Enhancements to the FAA Ground-Delay Program Under Collaborative Decision Making

KAN CHANG
Metron, Inc.
11911 Freedom Drive, Suite 800
Reston, Virginia 20190

KEN HOWARD
DTS-920
Volpe National Transportation Systems Center
55 Broadway
Cambridge, Massachusetts 02142-1093

RICK OIESEN
DTS-56
Volpe National Transportation Systems Center

LARA SHISLER
Metron, Inc.

MIDORI TANINO
Metron, Inc.

MICHAEL C. WAMBSGANSS
Metron, Inc.

When airport arrival capacity is reduced, it may not meet the demand placed by arriving aircraft. In these cases, the FAA enacts a ground-delay program (GDP) to delay flights before they depart from their origin airports, keeping traffic at an acceptable level for the affected arrival airport. However, air-traffic managers sometimes lacked current data and a common situational awareness when running a GDP. Working with the FAA and the airline community, Metron, Inc. and Volpe National Transportation Systems Center improved the process by using real-time data exchange between all users, new algorithms to assign flight-arrival slots, and new software at FAA facilities and airlines. This paper reflects the views and opinions of the authors and does not necessarily reflect that of the FAA.

Traffic in the US National Airspace System is expected to increase by three to five percent annually for the next 15 years [FAA 2000]. And while the number of aircraft in the sky grows larger, the National Airspace System remains a static resource. The system needs to accommodate the increasing number of aircraft without compromising safety and efficiency. This is the job of the air-traffic-management system.

Two entities comprise air-traffic management: air-traffic control, which ensures safe separation between aircraft, and

0092-2102/01/3101/0057/$05.00
1526–551X electronic ISSN

INDUSTRIES—TRANSPORTATION—SHIPPING
INFORMATION SYSTEMS—DECISION SUPPORT SYSTEMS

INTERFACES **31:** 1 January–February 2001 (pp. 57–76)

traffic-flow management, which balances demand and capacity to maintain safe and efficient traffic flow. Collaborative Decision Making (CDM) is a joint government-industry initiative aimed at improving the traffic-flow-management aspect of air-traffic management by increasing the exchange of information and improving decision-support tools.

CDM really began in 1991, when the Federal Aviation Administration's (FAA's) Air Traffic Management office commissioned an analysis to measure the effects of the airlines' flight-substitution process on the efficacy of ground-delay programs (GDPs). CDM started as a very small effort among a few committed individuals in the airline community who met on their own time outside of any office. The group encountered and, through the sheer persistence of its participants, overcame numerous cultural, political, and financial obstacles during the development of the CDM program [Wambsganss 1997]. The program is now the acknowledged means of developing and modifying the traffic-flow-management system, with participants from every arm of the aviation community involved in defining procedures and automation requirements.

We could not use sophisticated or complicated optimization models.

Scheduled to become part of the FAA operational system in June 2000, the Ground-Delay Program Enhancements (GDPE) effort is a major component of CDM that began prototype operations in January 1998 at San Francisco and Newark airports. In April 1998, the operation was extended to St. Louis and La Guardia, with expansion to all US airports in September 1998.

Problem Statement

Each GDP delays carrier operations by hundreds or thousands of minutes. With anywhere from 600 to 1,000 GDPs enacted each year, the resulting inefficiency has great potential for disaster. For this reason, the CDM group decided that improvements to the GDP process were needed immediately. GDPE addressed two problems: the initial problem of reduced airport capacity and the problems of handling the reduced airport capacity.

What is a GDP?

Every airport is constrained by the rate at which it can land arriving aircraft—the arrival rate. Airlines design flight schedules in accordance with the airport-arrival rates expected during normal operations. But bad weather and other factors can lower the arrival rate so that the expected number of arriving aircraft exceeds the airport capacity (Table 1). In these circumstances, the FAA Air Traffic Control System Command Center, commonly called the FAA Command Center, must intervene, often by enacting a ground-delay program (GDP). A GDP assigns departure delays to aircraft scheduled to arrive at the constrained airport. These ground delays are less costly and safer than airborne holding delays in the airspace at the arrival airport.

The FAA Command Center does not make the decision to enact a GDP lightly. A total of thousands of minutes of delay can be incurred at each airport affected. The true problem is more severe, since the

Airport	Original arrival rate	Reduced arrival rate	Total delay	Average delay
Newark (EWR)	50	30	11,627	82
Chicago O'Hare (ORD)	100	60	16,028	55
San Francisco (SFO)	60	28	6,915	63
Boston (BOS)	60	30	6,873	58

Table 1: Ground-delay programs (GDPs) are very disruptive. In this example, four airports frequently affected by GDPs have been given typical reduced arrival rates for four hours during irregular operational days. The reduced rates result in very high total delays for the airports and average delays per flight (in minutes).

arrival delays at the initial airports do not include down-line, or propagation effects. For example, an arriving flight may share resources with several outbound flights: the flight attendants may be assigned to one departure, the crew to another, and the pilots to yet a third. Therefore, a 100-minute delay on the arriving flight may propagate to 300 minutes of delay on the three departing flights. Those delayed departures in turn propagate even more delay and so on. A poorly operated GDP or an unnecessary GDP causes even greater consequences.

Problems with the Old GDP Process

The process of enacting GDPs existed long before the CDM group was formed. The FAA monitored airports for a capacity/demand imbalance. If it found a severe imbalance, it created a ground-delay program and issued program parameters to the airlines. The system was very much a command-and-control paradigm on the part of the FAA. However, that process had many shortcomings that led to inefficient use of valuable arrival resources. The GDPE team had to address many of these problems to improve the process.

The GDP process did not give everyone the same picture of current demand, the common situational awareness upon which the CDM program is based. Both the FAA and the users of the National Airspace System lacked quality data and had disincentives that discouraged their sharing data. The FAA Command Center had only the schedules published by the airlines in the *Official Airline Guide* (OAG). The airlines create their schedules long before the day of operations. This means that the FAA Command Center had no knowledge of such adjustments to the flight schedule as cancellations and delays on the day of operations. The FAA Command Center needed the airlines' updated schedule information. By the same token, even though the FAA's picture of demand was inaccurate, it had the only view of all the airports' arrival demand and capacity. The airlines operated blindly. They planned and adjusted their schedules according to weather reports without knowing how their schedule changes would affect demand and capacity at an airport.

Data quality also suffered because of functional problems and because of the "zone of ignorance" that existed in the FAA's Enhanced Traffic Management System (ETMS). ETMS is located at Volpe National Transportation Systems Center, the entity that is also responsible for ETMS maintenance and support. ETMS is the real-time system that the FAA uses to manage traffic flow. Through its traffic situation display (TSD), it gives traffic managers information on airborne flights and the weather. ETMS predicts system demand for the current time plus 15 hours into the future, using data from a variety of sources. The first three pieces of flight information that ETMS typically received included:

(1) OAG data, which ETMS loads 15 hours before each flight's departure;
(2) A flight plan for each flight roughly an hour before departure; and
(3) A departure message received at departure time.

ETMS received no data for flights during the 14-hour period between the OAG data time and the flight plan data time, the zone of ignorance. During this period, ETMS had no knowledge of flights canceled or delayed by their carriers, causing it to erroneously predict flight information. When cancellations and delays were particularly high, the ETMS database could be rendered useless just when traffic managers most needed good data. In addition, airlines were unable to send updated schedule data to ETMS even if the disincentives to their sharing data did not exist.

To improve data quality, the CDM group had to remove the disincentives to data sharing. Under the old system for running a GDP, traffic managers allocated arrival slots according to flights' latest estimated times of arrival. If a flight reported a delay and estimated a new and later time of arrival, it would be given a greater delay than if the airline had not shared that information. A GDP would inadvertently cause additional delay for any flights already delayed by their airlines. This concept is referred to as double penalty. In addition, a GDP would not allocate an arrival slot to canceled flights at all, leaving the airline no opportunity to substitute another flight for the canceled flight. By not reporting a cancellation, an airline could wait until the GDP was enacted, then cancel the flight, and substitute another flight to arrive in the vacant slot.

Because the old GDP process did not use a distributed planning system, resource rationing and interaction between the parties involved was inefficient or nonexistent. Even exploring different GDP strategies was cumbersome. The many parameters associated with a GDP included program time, geographical scope, and airport and flight exemptions. Exploring various combinations of these parameters to determine the best program was a tedious, manual, and time-consuming process. Once FAA traffic managers decided upon a GDP strategy, they still could not use airspace resources to the fullest extent. If an airline canceled flights without substituting other flights into the vacant slots, the FAA Command Center had no mechanism for adjusting flight delays to fill in these holes. It also had limited capabilities to monitor and adjust enacted GDPs. If either the arrival rate or the demand changed, it had no tool available to reallocate the flight delays.

Technical Approach and Accomplishments

To attack the problems that existed in the GDP process and the infrastructure used to manage that process, the GDPE team applied the main tenets of the CDM program, common situational awareness and distributed planning, to ground-delay situations. The GDPE team included representatives from the major CDM participants: the FAA, airlines, industry, and academia. Metron and Volpe National Transportation Systems Center both took lead roles on the GDPE development team.

Common Situational Awareness

Using the old GDP process, everyone involved did not have the same picture of demand. The FAA lacked current schedule data for the carriers. The carriers, in turn, had no idea how their updated schedules affected airport demand and capacity balance. The GDPE team first developed procedures and tools so that the FAA and the airlines could exchange current data on a real-time basis and quickly distribute the data to the necessary parties.

Thirty-three FAA facilities now use Flight Schedule Monitor.

The first step in providing a common situational awareness was to identify the needed data elements and to integrate data from multiple sources. ETMS already provided much of this. As the real-time system the FAA used for managing traffic flow, ETMS allows traffic managers to view current and future flight data. ETMS provides traffic managers with a current view of the National Airspace System and predicts flight events so traffic managers can deal with problems efficiently before safety is compromised.

CDM Flight Data Messages

CDM participants realized that the air carriers needed an improved mechanism to send real-time schedule changes to ETMS and improve ETMS flight predictions. CDM participants, including representatives from the FAA, air carriers, Metron, and Volpe, first defined what data air carriers would be expected to provide, under what circumstances, and in what format [Howard 1996; 1997]. Air carriers can send three CDM message types to ETMS:

(1) Cancel (FX): A flight will not fly.
(2) Create (FC): A new flight that will fly.
(3) Modify (FM): Change a flight; for example, delay the flight or change the aircraft type.

Within 12 hours of a flight's departure, airlines are required to send messages for all flights that are created, canceled, or delayed for more than 15 minutes.

Integration of CDM Messages into ETMS

Once the GDPE team defined what data air carriers should send, they developed a system the carriers could use to send the data to ETMS. The challenge for the GDPE development team, specifically Volpe, was to further enhance ETMS to accept new CDM messages. Volpe also had to develop a shadow system for use in prototype operations that would not adversely affect the operational ETMS system. To add stress, the operational ETMS was already undergoing a major conversion to an open systems platform. Because CDM program concepts, like GDPE, had not been proven in the field, only limited resources were available for developing the prototype system. Despite this, GDPE progress had to continue in order for participating airlines to invest the resources required to generate the CDM messages.

The set of machines that perform the

ETMS processing is called a string. The operational ETMS has two complete, redundant strings. For CDM prototype operations, we set up a new ETMS string in the summer of 1996 called the CDM string. In the summer of 1999, we set up a second CDM string to provide redundancy. A CDM string receives all the standard ETMS data and does all the standard ETMS processing. In addition, it receives CDM messages from the air carriers and updates the databases on the CDM string with these messages. The databases on the CDM string, then, contain the best data about what is expected in the aviation system.

As part of the CDM program, the air carriers agreed to invest in developing software that would allow them to send these messages. Participating carriers have written software that reads into their flight-control databases. The software detects schedule changes that necessitate CDM messages and automatically generates and sends messages to Volpe. In large part, these messages fill in the ETMS "zone of ignorance" and provide the data whose absence previously crippled ETMS.

Data Distribution

The purpose of integrating CDM messages with ETMS was to improve its overall predictions. We still had to distribute the data to FAA and airline users so that they achieved complete common situational awareness. To do this, we defined the airport demand list (ADL). An ADL contains data for all arrivals at one airport for a 16-hour period (one hour before the current time until 15 hours ahead). ETMS generates an ADL every five minutes for each airport and sends it to anyone registered to receive it. Every ADL provides the same data for all users (although some data is masked or filtered to hide information on military flights and to avoid antitrust problems).

Air carriers and FAA traffic-management centers request ADLs for specific airports. There have been many changes to the ADL since the CDM group began prototype operations because participants constantly discover new data requirements during CDM working-group meetings [Howard 2000]. For example, in April 2000, we included flight-departure data in the ADL.

Communications Infrastructure and the Flow of Data

For the CDM scheme to work, reliable communications paths are needed for both the CDM messages and the ADLs. Air carriers first started sending CDM messages to Volpe in the summer of 1996 via the ARINC teletype line. We chose this method because it was easy to use and universally available in the aviation community.

In 1996 and 1997, however, the air carriers helped to set up an air-carrier communications intranet called the AOCnet. This network, paid for by the air carriers but managed by ARINC, started operations in June 1997 when Volpe began to send test ADLs to the air carriers. As time passed, other vendors became eligible to participate in CDM. The enlarged network, which includes AOCnet as well as the other vendor networks, is called CDMnet (Figure 1). All of the vendors are on equal footing. The CDM group defined communications protocols so that carriers could use the ARINC teletype line or CDMnet to

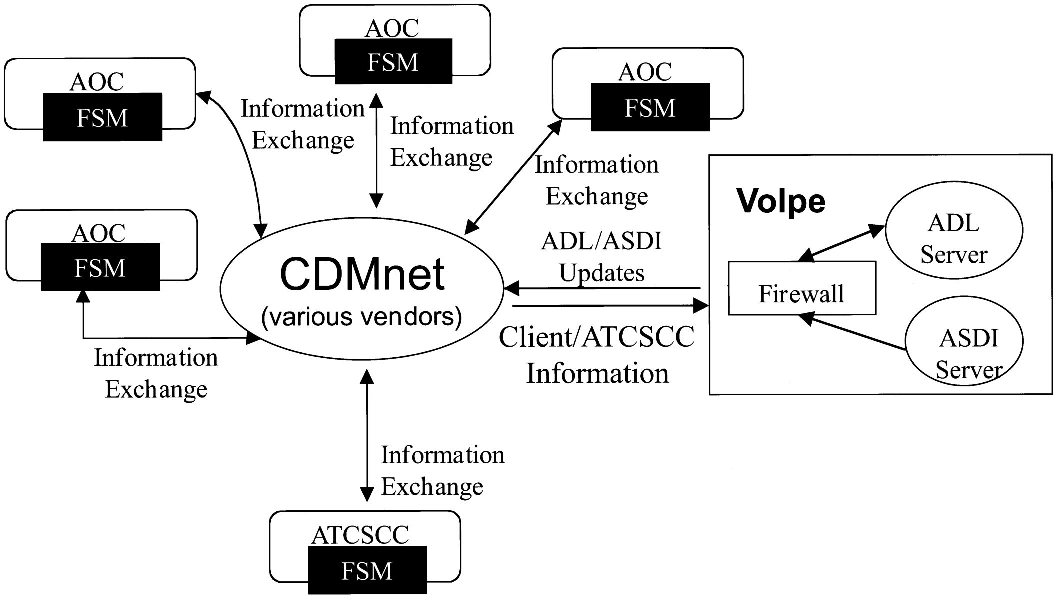

Figure 1: Through the CDMnet, airline operations-control centers (AOCs) and the FAA's Air Traffic Control System Command Center (ATCSCC) exchange pertinent air-traffic-management information through the Flight Schedule Monitor (FSM) interface. The system sends this information to Volpe, where the CDM string aggregates and issues the information as an airport demand list (ADL). Information exchanged includes carrier schedules, ground-delay-program parameters, and airport-arrival rates.

send CDM messages to Volpe.

Air carriers send CDM messages to Volpe, which incorporates the messages into the databases on the CDM string. Every five minutes, the CDM string collects all the data to make up the ADL, then sends the ADLs to those that requested them. FAA and air-carrier personnel monitor these ADLs. If traffic managers enact a ground-delay program, the program parameters are automatically sent to Volpe and forwarded to participants through the ADL. Anyone receiving the ADL and possibly affected by the proposed program has a chance to react to the program. The FAA Command Center, other FAA facilities, and the airlines use a software program, Flight Schedule Monitor, to display ADL information, monitor the airport-traffic situation, and collaborate on problems.

Flight Schedule Monitor

Our final step in providing common situational awareness was to develop a common display mechanism. Developed by Metron, Inc., Flight Schedule Monitor imports and displays ADL data, enabling all FAA and airline users to view airport demand and capacity, to list flights, to produce flight counts and statistics, and to color-code flights according to a variety of fields. Flight Schedule Monitor provides two displays: a very detailed time-line display and an aggregate bar graph. When the airport capacity line on the bar graph drops below a certain threshold, usually as a result of bad weather, it indicates a situation that could require a ground-delay program. Flight Schedule Monitor's ground-delay tools mode enables traffic managers to assess different ground-delay strategies and actually enacts ground-delay programs. Airline users apply these tools to perform what-if scenarios and to get early assessments of the effects of forecasted weather conditions on their operations.

Distributed Planning

In using the term *distributed planning*, we refer to the technical mechanisms used to allocate limited resources to the airlines and the response mechanisms by which airlines place their economically important flights within these resources. More philosophically, we are referring to the process of determining who is the appropriate decision maker and when decisions must be made with respect to ground-delay-program situations. The GDPE team had to improve the resource-rationing scheme the old GDP system used. The GDPE team also had to develop a system traffic managers could use to analyze a variety of GDP strategies quickly and enact the best program that used available resources fully. Common situational awareness could work only in conjunction with distributed planning.

Defining Roles and Responsibilities

Developing the associated procedures and technology to determine the decision-making process can be a challenging aspect of the collaborative process because it requires the users and service providers to examine the fundamental roles they have played for years. Through a series of difficult collaborative sessions and a joint experiment known as FAA-airline-data exchange (FADE) [FAA 1994], the CDM group produced the following statement of roles and responsibilities to form the basis of the GDPE program:

FAA Air Traffic Control-Traffic Flow Management will

(1) Monitor the National Airspace System for constraints that produce capacity and demand problems (for example, runway closures and weather fronts);

(2) Make these constraints known to the users of the National Airspace System (for example, air carriers, general aviation, and military aircraft); and

(3) In cooperation with the users, develop a base-line solution to the problem created by the constraint.

Each airline's operational control center will

(1) Keep FAA Air Traffic Control-Traffic Flow Management informed of the airline's current operational demand and intent; and

(2) Provide the airline's business plan within the general baseline solution provided by Air Traffic Control-Traffic Flow Management (for example, the cancellations and substitutions it plans to make in response to a ground-delay program).

Resource Allocation

Most members of the CDM community agreed to the following interpretation of the third duty of air-traffic control and traffic-flow management entities:

—It is the role of the FAA Command Center during GDP conditions to allocate arrival slots to the National Airspace System users.

Most CDM participants have agreed to the following interpretation of the second duty for airline-operations control centers:

—It is the role of National Airspace System users to determine how best to use the arrival slots they are assigned.

This interpretation leads to two classes of optimization. The FAA Command Center must allocate resources in the most efficient, yet equitable, manner possible. Meanwhile, National Airspace System users must seek to use their assigned slots optimally. They may do this through flight substitution, in which the airlines reassign flights to slots based on constraints or revenue factors or even cancel some flights to reduce the remaining delay of the other flights. Because these decisions are made under uncertain conditions, the processes of allocating resource and assigning flights to slots must be dynamic and iterative.

Crafting an efficient resource-allocation mechanism and ensuring that it meshed with the airline reassignment processes and was dynamic and iterative was only one of the challenges the GDPE team faced. The Volpe/Metron development team had to produce a mechanism that (1) removed the disincentive discouraging airlines from sending in real-time demand information, the double penalty, and (2) provided the airlines with a clear incentive to invest their resources and participate in the process.

Metron and Volpe overcame these challenges by developing two algorithms that were incorporated into the Flight Schedule Monitor software: ration by schedule (RBS) and compression. RBS eliminated any double penalty by prioritizing flights in a GDP according to their original schedule times, even when the flight was eventually canceled or delayed. Airlines no longer had to worry about a flight incurring GDP delay on top of delay already assigned to the flight. As well, RBS ensures that the airline would still have slots that it could use to substitute flights.

The intent of compression is to ensure that no arrival slots go unused and to provide an incentive for airlines to participate. Compression expands the idea of intra-airline substitution across all airlines. The algorithm processes each slot that is open because of a flight cancellation or delay. It first tries to find a flight operated by the same carrier to move up into the vacated slot. If one does not exist, it then opens the slot to the next available flight that can move up, regardless of which carrier operates that flight, but giving priority to airlines participating in CDM. The algorithm continues the process, always checking after each slot move to see whether the airline that owned the original slot can take advantage of an open slot. Compression can only either reduce or effect no change to each flight's delay. Compression has proven to be a win-win concept for both the FAA and the airlines. The airlines, especially carriers with a small presence at an airport who could not take advantage of the substitution process, can now strategically cancel flights and let compression move a late flight back on time. Their total delays can only be reduced. In addition, the FAA can now produce constant, smooth arrival rates at airports.

While we discussed and analyzed other methods, including optimization and simulation, our approach was proven to be the most appropriate for the problem. Two requirements determined the method to be used: any algorithm had to produce results almost instantaneously and had to be easy for users to understand to gain their participation in CDM. This meant we could not use sophisticated or complicated optimization models. We tested an optimization model called Optiflow. However, Optiflow was integer-programming (IP) based and took too long to run. As well, it included exactly the type of complex, modeling process that the user community did not trust and that might discourage participation. On the other hand, the RBS and compression algorithms were both simple and easily explained, and the airlines could judge whether the approach was equitable to all carriers. Both were easy to run and produced the quick results needed.

Decision Support

We designed the tool that displays the common picture of airport demand and capacity, Flight Schedule Monitor, to also act as a decision-support tool for traffic managers.

They can use the GDP setup panel in Flight Schedule Monitor to model and run ground-delay programs. The user must select many parameters to run a GDP: (1) the start and end times of the program; (2) the arrival rates; (3) any special conditions, such as whether specific equipment types need to be controlled; (4) the geographical scope, that is, which centers to include in the operation; (5) the run time of the operation; and (6) specific exemption criteria, such as exempting departure airports with severe ice conditions.

Selecting the appropriate GDP is an example of decision making under uncertainty, with each solution measured by multiple criteria. Because determining a GDP is not a straightforward single-criteria optimization problem, we concentrated on decision-support capabilities. The most notable of these is the Flight Schedule Monitor power-run feature.

Using the power-run feature, all FSM users can analyze GDPs for delay and equity. FAA traffic-management specialists use the power-run feature to analyze GDPs before enacting them. They can look at equity and delay statistics resulting from the use of different GDP parameters and determine the best program to enact. Air carriers have an economic interest in the amount of delay in a program and its fairness. Using the power run, air carriers can model a proposed GDP to study the delay and equity statistics that would result. If they disagree with the proposed program, they can use the power-run results to initiate discussion and collaboration with the FAA Command Center and other carriers. The carriers can also load data from previously enacted and completed GDPs into the power run to study historical trends of delay and equity in programs.

One GDP parameter an FAA traffic-management specialist must determine is which centers, or geographic locations, to include in a GDP operation. This is usually a function of the uncertainty associated with the weather forecast. If the forecast is extremely reliable, it is usually best to include all centers, as this usually produces the most even delay distribution among groups of users in the National Airspace System. On the other hand, if it is uncertain whether a forecast weather front will materialize and affect the airport in question, this solution would be inefficient. Flights from the East Coast to San Francisco, for example, would be delayed unnecessarily if the predicted bad weather did not materialize.

One Flight Schedule Monitor power run enables the exploration of multiple center strategies (Table 2).

Instead of using a weighted measurement of equity, the power-run analysis provides an easy-to-use list of various statistics and allows users to sort the list as desired. Weights are not associated with the power run because the statistic most important to one carrier might not be important to another.

In the power-run example in Table 2, the solution that includes all centers produces what apparently is the most equitable solution. The number of delayed flights (affected flights) is the highest at 142, and the average delay per flight is the least. Delay variability, the standard deviation of the average delays across carrier groups, is also lowest. However, with respect to other metrics, such as percent unrecoverable, this solution fairs rather

Power Run by Center-GDP
Airport: EWR
ADL Update Time: 2/15/2000 11:58Z
Start Time: 1900
Stop Time: 2259
Operation Type/Rule: RBS + +

Hour	AAR	All	First tier	Second tier	No west
1900	30	30/0/0	30/1/1	30/0/0	30/0/0
2000	30	30/0/0	32/0/3	32/0/2	31/0/1
2100	30	30/0/0	31/0/4	28/0/0	29/0/0
2200	30	31/0/1	26/0/0	31/0/1	31/0/1
#Total flights		180	180	180	180
#Affected flights		142	68	109	119
Total delay		11627	14626	12663	12108
Max delay		172	263	167	153
Avg delay		82	215	116	102
Stack		60	60	60	60
Unrecov. delay		4805	2730	3202	3406
% unrecoverable		41.33	18.67	25.29	28.13
Delay variability		38.23	54.11	42.43	42.58

Table 2: Flight Schedule Monitor's power-run feature allows users to analyze the effectiveness of ground-delay programs (GDPs) based on a variety of parameters. The table shows the results users might see after running a power run by center. In this case, a user has applied the power run to Newark airport (EWR) February 15, 2000 at 11:58 for four options: (1) all centers, (2) first tier: those centers adjacent to and including the New York center, (3) second tier: first tier plus the centers adjacent to the first tier centers, and (4) no west: all centers except those on the western part of the US. For each hour of a ground-delay program (GDP), the power run shows the flight demand, open slots, and stacked flights that would result from inclusion of a particular center group. As well, the user can weigh the various statistics to see how many flights would be included in and affected by the program, how much delay would result in a particular program, and how much of that delay would be unrecoverable should the program be canceled at the start time.

poorly. Unrecoverable delay, calculated through the time-modeling features of Flight Schedule Monitor, is the estimate of the amount of unnecessary delay that would be absorbed by the carriers if the forecasted weather conditions do not materialize. For example, fog forecast to hit San Francisco and reduce arrival capacity by 50 percent does not materialize and never affects the airport at all. If the percentage of unrecoverable delay is the most important metric, then the first-tier program is the best solution. However, the first-tier program is also the least equitable.

Other FSM power-run capabilities enable FAA specialists to consider the length of a program, the time a program is run, and various combinations of these. Delaying the run time of a program is another means of reducing its scope, since more flights will be airborne and uncontrolled by the program.

Flight Schedule Monitor has various graph and sort features that enable FAA traffic-management specialists to weigh the various criteria and select a single solution. They will often collaborate with carriers and factor in their input, such as carrier assessments of the certainty of the weather forecast or carrier plans that could significantly alter the demand at an airport. The carriers use power-run analyses as starting points for collaborating. For example, if a carrier disagrees with the weather forecast or the equity of a proposed program, it may call the FAA Command Center to voice its opinion. The power run helps everyone to be cognizant of such issues as equity and uncertainty. The carriers can replay past programs to find trends in inequity or the treatment of weather uncertainty.

In decision making under uncertainty, forecasted conditions may not materialize. The weather may improve, may be worse than forecast, or may hit earlier or later than originally predicted. Demand forecasts can be faulty. Traffic may materialize that was not predicted or scheduled. The GDPs may be executed poorly with flights departing earlier or later than their controlled departure times, creating unexpected peaks in arrival demand.

To handle such errors, users can employ several tools in Flight Schedule Monitor to adjust or reallocate arrival slots, notably the revise-GDP and adjust-delay functions. Air carriers must also possess dynamic substitution programs that enable them to efficiently reassign flights to slots once a reallocation takes place. Flight Schedule Monitor's dynamic, iterative nature and related procedures and infrastructure provide a way to deal with changing and unpredictable situations.

Performance Measurement

Performance measurement is the glue that holds the traffic-flow-management evolution process together. It provides feedback that helps users to identify and correct problems, and it also provides measures of user benefits, which are essential for retaining current participants and attracting future participants. The FAA's Free Flight Phase 1 Program Office is responsible for implementing the technologies it needs to achieve its own Free Flight initiative, including CDM. This office conducted an independent benefits analysis of the CDM GDPE project [Knorr and Wetherly 1999], from which we ob-

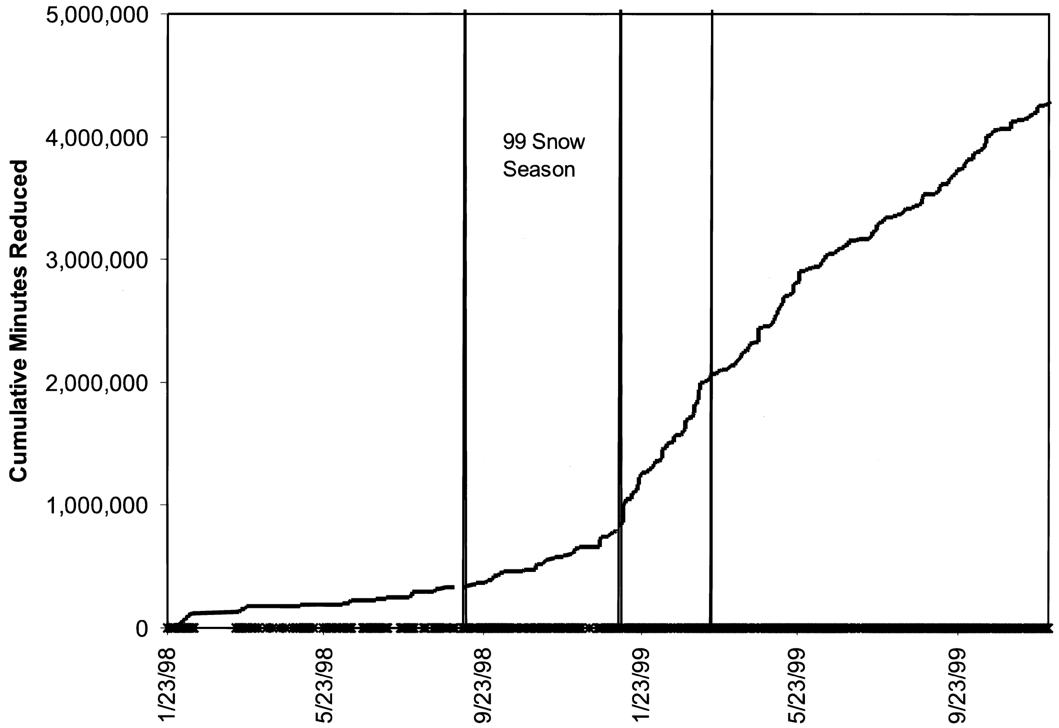

Figure 2: Over two years (January 20, 1998 to December 31, 1999) running the FSM compression function reduced delay by over 4.5 million minutes compared to the old system. These are savings in planned delay; actual savings depend on the quality of execution. The true savings may be greater, however, if we account for propagation effects.

tained many of the results we discuss.

Compression Benefits

Compression benefits are the CDM benefits easiest to quantify. Every time compression is run, an analysis program automatically computes all flight movements associated with compression and produces delay-reduction reports. The reports include information on total delay, delay by carrier, and delay by airport (Figure 2). The Air Transport Association estimates that every minute of delay costs an airline $29, thus compression yields significant savings.

EDCT Compliance

Whether flights depart at their controlled departure times or at the estimated departure clearance time (EDCT) determines the effectiveness of a GDP. Many late departures can produce peaks in arrivals that may require traffic managers to revise a GDP and assign additional delay. In the years that the FAA has tracked EDCT compliance, on average 50.85 percent of flights depart on time. Since the advent of CDM, we have seen a marked improvement in this percentage (Figure 3). As of November 1999, 65.87 percent of flights were departing on time, a percentage that continues to rise. This means that since the inception of CDM GDPE, 15.02 percent more flights maintain compliance with their controlled departure times, which implies that the number of on-time

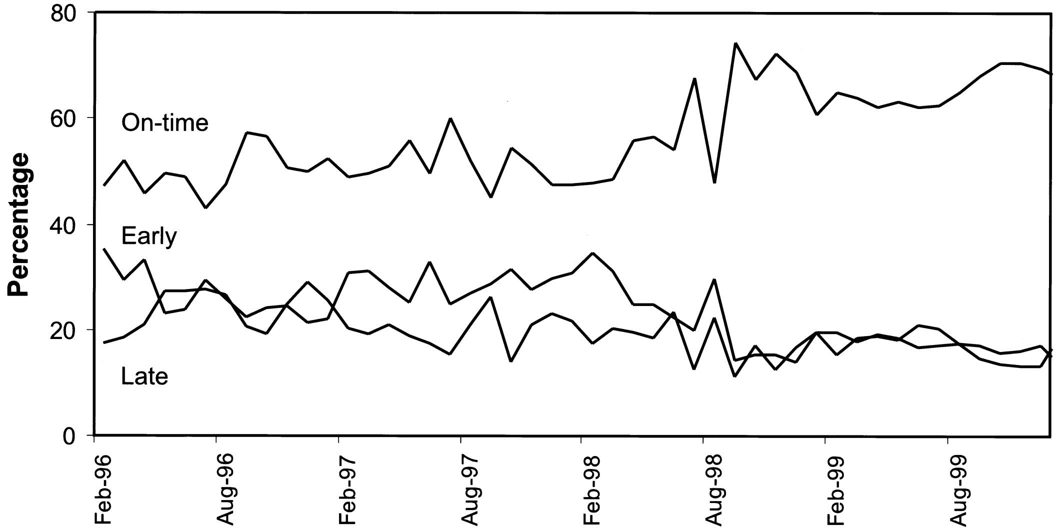

Figure 3: Since CDM was implemented at all airports in September 1998, airlines have increased their compliance with their controlled departure times. Improvements in flight departure compliance lead to improved GDP performance.

departures has grown by 31.74 percent [Knorr and Wetherly 1999].

The airline participants in CDM suggest that the factor contributing most to this improvement is the awareness achieved through a common view of the problem and their realizing the importance of adhering to the GDP plan. The airlines realize that poor departure compliance can lead to an increased number of GDPs or ground stops, in which flights are prohibited from departing for a certain period. The carriers have taken procedural steps to comply with their departure times to avoid unnecessary delay programs and thus to avoid further delays in their operations.

Integrated Predictive Error: The IPE Metric

The CDM community recognized that it needed to improve the quality of data about flight schedules to improve the execution of GDPs. Participating air carriers voluntarily augment ETMS flight data, providing such information as flight cancellations and estimated times of departure. Without this data, the FAA traffic managers were allocating arrival slots to flights that had been canceled or delayed beyond the arrival-slot time. This caused valuable arrival slots to go unused during GDPs.

Robert Hoffman, a graduate student working on CDM through the University of Maryland, defined the integrated-predictive-error (IPE) metric to monitor long-term trends in the accuracy of flight departure predictions. IPE is a weighted average of the errors in a stream of predictions made over time for a single event. Based on the data used in an initial analysis, we found that on average the accuracy of departure predictions increases (has less error) as a flight's departure time approaches. Since August 1997, the average departure-prediction error on GDP days at

San Francisco Airport (SFO) has dropped from 31.29 minutes per flight to 26.06 minutes per flight, for a net reduction of 5.23 minutes per flight. Comparable results have been found at Newark Airport (EWR).

The dramatic improvement on GDP days is noteworthy because accurate flight data is crucial during a GDP. Average IPE values for non-GDP days at the San Francisco and Newark airports have dropped as well because the participating air carriers are sharing their schedule data at all times, whether a GDP is in effect or not. For both airports, the departure-prediction error has been pushed to below 15 minutes [Knorr and Wetherly 1999].

Improved Data Quality

Overall data quality has significantly improved, in part because of the development of the CDMnet and CDM messages. Airlines provide early notice of flight cancellations and of new unscheduled flight operations. The CDM community continues to focus on improvements to data quality. Flight data is used to make many decisions about air traffic, such as GDPs. The better the data, the better the decisions and their execution. The Volpe Center has created a Web site that is available to CDM participants, the Volpe DataGate, which monitors and reports on data quality. Air carriers can use it to monitor their data quality and compare it to that of other air carriers.

The Rate-Control Index (RCI)

The rate-control index (RCI) measures the flow of air traffic into an airport and compares it to the targeted flow that the FAA Command Center traffic managers set during a ground-delay program. A single index, or percentage, is reported for the performance of a GDP on a single day. A high score (for example, 95 percent) corresponds to excellent performance, meaning the flow of traffic into the airport closely matched the targeted flow of traffic, both in quantity and in distribution. RCI is adept at flagging GDPs with particularly high or low performance.

As part of the FFP1 benefits analysis [Knorr and Wetherly 1999], the results of the RCI metric were tracked over a 30-month period for traffic arriving at San Francisco (SFO) and Newark (EWR) airports. Traffic flow into both airports improved slightly, more so at Newark than at San Francisco, meaning that the rate of flow tended to match the targeted flow more closely than it had in the past. In general, Newark tends to have more variation and lower performance than San Francisco. The FFP1 benefits-analysis team attributed this to the complexity of its terminal space (which borders on different traffic centers) and the less predictable nature of East Coast traffic. Also, the results for Newark are less conclusive than San Francisco because computing this metric depends upon modeling airborne holding, which is more difficult for Newark than for San Francisco.

Increased User Equity

The GDPE team introduced a new process, ration by schedule (RBS), for making the initial assignment of flights to arrival slots during a ground-delay program. The air carriers and the FAA worked hard to make this rationing process equitable to all parties affected by a ground-delay program. RBS rations arrival slots according to the scheduled arrival times listed in the

OAG, as opposed to amended estimated-arrival times. This removes disincentives that discouraged airlines from notifying the FAA Command Center of delays and establishes the concept of slot ownership. The group designed four metrics to assess the equity of the current arrival-slot allocation process. All of the metrics were variations on comparing the percent of total delay to the percent of total traffic for each carrier. Based on these metrics, the RBS algorithm has proven to be a fair and equitable mechanism for assigning arrival slots to flights during a GDP.

Airline Participation

Perhaps the most telling indicator of the impact of CDM is the current list of participating airlines (Table 3). The CDM process began with eight major airlines participating. We now have 36 member airlines, including all the major North American air carriers. Member airlines are those who have demonstrated the ability to send in quality real-time schedule information. A few additional carriers have begun participation by signing the MOA. While the CDM process benefits all carriers, small carriers may delay signing on because they must devote resources and programming time to initiate CDM participation.

Flight Schedule Monitor Utilization

Another indicator of the success of GDPE is its extension beyond the FAA

Collaborative Decision-Making Members (Including Subcarriers)

Commute Air
Air Canada
Air Ontario
Chataqua Airlines
Business Express
Atlantic Southeast Airlines
CC Air
ComAir
America West Airlines
American Airlines
Allegheny Airlines
Air Wisconsin Airlines
Canadian Regional
American Eagle Airlines
Air Shuttle
Atlantic Coast Airlines
Canadian Airlines
Continental Airlines
Delta Airlines
Federal Express Airlines
Great Lakes Aviation
Jet Link
Jetstream International Airlines
Mesaba Airlines
Midwest Express Airlines
Northwest Airlines
Piedmont Airlines
Southwest Airlines
Swiss Air
Trans World Airlines
Trans World Express
Union Flights
United Airlines
United Parcel Service Airlines
US Airways

Not CDM Members, but Signees of the Memorandum of Agreement

AirTran Airways
Reno Air
Skywest Airlines
Spirit Airlines

Table 3: All the major North American airlines are members of the Collaborative Decision-Making program and the list of participating airlines is growing.

Command Center. Thirty-three FAA facilities now use Flight Schedule Monitor. GDPE enables better collaboration between the FAA Command Center and local center traffic-management units (TMUs) and between the airlines and local TMUs. NavCanada now uses Flight Schedule Monitor to run GDPs at congested airports in Canada (for example, Toronto and Vancouver), and it can now properly be called the North American Flight Schedule Monitor.

Exportability and Expandability

CDM is fundamentally about change. The CDM program seeks to shift control and decision making to the most appropriate points, to create an environment in which competing entities cooperate to advance their mutual interests, and to make service operators and providers aware that they must work together to manage air traffic.

As the GDPE program moved from conception to deployment, three features proved essential, and they could easily be transferred to other types of projects:

(1) To establish common situational awareness, all parties must share the same view of the problem or constraint.

(2) To create a distributed planning environment, all parties must identify their roles and responsibilities, develop mechanisms to ration limited resources, and create procedures in which parties have a say in the rationing scheme.

(3) To measure performance, participants must establish a method for assessing benefits and providing feedback so that they can refine and improve the system.

Barnhart, Ioannou, Nemhauser, and Richardson sum up the real contribution of CDM [1999, p. 35]:

> "The sharing of information and the attendant decentralization of decision-making in system operations have given rise to a host of fascinating and hitherto unexplored theoretical and practical problems. These problems center on approaches for optimizing large-scale, distributed systems in which participants act as semiautonomous agents who share some goals (such as taking maximum advantage of available system capacity and resources) while, at the same time, also are often in competition with one another (as in the case of airlines or freight carriers). Problems in this class are extremely complex. Their solution will require major advances in knowledge and a "new breed" of Transportation Science that would draw concepts and methodologies from many disciplines, including operations research, optimal control theory, game theory, economics and simulation."

Perhaps the most telling evidence of the expandability of the CDM paradigm is the plethora of academic research focused on CDM [Ball et al. 2000; Goodhart and Yano 1999; Hall 1999].

The CDM paradigm is also being extended in the GDPE effort. Volpe is adding departure data to the ADLs in the spring of 2000. And researchers are exploring whether airport-throughput-optimization models and airport-gridlock-prediction models can improve the efficient use of US airports, beyond what was done with GDPE.

Collaborative routing is perhaps the biggest effort underway to extend the CDM paradigm. Collaborative routing focuses on the problem of managing airspace congestion, which is multidimensional and more complex than any of the problems that GDPE had to overcome. For example, in addition to a temporal component, it has a spatial component (rerouting aircraft around a constraint). The CDM community is starting with the first pillar of the

CDM paradigm, common situational awareness, and it has developed information-exchange mechanisms and displays to provide a common view of airspace-congestion problems. The CDM community has employed analysis tools with sophisticated data-mining technology to identify patterns of inefficiencies in the airspace to determine how best to focus its efforts.

Our success in the US is influencing developments in Europe. Eurocontrol, the French equivalent of the FAA, has initiated CDM research activity following the precedents successfully demonstrated by CDM GDPE.

The US Department of Defense is trying to extend many of the elements of CDM GDPE to the problem of surge logistics (the movement of masses of troops and equipment to support war-fighting initiatives, such as Desert Storm). The Defense Advanced Research Projects Agency is sponsoring a project called Virtual Airline to develop methods that exploit existing airline-command-and-control systems and connect participating airlines through a collaborative distributed system. In concept diagrams, the agency used the CDMnet and Flight Schedule Monitor as examples of the planned technology.

We can conceive of many other applications. A collaborative system could connect the manufacturing, marketing, and operations activities of a large corporation. CDM could be extended to intermodal transportation (the connection of rail, road, sea, and the air transportation systems).

CDM GDPE has clearly demonstrated the value of shifting from the old industrial-age hierarchical mode of doing business to the new information-age distributed-decision-making mode. This may be CDM's greatest legacy.

Acknowledgments

We acknowledge all those people from the airlines, the FAA, industry, and academia who have contributed to the success of CDM GDPE. Special acknowledgments go to Steve Alvania, the FAA's CDM program manager, Jim Wetherly, the FAA's technical manager and creator of the Flight Schedule Monitor power-run concept, and the original CDM participants recently recognized by the FAA Free Flight Phase I program office. Further acknowledgments go to the Volpe/Metron development team, to NEXTOR for their benefits analyses and for encouraging academic research in the area of CDM, and to the approximately 500 participants in CDM from the FAA, the airlines, academia, and industry [Metron, Inc., CDM Web Site 2000] who have so enriched this activity. We also acknowledge Laura MacDonald, our editor and technical writer who brought this paper together.

Executive summaries of Edelman award papers are presented here. The complete article was published in the INFORMS journal Interfaces [2001, 31:1, 57-76]. Full text is available by subscription at http://www.extenza-eps.com/extenza/contentviewing/viewJournal.do?journalId=5