modification to the program is considered, the planned delay will be the same as the realized delay and the metric value is 1. If the normal capacity is not achievable at time T, as shown in Figure 3, the planned delay will be smaller than the realized delay and α_P is less than 1. It should be noticed that when the capacity is not enough to satisfy the scheduled flights, the GDP may get extension or airborne delay may occur. The cost of the additional delay will be more expensive than the planned delay.

This difference is not reflected in the metric for predictability but will be considered in the metric for efficiency.

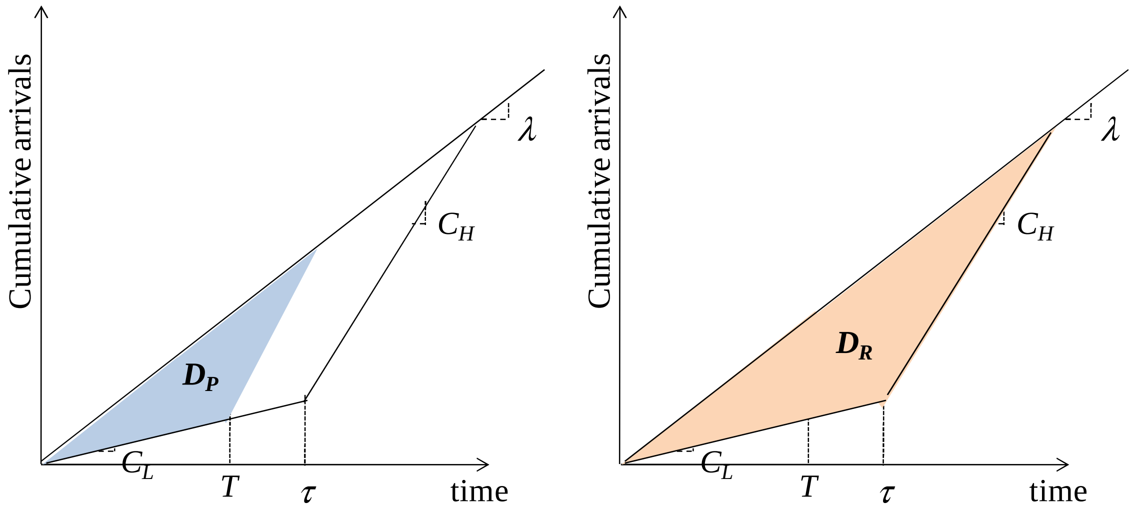


Figure 3: Planned Delay and Realized Delay when τ is larger than T

The predictability metric can be expressed as a function of τ:

$$\alpha_p(\tau) = \begin{cases} \frac{T^2}{\tau^2}, if\ \tau > T \\ 1, if\ \tau < T \end{cases}$$

Integrating it over τ, we may get the expectation of the predictability as a function of our decision variable T:

$$E[\alpha_p] = \frac{1}{t_{max} - t_{min}} \cdot \left[-\frac{T^2}{t_{max}} + 2T - t_{min} \right]$$

$$= \frac{1}{t_{max} - t_{min}} \left[-\frac{(T - t_{max})^2}{t_{max}} + t_{max} - t_{min} \right]$$

The feasible region for T is between t_{min} and t_{max}, so predictability is expected to increase with T.

4.3.3 Efficiency

As mentioned previously, the cost of the additional delay could be more expensive than the planned ground delay when the capacity goes up later than it is supposed to be. We assume the extra delay is twice as expensive as the ground delay. If perfect information is available when decision is made, then all the delay defined by τ will be ground delay. The cost of this delay is denoted as C_I. The total realized cost is defined as C_R, which consists of ground delay and additional delay when the capacity actually increases after the assumed time, as shown in Figure 4.

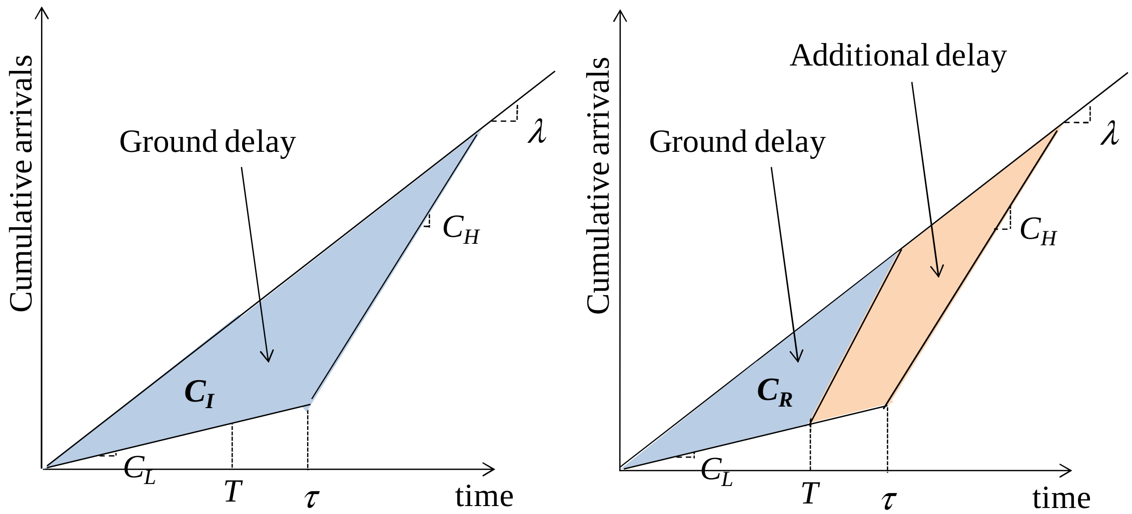


Figure 4: Ideal Delay and Realized Delay when τ is larger than T

The efficiency metric α_e then is defined as the ratio of the cost of ideal delay to that of realized delay and may be written as:

$$\alpha_e(\tau) = \begin{cases} \dfrac{\tau^2}{2\tau^2 - T^2}, if\ \tau > T \\ \dfrac{\tau^2}{T^2}, if\ \tau < T \end{cases}$$

Take the expectation and we can get

$$E[\alpha_e] = \frac{d + \frac{T}{4\sqrt{2}}e}{t_{max} - t_{min}}$$

where, $d = \frac{T^3 - t_{min}^3}{3T^2} + \frac{1}{2}(t_{max} - T)$ and $e = log\left(\frac{2t_{max} - \sqrt{2}T}{2t_{max} + \sqrt{2}T}\right) - log(\frac{2-\sqrt{2}}{2+\sqrt{2}})$.

4.4 Realization

The performance metrics are defined for capacity, predictability and efficiency in the previous section. Now, we will illustrate how values of these metrics will vary with the duration of the GDP with an example. The parameter set of the realization is summarized in Table 1.

Table 1: Parameter Set of Realization of the Ground Delay Program

Parameters	Notations	Values	Units
Demand arrival rate	λ	45	Flights per hour
Low capacity	C_L	30	Flights per hour
High (Normal) capacity	C_H	60	Flights per hour
Lower bound for τ	t_{min}	2	Hour
Upper bound for τ	t_{max}	4	Hour

The simulation results are shown in Figures 5 to 7. In Figure 5, it shows that with increasing T, capacity utilization will decrease but predictability ratio will increase. And predictability ratio is more sensitive to the increase. When the duration of the GDP increases from 2 hours to 4 hours, value of the capacity metric decreases only by 7% but the predictability ratio is double. Airlines may value capacity and predictability differently. If different unit costs are assigned to them, the sensitivity may change but the trend of change will not change. It means if only these two criteria are under consideration, there is always an optimum T which will maximize the output. Different from the metrics of capacity and predictability, the efficiency metric does not change with T monotonously, as shown in Figure 6. As T increases, value of efficiency metric first increases because of reduced chance of airborne delay. After a certain point, efficiency decreases with T because it has bigger chance that the capacity actually goes up before the assumed time and realized ground delay is much larger than it could be if we had perfect information at the beginning of the GDP. Variation of efficiency with capacity utilization in Figure 7 follows a similar pattern. But T decreases from the left to right in Figure 7.

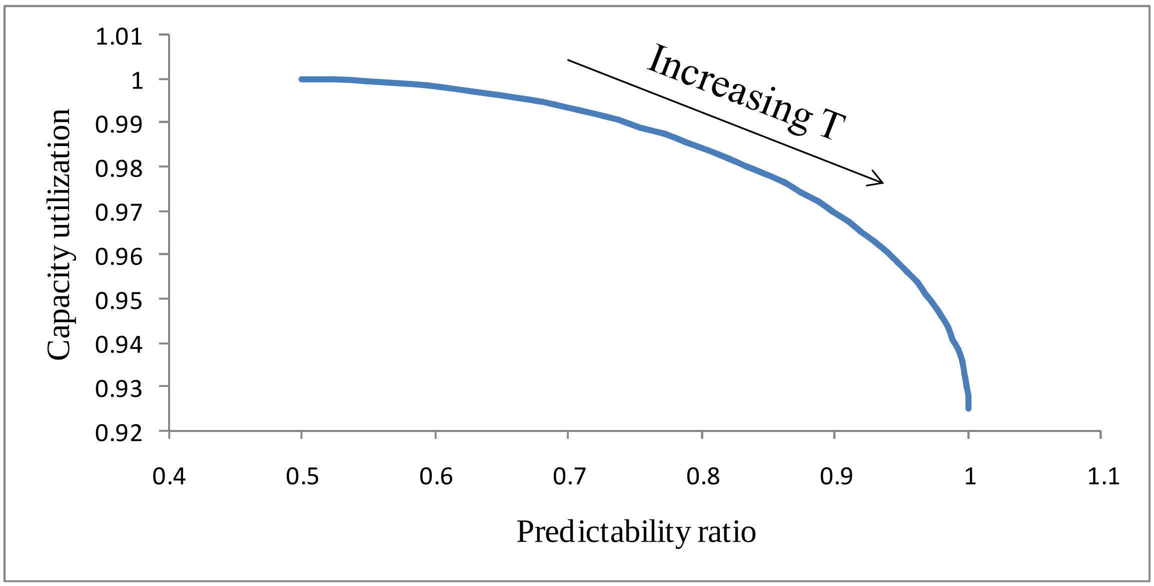


Figure 5: Expectation of Capacity Utilization against that of Predictability Ratio

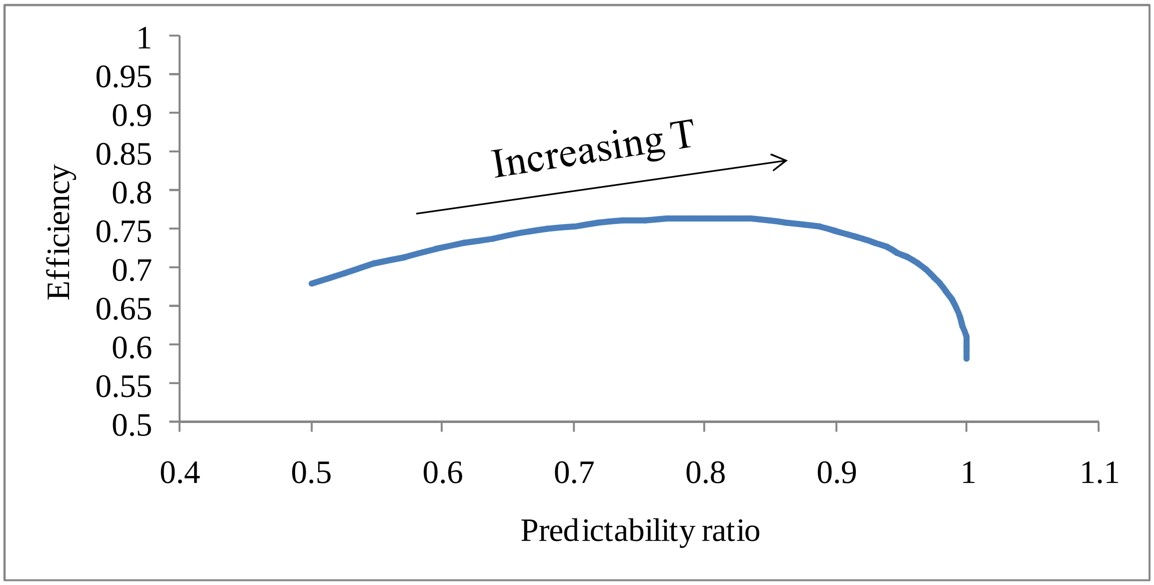


Figure 6: Expectation of Efficiency against that of Predictability Ratio

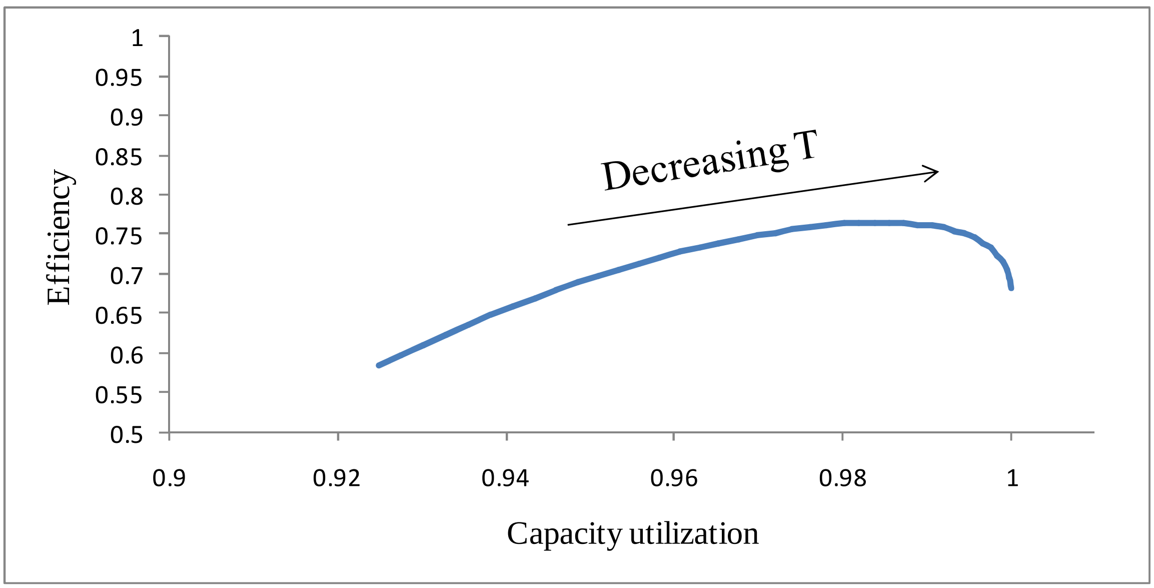


Figure 7: Expectation of Efficiency against that of Capacity Utilization

4.5 Conclusions

Performance metrics are constructed for capacity, predictability, and efficiency. By employing these metrics, the FAA and flight operator are able to see how operational performance will change with design parameters of the GDP under uncertainty. As shown in the realization example, conservative decision on the duration of the GDP will improve predictability but underutilize the runway capacity. On the contrary, a bold decision will diminish predictability but enable us to better utilize capacity. Metrics of predictability and capacity have monotonous relationship with the designed duration. However, efficiency metric is always below 1 due to uncertainty and has an intermediate peak. Different decisions on the duration will generate different performance goal vector.

In ongoing work, we are evaluating metrics for the other performance goals, flexibility and cost effectiveness, and adding another design parameter, scope of the GDP, to the existing model. Additionally, the impact of GDP modification in response to updated information is being taken into consideration.