RESPONSE MECHANISMS FOR DYNAMIC AIR TRAFFIC FLOW MANAGEMENT

Michael Ball, University of Maryland, mball@rhsmith.umd.edu, College Park, MD

Robert Hoffman, Metron Aviation, Inc., hoffman@metronaviation.com, Herndon, VA

David J. Lovell & Avijit Mukherjee, University of Maryland, lovell@eng.umd.edu, avijit@isr.umd.edu, College Park, MD

Abstract: Dealing with uncertainty and changing conditions represents a major challenge in air traffic flow management (ATFM). We use the newly developed slot credit substitution (SCS) mechanism as a prototypical example to understand how fast-response, dynamic mechanisms can improve ATFM performance. We develop a quantitative model that estimates the difference in impact of a batch-oriented periodic process and a fast-response asynchronous process and describe some key properties of fast-response mechanisms derived from the SCS case.

1. Introduction

SCS was developed by the Collaborative Decision Making (CDM) community in the U.S. to allow airspace users to quickly find replacement flights for airport arrival slots that are potentially unusable due to severely delayed or canceled flights during ground delay programs (GDPs). Two very important, somewhat novel features of SCS are: 1) it provides a near-real-time response to the requesting party and 2) the underlying request is conditional in that, the requesting party may specify that a slot will not be released unless an appropriate slot is provided in return. Among other topics, we will explore these two properties and try to understand their significance.

While we analyze SCS in detail in this paper, our broader goal is to identify areas of ATFM planning that could benefit from response mechanisms and to determine the key properties those mechanisms should have. We feel major improvements in ATFM performance can only be achieved by embracing more dynamic, fast-response approaches.

Section 2 of this paper gives an overview of the problem of uncertainty in ATFM and how to address it. Section 3 describes SCS and its batch-oriented predecessor algorithm, compression. Section 4 provides a quantitative model that describes the difference between batch-oriented periodic processes and fast-response asynchronous processes. Section 4 describes some key properties of fast-response mechanisms. Section 6 discusses the historical performance of SCS and Section 7 concludes by presenting potential new application areas for fast-response mechanisms in ATFM.

2. Background

Uncertainty represents a major challenge in the design of effective ATFM procedures. In fact it can be argued that uncertainty and variability are the driving forces behind the need for ATFM since, if all flights operated under a static schedule using NAS resources with pre-determined capacities, then there would be little need for ATFM. In the past several years a body of research has emerged that develops stochastic models for ATFM in order to address uncertainty. Yet, it is probably safe to say that the development of ATFM systems and procedures to deal with uncertainty still is a significant research challenge and the fielding of such systems poses an even greater challenge.

2.1 Sources of Uncertainty

It is instructive to break down uncertainty into four categories (the first three categories have been widely described in the literature; the fourth, to our knowledge, was first described in the ATFM context by Beatty in [1]):

Demand uncertainty occurs when flights fail to meet planned departure, arrival, and en route travel times. Contributing factors are aircraft mechanical problems, boarding passengers, en-route weather constraints, and en route or airport surface queues.

Capacity uncertainty of airports and airspace occurs due to natural fluctuations in the throughput levels that can be achieved by various airspace elements. Contributing factors are variable weather conditions and changes in flight sequences that alter flight departure or arrival spacing.

TFM control actions are taken by Federal Aviation Administration (FAA) traffic managers and air traffic controllers in response to the first two uncertainty types (demand and capacity). The human element of decision making adds yet another layer of planning uncertainty for the NAS users: they do not know in advance which TFM actions (such as ground delay programs, miles-in-trail restrictions and strategic rerouting initiatives) will be taken or when they will be taken.

Uncertainty relative to coordination and timing of activities is a subtler, yet equally important effect. Even if air traffic control and airspace users know what actions are being taken and their approximate timing, there can still be a large degree of uncertainty regarding the performance of specific flights related to the exact timing of events and how various activities interact. For example, time differences on the order of minutes or even seconds could impact the position a flight received in a queue, the timing-out of a crew or the need to divert to an alternate destination because of fuel shortage. All such events could have a severe impact on performance statistics for the flight.

2.2 Mitigating Uncertainty

The research and development (R&D) community has tried to reduce the detrimental impacts of uncertainty in several ways. The most natural approach is to reduce or eliminate uncertainty by improving the quality of future event prediction. A second approach, which has stimulated the most interest in the R&D community, characterizes alternate future scenarios so that a plan can be formulated taking into account all future possibilities. A third approach, the focus of this paper, develops plans that can dynamically adjust to changing conditions. This de-emphasizes the importance of correctly predicting the future. We now discuss each of these approaches in more detail.

2.2.1 Improving Information Quality

Probably the most successful and widespread approach to dealing with uncertainty is to reduce it by improving information quality. In fact, all R&D activities aimed at improving weather forecasts fall under this domain. Any improvement in the accuracy of a weather forecast or any increase in its effective lead time reduces uncertainty, especially for capacity of airspace elements.

One of the original CDM objectives in the U.S. was to improve the quality of airport arrival demand predictions by merging daily operational airline and FAA data. Dramatic improvement in demand predictions were achieved by new message capabilities that allowed the airlines to notify the FAA of intended flight cancellations and delays. Analysis of CDM performance data for San Francisco International Airport (SFO) from January, 1998 to May, 1999 showed that, using pre-CDM systems, a flight cancellation was recognized by GDP control systems an average of 29 minutes *after* the flight's original departure time estimate (OEDT), while, with the CDM systems, cancellations were known an average of 48 minutes *before* the OEDT [2]. These information quality improvements led to improvement in the planning and control GDPs.

2.2.2 Hedging against Multiple Possible Outcomes

There is an extensive body of literature on stochastic optimization. A very popular approach within this domain defines multiple possible future scenarios with associated probabilities, and then solves an optimization planning problem that maximizes the expected value of system performance under all scenarios. This approach has been applied to planning and controlling GDPs in the presence of capacity uncertainty (see [3],[4],[5]). In this case, each future scenario is represented by an airport acceptance rate (AAR) vector, with an associated probability of occurrence. Thus, a GDP plan is developed that hedges away from the most likely capacity scenarios in favor of some of the alternative possible outcomes. Despite a large body of research, this approach has enjoyed limited success in practice, due to difficulties in defining scenarios and estimating their probabilities. Some promising work in this direction exists for SFO airport [6], including recent work that has led to an operational system [7].

2.2.3 Dynamic Decision Making

In practice, nearly all decision-making processes become dynamic in the sense that they must adjust to environmental changes. For example in GDP planning, if the weather conditions that induced the GDP linger or improve unexpectedly, then the GDP is extended or canceled. In strategic routing, flights originally rerouted around thunderstorms might be able to make enroute adjustments if such thunderstorms clear or never materialize. The challenge in ATFM planning is to anticipate the possibility of such adjustments so that improving conditions can be taken advantage of to the greatest extent possible and the adverse impact of worsening conditions can be minimized. There is a body of research on dynamic planning that builds upon work described in Section 2.2.2. For example, in [8] an approach to GDP planning is described that builds AAR scenario trees that not only capture a range of possible future AAR scenarios but also capture both the timing of AAR changes and when information about future AARs becomes known. While we view such approaches very promising, they do suffer from rather extensive data requirements.

In this paper, we describe and analyze another class of dynamic ATFM mechanisms. Fast-response dynamic mechanisms should be able to respond to a range of changing conditions and, moreover, should be able to do so very quickly. We use SCS as a prototypical example of such mechanisms. Such approaches reduce the need to accurately predict future events and their likelihood. Rather, their flexibility and adaptively allow them to adjust to a broad range of future events when they become known.

3. Slot Exchange Mechanisms

In this section, we describe the operation of compression and SCS.

3.1 The Compression Concept

Many of the response mechanisms we will consider in this paper live in the context of ground delay actions taken by FAA traffic managers. A ground delay program (GDP) is a traffic management initiative designed to replace anticipated airborne delay at a single airport with ground delay at associated origin airports. Basically, flights bound for a common destination airport are held at their origin airports for a length of time sufficient to insure that the arrival rate at the destination airport does not exceed its forecasted acceptance rate. A GDP is implemented whenever predicted demand will exceed forecasted airport capacity for an extended period of time (e.g. under adverse weather conditions at the GDP airport). GDPs are planned and controlled by the Air Traffic Control System Command Center (ATCSCC or "command center").

When a GDP is enacted, a queue of virtual arrival slots is created in proportion to capacity. For instance, if the GDP airport can accept 30 flights per hour, then 30 two-minute slots are created. Flights are then allocated to slots on a first-scheduled, first-served basis, by a resource allocation algorithm called ration-by-schedule (RBS). See [9],[10],[11] for background on GDP resource allocation.

Another algorithm, called compression, restores system efficiency by moving flights earlier in time (and never later) to fill arrival slots vacated by canceled or delayed flights. Also, compression (and/or RBS) can be enacted at any time during the life of a GDP, to smooth out the planned arrival flow or to reduce overall delay.

An important feature of compression is that it compensates air carriers for forfeiture of arrival resources. Even air carriers that are not directly involved with the delay or cancellation can benefit from compression movements.

Example 1. To see how this works, consider the arrival slot assignments shown in Table 1. Each row corresponds to a single arrival slot, which has an associated flight and a controlled time of arrival, equal to the slot time. Suppose that carrier A cancels flight 7 in the 0800 slot. The compression algorithm searches the later flights (8, 9, 10, 11) for a flight from airline A that can trade slots with the canceled flight. The only candidate from carrier A is flight 11, but this is infeasible because its earliest runway time of arrival (ERTA) is 30 minutes too late. (The ERTA data field is usually set to the scheduled time of arrival.) So, the algorithm selects the earliest feasible flight, independent of its owner. In this case, it happens to be flight 8, from airline C, whose ERTA also happens to be compatible with an 0800 slot. Flights 7 and 8 trade arrival slots. This one-for-one slot exchange is the basis of the compression algorithm.

CTA	Owner	Flight	ERTA	Delay (min)
0800	A	7	CNX	-
0810	C	8	0720	50
0820	B	9	0740	40
0830	C	10	0740	50
0840	A	11	0830	10

Table 1: The 0800 slot is vacated by carrier A.

CTA	Owner	Flight	ERTA	Delay (min)
0800	C	8	0720	40
0810	B	9	0740	30
0820	C	10	0740	40
0830	A	11	0830	0
0840	A	7	CNX	-

Table 2: The final assignment, after all slot exchanges are made.

Compression then searches for a flight that can fill the newly-created vacancy in the 0810 arrival slot (occupied by a canceled flight). Again, priority is given to airline A because it acquired the slot from the prior exchange. The algorithm searches the remaining flights (9, 10, 11) for a feasible flight from airline A. Again, flight 11 is infeasible because of its ERTA. So, the 0810 slot is filled with flight 9, from airline B. This process of slot exchanges continues until flight 11 can be moved up, at which point flight 7 takes over its slot. The final slot assignments are shown in Table 2. The canceled flight moved downward in the queue, passing over flights belonging to competitors B and C. These flights (8, 9, and 10) are called *bridging flights*, while the two flights belonging to A (7 and 11) are called *substitution flights*. In all, four exchanges were made, but for the purposes of this paper, we will consider this one *transaction*.

Effectively, carrier A receives 10 minutes of delay savings for having posted the cancellation. Note that carriers B and C have received 10 and 20 minutes of savings, respectively, even though they posted no delays or cancellations. The overall system efficiency is a savings of 40 minutes. +

The compression algorithm is housed in the flight schedule monitor (FSM), the ground delay modeling and monitoring software tool distributed to the FAA and the air carriers. Due to the dynamic use of compression, overall GDP delay has been significantly reduced [2].

3.2 Airline Slot Substitutions

At any time during a GDP, a carrier can perform a one-for-one slot exchange between two of its own flights. One can think of this as a voluntary delay or cancellation of one flight designed to benefit another flight. On the surface, the net savings to the airline seems to be zero. However, delay cost as a function of time varies by flight. A carrier may be willing to add 30 minutes of delay to a terminal flight (or cancel it) to save 30 minutes of delay on a flight that has critical downstream connections through pilots, crew, or passengers.

In Example 1, if the ERTA of flight 11 had been 0800 or earlier, then carrier A could have directly traded the arrival slots of flights 7 and 11, without involving its competitor's flights. Without this type of real-time schedule manipulation, the delays imposed by a GDP can ripple throughout an air carrier network for hours, or even into the next day. This process of unilateral cancellations and substitutions within an airline's own resources does not require or allow input from the FAA or other carriers, and is therefore an attractive first choice for resource management during GDP's, if it is feasible.

3.3 Slot Credit Substitution

There are situations, however, in which a carrier would like to cancel or delay an earlier flight in order to advance another one of its flights, but the time gap between the flights is so great that the later flight would not be able to 'reach' the vacated slot; thus, unilateral transactions such as described above would not work. Since there is (as of yet) no mechanism by which two carriers can conduct a direct trade of GDP arrival slots, the carrier must either accept their current predicament or proceed with the cancellation of the earlier flight, hoping that their critical flight gets moved earlier in the arrival queue by an execution of the compression algorithm.

The problem is that an air carrier never knows when (or if) the next compression will occur. Compression is executed entirely at the discretion of the FAA traffic mangers and, although it was originally thought that this would occur every half-hour or so, traffic managers are often reluctant to compress for fear of adding demand to an already-stressed airport. This is particularly problematic when 'popup' flights appear in the demand list after all the arrival slots have been allocated. The lack of demand predictability can lead to unpredictability in the frequency of compression execution (for background on GDP demand uncertainty see [12]).

Realizing that constructive flight delay and cancellations by the air carriers is the only way to reduce system demand (and therefore system delay) during these critical periods, the CDM community has instituted a procedure known as *slot credit substitution (SCS)* [13]. Just as with a slot exchange performed under compression, an earlier time slot is traded for a later slot that is more beneficial to the carrier, and a number of bridging flights can be caught up in (and benefit from) the transaction. However, under SCS, the forfeiture of a slot is purely *conditional*. In essence, the downward movement of a flight will take place if and only if the carrier proposing the trade can be compensated with a later time slot that it desires. There is no penalty to a carrier who proposes a slot credit substitution that cannot be accommodated. Furthermore, the response to an SCS message occurs in near-real-time. The conditional nature of SCS and its fast response encourage carriers to explore and conduct a richer set of trading options than could be performed under prior procedures.

4. Fast-Response Asynchronous Mechanisms vs. Periodic Batch Processes

We can view both SCS and compression as mechanisms for responding to airline slot exchange requests. Their essential difference is that SCS provides immediate feedback, and compression is a batch process executed at a small number of time points designated by the command center. In this section, we provide a model that compares, on a very generic level, a fast-response asynchronous process (Model 1, which represents SCS) with a periodic batch process (Model 2, which represents compression). Of course, the conditional nature of SCS represents another difference, which we ignore in the analysis in this section. Our primary goal is not to create a near-exact representation of the performance impact of SCS vs. compression but rather to capture the essential difference between a fast response mechanism and a periodic batch process. We would argue that this "essential difference" is a principal reason that air carriers have embraced SCS, even with compression continuing usual operation.

Under the generic model, action events (AEs, which represent slot exchange requests) arise over time. Each AE requires a system response and the nature of that response impacts overall system performance. Model 1 generates a near-real-time response to each request when it arises. Model 2 allows AEs to collect into a batch and processes all AEs periodically when the batch process is executed.

Each AE has an effective lifetime such that if a response is not given within the lifetime then the opportunity represented by that event is no longer valid. For the SCS example, an SCS message is executed by exchanging arrival slots among a group of flights. This arrival slot exchange implies corresponding departure time modifications. The earliest revised departure time represents a deadline for implementation of the slot exchange request, since flights must depart by the revised departure times in order to meet the appropriate arrival slots. The difference between this deadline and the current time is the effective lifetime of an SCS slot exchange request (see Figure 1).

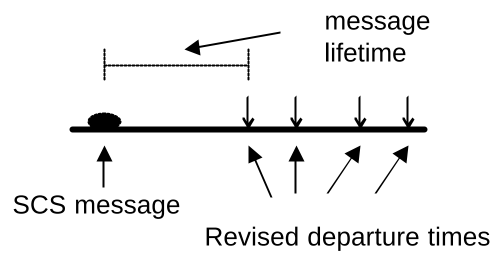

Figure 1: SCS message lifetime

Since Model 2 (compression) is only run at periodic intervals, it is entirely possible that certain slot exchange requests that are captured by Model 1 (SCS) would not be captured by Model 2 (see Figure 2). That is, a run of Model 2 is not able to respond to any AEs whose lifetimes have expired at the time of the execution of Model 2.

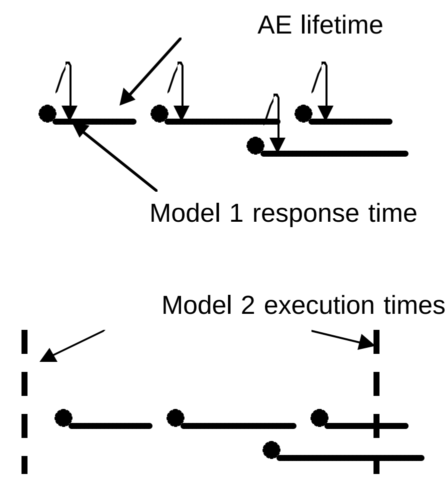

Figure 2: Under Model 2, the first two AEs are not executed since Model 2 is run after the end of their lifetime.

Thus, the principal advantage of Model 1 over Model 2 is that Model 2 may fail to execute several AEs because their lifetimes expire before the next execution of Model 2.

In general, at each periodic interval when Model 2 is executed, a batch of AEs has been collected. It is then able to view a set of AEs and make tradeoffs amongst them when performing its function.

Thus, the principal advantage of Model 2 is that it can optimize the resource allocation problem performed relative to this set of AEs.

In our example case, each execution of compression, in general, must consider a set of exchange requests, i.e. a set of unusable slots, and optimize over the manner in which the entire set of slot exchanges is performed. We note, however, there is no formal optimization performed and the execution of compression is not different in any significant way from the sequential execution of the SCS transactions. We will not provide a formal "proof" of this statement; in fact, it is certainly not the case that the execution of n SCS transactions normally produces the same result as the execution of compression given the slot vacancies created by the SCS messages. However, the differences in impact are relatively minor, involving the order in which each slot vacancy is treated and other details. Some relevant analysis is provided in [10].

4.1 Stationary stochastic comparison model

We now develop a model that quantifies for a range of parameters the advantages of a fast response mechanism (Model 1) over a periodic batch mechanism (Model 2). This model helps illustrate and explain why SCS is used, and provides benefits, over and above those provided by compression. On the other hand, we admit that this analysis is highly generic and does not treat many of the specific realistic details related to SCS and compression. Further, we have made no attempt to estimate actual values of the parameters used in this model for the SCS/compression setting.

The arrival process of action events is modeled as a stationary stochastic process. In reality, SCS messages can come from any number of different airlines, and relate to a wide variety of flights at a large number of airports. For these reasons, it is reasonable to assume that the Markov property holds for this process, and therefore that we can model it as a stationary Poisson point process. Not all AEs would be feasible; for Model 1, we assume that each has some independent likelihood of being successful, keeping in mind that an independently filtered Poisson process is itself Poisson. We also assume that the feasible AEs are executed under Model 1 immediately.

If Model 2 were used instead, it would be run periodically, each time subsuming those feasible AEs whose lifetimes had not yet expired. Traffic flow managers might decide to run compression on constant intervals, or perhaps less regularly in response to some conditions that we are not aware of. A Markov assumption is certainly inappropriate, and distribution assumptions of any nature are not supported by the resolution of this problem. As a result, we assume that Model 2 is executed at constant time intervals of duration T, and we assume that the benefit to a carrier from any unexpired SCS transaction could be realized by compression instead. The measure of comparison, then, is the expected proportion of AEs with lifetimes long enough to be captured by these regular runs of Model 2. This is equal to the probability that any single feasible AE would be subsumed by the subsequent run of Model 2. In the analysis that follows, we consider a single (randomly chosen) action event.

The lifespan of the AE is denoted by s. Lacking specific information on bridging flights and their departure times, s is modeled as uniformly distributed on the interval $[a,b]$. The analysis then depends only on the parameters a, b, and T. If $a > T$, it will always be the case that compression occurs before the AE expires, hence the fraction subsumed by compression would be 1. Now consider the case $a \le T$. Without loss of generality, assume that consecutive compression run times are 0 and T. Given that an AE has arrived in the interval $[0,T]$ according to a Poisson process, its (unordered) arrival time is uniformly distributed on that same interval. The time t from the message arrival until the end of the interval is also uniform on $[0,T]$. If $s < t$, then the first affected flight has a scheduled departure before the next compression run, so the SCS message would effectively expire before compression would have a chance to subsume it. If, on the other hand, $s \ge t$, the next compression run will capture this message, and it is the probability of this latter event that we are interested in. It is shown in Appendix A that this is given by:

$$\Pr\{s \ge t\} = \begin{cases} 1, & \text{if } a > T \\ \dfrac{a+b}{2T} - \dfrac{(b-T)^2}{2T(b-a)}, & \text{if } a < T < b \\ \dfrac{a+b}{2T}, & \text{if } a < b \le T \end{cases} \qquad (1)$$

Figure 3 shows the percentage of SCS messages subsumed by running compression at regular time intervals for different values of the parameter b. For this figure, the parameter a is set to 15 minutes in all cases. For each possible value of b, a curve results that is concave until the inflection point $T = b$, at which point the graph changes to a simple convex function proportional to T^{-1}.

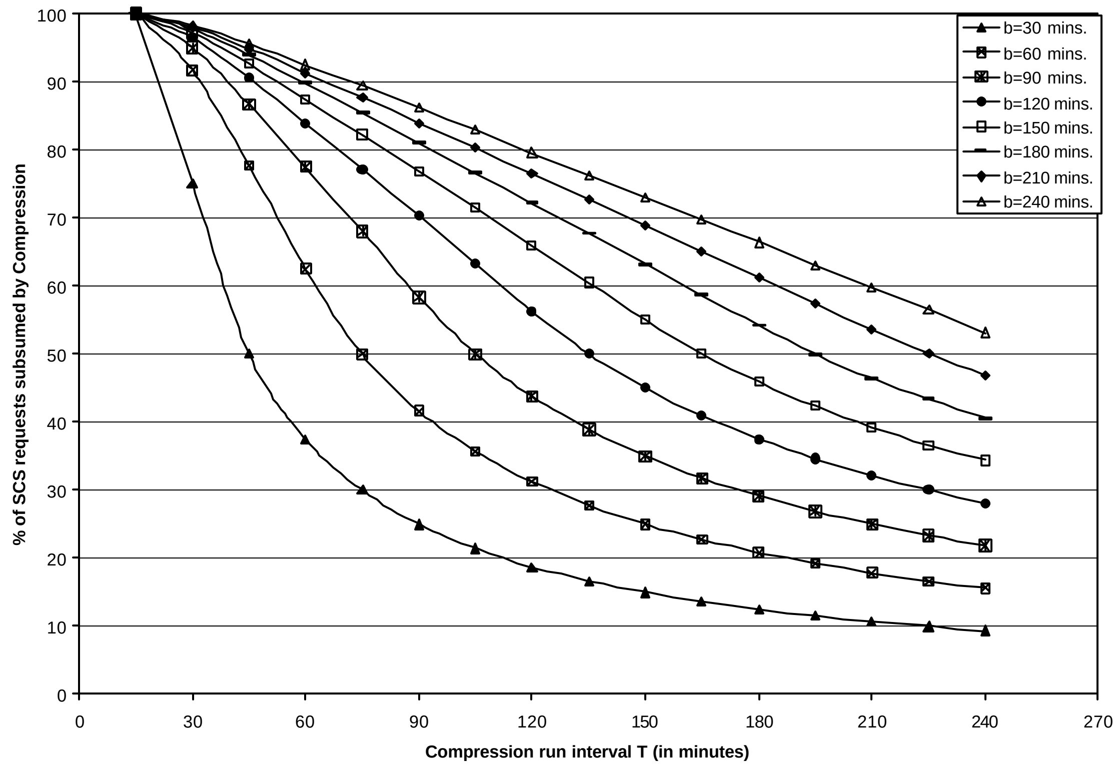

Figure 3: Percentage of SCS Messages Subsumed by Running Compression at Regular Time Intervals (*T*)

The concavity of the left pieces might not be obvious from the equation; the 2nd derivative of (1), as a function of *T*, in the domain $a < T < b$, can be shown to be $\frac{-a^2}{T^3(b-a)}$, which is strictly negative for appropriate values of a, b, and T.

By itself, the parameter a represents, in a sense, the amount of lead time by which the SCS proposal is submitted, since it represents the smallest possible duration until a flight relevant to that proposal departs. The parameter b, relative to a, represents the uncertainty with which the makeup of the set of affected flights and their relative departure times is known. Clearly, this has an important impact on the shape of the function generated. It makes sense that as this parameter gets smaller, the effectiveness of compression at capturing that SCS transaction in the next cycle is reduced. It is also perfectly reasonable for all of these functions to be decreasing in T; the more interesting observation is that the decrease is sharpest for the lowest values of T.

5. Key Features of Fast Response Mechanisms

In this section, we again use SCS as a prototypical application to investigate key features of fast-response mechanisms.

5.1 Flexibility in Original Plan

Although fast-response mechanisms must be highly adaptive, it is generally true that the initial plan or control action for an individual flight should always fit within a coordinated comprehensive plan. In the case of SCS, CDM processes such as RBS give each flight a controlled time of arrival (CTA). The set of CTAs represents the comprehensive plan. Each SCS transaction modifies this comprehensive plan by changing CTAs. A subtle aspect of the modification ability is that there is flexibility built into the initial plan. When an airline sets the ERTA field for a given flight, it effectively specifies a range of acceptable slot times, in the interval [ERTA, CTA]. This gives implicit permission for the arrival times of “bridging flights” to be changed in an SCS transaction without explicit approval of the owner airline.

This type of flexibility is a necessary ingredient for successful response mechanisms and ATFM plans in general. Movement in this direction is underway in route planning. For example, the coded departure route (CDR) and low altitude arrival and departure route (LADDR) initiatives (for information on these initiatives see the CDM web site[14]) both are based on the concept that flights have alternative flight plans available at the time of departure. The choice between these plans is made in near-real-time in response to realized conditions.

5.2 Distribution of Decision Responsibility

In order to achieve rapid response, decisions should generally be made by the party closest to the events that induce the decision need. "Distributed Decision-Making" has become something of a cliché, but it is often times difficult to achieve in practice. In fact, CDM processes rely heavily on effective use of distributed decision-making. The FAA command center initially sets CTAs using RBS, the airlines then adjust the CTAs through the cancellation and substitution process and the command center can make further adjustments using compression. SCS provides a further refinement by allowing airlines to control some slot exchanges (normally the domain of compression) in order to achieve immediate response. We note that, in fact, an unanticipated use of SCS involves SCS transactions initiated by the command center in order to find replacement slots for flights that depart from the gate late and miss their CTAs. In this case, since the command center and tower controllers have the most direct knowledge of the situation, they are in the best position to make the decision. The transactional nature of SCS made this easy to accomplish. In fact, in 2004, 50% of the delay savings produced by SCS were associated with command center initiated messages (data source: Metron Aviation).

Today, airline operations personnel make all decisions regarding GDP arrival slot modifications. However, airline gate personnel have the most up-to-date information on aircraft status, such as how long it would take to board passengers for departure. In the future, it may be beneficial to allow airline gate personnel to participate in GDP arrival slot modifications. This could be used to more quickly capitalize on unexpected GDP airport capacity surges (such as the fog lifting at SFO).

5.3 Conditional Decisions and Decision-Impact Queries

As mentioned earlier, SCS differs from compression by the conditional nature of its transactions: a slot will be released only if a slot within a certain range is received in return. Compression works with slots that have already been vacated. The conditional nature of SCS allows airlines to base their decisions on which flights to cancel on the benefits they can receive in return. This is a very valuable capability and certainly has the potential to substantially improve airline decision-making. A related capability is a decision-impact query, which asks for a projection of the impact of a decision prior to making that decision, i.e. a type of "what-if" capability. While both of these capabilities are very useful in evaluating any decision, they are particularly important for fast-response mechanisms, where decisions must be made with little time for analysis.

6. SCS in Practice

Since its adoption in May 2003, the use of SCS has increased substantially. On an average, the number of SCS messages that arrive during an hour of GDP has increased from 0.32 during the year 2003 to 1.25 during 2004, an increase of approximately four times (Data Source: Metron Aviation Inc.). In 2004, SCS activity resulted in approximately 45 minutes of ground-delay reduction per hour of GDP in the NAS. The significant use of SCS and the growth of that use give evidence that it provides a substantial benefit over and above compression. We would argue that a principal reason for the added benefit is based on the effects demonstrated in Section 3.

When scanning SCS transaction data it becomes clear that SCS activity varies substantially among airlines. This is not surprising, since SCS, in a sense, is a "precision" tool that can be used to implement specific changes to an individual flight. As such, each airline can tune its use to their internal processes. It is also noteworthy that SCS requires that the airlines be proactive. That is, airlines must initiate messages in order for any activity to take place. In contrast, compression is run by the command center based on flight data routinely generated by the airlines. This is another respect in which the two procedures can be seen to complement each other.

7. Conclusions

Our objective in analyzing SCS and generic fast-response mechanisms was to provides incentives and concepts related to more broad implementation of such mechanisms within ATFM.

Probably the most promising application area is flexible flight planning. The general paradigm should be one in which the static flight plan is replaced by a dynamic flight plan that can rapidly adjust to changing conditions. In fact, several steps have already been taken in this direction. A very early example is the Pacific Track Program, under which airlines submitted alternative flight plans, with criteria for switching from one to the other depending on estimated route delay. More recent CDM initiatives include CDRs and LAADR, which we have discussed briefly in Section 5. Some on-going CDM activities also reflect these themes. However, following the principles from Section 5, we see that flight plan flexibility must tie into a new system-wide approach to flexible flight planning. For example, when the choice is made to use an alternate flight plan, the implications of this change on overall system performance should be visible. Fast response clearly would reap substantial benefits. For example, the ability to launch aircraft quickly during "windows of opportunity" made possible by changes in weather formations or unexpected demand reductions, could reap substantial benefits, but, at the same time, requires a high degree of route flexibility. We view flexible flight planning as a rich research domain with the potential to provide very significant benefits.

Acknowledgements

This work was supported by the National Center of Excellence for Aviation Operations Research (NEXTOR), under FAA cooperative agreement number 01CUMD1. Opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation.

References

[1] Beatty, R. (2004) Measuring uncertainty: a core problem for TFM, *NEXTOR Performance Metrics, Performance Organization and Air Transport Workshop*, January, 2004, Asilomar, California

[2] Ball, M.O., R. L. Hoffman, D. Knorr, J. Wetherly and M. Wambsganss, 2000, Assessing the benefits of collaborative decision making in air traffic management, in *Proceedings of 3rd USA/Europe 'Air Traffic Management R&D Seminar*, Naples, June, 2000.

[3] Richetta, O. and A. Odoni, 1993, Solving optimally the static ground-holding policy problem in air traffic control. *Trans. Sci.* 27, 228-238.

[4] Ball, M.O., Hoffman, R., Odoni, A. and Rifkin, R., 2003, A stochastic integer program with dual network structure and its application to the ground-holding problem, *Ops. Res.* **51**, 167 – 171.

[5] Willemain, T.R., 2002, Contingencies and cancellations in ground delay programs. *Air Traffic Control Quarterly* 10, 43-64.

[6] Inniss, T. and M. O. Ball, 2002, Estimating one-parameter airport arrival capacity distributions for air traffic flow management. *Air Traffic Control Quarterly*, 12, 223-252.

[7] Wilson, F.W., 2004, A stochastic air traffic management model that incorporates probabilistic forecasts. to appear in *Proceedings of 20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology.*

[8] Mukherjee, A and M. Hansen, 2004, Dynamic stochastic model for single airport ground holding problem, submitted for publication.

[9] Ball, M., R. Hoffman and T. Vossen. 2002, An analysis of resource rationing methods for CDM, ATM-2002, Capri, Italy.

[10] Vossen, T. and Ball, M.O., 2003, Optimization and Mediated Bartering Models for Ground Delay Programs", submitted for publication, available at http://bmgt1-notes.umd.edu/faculty/km/papers.nsf.

[11] Wambsganss, M., 1996, Collaborative decision making through dynamic information transfer, *Air Traffic Control Quarterly*, **4**, 107-123.

[12] Ball, M.O., Vossen, T. and Hoffman, R., 2001, Analysis of demand uncertainty in ground delay programs, in *Proceedings of 4th USA/Europe Air Traffic Management R&D Seminar*, 2001.

[13] Howard, K. 2002, ETMS/ATMS system requirements: slot credit substitutions (SCS) version 1.2, the John A. Volpe Res. Ctr., Cambridge, MA.

[14] Metron Aviation, CDM Web Site: http://www.metronaviation.com/cdm/index.html

Appendix A:

The random variable t has density function $f_t(z) = 1/T$. Since $s \sim U(a,b)$, then for any constant t, the complementary distribution function of s is:

$$F_s^C(t) = \Pr\{s \geq t\} = \begin{cases} 1, & \text{if } t < a \\ \dfrac{b-t}{b-a}, & \text{if } a \leq t < b \\ 0, & \text{if } t \geq b \end{cases} \qquad (2)$$

We must consider separately the cases $a < T < b$ and $a < b \leq T$:

Case 1: $a < T < b$

In this case, it is not possible for $t \geq b$; thus there are only two possibilities from (2) for $\Pr\{s \geq t\}$:

$$\begin{aligned}
\Pr\{s \geq t\} &= \int_0^T \Pr\{s \geq t \mid t = z\} f_t(z)\, dz \\
&= \int_0^a \Pr\{s \geq t \mid t = z, t < a\} f_t(z)\, dz \\
&+ \int_a^T \Pr\{s \geq t \mid t = z, a \leq t \leq b\} f_t(z)\, dz \\
&= \int_0^a \frac{1}{T} dz + \int_a^T \frac{b-z}{b-a}\left(\frac{1}{T}\right) dz = \frac{a}{T} - \left.\frac{(b-z)^2}{2T(b-a)}\right|_{z=a}^{z=T} \\
&= \frac{a}{T} + \frac{b-a}{2T} - \frac{(b-T)^2}{2T(b-a)} = \frac{a+b}{2T} - \frac{(b-T)^2}{2T(b-a)}
\end{aligned} \qquad (3)$$

Case 2: $a < b \leq T$

In this case, there are three possibilities from (2) for $\Pr\{s \geq t\}$, although the third vanishes:

$$\begin{aligned}
\Pr\{s \geq t\} &= \int_0^T \Pr\{s \geq t \mid t = z\} f_t(z)\, dz \\
&= \int_0^a \Pr\{s \geq t \mid t = z, t < a\} f_t(z)\, dz \\
&+ \int_a^b \Pr\{s \geq t \mid t = z, a \leq t < b\} f_t(z)\, dz \\
&+ \int_b^T \Pr\{s \geq t \mid t = z, b \leq t \leq T\} f_t(z)\, dz \\
&= \frac{a}{T} + \int_a^b \frac{b-z}{b-a}\left(\frac{1}{T}\right) dz = \frac{a}{T} - \left.\frac{(b-z)^2}{2T(b-a)}\right|_{z=a}^{z=b} \\
&= \frac{a}{T} + \frac{b-a}{2T} = \frac{a+b}{2T}
\end{aligned} \qquad (4)$$

Keywords

Traffic flow management, collaborative decision making, stochastic model, distributed decision making.

Biographies

Michael O. Ball: Michael Ball is the Orkand Corporation Professor of Management Science in the Robert H. Smith School of Business at the University of Maryland. He also holds a joint appointment within the Institute for Systems Research (ISR) in the Clark School of Engineering. Dr. Ball received his PhD in Operations Research in 1977 from Cornell University. He is co-Director of NEXTOR, the National Center of Excellence for Aviation Operations Research, and he leads the NEXTOR Collaborative Decision Making project.

Robert Hoffman: Robert Hoffman manages the Long-term Research Group of Metron Aviation's Research and Development Division. He received his Ph.D. from the University of Maryland at College Park (1997, Applied Mathematics). Currently, he is involved in modeling and analysis of strategic ATM initiatives - specializing in resource allocation and equity issues. Dr. Hoffman is also a Visiting Research Scientist at the University of Maryland's Institute for Systems Research. He has received an award from the INFORMS Transportation Science Section Dissertation Competition and a Best Paper award from the USA/Europe ATM R&D Seminar in 2003 for his work in traffic flow optimization.

David J. Lovell: David Lovell is an Associate Professor in the Department of Civil and Environmental Engineering at the University of Maryland. He holds a joint appointment with the Institute for Systems Research. Dr. Lovell received his Ph.D. in Civil Engineering from the University of California, Berkeley, in 1997.

Avijit Mukherjee: Avijit Mukherjee is a Research Associate in the Institute for Systems Research at University of Maryland, College Park. He received his PhD in Civil and Environmental Engineering from the University of California, Berkeley in 2004.