How to Fix a Broken Timetable: Collaborative Mechanisms for Managing Airport Capacity Reductions in the U.S.

Michael O. Ball
Robert H. Smith School of Business & Institute for Systems Research
University of Maryland, College Park 20742, USA
mball@rhsmith.umd.edu

June 24, 2008

Abstract

Scheduled commercial passenger airlines operate based on published itineraries between designated city-pairs. Due the dependence of air transportation system capacity on weather conditions, it is very common for there to be significant disruptions in airline schedules. This creates a strong need for systems to optimize system performance under degraded conditions. Typically, one must minimize the deviation between the original timetable and the schedule actually used. In this paper we describe the set of resource allocation methods and practices used jointly by airlines and the Federal Aviation Administration within the U.S. to address these types of problems.

keywords: air traffic management, collaborative decision making, scheduling.

1 Introduction

Commercial airlines present their services to passengers by means of published schedules offered between select city-pairs. Underlying this service is perhaps the most complex transportation system and possibly the most complex man-made system in the world. Airports make up the fixed "nodes" on which the system is built. Aircraft represent the very valuable assets that provide the basic transportation service. Crews of pilots and flight attendants operate the aircraft and provide service to passengers. These disparate entities are coordinated through a flight schedule, comprised of flight legs between airport locations. The associated crew schedule is an assignment of the legs in the flight schedule to pilots and flight attendants, ensuring that all crew movements and schedules satisfy collective bargaining agreements and government regulations (See [1] for more details).

The fact that a single flight leg is a component of several different types of schedules implies that a perturbation in the timing of one leg can have significant "downstream" effects leading to delays on several other legs (see [8] for an analysis of delay propagation).

This "fragility" is exacerbated by the fact that most of the largest carriers rely heavily on hub-and-spoke network configurations that tightly inter-connect flights to/from many different "spokes" at the network's hubs. Thus, any significant disturbance at a hub, rapidly leads to disruptions of extensive parts of the carrier's schedules. The fact that the number of scheduled operations has grown very close to the capacity of many airports together with the relative frequent occurrence of capacity-reducing weather events means that the system goes into a mode of "irregular operations" quite frequently. Under such conditions the timetables set up between specific city-pairs become severely disrupted. It is not unusual for many flights to arrive very late and also flight cancelations can be common.

In response to the strong economic incentives to perform as well as possible under degraded conditions, there has been significant research (e.g. see [1], [12], [19], [20]) and commercial products aimed at improving performance in such situations.

There is a particular aspect to the problem of providing good operational performance within aviation that is somewhat unique among transportation sectors. In aviation, an air navigation service provider (ANSP) plays a very strong role in both insuring the safety of aviation operations and also in their overall performance. Specifically, ANSP's, such as the Federal Aviation Administration (FAA) in the U.S. and Eurocontrol, in Europe, have responsibility for overall airspace management and as such are interested in achieving high levels of system-wide performance. These entities must work collaboratively with flight operators, including large scheduled air carriers, various business jet operators offering a range of services and also, individual general aviation pilots. In this paper we focus on a set of systems and concepts that very specifically tries to "optimize" the relationship between an ANSP and the flight operators. Collaborative Decision Making (CDM)(also known as Collaborative Air Traffic Management) has developed over the past 15 years both in the U.S. and Europe (see [3], [9], [11],[17]). CDM makes up a body of knowledge that consists of a set of operational concepts, decision principles and algorithms and most broadly an overarching philosophy. Several operational systems and computer-based decision support tools have been based upon it. At this point, any "modern" air traffic management solution must embody certain CDM principles.

The second aspect of the focus of this paper involves the general problem area of ground delay program (GDP) planning within the U.S. GDP planning and control represents the problem domain within which CDM principles were initially developed. As such it has the richest history and most completely developed mechanisms. Further, many aspects of this problem deal very explicitly with the timetable structure of airline schedules and the constraints this structure places on how services can be dynamically adjusted based on adverse conditions.

Section 2 of this paper describes the basic principles of CDM-based GDP planning and control as well as the impact the move to CDM principles produces. Section 3 reviews more recent research and emerging principles and systems.

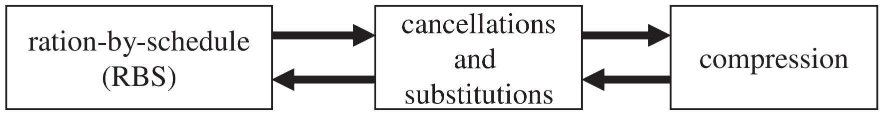

Figure 1: CDM Resource Allocation

2 Collaborative Ground Delay Program Planning and Control

The FAA continuously monitors airports throughout the US for capacity-demand imbalances. Whenever it predicts that the number of flights arriving at an airport within a 15-minute interval exceeds the capacity of the current runway configuration, FAA directives mandate a response [10]. Short periods of congestion are usually resolved by airborne tactics, such as re-routing and variations in airborne speed. GDPs on the other hand primarily address longer periods of congestion at an airport. Usually, a GDP spans a period of 4 hours or longer, and is initiated 3-4 hours in advance.

Prior to the implementation of GDP enhancements under CDM, arrival slots were assigned to flights by a first-come, first-served algorithm affectionately known as *Grover Jack*. The affected flights were ordered according to their most recent estimated arrival times, so the net effect of the Grover Jack algorithm was more or less to stretch out the incoming flights over time. Though intuitively appealing, this method of assigning flights to slots penalized the airlines for providing accurate information. This effect, known as the double penalty issue, is best explained by the following example. Suppose a flight was originally scheduled to arrive at 10:00, but experienced a delay of 30 minutes due to mechanical problems. If a GDP was implemented in which the flight was delayed for another 30 minutes, its total delay would therefore be 60 minutes. However, had the airline not notified the FAA of its mechanical delay, it would have only been assigned a delay of 30 minutes! As a result, airlines were hesitant to provide the FAA with accurate estimates of their flight delays. In addition, airlines had no incentives to report flight cancelations in a timely manner. As a result, GDP decisions were based on poor data, which led to inefficient programs.

The GDP enhancements implemented under CDM address these issues with a fundamental change in the allocation of capacity by the FAA. Instead of an assignment of flights to slots, the CDM "philosophy" considers the allocation of capacity to be an assignment of slots to airlines. The overall allocation process is illustrated in Figure 1.

Ration-by-schedule (RBS), provides an initial assignment of slots to flights. The assignment of time slots by RBS can be viewed as a simple priority rule. Using the *scheduled arrival order* as a priority order, each flight is assigned the next available arrival slot. Note the conceptually simple but practically very significant difference between RBS and Grover Jack, i.e. priority based on expected arrival order is changed to priority based on scheduled arrival order. This change implies that resource allocation no longer depends on the information provided. Thus, the disincentive to provide current information is eliminated.

This change, however, introduces a potential problem: the RBS slot-to-flight allocation

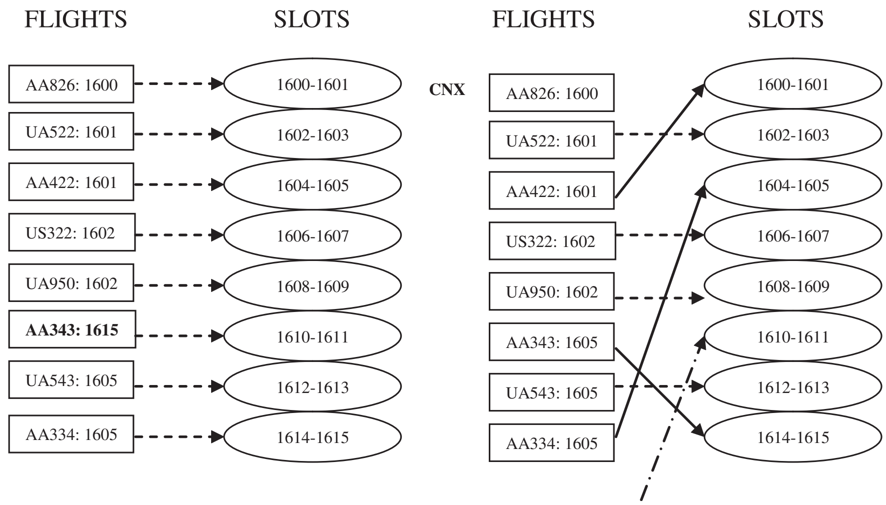

Figure 2: Cancelation and Substitution Process

may no longer be feasible. However, the cancelation-and-substitution process effectively converts the slot-to-flight allocation into a slot-to-airline allocation. Figure 2 illustrates the cancelation-and-substitution process.

In this example, the left hand side of the diagram gives the allocation produced by RBS. Flight AA343, which is in bold has an expected (earliest) time of arrival of 1615, which would seem to be out of sequence based on the order in which slots were allocated. In fact, under RBS, this could happen if its scheduled time of arrival was earlier, say 1604. The right hand side illustrates the cancelation and substitution strategy employed by AA. AA cancels its earliest flight AA826 and is able to assign AA343 to a feasible slot (1614-1615). In addition, it reassigns flights AA422 and AA334 eliminating all of their delay. One can view the process AA employed as substituting one flight assignment for another. Alternatively, a more global view is that AA "owns" a set of slots and assigns its flights to those slots in a way that minimizes an appropriate cost function.

The final step, compression, which is carried out by the FAA, maximizes slot utilization by performing an inter-airline slot exchange in order to ensure that no slot goes unused. Compression is warranted because after the round of substitutions and cancelations, the utilization of slots can usually be improved. The reason for this is that an airline's flight cancelations and delays may create "holes" in the current schedule; that is, there may be arrival slots that an the owner-airline is unable to assign a flight to. The purpose of the compression step is to move flights up in the schedule to fill these slots. The basic premise behind the algorithm currently used to perform compression is that airlines should be "paid back" for the slots they release, so as to encourage airlines to report cancelations.

To illustrate the compression algorithm, consider the example shown in Figure 3.

The left side of the figure represents the flight-slot assignment prior to the execution of the compression algorithm. Associated with each flight is an earliest time of arrival, and each slot has an associated slot time. Note that there is one canceled flight. The right side

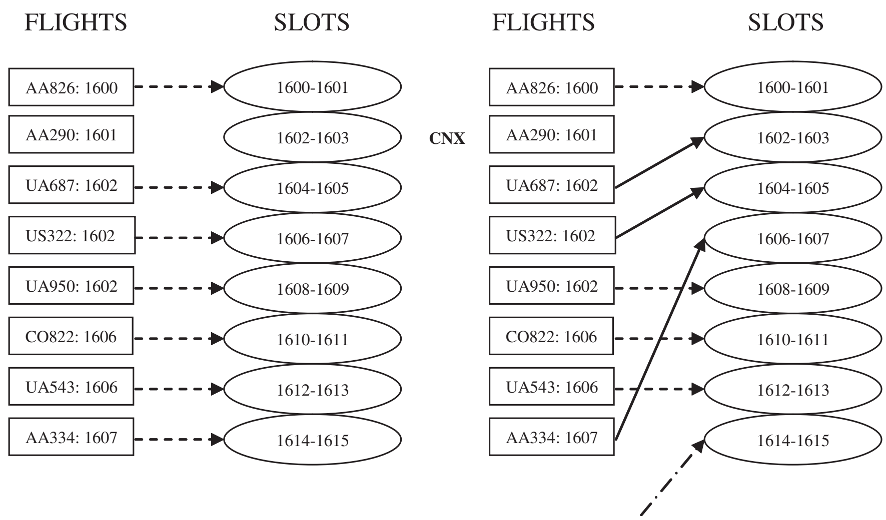

Figure 3: Compression Algorithm

of the figure shows the flight schedule after execution of the compression algorithm: as a first step, the algorithm attempts to fill AA's open slot at 16:02-16:03. Since, there is no flight from AA that can use it, the slot is allocated to UA, and the process is repeated with the next open slot, which, using the same logic, is assigned to US. The process is repeated for the next open slot, which is now assigned to AA. AA thus receives the earliest slot that it can use. The net result of compression in this case can be viewed as an exchange among airlines of the slots distributed through the initial RBS allocation.

We now investigate the special properties of the CDM resource allocation procedures and describe how these are being extended and enhanced.

RBS, Compression and Information Sharing: One of the principal initial motivations for CDM was that the airlines should provide updated flight status information. As discussed above, it was quickly discovered that the existing resource allocation procedures, which prioritized flights based on the current estimated time of arrival, actually discouraged the provision of up-to-date information. Since the RBS priority does not vary with changes in flight status, this problem was eliminated. Further, as in the example of Figure 3 illustrates, compression encourages airlines to announce their flight cancelation intentions early since they are compensated for unusable slots that can be created by canceled flights.

Properties of RBS: We summarize some basic properties of RBS derived in [14]. RBS assigns to each flight, f, a controlled time of arrival, $CTA(f)$. This is equivalent to assigning a delay, $d(f)$, to flight, f, given by $d(f) = CTA(f) - OAG(f)$, where $OAG(f)$ is the scheduled arrival time for f. All time values are rounded to the nearest minute under RBS, hence, each delay value $d(f)$ is integer. If we let D equal the maximum delay assigned to any flight and $a_i = |f : d(f) = i|$ for $i = 0, 1, 2, \ldots, D$, then the important properties of unconstrained RBS (RBS with no flight exemptions) can be defined by,

Property 1: RBS minimizes total delay $= \sum_f d(f)$.

Property 2: RBS lexicographically minimizes $(a_D, , a_1, a_0)$. That is, a_D is minimized; subject to a_D being fixed at its lexicographically minimum value, a_{D-1} is minimized; subject to (a_D, a_{D-1}) being fixed at its lexicographic minimum value, a_{D-2} is minimized; and so on.

Property 3: For any flight f, the only way to decrease a delay value, $d(f)$, set by RBS is to increase the delay value of another flight g to a value greater than $d(f)$.

CDM and Equity: Property 3, which follows directly from Property 2, expresses a very fundamental notion of equity that has been applied in a number of contexts [18]. It is remarkable that procedures, such as RBS, which were developed in very practical war-gaming and consensus-building exercises have such elegant and desirable properties. On the other hand, this may not be surprising in that these properties represent the basis for reaching a consensus in the first place.

Compression as Trading: Compression as Trading: Although the initial interpretation of compression is as a slot reallocation procedure that maximizes slot utilization, Vossen and Ball in [14] provide a natural interpretation of compression as an inter-airline trading or bartering process . For example, in Figure 3, American Airlines "traded" the 1602 - 1603 slot, which it could not use for the 1606 - 1607 slot, which it could use, and United Airlines reduced its delay by trading the 1604 - 1605 slot for the earlier 1602 - 1603 slot. Vossen and Ball show that a bartering process can be structured so as to produce a result essentially equivalent to compression. This view of compression suggests many possible extensions. For example, in [15], they define a more complex 2-for-2 bartering mechanism and show that using this mechanism offers a substantial potential for improved economic performance. Probably the most intriguing enhancement is allowing "side payments" with any exchange as well as the buying and selling of slots. Ball, Donohue and Hoffman in [2] provide a discussion of this and other aviation-related market mechanisms.

3 Recent CDM Research and Trends

3.1 Real-Time Versions of Compression

When one interprets compression as trading, as discussed above, it is natural to consider "transaction-oriented" trading as opposed to the "batch-oriented" trading underlying the original implementation of compression. Specifically, the original compression paradigm called for the periodic execution of compression so that "trades" were executed in batches, say about once per hour. Here "periodic" is actually an inaccurate description in that the execution of compression was at the discretion of the FAA traffic management specialist managing the GDP. In fact, compression was typically not executed on a reliable schedule and, in any event, there was a significant time lag between executions. This often led to missed opportunities (see [6] for a quantification of this phenomenon). The CDM community developed a transaction oriented version of compression called *slot credit substitution (SCS)*. Under SCS an airline can submit a proposed exchange in which an airline proposes to release a slot, e.g. due to a cancelation, in exchange for a later slot usable by that airline. The FAA's Enhanced Traffic Management System (ETMS) examines this request and provides a real-time response indicating whether it can be honored. If it can, it is then implemented. If not, the requesting airline retains “ownership” of the slot it proposed to release. In this way, an airline does not need to commit to canceling a flight unless it is sure of what it will get in return.

A second transaction-oriented version of compression called *adaptive compression* has recently been implemented [11]. While each SCS transaction is initiated by an airline, adaptive compression transactions are initiated by ETMS. Specifically, ETMS continuously monitors the status of flights and slots and identifies slots that likely will go unused, e.g. an arrival slot would be identified if the flight assigned to it had not departed in sufficient time to arrive at the designated slot arrival time. In such, cases an exchange transaction is implemented, which moves a flight into the unused slot and (possibly after a sequence of exchanges) moves the flight originally assigned to it to an arrival slot that it can still meet.

3.2 Taking into Account Uncertainty

GDP’s are most often motivated by reductions in arrival capacity due to poor weather conditions. Since a ground delay must be “served” before a flight departs and flight-times are typically one or more hours, GDP plans must be based on weather conditions forecasted 2,3 or even 4 or more hours in advance. Clearly, there will be a high level of uncertainty associated with such forecasts and ideally GDP planning models will take this uncertainty into account. In fact, there is a literature on GDP planning under uncertainty, e.g. see [5], [13]. Formal stochastic integer programming approaches have yet to make there way into practice. On the other hand, a simple exemption policy, [7], which helps mitigate uncertainty is now routinely used. This policy states that a GDP should be applied to an “included set” of flights. The included set is defined as those flights, whose origin airport is within a certain specified radius defining a circle surrounding the GDP (destination) airport. This policy has the effect of focusing delays on “short-haul” flights. To see the motivation for this policy, suppose that one had the choice between assigned a delay to a flight 4 hours away from the GDP airport vs a flight 2 hours away. The flight 4 hours away would have to serve its ground delay more than 4 hours in advance of its arrival at the airport. If the weather cleared (and capacity returned to normal), say 2 hours before the flight’s arrival at the airport, then the delay would have been served unnecessarily. On the other hand, if the delay was assigned to the flight 2 hours away, there is more flexibility so that if the weather cleared early, it might be possible to rescind the ground delay before it was served and allow the flight to depart on time. While this policy definitely does help to mitigate the effect of uncertainty in weather forecasts, it can introduce inequity by focusing delays on short-haul flights (see [16]). Recent research [4], [16] has proposed methods that tradeoff the efficiency benefits and the loss in equity.

3.3 Airspace Flow Programs

While it is commonly accepted that the principal capacity bottlenecks within the U.S. air transportation system occur at airports, airspace congestion is also a problem particularly in the presence of convective weather. The CDM community has developed a set of procedures to respond to anticipated airspace congestion. The associated traffic management initiative is called an airspace flow program (AFP) [11]. The AFP procedures generally represent a fairly direct translation of the GDP procedure to the enroute environment. Specifically, an AFP starts with the identification of a *flow constrained area (FCA)*, which is a volume of airspace where demand is predicted to exceed capacity for an extended period of time. FAA specialists define a degraded capacity for the airspace in terms of a a maximum rate of flow across the FCA boundary. With this (relatively simple) definition of capacity, arrival slots can be associated with the boundary of the FCA. By requiring flights passing through the FCA to be assigned to a specific slot the rate of flow across the FCA boundary is kept within the designated capacity limit. Once the FCA boundary slots are defined, the GDP planning procedures can be applied in a reasonably direct manner. AFP's have been in use for almost 2 years and the AFP mechanisms should serve as a solid foundation for implementation of more sophisticated processes.

4 Conclusions

The CDM principles and philosophy now represent a required component of sound air traffic flow management. In this paper, we have outlined the general CDM principles by describing the CDM mechanisms used for GDP planning and control. In fact, there are several other successful application examples both in the U.S. and Europe. We also feel that the concepts described here should have applicability beyond air traffic flow management.

5 Acknowledgements

This work was supported in part by NEXTOR, the National Center for Excellence in Aviation Operations Research, under Federal Aviation Administration cooperative agreement number 01CUMD1.

References

[1] Ball, M., C. Barnhart, G. Nemhauser and A. Odoni. 2007. Air transportation: irregular operations and control, Handbook of Operations Research and Management Science: Transportation, C. Barnhart and G. Laporte, eds, Elsevier, Amsterdam, pp 1-67.

[2] Ball, M., G. Donohue and K. Hoffman. 2005. Auctions for the safe, efficient and equitable allocation of airspace system resources, in Cramton, P., Y. Shoham and R. Steinberg, eds. Combinatorial Auctions, MIT Press, Cambridge, pp 507-538.

[3] Ball, M., R. Hoffman, D. Knorr, J. Wetherly and M. Wambsganss. 2001. Assessing the benefits of collaborative decision making in air traffic management, in Donohue G. and A. Zellweger, eds., *Air Transportation Systems Engineering*, American Institute of Aeronautics and Astronautics, Inc., Reston, VA., pp 239-250 (a preprint appeared in *Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar*, 2000).

[4] Ball, M., R. Hoffman and A. Mukherjee. 2008. Ground delay planning under uncertainty based on the raion-by-distance principle. submitted for publication.

[5] Ball, M., R. Hoffman, A. Odoni, and R. Rifkin. 2003. A stochastic integer program with dual network structure and its application to the ground-holding problem. Operations Research, 51, pp. 167 - 171.

[6] Ball, M., Lovell, D., Hoffman, R. and Mukherjee. A. 2005. Response mechanisms for dynamic air traffic flow management, Proceedings of 6th USA/Europe Air Traffic Management R& D Seminar.

[7] Ball, M. and G. Lulli. 2004. Ground delay programs: optimizing over the included flight set based on distance. Air Traffic Control Quarterly 12, 1-25

[8] Beatty, R., R. Hsu, L. Berry and J. Rome 1998. Preliminary evaluation of flight delay propagation through an airline schedule. Proceedings of the Second USA/Europe Air Traffic Management R & D Seminar, Orlando, Fl.

[9] Chang, K., K. Howard, R. Oiesen, L. Shisler, M. Tanino and M. C. Wambsganss. 2001. Enhancements to the FAA ground-delay program under collaborative decision making. Interfaces 3, 57-76.

[10] Federal Aviation Administration. 2004. FAA Order 7210.3T available at http://www.faa.gov/atpubs.

[11] Federal Aviation Administration. 2008. CDM website. http://cdm.fly.faa.gov/

[12] Niznik, T. 2001. Optimizing the airline response to ground delay programs. Presented at AGIFORS Symposium, Ocho Rios, Jamaica.

[13] Richetta, O. and A. R. Odoni. 1993. Solving optimally the static ground-holding problem in air traffic control. Transportation Science 27, pp. 228-238.

[14] Vossen, T., M. Ball. 2006. Optimization and mediated bartering models for ground delay programs. Naval Research Logistics, 53, 75–90.

[15] Vossen, T. and M. Ball. 2006. Slot trading opportunities in collaborative ground delay programs. Transportation Science, 40, 29–43.

[16] Vossen, T., M. O. Ball, R. Hoffman, and M. Wambsganss. 2003. A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs. Air Traffic Control Quarterly, 11, pp. 277-292. (a preprint was published in Proceedings of 5th USA/Europe Air Traffic Management R& D Seminar, 2003).

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

[18] Young, H.P. 1994. Equity in theory and practice. Princeton University Press, Princeton, NJ.

[19] Yu, G. 1995. An optimization model for airlines' irregular operations control. Proceedings of the International Symposium on Operations Research with Applications in Engineering, Technology, and Management, Beijing, China.

[20] Yu, G., ed. 1997. Operations research in the airline industry. Kluwer Academic Publishers, Boston, MA.