Distributed Mechanisms for Determining NAS-Wide Service Level Expectations: Concept Description

prepared by Michael Ball

project Team: Michael Ball, U of Maryland; Cindy Barnhart, MIT; Mark Hansen, UC Berkeley; Yi Liu, UC Berkeley; Prem Swaroop, U of Maryland; Vikrant Vaze, MIT; Chiwei Yan, MIT

project Sponsor: Richard Jehlen, ATO System Operation Services

Introduction
Over the past 15 or more years collaborative air traffic management (CATM) has become a fundamental principle underlying all new air traffic management (ATM) system development in the U.S. In fact, this trend goes beyond the U.S. to Europe and increasingly the rest of the world. Its origins go back to the deployment of new information exchange and resource allocation mechanisms for planning and controlling ground delay programs (GDPs) in the U.S. CATM originally, and for most its life, was known as collaborative decision making (CDM). GDP decision support tools evolved and became more sophisticated and the underlying GDP ideas were transferred to the enroute environment with the development of airspace flow programs. A variety of other tools, based on CATM paradigms, have been developed and adopted or are on their way to adoption. In Europe, where a somewhat different set of ATM challenges exist, CATM has been most vigorously adopted in the context of the airport CDM decision support tools.

It is probably safe to say that the bulk of the CATM-based development has focused on tools and processes to support very specific ATM operational decisions, e.g. assigning ground delay to a specific flight during a GDP. At the same time there is a very important strategic planning aspect to the daily execution of ATM. Specifically, FAA traffic managers consult with airline/flight operator operational personnel at both the local and national levels in planning operational strategies for the day. These take the form of strategic planning teleconns (SPTs). To be sure, the SPTs should be considered a very important part of the general trend toward the widespread use of CATM. However, while the various CATM-based decision support tools employ novel resource allocation and information exchange principles, the SPTs do not employ any new collaborative principles or technologies and are, by-and-large, highly unstructured.

It should be emphasized that the SPTs perform a very legitimate and even vital function in the overall traffic management process. Specifically, flight operators have key information not known by the FAA, including air carrier business objectives and economic tradeoffs and the status of aircraft and personnel, just to name a few. However, while the CATM initiative has produced a host of innovations in the
manner in which specific traffic management initiatives (TMIs) are planned and controlled, very little innovation has been directed toward the operation of SPTs. While this per se may not necessarily be bad, there are several concerns and issues related to SPTs and more generally strategic planning on the day-of-operations that merit research attention:

1. The SPTs are free form and highly unstructured and so, at times, can devote an inordinate amount of time to unimportant topics.
2. Again due to their free-form nature, the SPTs do not attempt to assign priority to the various flight operators based on objective measures. Thus, the more persistent and/or “loudest” flight operators tend to have the most influence.
3. The operational concept for the Next Generation Air Transportation System (NextGen) calls for a performance-based ATM system. One embodiment of this concept calls for the separation of strategic ATM planning into i) service level expectation setting and the ii) planning of an operational response. Flight operator input should be provided in i) and the air navigation service provider (ANSP) should then optimize ii) based on the output of i). In fact, today’s SPT’s totally focus in ii).

The proposed system – COuNSEL, CONsensus Service Level Expectations -- described in this white paper addresses the service level expectation (SLE) setting problem and, in so doing, seeks to eliminate the deficiencies discussed in 1) and 2).

Desirable Properties
The starting point for the COuNSEL design process undertaken was to lay out a set of desirable properties that a good design should have. These are listed below.

1) **Consensus Building:** the system should take into account the input of all involved flight operators and should generate an output that represents a consensus of those flight operators.
2) **Equitability:** the system should treat the involved flight operators equitably. It should be noted, however, that the notion of equitability should take into account appropriate measure of flight operator interests. For example, it can be expected that the influence of a flight operator over the consensus solution should increase as the number of impacted flight operations belonging to that flight operator grows.
3) **Practicality:** the system should be easy to administer and efficient in terms of effort and time commitment required by both the ANSP and the flight operators. It should also be efficient in terms of required computing resources.
4) **Confidentiality:** the system should seek to minimize input information required from the flight operators and should keep confidential any information provided.
5) **Strategy Proofness:** To the extent possible the system should encourage truth-telling on the part of the flight operators and prevent or minimize the opportunities for “strategic behavior”, i.e. system gaming.
6) **Single Winner Determination:** the system should provide a single recommended consensus output.
System Concepts and Operation
The basic mode of operation for COuNSEl is fairly straight-forward, however, it is quite different from the SPTs because its basic output is different. As discussed above the system seeks to set service level expectations. Another process, not the subject of this white paper, takes the further step of converting the service level expectations into planned TMIs. Note that SPTs talk directly in terms of TMIs, e.g. discussing which TMIs should be run and which parameter setting should be used. The SLE problem can be viewed as one of setting constraints or guidelines to be used to determining those TMIs and their parameters. A basic question then is what form should the output of a SLE setting process take.

The TMI planning process is viewed as a design problem that requires performance goals. The output of COuNSEl is a set of such goals. In the follow-on step, traffic management specialists carrying out the design process are faced with decisions that require trading off one performance criterion with another. The performance goals provide the designers with the necessary information to do this. As discussed above, this second step is not discussed in this white paper: only the first goal-setting step is addressed.

Performance Metrics
Before describing the exact nature of the output, it is worthwhile to consider an appropriate set of performance criteria. The global ATM community working through ICAO has agreed upon a set of eleven service expectation categories. These were considered carefully in the context of the SLE setting problem and a set of three was chosen to be relevant to the TMI design and control problem. These are discussed below.

*Capacity* measures the number of flight operations that the overall system or constituent subsystems can process safely over a specified time period. In the context of a GDP, an important capacity metric is the number of arrivals that can be accepted by an airport per hour. Capacity is perhaps the most visible and important performance category as it directly relates to flight delays and more generally the ability of an air carrier to maintain its schedule.

*Predictability* has multiple interpretations depending on the time frame in question. For planning specific TMIs, predictability refers to the degree to which flight operators know in advance resources available to them and, more generally, the intentions and planned actions of the ANSP. An ANSP could increase predictability by announcing farther in advance its intention to carry out specific TMIs, and giving earlier indications of the ground delays assigned to flights, earlier announcements of the open/closed status of airways, etc.

*Efficiency* refers to the cost-effectiveness of individual flight operations from the perspective of the flight operator. During a GDP, a policy that leads to high amounts of airborne holding would be less efficient than one that converted that airborne delay into less costly ground delay.

TMI design strategies very often trade off these performance criteria either explicitly or implicitly. For example, one GDP strategy might limit the amount of assigned ground delay, when compared to others. Such a strategy would tend to send a larger number of flights to the airport earlier in hopes that the weather would clear earlier than expected or that a slightly higher than planned acceptance rate could
be accommodated. Such a strategy on the average would lead to higher rates of arrival throughput, increasing the capacity metric, but larger amounts of airborne holding, decreasing the efficiency metric.

Another strategy might announce and implement early in the day certain TMI actions, such as ground delays and reroutes. These would provide ample time for the flight operators to plan for the day’s operations, allowing them, for example, to cancel strategic flights and to take early steps to re-accommodate passengers. On the other hand, such a strategy would have a tendency to impose unnecessary ground delays or reroutes. Thus, it would tend to have a higher level of predictability but lower levels of capacity and efficiency.

The SLE setting problem is to provide guidance to TFM specialist on how to trade off TMI performance in the three performance categories given above. The approach chosen to do this involves choosing a specific metric for each of the three categories and specifying a goal for each of those metrics. Thus, the output of COuNSEl is a vector of size three that contains a goal for each of the metrics. The metrics chosen are normalized to be between 0 and 1, with 1 being the best possible value and 0 the worst. One can view a value of 1 indicating the best performance level for that performance category on a perfect-weather day. Of course, a very simplistic solution to this goal setting problem would be to choose a goal of 1 for each metric. However, a vector of three 1’s provides little insight or tradeoff guidance. Rather one should view the process as starting with an assessment of the weather and traffic conditions. This in turn implies constraints on the set of feasible goal vectors. For example, it would generally be the case that on a poor weather day, it would be impossible to achieve a vector of three 1’s. In general, the constraints implied by the day’s conditions would generate an efficient frontier of possible vector values. Conceptually any such vector could be achieved on the day given an appropriate TMI. In fact, the choice between these vectors represents the choice among TMI strategies and provides exactly the tradeoff information that is sought. For example, suppose that the SLE vector was ordered as follows:

(capacity metric, predictability metric, efficiency metric)

Consider the following possible vectors chosen from the efficient frontier:

A: (.95, .90, .91), B: (.90, .94, .89), C: (.97, .87, .89)

Suppose particular flight operator had a very heavy emphasis on capacity. That flight operator when given the choice between A and B might choose A, indicating a willingness to increase capacity and to a less extent efficiency, while sacrificing predictability. That flight operator might further be given the choice between A and C and choose C again in order to increase capacity while further sacrificing predictability and efficiency. In this way, by choosing a particular vector, a flight operator is forced to make key performance tradeoffs.

Given this choice of three performance categories one is still left with the problem of choosing three specific metrics. Obviously, the choice of specific metrics is very fundamental and a key driver to the effectiveness of the system. However, the development and/or choice of metrics is not a focus of the research activity summarized here and so specific metric definitions will not be provided in this white paper. Henceforth, it is assumed that metrics for capacity (C), predictability (P) and efficiency (E) have
The output of COuNSEL is a vector of metric values: \((m_C, m_P, m_E)\), where each of \(m_C\), \(m_P\), and \(m_E\) are between 0 and 1.

The output vector represents goals for the metric values that the ANSP should seek on the day in question. The “feasible” values for \((m_C, m_P, m_E)\) depend on the conditions of the day so that on poorer weather days, the possible values will tend to be lower (closer to 0) than on better weather days.

**Feasible Metric Vectors**

COuNSEL seeks to generate a consensus among the flight operators. As such an iterative process is required where each flight operator evaluates and compares possible vectors. Flight operators are also given the opportunity to generate candidate vectors. Figure 1 illustrates the domain of feasible vectors, flight operator preferences and a consensus vector, in the case where there are two (rather than three) metrics.

![Figure 1](image)

For any given day of operations there would be many feasible vectors. However, the flight operators and the ANSP should only consider vectors on the efficient frontier. These dominate the others in the sense that for a vector on the efficient frontier, it is not possible to increase one metric value without decreasing another. Generally, it is the case that each flight operator would have a preferred vector. The consensus vector would tend to represent a compromise among the vectors preferred by each flight operator.

**System Operation**

Figure 2 illustrates the basic operation of COuNSEL. It proceeds using a process in which candidate vectors are generated and evaluated by the flight operators until one is found that represents a consensus. The “evaluation” of a metric vector on the part of a flight operator involves assigning a “grade” to the vector. A grade is a value between 0 and 100, 100 being the best possible and 0 the worst. The flight operators are free to interpret and assign grades as they see fit. However, in concept,
grades should vary in proportion to the value, or inverse of cost, that a vector brings to the flight operator. The flight operators will be asked to grade many vectors and given the opportunity to generate new vectors, including providing their most preferred vector. Over time, it is expected that flight operators will develop formal approaches for grading vectors and eventually automated systems for grading. The use of flight-operator-assigned grades is very robust in the sense that it can evolve over time as all stakeholders become more familiar and expert with COuNSEL concepts. For example, when the system is first released flight operators might use a manual process driven by judgment and heuristic rules; however, in the long run it is anticipated that the grading process would become fully automated, making it a low-overhead and very fast process.

Similarly, the use of flight-operator-generated candidate vectors would likely evolve. The ability for a flight operator to generate a candidate vector will require that the ANSP provide some additional information or capabilities to the flight operators. As discussed, the feasibility of a vector will depend on the conditions of the NAS on the day-of-operations and so the ANSP must provide to the flight operators either some representation of the constraints defining the feasible region of vectors or the capability to generate feasible vectors, e.g. via a web-based application. It is not necessary for the flight operators to generate vectors, as the ANSP will do this in any event. However, a flight operator may find it advantageous to generate vector(s) as this may make it more likely that the consensus vector will be closer to its most preferred vector. Thus, a possible scenario might be that in the initial deployment of COuNSEL no flight operators would generate vectors but over time flight operators would develop the capability and use it effectively.

Since the COuNSEL process produces the equivalent of a consensus strategic plan on the day-of-operation it is important that it have very fast response time. As discussed above, in the long run, it
should be the case that all processes both on the ANSP and flight operator sides should be automated so that the entire process, including multiple iterations should be completed in a matter of minutes, if not seconds. In the initial stages of implementation, there may be human involvement in the grading so that response time should be slower. However, there is no reason that total response time of less than 30 minutes should not be possible.

Consensus Vector Definition
A fundamental question to ask is what is the definition of a consensus vector. The theory that underlies COuNSEL is the majority judgment voting procedure. The winner is the candidate vector that has the highest majority grade. The majority grade of a vector is the grade G, such that a majority of flight operators have assigned a grade of G or higher. Figure 3 illustrates three candidate vectors together with the grades assigned by seven flight operators. The assigned grades are ordered from lowest to highest so that the majority grade for each is the one that appears in the third column.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector 1</td>
<td>55</td>
<td>60</td>
<td>76</td>
<td><strong>78</strong></td>
<td>88</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Vector 2</td>
<td>50</td>
<td>59</td>
<td>65</td>
<td><strong>70</strong></td>
<td>70</td>
<td>85</td>
<td>91</td>
</tr>
<tr>
<td>Vector 3</td>
<td>60</td>
<td>60</td>
<td><strong>70</strong></td>
<td><strong>75</strong></td>
<td>84</td>
<td>87</td>
<td>89</td>
</tr>
</tbody>
</table>

Figure 3

Thus, the majority grades for Vectors 1, 2 and 3 are 78, 70 and 75 respectively and the winning vector is Vector 1. It is possible for ties to occur and there are a set of tie-breaking rules, which will not be discussed here.

Flight Operator Weights: Geographic and Temporal Aspects
Under the process as described, all flight operators have equal influence over the outcome. In fact, few would disagree that it is appropriate that flight operators with larger numbers of impacted operations should have more influence than those with smaller numbers of operations. A simple modification of the definition of consensus vector could be made to take into account flight operator weights. That is, flight operators with larger numbers of operations would be given a high weight (more influence) than those with small numbers of operations. Now rather than a majority of flight operators being required, the condition would be that a set of flight operators whose combined weight was greater than ½ the total. If one views the process as voting, then this approach could be interpreted as giving larger flight operators a larger number of votes. Probably the most natural weighting would be to simply use the number of involved operations as the weight. This could give larger operators excessive influence so a more moderate approach, e.g. weights equal to the square root of the number of operations, could be appropriate. It seems clear that weight should increase with the number of operations, but the best function to be used remains an open question.
Consideration of the flight operator weighting question leads one naturally to notion that influence / weight should vary by geography. For example a carrier with a large hub operation at EWR should have more influence over the strategy or part of the strategy that applies to EWR while a carrier with a large hub operation at ORD should wield more influence at ORD. This consideration in turn leads to the question of how COuNSEL should be applied across the large and complex NAS. Certainly the target application is to develop a daily national NAS strategy. To the extent that the strategy has components that specifically apply to ORD or EWR the carriers with large operations at one or the other airport should have a greater influence over that portion of the strategy.

It is also anticipated that COuNSEL could be applied on a regional basis. For example, it could be used to develop a TMI strategy for the Northeast. In that case, the flight operator with larger numbers of operations in the Northeast would have the most influence.

Another related issue is when and how often COuNSEL might be executed. Probably the “classic” application would be to develop a strategic plan at the beginning of the day. However, given the uncertainty of weather and the dynamic nature of the ATM, it is certainly the case that it would be desirable to modify the strategy over the course of the day. Such modifications would probably occur a few times each day. As discussed above, COuNSEL could be used to develop a regional plan so there might also be multiple executions related to regional planning needs, e.g. the development of plans for different regions of the country.

**Perspectives and System Impact**

COuNSEL represents a very different approach to strategic planning for NAS operations. As such much work will be required to move it toward implementation. For example, the concept of providing strategic advice by prioritizing performance criteria and grading performance vectors represents a very different way of doing business for flight operators. Work is on-going to provide intuitive mechanisms for flight operators to provide the necessary information. Human-in-the-loop simulations are also planned. Also, as discussed above, using the output of COuNSEL requires new TMI planning tools. A parallel research activity is addressing this challenging problem.

While challenges remain, the benefits of this new approach could be quite significant. The most obvious benefit will be a reduction in the significant time expended on the SPTs. In fact, this is probably a relatively minor benefit compared to the improvement in overall NAS operations. Specifically, by using NAS strategies that represent a consensus of operator preferences, TMIIs will be developed that lead to better overall flight operator performance leading to an overall reduction in flight operator costs. Further, by balancing flight operator input in a formal way, more equitable strategies and TMIIs will result. Historically, equitable treatment of NAS users had led to greater cooperation in CATM, e.g. through high quality information exchanges. Work is ongoing to quantify potential benefits.