

# A Dual Decomposition Method for Sector Capacity Constrained Traffic Flow Optimization

Dengfeng Sun, *Member, IEEE*, Alexis Clinet, *Member, IEEE*, Alexandre M. Bayen, *Member, IEEE*

**Abstract**—An aggregate air traffic flow model based on multicommodity network is used for traffic flow management in the en route airspace. A problem of minimizing the total travel time of flights in the National Airspace System of the United States is formulated as an Integer Program with billions of variables and constraints, which is relaxed to a Linear Program for computational efficiency. A dual decomposition method is applied to solve the large scale Linear Program in a computationally tractable manner. The resulting solution method achieves optimal en route delay control.

**Index Terms**—Traffic flow management, dual decomposition, large scale linear program.

## I. INTRODUCTION

THE *National Airspace System* (NAS) in the United States is a large scale, nonlinear dynamic system with a control authority which is organized hierarchically. A single *Air Traffic Control System Command Center* (ATCSCC) in Herndon, VA, supervises the overall traffic flow. Organized by geographical region, the airspace is divided into 22 (20 in the continental US) *Air Route Traffic Control Centers* (ARTCCs, or simply, Centers), controlling the airspace up to 60,000 feet [1]. Each Center is sub-divided into smaller regions, called *Sectors*, with at least one Air Traffic Controller responsible for each sector. The last few decades have witnessed the almost uninterrupted growth of air traffic. Except for a dip immediately after the tragic events of September 11, 2001, air traffic in the United States continues to grow at a steady pace. There are different growth scenarios associated both with the magnitude and the composition of the future air traffic. *Terminal Area Forecast* (TAF), prepared every year by the *Federal Aviation Administration* (FAA), projects the growth of traffic in the United States [2]. Since the main goal of Air Traffic Controllers is to maintain safe separation between aircraft while guiding them to their destinations, an imbalance between the growth of air traffic and Air Traffic Controllers' capacity (more generally, airspace capacity) poses potential issues to the air traffic control systems. Thus, it has motivated the design of a semi-automated *Air Traffic Control* (ATC) system to help Air Traffic Controllers manage the increasing complexity of traffic flow in the en route airspace.

D. Sun is with the School of Aeronautics and Astronautics Engineering, Purdue University, West Lafayette, IN, 47907, USA e-mail: dsun@purdue.edu.

A. Clinet is working for the Direction générale de l'aviation civile (DGAC). This work was done when he was a Visiting Scholar in Department of Civil and Environmental Engineering, University of California at Berkeley.

A.M. Bayen is an Assistant Professor in Systems Engineering, Department of Civil and Environmental Engineering, University of California at Berkeley.

Manuscript received February 24, 2009; revised XXX, 2009.

Using the framework of aggregate air traffic flow models, optimization for *Traffic Flow Management* (TFM) can be cast in the form of control and optimization of a dynamical system evolving on a network. In this article, given flight departures and destinations, we solve the problem of finding optimal en route delay control actions which respect sector capacity constraints while minimizing the total flight time for all the flights in the *National Airspace Systems* (NAS). Because of the size of the system and the accuracy of the model, the problem is formulated as an Integer Program with billions of variables and constraints, which is relaxed to a *Large Scale Linear Program* (LSLP) for computational efficiency. Solving the Linear Program directly is prohibitively complex due to its high dimension. A method based on dual decomposition is applied to solve the LSLP [3].

The method of dual decomposition has been used since the 1960's with the historical work of Dantzig and Wolfe [4]. A good, modern reference of dual decomposition is Chapter 6 in the book of Bertsekas [3]. The dual decomposition method has been applied in engineering, such as in rate control for communication networks [5], and to networking problems for simultaneous routing and resource allocation [6]. Recently, the dual decomposition method was presented in the article [7] in an effort towards a systematic understanding of "layering" as "optimization decomposition," where the overall communication network is modeled by a generalized network utility maximization problem, and each layer corresponds to a decomposed subproblem. Alternative decomposition methods were also applied in network utility maximization problems to obtain specific distributed algorithms [8], [9]. The dual decomposition method was also recently used to solve large computationally intractable problems for formation flight with multiple cooperative agents, which resulted in an algorithm that is easily implementable in a decentralized manner [10].

In this article, it is shown that the dual decomposition method is well suited to the multicommodity network structure of the aggregate traffic flow model, *Large-capacity Cell Transmission Model*, in short CTM(L), developed in earlier work [11]. It breaks the LSLP into a sequence of smaller Linear Program problems (subproblems), which are tractable and can be solved very efficiently in real time. It consists of an iterative algorithm, in which the subproblems are solved involving their own local variables as well as the variables of subproblems they are coupled with.

This article is organized as follows. The second section introduces the aggregate air traffic flow model. The third section formulates the optimal traffic flow management problem as a LSLP. In the fourth section, the LSLP is solved using a

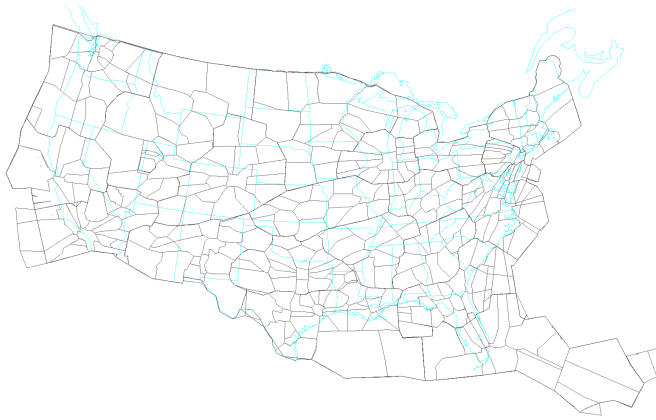


Fig. 1. Map of airspace considered in this study: continental NAS of the U.S. Small polygons represent sectors. Figure obtained using FACET [16].

dual decomposition method. Finally, in section five, we present some numerical results.

II. LARGE-CAPACITY CELL TRANSMISSION MODEL

This section briefly summarizes two earlier articles [12], [11], in which we constructed a traffic flow model used for the present article, the *Large-capacity Cell Transmission Model*, in short CTM(L) [11]. It is based on a network flow model constructed from historical *Enhanced Traffic Management System* (ETMS) and *Aircraft Situation Display to Industry* (ASDI) air traffic data [13]. This model is called Large-capacity Cell Transmission Model in reference to the *Cell Transmission Model* in highway traffic [14], [15]. The key element in the CTM(L) most relevant to the *dual decomposition* method is the underlying flow network, which is illustrated in the rest of this section, and uses a multicommodity routing network.

A. Definitions

The system modeled in this article is the continental en route U.S. NAS, of the size of 20 ARTCCs, for altitudes 24,000 feet and above (Fig. 1). The ETMS/ASDI data used in this study, provides the position and altitude of all airborne aircraft in the U.S., updated every minute. Additional information related to flight plans or other flight parameters, such as speed and heading, are also provided in the data, but were not used to build the present CTM(L).

B. Construction of the network

Network models are standard in science and engineering. A network model is composed of fundamental components called vertices or nodes; each node is connected to other nodes by links or edges. The network modeling traffic flow in the NAS in CTM(L) is constructed as follows.

**Vertices (nodes)** The vertices in the network are constructed following a paradigm developed in earlier work [11]. For any two neighboring sectors  $s_1$  and  $s_2$ , the vertices at the boundary of  $s_1$  and  $s_2$  are denoted by  $v_{\{s_1,s_2\}}$  and  $v_{\{s_2,s_1\}}$ . Vertex  $v_{\{s_1,s_2\}}$  represents the entry point used by flights going from

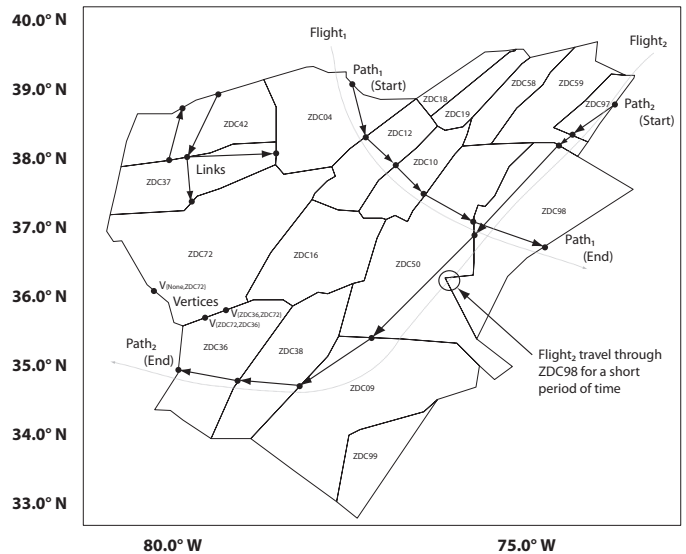


Fig. 2. Examples of vertices, links, trajectories and paths for a subset of the airspace depicted in Fig. 1.

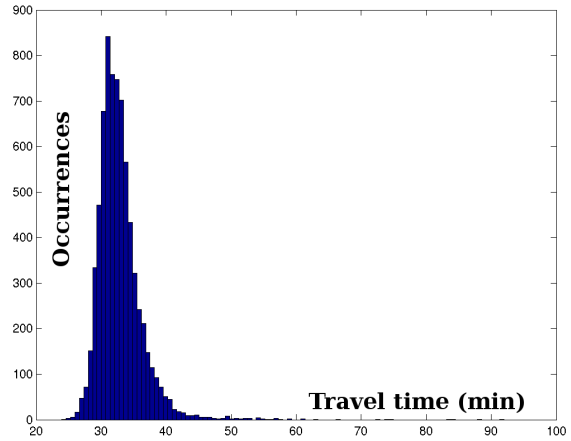


Fig. 3. Distribution of travel time on one link (ZLC45-ZOA33-ZOA34). One full year of aggregated data.

$s_1$  to  $s_2$ , while vertex  $v_{\{s_2,s_1\}}$  represents the exit point used by flights going from  $s_2$  to  $s_1$ .

**Links (edges)** For any sectors  $s_1$ ,  $s_2$  and  $s_3$ , if  $s_1$  and  $s_2$  share a boundary and if  $s_2$  and  $s_3$  are neighbors, two *directed links* are created: one from vertex  $v_{\{s_1,s_2\}}$  to vertex  $v_{\{s_2,s_3\}}$  and one from vertex  $v_{\{s_3,s_2\}}$  to vertex  $v_{\{s_2,s_1\}}$ . In the rest of this work, the term *link* refers to a directed link [17]. Fig. 2 illustrates the concept of a link.

For each link of the network, the flight times for a full year (Oct. 1st, 2004 to Sept. 30, 2005) extracted ASDI/ETMS data are aggregated into travel time distributions. The mean of this distribution is computed, and its value is chosen to represent the “time length” of the link, i.e. the aggregated travel time along the link.

Fig. 3 shows a typical travel time distribution. The expected travel time of a flight through a link is used to determine the

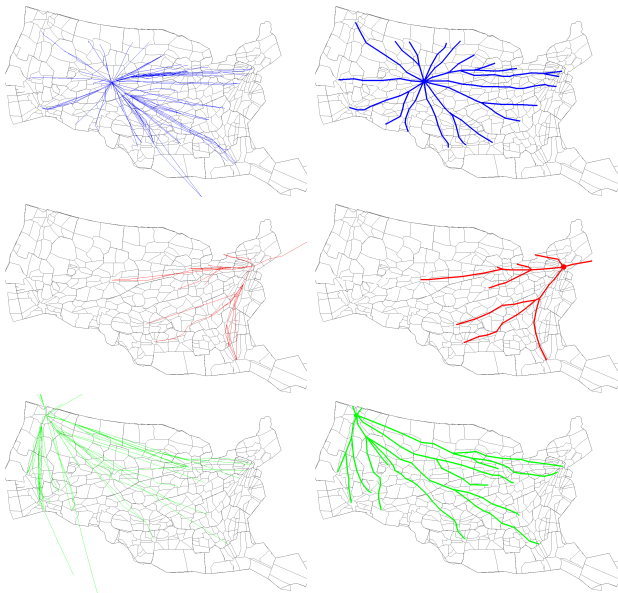


Fig. 4. An illustration of decoupled multicommodity network models by destination for airports (DEN, LGA, SEA). An aggregation of the trees corresponding to destination airports provides a complete multicommodity network level model. **Left:** recorded flight tracks; **Right:** corresponding air traffic flow trees.

length of the link. As will be seen in the subsequent sections describing the CTM(L), each link is divided into several *cells*. The number of aircraft in a cell will be used as a component of the state in the model derived below. In the present setting, cells correspond to one minute of flight time.

**Multicommodity network** For a complete network model including the whole continental NAS of the US, a multicommodity flow structure [18] is used in CTM(L). Flights are clustered based on their entry-exit node pairs (origin-destination pairs) in the network. Each pair corresponds to a *path* consisting of links between these nodes. If two or more paths have one link in common, this link will be duplicated, using a multicommodity flow structure. Fig. 4 illustrates decoupled multicommodity network models for several destination airports. These decoupled multicommodity networks are in fact trees, with air traffic flows originating from the airports in the continental US. Note that only a portion of the origin airports are shown in the figure for clarity. An aggregation of the trees for all destination airports provides a complete (NAS-wide) network model.

The complete NAS-wide network used for this study is shown in Fig. 5.

### C. Dynamics

The CTM(L) model, based on the network constructed above, is developed inspired by the *Lighthill-Whitham-Richards* theory [19], [20], and the *Daganzo's Cell Transmission Model* (CTM) [14], [15], commonly used in highway traffic modeling. The model is reduced to a linear time invariant dynamical system, in which the state is a vector of aggregate aircraft counts. The controlled input to the model is

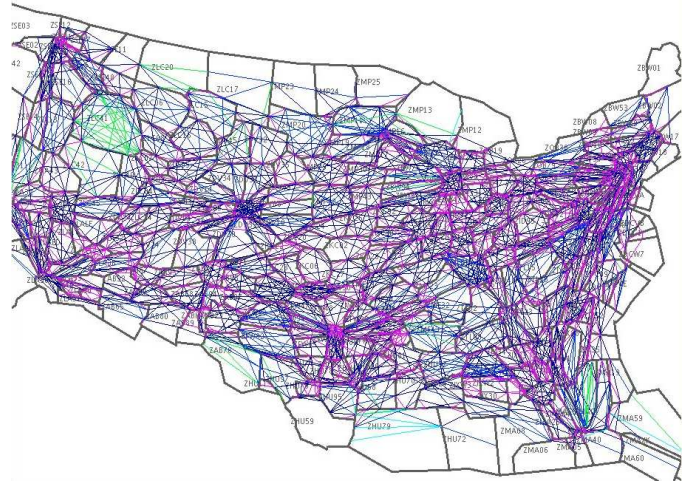


Fig. 5. Continental NAS-wide network model constructed by aggregation of the trees of Fig. 4.

airborne delay control, which can take several forms such as speed change, *vector for spacing* (VFS), *holding pattern* (HP).

The behavior of air traffic flow on a single link indexed by  $i$  can be modeled by a deterministic linear dynamical system with a unit time delay, defined as follows [11]:

$$x_i(k+1) = A_i x_i(k) + B_1^i u_i(k) + B_2^i f_i(k), \quad (1)$$

$$y(k) = \tilde{C}_i x_i(k), \quad (2)$$

where  $x_i(k) = [x_i^{m_i}(k), \dots, x_i^1(k)]^T$  is the state vector whose elements represent the corresponding aircraft counts in each cell of link  $i$  at time step  $k$ , and  $m_i$  is the number of cells in link  $i$ . The forcing input,  $f_i(k)$ , is a scalar that denotes the entry count onto link  $i$  during a unit time interval from  $k$  to  $k+1$ , and the control input,  $u_i(k)$ , is an  $m_i \times 1$  vector, representing delay control. The output,  $y(k)$ , is the aircraft count in a user-specified set of cells at time step  $k$ . The nonzero elements of the  $m_i \times 1$  vector  $\tilde{C}_i$  correspond to the cells in the user-specified set, and are equal to one.  $A_i$  is an  $m_i \times m_i$  nilpotent matrix with 1's on its super-diagonal.  $B_2^i = [0, \dots, 0, 1]^T$  is the forcing vector with  $m_i$  elements, and  $B_1^i$  is the  $m_i \times m_i$  holding pattern matrix, in which all nonzero elements are 1 on the diagonal and  $-1$  on the super-diagonal.

Based on the link level model, it is straightforward to build a sector level model using the same technique. Suppose that there are  $n$  links in a sector, then the dynamics for the sector level model can be described as:

$$x(k+1) = Ax(k) + B_1 u(k) + B_2 f(k), \quad (3)$$

$$y(k) = \tilde{C}x(k), \quad (4)$$

where  $x(k) = [x_n(k), \dots, x_1(k)]^T$  denotes the state, and  $f(k) = [f_n(k), \dots, f_1(k)]^T$  is the forcing input vector (the entry count onto the sector). The control input vector is denoted as  $u(k) = [u_n(k), \dots, u_1(k)]^T$ . The vector  $y(k)$  represents the aircraft count in a user-specified set of cells at time step  $k$ . The matrices  $A$ ,  $B_1$ , and  $B_2$  are block diagonal, with block elements associated with each link in the sector. For example,  $A = \text{diag}(A_n, \dots, A_1)$  with  $A_i$ 's defined by

equation (1). In the above model, the matrices  $A$ ,  $B_1$  and  $B_2$  are sparse and highly structured (e.g.,  $A$  is nilpotent), which can be exploited to develop efficient algorithms for optimization using this model.

When building a NAS-wide model (at the ARTCC level), flights are first clustered based on their origin-destination (source-sink) pairs in the network. Each pair corresponds to a *path* consisting of links between these nodes. If two or more paths have one link in common, this link will be duplicated. Therefore, the NAS-wide model can also be cast in the same framework of (3-4) and the corresponding  $x(k)$  includes all cells of the complete network. The forcing input,  $f(k)$ , is now the entry count into the NAS. The output,  $y(k)$ , denotes the aircraft count in a user-specified set of cells at time step  $k$ . Further details about the CTM(L) are described in [11].

### III. PROBLEM FORMULATION

#### A. Notations, nomenclature

The following notations are used to formulate the optimization problem presented in the rest of the article.

- $S = \{s_1, s_2, \dots, s_{|S|}\}$ : set of sectors in the NAS.  $|S|$  denotes total number of sectors in the NAS.
- $V = \{v_1, v_2, \dots, v_{|V|}\}$ : set of vertices, defined in Section II-B.  $|V|$  denotes total number of vertices in the NAS model constructed earlier.
- $E = \{e_1, e_2, \dots, e_{|E|}\}$ : set of links (Section II-B). Link  $e_m = (v_i, v_j)$ , or simply  $e_m = (i, j)$ , corresponds to an ordered pair of vertices.
- $K = \{k_{o_1, d_1}, k_{o_2, d_2}, \dots, k_{o_{|K|}, d_{|K|}}\}$ , or simply  $K = \{k_1, k_2, \dots, k_{|K|}\}$ : set of origin-destination (OD) pairs: origin (source)  $o_k$ , destination (sink)  $d_k$ .  $|K|$  denotes total number of OD pairs. In this article, OD pairs can also be understood as “paths,” because given an OD pair, a path is uniquely defined: no multiple paths exist between a single OD pair.<sup>1</sup>
- $T$ : time horizon of the optimization.
- $Q_s \subset E$ : set of cells in sector  $s \in S$ .  $(j, k) \in Q_s$  means the  $j$ -th cell on path  $k$  is in  $Q_s$  (therefore in sector  $s$ ).
- $x(i, j, t)$  or  $x_t^{j,i}$ : state of cells in the dynamical system.  $x_t^{j,i}$  represents the number of aircraft in the  $j$ -th cell on path  $i$  (namely cell  $(i, j)$ ; defined in Section II-B) at time  $t$ . The vector  $x_t^k$  or  $x^k(t)$  is used to denote the aggregation of states on path  $k$  at time  $t$ :  $x_t^k = [x(k, 1, t), x(k, 2, t), \dots, x(k, n(k), t)]^T$ , where  $n(k)$  is the total number of cells on path  $k$ . The vector  $x^k$  is the aggregation of states on path  $k$ :  $x^k = [x_1^k; x_2^k; \dots; x_T^k]$ . The vector  $x(t)$  or  $x_t$  is the aggregation of states at time  $t$ :  $x_t = [x_1^1; x_2^1; \dots; x_1^{|K|}; x_2^{|K|}; \dots; x_T^{|K|}]$ .  $x$  represents the vector of all the states:  $x = [x(1); x(2); \dots; x(T)]$ .
- $u(i, j, t)$  or  $u_t^{j,i}$ : delay control in cell  $(i, j)$  at time  $t$ , representing number of delay controlled aircraft in the  $j$ -th cell on path  $i$  at time  $t$ . The vector  $u_t^k$  or  $u^k(t)$  is used to denote the aggregation of controls on path  $k$  at time  $t$ :  $u_t^k = [u(k, 1, t), u(k, 2, t), \dots, u(k, n(k), t)]^T$ ,

where  $n(k)$  is total number of cells on path  $k$ . The vector  $u^k$  is the aggregation of the controls on path  $k$ :  $u^k = [u_1^k; u_2^k; \dots; u_T^k]$ . The vector  $u(t)$  or  $u_t$  is the aggregation of controls at time  $t$ :  $u_t = [u_t^1; u_t^2; \dots; u_t^{|K|}]$ .  $u$  represents the vector of all the controls:  $u = [u(1); u(2); \dots; u(T)]$ .

- $c(i, j)$ ,  $i = 1, \dots, |K|$ ,  $j = 1, \dots, n_i$  ( $n_i$  is the number of cells on path  $i$ ), means the cost associated with flying through cell  $(i, j)$ , which is the travel time of a flight through cell  $(i, j)$ .  $c(i, j) = 1$  represents one minute travel time in the present CTM(L). Let  $c^k$  represent an aggregation of costs on path  $k$ :  $c^k = [c(k, 1), c(k, 2), \dots, c(k, n(k))]^T$ , and let  $c = [c^1; c^2; \dots; c^{|K|}]$ .
- $C_s(t)$ ,  $s = 1, \dots, |S|$ : sector capacity for sector  $s$  at time  $t$ . The sector capacities are time dependent because the usage of airspace is dynamic: capacity can change due to weather or operations. Denote  $C(t) = [C_1(t), C_2(t), \dots, C_{|S|}(t)]^T$ , was the aggregation of capacity constraints at time  $t$ , and  $C = [C(1); C(2); \dots; C(T)]$  as the aggregation of all sector capacities at all times.
- Slack variables  $Z_s^k(t)$ ,  $s = 1, \dots, |S|$ ,  $k = 1, \dots, |K|$ , represent the number of aircraft in sector  $s$  on path  $k$  at time  $t$ :  $Z_s^k(t) = \sum_{(j,k) \in Q_s} x(k, j, t)$ , where  $(j, k)$  is the  $j$ -th cell on path  $k$  and  $Q_s$  is the set of links in sector  $s$ . Let  $Z^k(t)$  denote the aggregation of slack variables  $Z_s^k(t)$  on path  $k$  at time  $t$ :  $Z^k(t) = [Z_1^k(t), Z_2^k(t), \dots, Z_{|S|}^k(t)]^T$ , and  $Z^k = [Z^k(1); Z^k(2); \dots; Z^k(T)]$  are slack variables associated with path  $k$ . Let  $Z(t)$  denote the aggregation of all slack variables at time  $t$ :  $Z(t) = [Z^1(t); Z^2(t); \dots; Z^{|K|}(t)]$ . Let  $Z = [Z(1); Z(2); \dots; Z(T)]$  be the aggregation of all slack variables.
- $\mathbb{T}_0 = \{0, \dots, T-1\}$ : set of time indices from 0 to  $T-1$ .
- $\mathbb{T}_1 = \{1, \dots, T\}$ : set of time indices from 1 to  $T$ .
- $\mathbb{T} = \{0, \dots, T\}$ : set of time indices from 0 to  $T$ .
- $\mathbb{S} = \{1, \dots, |S|\}$ : set of sector indices.
- $\mathbb{K} = \{1, \dots, |K|\}$ : set of path indices.
- $\mathbb{Z}_+$ : the non-negative integer set.

#### B. Formulation of en route airspace problem

The problem of minimizing the total travel time of the flights in the NAS can be formulated as follows:

$$\min_{x, u} \sum_{t=0}^T c^T x_t \quad (5)$$

$$\text{s.t.} \quad x_0 = B_2 f_0 \quad (6)$$

$$x_{t+1} = Ax_t + B_1 u_t + B_2 f_t, \quad t \in \mathbb{T}_0 \quad (7)$$

$$\sum_{(i,j) \in Q_s} x_t^{i,j} \leq C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T} \quad (8)$$

$$u \leq x \quad (9)$$

$$x \subset \mathbb{Z}_+ \quad (10)$$

$$u \subset \mathbb{Z}_+ \quad (11)$$

<sup>1</sup>This feature can easily adapt to cases of multiple paths between an OD pair. It is omitted in this article for simplicity of the description.

The objective function (5) encodes the minimization of the total travel time for all the flights in the NAS for the time horizon of interest. Equation (6) represents the initial condition, i.e., the airborne flights at the beginning of optimization. Equation (7) encodes the dynamics of the system. Equation (8) enforces the capacity constraint for every sector, meaning that the number of aircraft in the sector cannot exceed the sector capacity. Equation (9) is the control constraint for every cell: the number of delay controlled aircraft cannot exceed the total number of aircraft in the cell. Equations (10) and (11) represent the non-negativity integer constraint on the states and controls, respectively.

*Remark:*

Using the framework above, different objective functions can be used for other optimization purposes. In particular, as long as the objective function is convex, one can find efficient algorithms to solve the optimization problem [21], [3]. Moreover, when the terms in the objective function are seperable path by path, the dual decomposition method described in the rest of this article can be applied following the same algorithm developed in Section IV. This is the main contribution of the article, which makes this large scale optimization computationally tractable.

#### IV. DUAL DECOMPOSITION

A realistic and accurate NAS model approximately includes 100,000 paths to accurately represent the high altitude flow structure of air traffic. Every path usually has hundreds of cells. In order to perform two-hour TFM ( $t = 1, \dots, 120$ ), the problem consists of about five billion states and controls ( $x(i, j, t)$  and  $u(i, j, t)$ ); the number of constraints is of the same order (billions). The formulation in Section III-B is an *Integer Program* (IP), which is computationally challenging to solve efficiently. To make the problem tractable, we relax the last two constraints (10) and (11) to  $u \geq 0$ , and as a consequence,  $x \geq u \geq 0$ , by constraint (9). The formulation now becomes a *Linear Program* (LP):

$$\min_{x,u} \sum_{t=0}^T c^T x_t \quad (12)$$

$$\text{s.t.} \quad x_0 = B_2 f_0 \quad (13)$$

$$x_{t+1} = Ax_t + B_1 u_t + B_2 f_t, \quad t \in \mathbb{T}_0 \quad (14)$$

$$\sum_{(i,j) \in Q_s} x_t^{i,j} \leq C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T} \quad (15)$$

$$0 \leq u_t \leq x_t, \quad t \in \mathbb{T}. \quad (16)$$

However, LP relaxation does not change the size of the problem: we still have the same number of variables and constraints as in the IP formulation. Now, we apply the dual decomposition method [3] to solve the large scale LP.

**Step 1** Decompose the terms path by path. The objective function can be rewritten as a summation of the total travel time of flights along each path, where the path index is denoted

by  $k$ . Each constraint can also be written path by path, which is also indexed by  $k$  for the  $k$ -th path.

$$\begin{aligned} \min_{x,u} \quad & \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^{kT} x_t^k \right) \\ \text{s.t.} \quad & x_0^k = B_2^k f_0^k, \quad k \in \mathbb{K} \\ & x_{t+1}^k = A^k x_t^k + B_1^k u_t^k + B_2^k f_t^k, \quad t \in \mathbb{T}_0, k \in \mathbb{K} \\ & 0 \leq u_t^k \leq x_t^k, \quad t \in \mathbb{T}, k \in \mathbb{K} \\ & \sum_{k=1}^{|K|} \sum_{(i,k) \in Q_s} x_t^{i,k} \leq C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T}. \end{aligned}$$

**Step 2** Introduce slack variables  $Z_s^k(t)$ ,  $Z(t)$  and  $Z$ , as defined in Section III-A.

$$\begin{aligned} \min_{x,u,Z} \quad & \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^{kT} x_t^k \right) \\ \text{s.t.} \quad & x_0^k = B_2^k f_0^k, \quad k \in \mathbb{K} \\ & x_{t+1}^k = A^k x_t^k + B_1^k u_t^k + B_2^k f_t^k, \quad t \in \mathbb{T}_0, k \in \mathbb{K} \\ & 0 \leq u_t^k \leq x_t^k, \quad t \in \mathbb{T}, k \in \mathbb{K} \\ & \sum_{(i,k) \in Q_s} x_t^{i,k} = Z_s^k(t), \quad s \in \mathbb{S}, k \in \mathbb{K}, t \in \mathbb{T} \\ & \sum_{k=1}^{|K|} Z_s^k(t) \leq C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T} \end{aligned}$$

**Step 3** Form the partial Lagrangian for the last constraints,

$$\sum_{k=1}^{|K|} Z_s^k(t) \leq C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T},$$

and express the problem in an equivalent partial-Lagrangian form as

$$p^* := \min_{x,u,Z} \max_{\lambda \geq 0} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^{kT} x_t^k \right) + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) \left( \sum_{k=1}^{|K|} Z_s^k(t) - C_s(t) \right) \quad (17)$$

$$\begin{aligned} \text{s.t.} \quad & x_0^k = B_2^k f_0^k, \quad k \in \mathbb{K} \\ & x_{t+1}^k = A^k x_t^k + B_1^k u_t^k + B_2^k f_t^k, \quad t \in \mathbb{T}_0, k \in \mathbb{K} \\ & 0 \leq u_t^k \leq x_t^k, \quad t \in \mathbb{T}, k \in \mathbb{K} \\ & \sum_{(i,k) \in Q_s} x_t^{i,k} = Z_s^k(t), \quad s \in \mathbb{S}, k \in \mathbb{K}, t \in \mathbb{T}, \end{aligned}$$

where  $p^*$  denotes the primal optimal value of the problem. The  $\lambda_s(t)$ ,  $s = 1, \dots, |S|$ ,  $t = 0, \dots, T$ , are Lagrange multipliers. The aggregation of all Lagrange multipliers (over space and time) is denoted by  $\lambda$ .

**Step 4** Switch the min and max operators, and obtain the dual problem:

$$d^* := \max_{\lambda \geq 0} \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^{kT} x_t^k \right) + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) \left( \sum_{k=1}^{|K|} Z_s^k(t) - C_s(t) \right) \quad (18)$$

$$\begin{aligned} \text{s.t. } \quad & x_0^k = B_2^k f_0^k, & k \in \mathbb{K} \\ & x_{t+1}^k = A^k x_t^k + B_1^k u_t^k + B_2^k f_t^k, & t \in \mathbb{T}_0, k \in \mathbb{K} \\ & 0 \leq u_t^k \leq x_t^k, & t \in \mathbb{T}, k \in \mathbb{K} \\ & \sum_{(i,k) \in Q_s} x_t^{i,k} = Z_s^k(t), & s \in \mathbb{S}, k \in \mathbb{K}, t \in \mathbb{T}, \end{aligned}$$

where  $d^*$  denotes the optimal value of the dual problem. Notice that the primal and dual functions are both linear. We assume that Slater's condition for constraint qualifications [21], [3] is satisfied, i.e., there exists a feasible solution  $x, u$  and  $Z$  such that the capacity constraints hold with strict inequality:

$$\sum_{k=1}^{|K|} Z_s^k(t) < C_s(t), \quad s \in \mathbb{S}, t \in \mathbb{T}.$$

This is always true in practice: when the initial airborne aircraft strictly satisfy the capacity constraints, i.e.,  $\sum_{k=1}^{|K|} Z_s^k(0) < C_s(t)$  for all  $s = 1, \dots, |S|$ ,  $t = 0, \dots, T$ , we can hold all of them in their sectors by delay controls, and apply "do-not-fly" to all other departure flights by holding them on the ground. With this assumption, the optimal values of the dual problem (18) and the primal problem (17) are equal [21], [3]. This allows us to solve the primal (17) via the dual (18).

**Step 5** Re-arrange the terms in the objective function of the dual problem (18) to group the terms path by path (details in Appendix):

$$\begin{aligned} & \max_{\lambda \geq 0} \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^{kT} x_t^k \right) \\ & + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) \left( \sum_{k=1}^{|K|} Z_s^k(t) - C_s(t) \right) \\ & = \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) + \sum_{k=1}^{|K|} d^{k*}(\lambda) \right\}, \end{aligned}$$

where

$$d^{k*}(\lambda) = \min_{x, u, Z^k} \sum_{t=0}^T c^{kT} x_t^k + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) Z_s^k(t),$$

which is actually the subproblem for path  $k$ .

**Step 6** Iterations.

- Subproblems  $d^{k*}(\lambda)$ ,  $k \in \mathbb{K}$

$$\begin{aligned} \min_{x, u, Z^k} \quad & \sum_{t=0}^T c^{kT} x_t^k + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) Z_s^k(t) \\ \text{s.t.} \quad & x_0^k = B_2^k f_0^k \\ & x_{t+1}^k = A^k x_t^k + B_1^k u_t^k + B_2^k f_t^k, \quad t \in \mathbb{T}_0 \\ & 0 \leq u_t^k \leq x_t^k, \quad t \in \mathbb{T} \\ & \sum_{(i,k) \in Q_s} x_t^{i,k} = Z_s^k(t), \quad s \in \mathbb{S}, t \in \mathbb{T}. \end{aligned} \quad (19)$$

There are  $|K|$  (number of paths) subproblems.

- Master problem

$$d^*(\lambda) = \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) + \sum_{k=1}^{|K|} d^{k*}(\lambda) \right\} \quad (20)$$

To solve the dual problem (18), or (20), we need to compute the subgradient of  $d^*(\lambda)$ . In the subgradient method used in this article, we start with initial  $\lambda = \lambda_0 > 0$ . At each iteration step  $i = 1, 2, 3, \dots$ , we compute a subgradient of the dual function:

$$g(t) = - \left( \sum_{k=1}^{|K|} Z_s^k(t) - C_s(t) \right), \quad t = 0, \dots, T.$$

Then we update the dual variable (Lagrange multiplier, in this formulation) by

$$\lambda(t) := (\lambda(t) - \alpha_i g(t))_+, \quad t = 0, \dots, T,$$

where  $(\cdot)_+$  denotes the non-negative part of a vector (i.e., projection onto the non-negative orthant), and  $\alpha_i$  is the subgradient step size rule, which is any nonsummable positive sequence that converges to zero [21]:

$$\alpha_i \rightarrow 0, \quad \sum_{i=1}^{\infty} \alpha_i = \infty.$$

In this study, we use the diminishing step size rule  $\alpha_i = 1/(i+1)$ . A proof of convergence of the above algorithm can be found in Chapter 2 of [22]. Numerous methods to accelerate the convergence can be found in literature [22], [3].

The dual decomposition method, with the subgradient method for the master problem outlined above, gives Algorithm 1 (see next page).

## V. RESULTS

The dual decomposition method was implemented in C++, with subproblems (19) solved using ILOG CPLEX Concert [23]. A one-hour TFM problem was solved for the whole continental NAS in the United States.

Figures 6–11 show the delay control situation of the sectors using the dual decomposition method at different times. Sectors with a larger controlled count are colored with a darker color.

Figures 12–14 show a comparison between sector counts for three sectors. For each of them, the counts are displayed for an

**Algorithm 1** Dual decomposition algorithm.**Inputs:**

Initial state  $x_0$ .  
 Time horizon  $T$ .  
 Inputs  $f_t, t \in \mathbb{T}$ .  
 Required sector capacity constraints  $C_s(t), s \in \mathbb{S}, t \in \mathbb{T}$ .  
 Initial  $\lambda := \lambda_0 \geq 0$ .

**Start:** Iteration number  $i = 0$ .

**repeat**

$i := i + 1$ .

Solve the subproblems (19) for  $d^{k*}(\lambda)$ , obtain  $x_t^k, u_t^k, Z^k(t), k \in \mathbb{K}, t \in \mathbb{T}$ .

Master algorithm subgradients

$$g(t) = - \left( \sum_{k=1}^{|\mathbb{K}|} Z^k(t) - C(t) \right), t \in \mathbb{T}.$$

Master algorithm update

$$\lambda(t) := (\lambda(t) - \alpha_i g(t))_+, t \in \mathbb{T}.$$

**until**  $d^*$  converges or  $i = \max$  iterations.

**Output:**  $x_t^k, u_t^k, t \in \mathbb{T}, k \in \mathbb{K}$ .

uncontrolled scenario, in which aircraft are flying according to their flight plan (when the controls  $u$  in system (3) are set to be zero) and a controlled scenario (when the controls  $u$  are optimized by the dual decomposition method). The sector capacities are also represented in the figures as a reference. As can be seen from the solution, by generating an optimal delay allocation in the NAS, the dual decomposition algorithm minimizes the total travel time of the flights in the entire airspace while respecting sector capacity constraints. For example in Figure 13, the capacity of sector ZOB26 (a high altitude sector in Cleveland ARTCC) is 18; when no delay control is applied, the number of flights in ZOB26 is above 18 after 52 minutes and can reach 22 (at 55 and 56 minutes), which exceed the sector capacity (18 flights at a time), while the number of flights stays below the sector capacity all time when delay control is applied. Figure 14 shows the situation for sector ZOB29, a neighbor sector of ZOB26 in Cleveland ARTCC, whose capacity is 17. When no delay control is applied, the number of flights in ZOB29 is always below the capacity (under-utilized). However, with an optimal delay control, the dual decomposition algorithm allocates delays in ZOB29, increasing its sector load, which reaches the sector capacity at 35-37 minutes. This is exactly how dual decomposition helps: reducing sector loads by allocating delays in under-utilized sectors (usually in under-utilized neighboring sectors).

The computation is done on a Dell Server PE1900 with a 2.33GHz Intel(R) Xeon(R) CPU E5345, 8GB RAM, running Microsoft Windows Server 2003 R2, Standard x64 Edition with Service Pack 2. The computing time for a one-hour TFM problem is about 90 minutes. With a fixed network model of the NAS, the computing time does not significantly change with different scenarios (amount of inputs, level of sector capacities, etc.). Using the fact that the dual decomposition method is highly suitable for parallel computing, the computing time can be greatly reduced when parallel computing facilities are available, which makes real-time NAS-wide TFM possible.

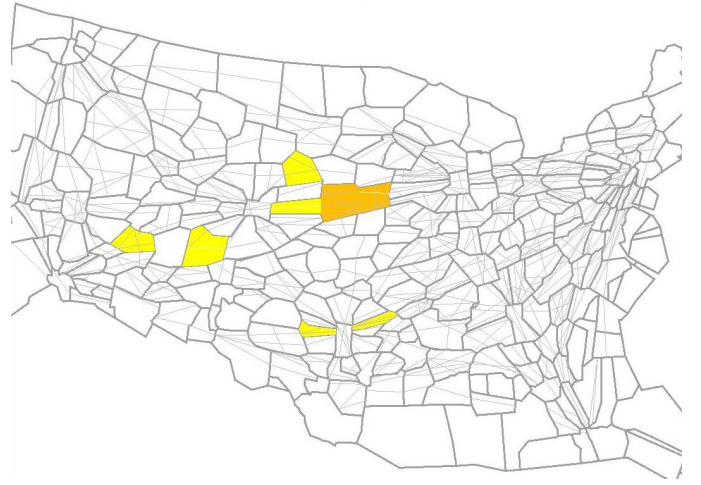


Fig. 6. Solution 9 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

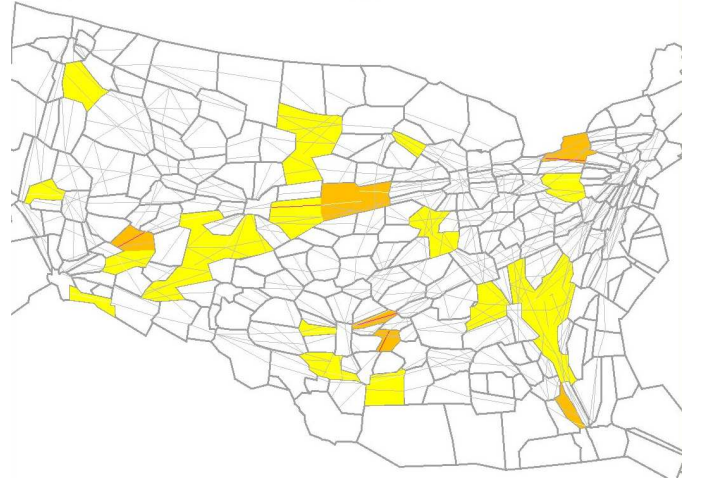


Fig. 7. Solution 19 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

## VI. CONCLUSION

In this paper, a linear time-varying aggregate traffic flow model based on multicommodity flow network is developed and implemented. Using the aggregate flow model, an optimization framework is proposed for traffic flow management problems, which are formulated as an Integer Program. The Integer Program is relaxed to a Linear Program for computational efficiency, which is solved by the proposed dual decomposition method. The program is implemented for the whole National Airspace System of the United States, which results in billions of variables and constraints. Optimal en route delay controls are obtained based on the resulting optimization solutions, which are tractable for the problem of interest.

## ACKNOWLEDGMENT

This work was supported by NASA under Task Order TO.048.0.BS.AF and NASA NRA Subtopic 4. The authors

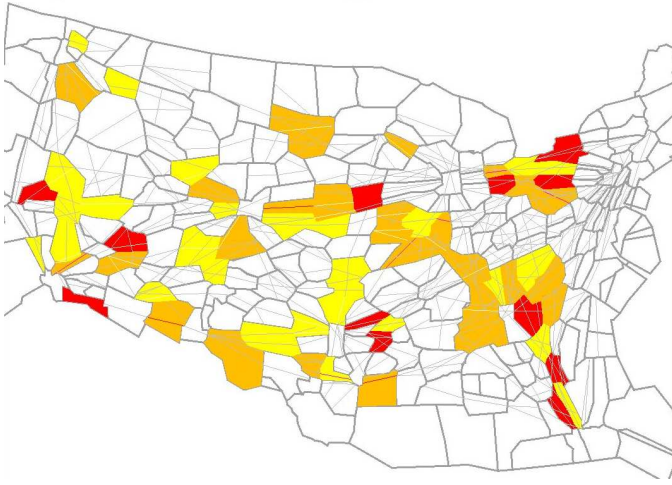


Fig. 8. Solution 29 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

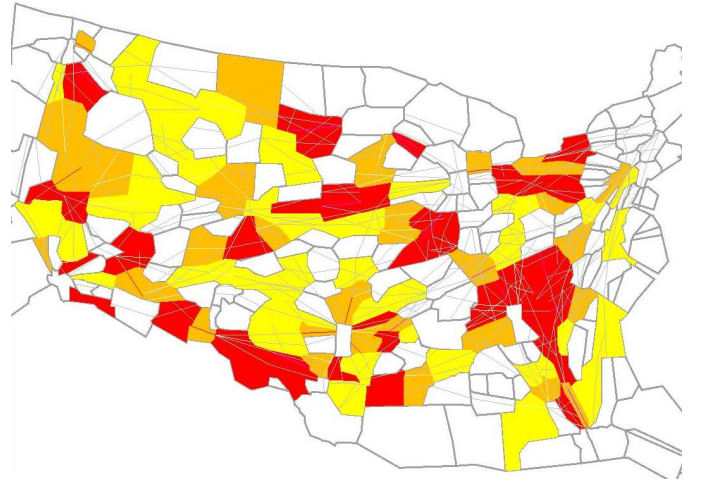


Fig. 10. Solution 49 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

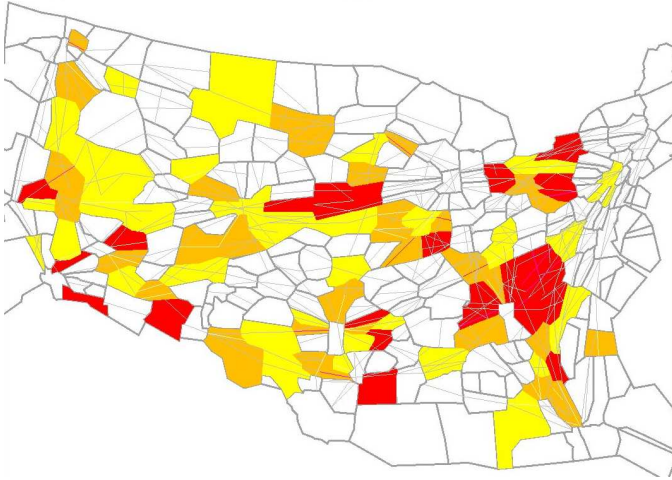


Fig. 9. Solution 39 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

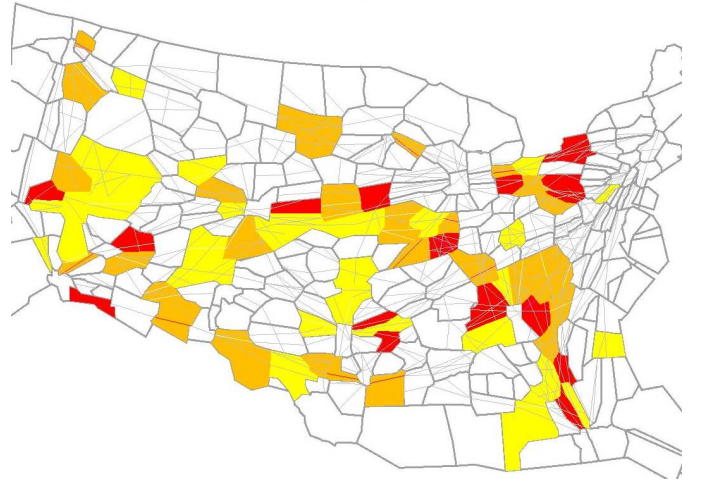


Fig. 11. Solution 59 minutes after the start time, with delay control computed by the dual decomposition algorithm. Shading of sectors indicates aircraft occupancy in each sector with control.

want to acknowledge Banavar Sridhar for his vision on aggregate traffic flow modeling, and fruitful conversations on air traffic control. The authors are thankful to Shon Grabbe and Kapil Sheth for their help with ASDI/ETMS data and FACET. The authors are grateful to Gano Chatterji for his suggestions regarding the models of the National Airspace System. Carlos Daganzo is gratefully acknowledged for the insightful discussions about his Cell Transmission Model and his suggestion to name the model presented in this study the Large-capacity Cell Transmission Model. Laurent El Ghaoui helped the authors with large scale optimization and the dual decomposition method, and is warmly acknowledged. Metron Aviation contributed to the integration of our algorithm to the software FACET developed at NASA Ames. The authors are grateful to Larry Hogle for helping with the collaboration between UC Berkeley and NASA Ames through the UARC.

## APPENDIX

Step 5 of the dual decomposition algorithm (Section IV) includes some algebraic transformations of the optimization problem, outlined here.

$$\begin{aligned}
 & \max_{\lambda \geq 0} \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^k T x_t^k \right) \\
 & + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) \left( \sum_{k=1}^{|K|} Z_s^k(t) - C_s(t) \right) \\
 = & \max_{\lambda \geq 0} \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^k T x_t^k \right) \\
 & + \sum_{t=0}^T \sum_{s=1}^{|S|} \sum_{k=1}^{|K|} \lambda_s(t) Z_s^k(t) - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t)
 \end{aligned}$$

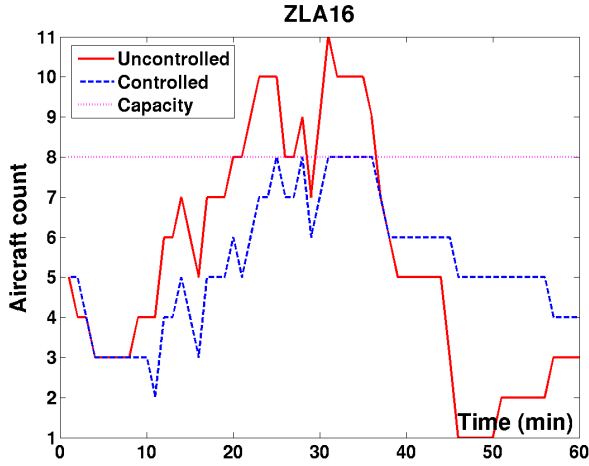


Fig. 12. Comparison of controlled and uncontrolled sector counts in Sector ZLA16. The sector capacity is set to be eight.

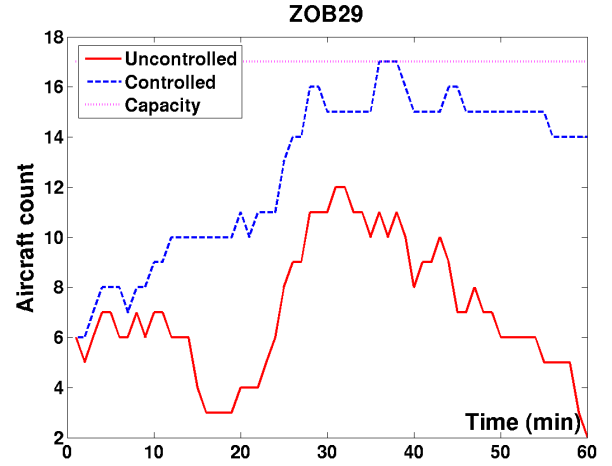


Fig. 14. Comparison of controlled and uncontrolled sector counts in Sector ZOB29. The sector capacity is set to be 17.

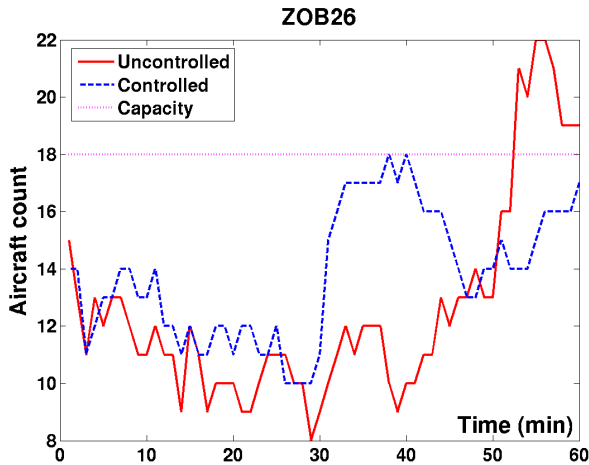


Fig. 13. Comparison of controlled and uncontrolled sector counts in Sector ZOB26. The sector capacity is set to be 18.

$$\begin{aligned}
 &= \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) \right. \\
 &\quad \left. + \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^k x_t^k \right) + \sum_{k=1}^{|K|} \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) Z_s^k(t) \right\} \\
 &= \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) \right. \\
 &\quad \left. + \min_{x, u, Z} \sum_{k=1}^{|K|} \left( \sum_{t=0}^T c^k x_t^k + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) Z_s^k(t) \right) \right\} \\
 &= \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) \right. \\
 &\quad \left. + \sum_{k=1}^{|K|} \left( \min_{x, u, Z^k} \sum_{t=0}^T c^k x_t^k + \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) Z_s^k(t) \right) \right\}
 \end{aligned}$$

$$= \max_{\lambda \geq 0} \left\{ - \sum_{t=0}^T \sum_{s=1}^{|S|} \lambda_s(t) C_s(t) + \sum_{k=1}^{|K|} d^{k*}(\lambda) \right\}.$$

## REFERENCES

- [1] M. S. Nolan, *Fundamentals of Air Traffic Control*, 4th ed. Reading, MA: Brooks Cole, 2003.
- [2] Anonymous, "Terminal area forecast summary, fiscal year 2004-2020," U.S. Department of Transportation, Federal Aviation Administration, Tech. Rep. FAA-APO-05-1, March 2005.
- [3] D. Bertsekas, *Nonlinear programming*, 2nd ed. Belmont, MA: Athena Scientific, 2002.
- [4] G. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, pp. 101-111, 1960.
- [5] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237-252, 1997.
- [6] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Transactions on Communications*, vol. 52, no. 7, pp. 1136-1144, July 2004.
- [7] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: a mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255-312, January 2007.
- [8] C. Tan, D. Palomar, and M. Chiang, "Distributed optimization of coupled systems with applications to network utility maximization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, Chicago, IL, May 2006, pp. 981-984.
- [9] D. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: framework and applications," *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2254-2269, December 2007.
- [10] R. Raffard, C. Tomlin, and S. Boyd, "Distributed optimization for cooperative agents: application to formation flight," in *Proceedings of the IEEE Conference on Decision and Control*, Nassau, Bahamas, December 2004, pp. 2453-2459.
- [11] D. Sun and A. Bayen, "Multicommodity Eulerian-Lagrangian Large-capacity Cell Transmission Model for en route traffic," *AIAA Journal of Guidance, Control and Dynamics*, vol. 31, no. 3, pp. 616-628, 2008.
- [12] D. Sun, I. Strub, and A. Bayen, "Comparison of the performance of four Eulerian network flow models for strategic air traffic management," *AIMS Journal on Networks and Heterogeneous Media*, vol. 2, no. 4, pp. 569-595, 2007.
- [13] A. M. Bayen, R. L. Raffard, and C. J. Tomlin, "Adjoint-based control of a new Eulerian network model of air traffic flow," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 5, pp. 804-818, 2006.
- [14] C. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research*, vol. 28B, no. 4, pp. 269-287, 1994.

- [15] ———, “The cell transmission model, part II: network traffic,” *Transportation Research*, vol. 29B, no. 2, pp. 79–93, 1995.
- [16] K. Bilimoria, B. Sridhar, G. Chatterji, K. Sheth, and S. Grabbe, “FACET: Future ATM concepts evaluation tool,” in *Proceedings of the 3<sup>rd</sup> USA/Europe ATM 2001 R&D Seminar*, Naples, Italy, June 2001.
- [17] R. K. Ahuja, T. L. Magnati, and J. B. Orlin, *Network flows: theory, algorithms, and application*. Upper Saddle River, NJ: Prentice Hall, 1993.
- [18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [19] M. J. Lighthill and G. B. Whitham, “On kinematic waves. II. A theory of traffic flow on long crowded roads,” *Proceedings of the Royal Society of London*, vol. 229, no. 1178, pp. 317–345, 1956.
- [20] P. I. Richards, “Shock waves on the highway,” *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. New York, NY: Cambridge University Press, 2004.
- [22] N. Shor, K. Kiwiel, and A. Ruszcayński, *Minimization methods for non-differentiable functions*. New York, NY: Springer-Verlag, 1985.
- [23] <http://www.ilog.com/corporate/training/acrobat/CPLEX.pdf>.