A Hierarchical Flight Planning Framework for Air Traffic Management
Wei Zhang, Member, IEEE, Maryam Kamgarpour, Student Member, IEEE, Dengfeng Sun, Member, IEEE, and Claire J. Tomlin, Fellow, IEEE

Abstract—The continuous growth of air traffic demand, skyrocketing fuel price, and increasing concerns on safety and environmental impact of air transportation necessitate the modernization of the air traffic management system (ATM) in the United States. The design of such a large-scale networked system that involves complex interactions among automation and human operators poses new challenges for many engineering fields. This article investigates several important facets of the future ATM system from a systems-level point of view. In particular, we develop a hierarchical decentralized decision architecture that can design 4D (space + time) path plans for a large number of flights while satisfying weather and capacity constraints of the overall system. The proposed planning framework respects preferences of individual flights and encourages information sharing among different decision makers in the system, and thus has a great potential to reduce traffic delays and weather risks while maintaining safety standards. The framework is validated through a large-scale simulation based on real traffic data over the entire airspace of the contiguous United States. We envision that the hierarchical decentralization approach developed in this article would also provide useful insights into the design of decision and information hierarchies for other large-scale infrastructure systems.

Index Terms—Air traffic control, Cyber-physical systems, Path planning, Hierarchical systems, Decentralized optimization.

I. INTRODUCTION

The Air Traffic Management (ATM) System of the United States is a large-scale, safety-critical, Cyber-Physical System (CPS), which involves more than 50,000 flights daily, being monitored and regulated by thousands of air traffic controllers based on weather and traffic measurements obtained through numerous weather stations and radar facilities located across the entire country. The physical component in the system consists of a network of fast-moving aircraft, whose coordination and control rely crucially on many cyber components such as weather and traffic prediction algorithms, decision-support software, and radio communications among pilots, dispatchers in Airline Operation Centers (AOC), and air traffic controllers.

Since its birth in 1920s, the ATM system has gradually evolved from its primitive form that consisted of a set of simple operation rules to its current version that is a complex network of sensing, communication and control subsystems. Although various automation systems have been continuously introduced, the backbone of the current system was formed during the 1950s when the introduction of radar surveillance and radio communication technologies revolutionized the way the system was operating [1]. After more than half a century, another major evolution of the ATM system to incorporate modern sensing and information technologies is currently underway [2] in order to address the growing concerns about its operational and energy efficiency, environmental impacts, and safety.

A. Notable Issues in Current ATM

Air traffic demand in the past 20 years has grown by over 64%, while human traffic controllers and airspace resources such as airports and runways have not kept up with this growth rate [3]. It has been estimated that domestic air traffic delays in 2007 cost the US economy about $41 billion, including more than $19 billion in direct operating costs [4]. The delays also contributed to about 740 million extra gallons of jet fuel, and an additional emission of about 7.1 million tons of carbon dioxide. The situation will be further aggravated by the expected two- to three-fold increase in air traffic demand over the next two decades [5]. Meanwhile, the need to constrain the rapid growth of aviation’s impact on the global climate is becoming increasingly clear. Global carbon dioxide emissions from aircraft grew about 45% between 1992 and 2005. It has also been forecasted that aviation emissions will increase an additional 150% above the 2006 level by 2036 [6], [7].

Aside from being strained by the current levels of demand, air transportation safety has recently been compromised to a level requiring immediate attention. It was reported that the number of certified air traffic controllers in 2008 reached the lowest level in 15 years, causing many major air traffic control (ATC) facilities to declare staffing emergencies. The lack of enough air traffic controllers has caused an upsurge of operational errors, many of which could have turned into major accidents. For example, the plane crash at Lexington, Kentucky, on August 27, 2006, was partly due to the fatigue of the sole Tower controller causing him to not respond promptly to the incorrect runway use. As a more recent example, the sole Tower controller in the Ronald Reagan Washington National Airport (DCA), who was on his fourth straight overnight shift on Mar 23, 2011, fell asleep, forcing two commercial aircraft to land without assistance.

The efficiency and safety issues make the system even more vulnerable to weather disruptions. It was found that hazardous weather such as storms or high winds accounted for more than 70% of the air traffic delays in 2007 [4] and about 30% of all aviation accidents [8]. Although various weather products with increasing accuracy and resolution are...
continuously being developed, their use in ATM is still in its infancy and depends heavily on the experience of human controllers. A more efficient and systematic way to utilize weather forecast to support operation planning and traffic control requires substantial research efforts.

B. Next Generation of ATM as a Cyber-Physical System

Clearly, the current air transportation system is approaching its capacity and safety limits. To resolve its existing issues as well as to be able to support the rapid growth of traffic demand, we need to automate the system at a faster pace. It is a common vision shared by FAA, air traffic controllers, and airline companies that the efficiency, environmental and safety concerns of the current ATM system can be considerably alleviated by properly incorporating modern sensing and information technologies to enable reliable communications, real-time common situation awareness, and prompt safety-guaranteed decision supports. Such a vision clearly coincides with the CPS viewpoint of modern engineering systems, and is currently being implemented through the concept of Next Generation Air Transportation System (NextGen) in the United States [2]. The NextGen concept advocates for an evolution from the current ground-based navigation system to a satellite-based ATM system, where verbal communications and ground radar systems are replaced with more reliable and accurate data-link communications and Global Positioning Systems (GPS), so that many traffic control tasks can be handled (semi)-automatically. In addition, the increased automation in conflict detection and resolution would facilitate 4D (space-time) Trajectory Based Operations (TBO), in which individual flights would have freedom to adjust their trajectories according to real time traffic and weather conditions, rather than having to follow fixed nominal flight plans as in the current clearance-based operations [9].

The implementation of NextGen not only calls for an integrated engineering effort, but also poses new challenges in many research disciplines as listed in Figure 1, including transportation engineering, human factors, communications, sensor networks, cyber security, operation research, and machine learning [2]. Information sharing has already played an important role in the current system. For example, many pilots report the actual weather conditions experienced in flight to air traffic controllers through verbal communications. These reports are being updated continuously (see Figure 2 for a snapshot of these reports) and used heavily for making aviation decisions [10]. With the vision of NextGen, one can imagine that the future air transportation system will soon evolve from this verbal information sharing platform to an automatic “sensing and action web” in the sky, in which on-board sensed data are disseminated through secure communications, and are processed and analyzed in real time to provide reliable decision support for ATM.

It is also important to notice that the modernization of this kind of legacy and safety-critical system will have to take many intermediate steps. Any proposed solution set for NextGen must respect the way the system is currently operating, and be able to work with both old and new frameworks and technologies.

C. From Centralized to Hierarchical Decentralized Planning

Among the various visions of NextGen, this work investigates the en-route flight planning problem of the future ATM system. The problem is concerned with modifying, adjusting, or even fully designing scheduled flight plans (represented by high-altitude cruise waypoints) to meet en-route airspace capacity constraints and weather restrictions. The planning decisions rely crucially upon two classes of information: traffic prediction and weather forecast. The highly-regulated nature of the air transportation system enables reliable predictions for the future traffic distributions based on the approved flight path plans. The predictions can be facilitated by Partial Differential Equation (PDE) models [12], [13], [14] that explicitly consider the spatial-temporal evolution of the traffic flow dynamics. On the other hand, numerous existing weather products can be used (mostly manually) to identify hazardous regions or quantify capacity drops over the affected parts of the airspace [15], [16]. With these traffic and weather outlooks, the traffic management decisions are often made through assistance from centralized optimization algorithms subject to appropriate constraints.
Many centralized traffic management methods suffer from complexity and scalability issues. They are often only capable to determine ground delays [17], [18], [19] rather than designing the entire flight path. As such, modifications to flight plans to better utilize airspace resources are not explored. In addition, a centralized approach in general assumes a universal cost metric such as the total delay of all flights, which ignores the operation preferences of individual flights. For example, a large commercial flight with many passengers onboard may decide to take a relatively long detour to reduce departure delays as much as possible, while a small flight or a private plane would rather delay the departure until a shorter path becomes available.

Moreover, it is difficult for a centralized traffic planning approach to explicitly consider different aircraft limitations in handling hazardous weather. A typical idea shared by most weather-aware flight planning strategies [20], [21], [22], [23] is to treat hazardous weather regions from certain weather products and treat these regions as common obstacles to be avoided by all flights during the planning process. However, in reality, the effects of weather on different flights could be substantially different [24]. For example, light or moderate icing conditions are important for intermediate-size aircraft with reciprocating engines and straight wings, but will in general not affect large commercial aircraft. Moreover, even with the same type of aircraft, experienced pilots may decide to fly through regions with adverse weather conditions while inexperienced ones would not.

As the demand continuously increases, the need for a shift from the current centralized ATM system to a more distributed traffic management architecture becomes more apparent. To enable such a shift, many research agendas have been proposed. For example, the traffic flow management problem is formulated as a multi-player game played by airline companies in [25], and a market mechanism is then designed with a provable convergence to an equilibrium depending on the cost metrics of the airline companies. Alternatively, the concept of credit points is introduced in [26] to specify flight priorities, which allow the ATM system to assign delays to flights that are relatively less important to an airline. Although both approaches incorporate airlines’ preferences on existing paths, optimizing individual path plans to further improve the overall performance is not considered.

The increased information exchange in NextGen would provide numerous opportunities to improve efficiency and address individual preferences. To design this future ATM system, a critical step is to determine how much responsibility should be distributed to the users¹ and how to achieve such a transition in a reliable way. This paper presents one set of preliminary results towards a better understanding of these important questions. Our main contribution is the development of a hierarchical decision and information architecture for large-scale en-route traffic planning, which fully respects preferences of individual flights and systematically considers both weather risks and en-route capacity constraints. With the proposed architecture, the overall functionality of the ATM system is decomposed into two interactive decision layers: traffic regulation and performance optimization. The regulation layer is responsible for computing traffic and weather predictions and setting up traffic regulation rules based on these predictions, while the optimization layer optimizes the cost functions of individual flights subject to the regulation rules imposed by the regulation layer. Through this hierarchical decomposition, the performance optimization task can be accomplished in a fully decentralized way and can be distributed to individual users without violating safety constraints. This gives each user full freedom to make its preferred decisions subject to traffic regulations, which may greatly improve its operational efficiency and passenger satisfaction. In addition, the proposed flight planning framework can design the entire 4D flight path plans, represented by sequences of waypoints and the corresponding time stamps, instead of just computing the ground delays. This is certainly in line with the TBO initiative in NextGen.

The rest of the paper proceeds as follows. The general 4D flight planning problem is formulated in Section II. A decentralized solution to this problem is developed in Section III and its advantages and practical implications are carefully discussed in Section IV. Simulation results based on real air traffic data of the US airspace is presented in Section V and some concluding remarks are given in Section VI.

II. 4D Flight Planning Problem

A. Background

The en-route airspace of the US is covered by a network of airways as shown in Figure 3, which were historically designed to connect ground-based navigational aids, such as radio beacons so that pilots could easily check where they are. With technological advances, the pilot may request to fly directly over geographical waypoints, which are points in airspace identified only by their latitude, longitude, and altitude, while reporting their locations periodically to the air traffic controllers. To simplify traffic control tasks and divide responsibility, the en-route airspace is divided geographically into 22 Air Route Traffic Control Centers (hereby referred to as Centers), and each Center is further divided into approximately 20 en-route sectors. The maximum number of aircraft allowed in a given sector is referred to as sector capacity, which depends on the weather conditions and the number of human controllers. For high-level path planning purposes, it is often assumed that once the sector capacity constraints are satisfied, local separation of aircraft can be accomplished by the human controllers in the corresponding sector.

B. Graph Model of En-Route Airspace

We consider a flight path planning problem over a bounded and connected subregion $\mathcal{X} \subset \mathbb{R}^3$ of the en-route airspace, which represents either a Center, a collection of Centers, or the entire en-route airspace of the US. The airway structure within $\mathcal{X}$ is described as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node on the graph, $v_i \in \mathcal{V}$, represents a waypoint with a given

¹In the rest of this paper, the term “user” is used interchangeably with “flight”, which refers to the planning decision maker for a flight, including the pilot in command as well as the flight dispatcher at the corresponding airline operation center.
latitude, longitude and altitude, while each link $\text{link}(v_i, v_j) \in \mathcal{E}$ represents a directed airway, also referred to as jet route at high altitudes, from waypoint $v_i$ to $v_j$. The waypoints we consider could represent both locations of navigational aids as well as geographical (virtual) waypoints with temporary positions introduced to assist flight planning or monitoring.

Suppose that the airspace region $\mathcal{X}$ consists of $n_s$ non-overlapping sectors $\{S_m\}_{m \leq n_s}$. Denote by $\mathcal{I}_S = \{1, \ldots, n_s\}$ the index set for the sectors. Let $\beta : \mathcal{V} \rightarrow \mathcal{I}_S$ be the function that assigns each node on the graph to its corresponding sector, i.e., for any $v \in \mathcal{V}$, $\beta(v) = m$ if and only if $v \in S_m$. Each sector $S_m$ is a bounded subset of $\mathcal{X}$ and is associated with a maximum capacity $c_m \in \mathbb{Z}_+$. Although the capacity $c_m$ may be time-varying based on conditions such as weather, for simplicity, we do not make this time variation explicit in the notation. It is required that the number of aircraft within each sector be less than or equal to the corresponding sector capacity at any time instant.

Fig. 3. Existing navigational aids, represented by solid circles, and a subset of the airways, represented by line segments connecting the solid circles, over the contiguous United States. Here the background polygons represent high-altitude air traffic control sectors. Data courtesy: Prof. Dominick Andrisani, School of Aeronautics and Astronautics, Purdue University.

C. Flight Model for High-Level Planning

The main function of most commercial flight planning software is to compute several candidate paths, each represented by a sequence of waypoints and airways connecting the departure and destination airports, and an estimate of the corresponding fuel consumptions. Before departure, one of the candidate paths is chosen by the pilot according to certain criteria and downloaded to the onboard flight management system (FMS) to generate a detailed trajectory to guide the pilots or autopilots. Through this hierarchical procedure, the planning algorithm need not consider the detailed aircraft dynamics when planning the high level path. This is no longer the case for 4D path planning, as one needs to specify not only the sequence of waypoints along the path, but also the time stamps at which the flight reaches these waypoints. In this case, certain knowledge of the aircraft physical model is needed to ensure that the aircraft will indeed be able to reach each waypoint at the desired time instant. For this reason, we assume that the flight planning algorithm knows the aircraft type and its associated speed range.

Let $\mathcal{A} \triangleq \{1, \ldots, n_a\}$ be the set of available aircraft types. Each type of aircraft $\alpha \in \mathcal{A}$ is characterized by its corresponding maximum/minimum speeds $s_\alpha^+ / s_\alpha^-$. Consider $N \in \mathbb{Z}_+$ flights to be completed within the overall discrete planning horizon $\mathcal{T} \triangleq \{1, \ldots, T\}$, with $t \in \mathcal{T}$ corresponding to the real time $t \cdot T_s$ for some discrete interval $T_s$. Let $\mathcal{I}_F = \{1, \ldots, N\}$ denote the index set for the flights. Each flight $i \in \mathcal{I}_F$ is associated with the following parameters:

<table>
<thead>
<tr>
<th>Flight $i$ Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>- aircraft type $\alpha^i \in \mathcal{A}$</td>
</tr>
<tr>
<td>- scheduled departure time $t_0^i \in \mathcal{T}$</td>
</tr>
<tr>
<td>- maximal allowable flight time $\tau^i \in \mathcal{T}$</td>
</tr>
<tr>
<td>- initial location $x_0^i \in \mathcal{V}$</td>
</tr>
<tr>
<td>- destination location $x_f^i \in \mathcal{V}$</td>
</tr>
<tr>
<td>- planning horizon $\mathcal{T}_i \triangleq [t_0^i, t_0^i + \tau^i]$</td>
</tr>
</tbody>
</table>

Let $x(t) \in \mathcal{V}$ denote the location of flight $i$ at discrete time $t \in \mathcal{T}$. For the flight planning problem, the decision to be made at each time step is the waypoint to reach at the next time step. Different types of aircraft may have different sets of reachable waypoints over one unit of time. Let $\mathcal{U}(v, \alpha)$ be the set of reachable waypoints at the next time step if the current flight location is $v \in \mathcal{V}$ and the aircraft type is $\alpha \in \mathcal{A}$. Assume that $v \in \mathcal{U}(v, \alpha)$, that is, an aircraft can always stay at the same waypoint over two consecutive time steps. This corresponds to
a delayed departure if \( v = x^i_0 \), and a holding pattern otherwise. In addition, assume that \( U(v, \alpha) \) always contains waypoints other than \( v \), that is, \( U(v, \alpha) \setminus \{ v \} \neq \emptyset \). This can always be satisfied by inserting virtual waypoints when needed. For example, suppose that flight \( i \) is at waypoint \( v \) at time \( t \) and the airway structure around \( v \) is shown in Figure 4 with circles representing the named waypoints and triangles representing the virtual waypoints. Then due to the maximum speed of aircraft \( a^i \), the set of reachable waypoints at time step \( t \) may only contain the ones corresponding to the solid circles and triangles.

With this notation, the evolution of the trajectory of flight \( i \) over horizon \([t_0, t_0 + \tau] \) is given by:

\[
x^i(t + 1) = f^i(x^i(t), u^i(t)) \triangleq \begin{cases} 
  u^i(t), & \text{if } x^i(t) \neq x^i_f \\
  x^i_f(t), & \text{if } x^i(t) = x^i_f
\end{cases},
\]

\[
x^i(t_0) = x^i_0 \text{ and } u^i(t) \in U(x^i(t), \alpha^i).
\]

We call \( u^i = [u^i(t_0), u^i(t_0 + t), \ldots, u^i(t_0 + \tau)] \) the control sequence for flight \( i \). At each time before arriving at the destination, the control \( u^i(t) \in U(x^i(t), \alpha^i) \) specifies the next waypoint along the path.

### D. Weather Information Representation

Weather forecasts are critical for making flight planning decisions. There are numerous weather (nowcast and forecast) products available to assist flight planning, and these products are continuously improving with the development of new sensor systems, meteorological models, and forecast algorithms. Weather products are roughly divided into two categories: primary weather products that meet the regulatory requirements and can be safely used to make flight-related aviation decisions, and supplementary weather products that are often used with one or more primary weather products to enhance situational awareness [24]. The data available from these products consists of information about the current weather conditions and forecasts at regular sampling times for a finite forecast horizon of typically few hours. The detailed descriptions of some commonly used weather products can be found in [24]. Many of these products are also publicly accessible [11], [27].

As mentioned in Section I-C, the effect of weather on different flights could be substantially different. To account for the different capabilities of flights in handling various weather conditions, we adopt a new viewpoint of the weather information in the planning process. Let \( W \) denote all the weather products to be used for the flight planning purpose. Different users (flights) may interpret the information in \( W \) differently. The interpretation of flight \( i \) is represented by a random variable \( h^i(x, t) \in [0, 1] \) that specifies the probabilistic weather hazard level for flight \( i \) at location \( x \in \mathcal{X} \) and time \( t \in \mathcal{T} \). Each flight is also associated with a hazard threshold \( h^i_0 \), which determines the maximum acceptable hazard level. Such a representation of weather information includes the deterministic viewpoint as a special case. In particular, it also defines a flight-dependent weather avoidance region:

\[
W^i(t) = \{ x \in \mathcal{X} : \text{Prob}(h^i(x, t) >= h^i_0) > 0 \}.
\]

The region \( W^i(t) \) characterizes the hard planning constraint for flight \( i \), that is, constraints that should not be violated, while the random variable \( h^i \) provides additional soft constraints that penalize flying over regions with high weather risks.

The above weather representation agrees well with the current practice in the sense that the common information \( W \) can be made available to all flights, while it may be interpreted and used differently by each one. The way to obtain the probability distribution of \( h^i \) depends on the weather products being used and the preferences of individual users; yet it should be consistent with safety regulatory rules in the sense that any unsafe region for flight \( i \) should be assigned with a hazard level larger than \( h^i_0 \). The technical details for computing a probability distribution is beyond the scope of this paper. Nevertheless, our proposed planning strategy incorporates this general weather representation.

### E. Problem Statement

Given airway graph structure \( \mathcal{G} \), there are usually multiple paths connecting the origin and destination airports. Each flight \( i \) may incorporate its preferences in choice of the path by using a cost function. The cost due to either traveling time or fuel consumption from waypoint \( v \) to the next waypoint chosen by input \( u \), namely \( f^i(v, u) \), can be represented by a function \( l^i_w(v, u) \). In addition, a function \( l^i_w(v, u) \) can penalize the weather risks at time \( t \) for traveling between the waypoints \( v \) and \( f^i(v, u) \). An example of \( l^i_w \) is the expected hazard level at the next waypoint \( f^i(v, u) \):

\[
l^i_w(t, v, u) = \mathbb{E}(h^i(f^i(v, u), t)).
\]

Then, the cost function of flight \( i \) for traveling between the departure and destination waypoints can be characterized by

\[
J^i(x^i_0, u^i) = \phi^i(x^i(t^i_0 + \tau)) + \sum_{t=t_0}^{t_0+\tau-1} L^i(t, x^i(t), u^i(t)),
\]
where $L^i$ is the running cost function, given by

$$L^i(t, v, u) = \begin{cases} 0 & \text{if } v = x^i_f \\ L_i^u(t, v, u) & \text{otherwise}, \end{cases}$$

and $\phi^i(v)$ is the terminal cost function, defined by

$$\phi^i(v) = \begin{cases} 0 & \text{if } v = x^i_f \\ \infty & \text{otherwise}. \end{cases}$$

While there may be alternative ways of quantifying the “soft” weather risks, incorporating the cost term $L_i^u$ takes advantage of the stochastic nature of the predicted weather data and allows us to penalize the path that is close to but not inside the weather avoidance region. The overall flight planning problem can be formulated as the following constrained optimal control problem.

**Problem 1 (Centralized Planning Problem)**

$$\min_{u^i} \sum_{i=1}^{N} J^i(x^i_0, u^i)$$

subject to

$$\begin{align*}
(Dynamics) & \quad x^i(t+1) = f^i(x^i(t), u^i(t)) \\
(Weather) & \quad x^i(t) \not\in W^i(t), \quad t \in T_i
\end{align*}$$

$$\begin{align*}
(Traffic) & \quad \sum_{i=1}^{N} \mathbf{1}_{S_m}(x^i(t)) \leq c_m, \quad m \in \mathcal{I}_s
\end{align*}$$

In the above, the constraints are imposed for all flights $i \in \mathcal{I}_F$ and $\mathbf{1}_{S_m}(\cdot)$ denotes the indicator function which equals to 1 if its argument is inside $S_m$ and equals to 0 otherwise.

**III. DECENTRALIZED PLANNING ALGORITHM**

The optimal centralized solution to Problem 1 is intractable when the number of flights $N$ is large. More importantly, even if such a solution is available, its application in the ATM system would be rather limited because the optimality and safety in terms of constraint satisfaction for one flight would be immediately lost if some other flights deviate from their optimal paths, or a new flight enters the system. In addition, a centralized solution to Problem 1 would either require the knowledge of the cost preferences of all individual users or assume a universal cost metric across all the users, neither of which is practically reasonable.

With these concerns in mind, we propose a hierarchical decentralized solution to Problem 1, which, though may not achieve the global minimum, can handle a large number of aircraft and respect the decision hierarchies in the current ATM system. The main idea of our approach is to decompose the overall functionality of the ATM system into two interactive decision layers: traffic regulation and performance optimization. In the first layer, the ATM system sets up traffic rules based on existing flight plans, namely, decides which sectors are open to use over the future time slots. In the second layer, the path plans for new flights are optimized, by the users, subject to the traffic rules set in the first layer.

A. Decentralization Through Traffic Regulation Function

The main challenge for solving Problem 1 lies in the traffic constraints (2c) which involve couplings among the paths of different flights. One way to address this challenge is through dual decomposition [28], which is a powerful tool to tackle large-scale constrained optimization problems by introducing Lagrange multipliers. Such an approach has been successfully applied to study congestion control and routing problems for both data communication and ground transportation networks [29], [30], [31], [32], [33]. In these applications, the Lagrange multipliers are naturally interpreted as prices for using constrained resources, and the solution algorithms can often be thought of as certain market mechanisms with provable convergence to desired equilibria.

Despite the tremendous success of the dual decomposition method in flow-level traffic control problems, that is, optimization of flow rates across large scale networks, its application in trajectory-level path planning problems, especially in the context of air transportation system, has not been adequately investigated. Disassociating traffic flows into individual flight plans changes the nature of the problem significantly. For example, while the utility function of traffic flows can reasonably be assumed to be concave, the cost functions of individual flights are usually not convex. In addition, the “price” interpretation of the Lagrange multipliers and the corresponding market mechanism proposed for the flow-level traffic management [34], [25], [35] may no longer be appropriate, because negotiating prices among thousands of flights in real time would cause too much operational uncertainties, especially for the safety-critical system under study here. Moreover, operational preferences of individual flights could be substantially different due to the diverse range of seating capacities and aircraft capabilities, which disallows a unified cost (price) metric across individual flights.

To take advantage of the idea of decomposition, while avoiding the concerns mentioned above, we introduce a so-called traffic regulation function

$$\lambda : \mathcal{I}_F \times \mathcal{I}_S \times T \to \{0, \infty\},$$

where $\lambda(i, m, t) = 0$ permits flight $i$ to use sector $m$ over time slot $[t, t+1]$, while $\lambda(i, m, t) = \infty$ disallows this use. A valid traffic regulation function will not allow more than $c_m$ flights to use sector $m \in \mathcal{I}_S$ at any time, that is

$$\sum_{i \in \mathcal{I}_F} \mathbf{1}_{\{0\}}(\lambda(i, m, t)) \leq c_m, \quad \forall t \in T. \quad (3)$$

The traffic regulation function $\lambda$ will be used as a regulation tool instead of a resource price as in the classical dual decomposition approach. The particular form of this function is partly motivated by practical air traffic control procedures. In the current ATM system, when a potential traffic jam is identified, certain actions will be taken, e.g., the en-route flights can be controlled through speed variation, vector for spacing (VFS), holding pattern (HP) or redirecting to other sectors, to avoid entering the overly-used sectors, while the flights that are still on the ground may be delayed or required to modify their paths. All of these forms of control can be viewed as particular ways of preventing affected flights from entering the congested
sectors, which can be mathematically characterized by the traffic regulation function introduced above.

The design of $\lambda$ needs to respect constraint (3), and can be accomplished through an iterative procedure to be discussed in detail in Section III-C. For now, assume that it has already been specified by the ATM system and must be obeyed by all the flights. Consequently, the plan of flight $i$ subject to the traffic rule $\lambda$ must satisfy

$$\lambda(i, \beta(x^i(t)), t) = 0, \quad \text{whenever } x^i(t) \neq x^i_f. \tag{4}$$

Once a valid regulation rule is given, the constraints in (2c) can be decoupled since they are encoded with $\lambda$ and imposed on each flight in such way to respect the sector capacity constraints. Hence, under a given regulation rule, the best flight $i$ can do is to solve the following decentralized planning problem.

**Problem 2 (Decentralized Planning Problem)**

$$\min_{u^i} J^i(x^i_0, u^i)$$

subject to constraints (2a), (2b), and (4).

B. Solution to the Decentralized Problem

In order to solve the constrained optimal control problem in Problem 2, we formulate an equivalent unconstrained problem through the penalty function method. First, the weather constraint (2b) is addressed by introducing a penalty term in the running cost function. Let $L_w: \mathcal{T} \times \mathcal{E} \to \{0, \infty\}$ be the weather penalty function for aircraft $i$ defined by

$$L^i_w(t, v_1, v_2) = \begin{cases} 0 & \text{if } v_2 \notin W^i(t) \\ \infty & \text{otherwise} \end{cases}.$$

For flight $i$, if its location and control at time $t$ are $x^i(t)$ and $u^i(t)$, respectively, then the weather penalty incurred over $[t, t+1]$ is $L_w(t, x^i(t), f^i(x^i(t), u^i(t)))$. Next, to respect both the weather constraint and the traffic regulation rule $\lambda$, we define a new running cost function for flight $i$ as

$$\tilde{L}^i(t, x^i(t), u^i(t)) = L^i(t, x^i(t), u^i(t)) + \sum_{m=1}^{n_s} \lambda(i, \beta(x^i(t)), t),$$

and a new overall cost function

$$\tilde{J}^i(x^i_0, u^i) = \phi^i(x^i(t_0^i + \tau^i)) + \sum_{t=t_0^i}^{t_0^i + \tau^i - 1} \tilde{L}^i(t, x^i(t), u^i(t)).$$

Problem 2 is then transformed into the following unconstrained optimal control problem.

**Problem 3 (Unconstrained Planning)**

$$\min_{u^i} \tilde{J}^i(x^i_0, u^i)$$

It is clear that the set of control sequences $u^i$ with finite $\tilde{J}^i(x^i_0, u^i)$ coincides with the set of feasible solutions to Problem 2, and $\tilde{J}^i(x^i_0, u^i) = \tilde{J}^i(x^i_0, u^i)$ for all feasible controls $u^i$. Therefore, Problems 3 and 2 must have the same set of optimal solutions.

**Proposition 1:** An optimal solution $u^i$ to Problem 3 is also optimal for Problem 2.

Problem 3 can be viewed as a shortest path problem with time-dependent link cost. Such a problem has been studied extensively for vehicle transportation applications and is often referred to as the Time-Dependent Shortest Path (TDSP) problem [36], [37]. A standard way to solve the TDSP problem is to expand the state space to include the time as a state variable. Following this idea, for each $i \in \mathcal{I}_F$, we extend the spatial graph $\mathcal{G}$ to a spatial-temporal graph $\tilde{\mathcal{G}}_i = \{\tilde{V}_i, \tilde{\mathcal{E}}_i\}$, where

$$\tilde{V}_i = \{(v, t) : v \in \mathcal{G}, t \in \{t_0^i, \ldots, t_0^i + \tau^i\}\}$$

$$\tilde{\mathcal{E}}_i = \{(v_1, t), (v_2, t+1) : v_1 \in \mathcal{G},$$

$$t \in \{t_0^i, \ldots, t_0^i + \tau^i\}, v_2 \in \mathcal{U}(v_1, \alpha^i)\}.$$  

The construction of graph $\tilde{\mathcal{G}}_i$ is illustrated in Figure 5-(a), where a copy of the spatial graph $\mathcal{G}$ is made at each time step, and every link starts and ends at two adjacent time layers. The set of links $\tilde{\mathcal{E}}_i$ may vary with aircraft type $\alpha^i$. For example, as shown in Figure 5-(b), the node at time $t$ may be associated with 3 downward links (solid lines) for one type of aircraft, but 5 downward links (solid and dashed lines) for another type of aircraft with a larger maximum speed.

Once the spatial-temporal graph is constructed, Problem 3 can be solved using dynamic programming [38]. The detailed procedure is illustrated in Figure 6. The resulting solution $x^i$ provides not only the 4D flight path plan, but also the corresponding departure delay under the given traffic regulation $\lambda$. For instance, if $t_0^i$ is the first time instant for which $x^i(t)$ is away from the initial location $x^i_0$, then the ground delay will simply be $t_0^i - t_0^i$. Notice that such a delay will be affected by the particular cost function specified by the user. By adjusting the running cost function $L^i$, the user can achieve a desired tradeoff among multiple cost factors such as departure delay, total traveling time, fuel consumption, expected turbulence, among others.

C. Computation of the Traffic Regulation Function

While there are numerous ways to obtain a traffic regulation function $\lambda$ satisfying (3), here we propose a particular approach to iteratively construct $\lambda$ and the path plans. The proposed approach requires an auxiliary function $\Gamma: \mathcal{I}_S \times T \to
orders on different days because most flight schedules repeat daily. However, to respect aviation practice, all these orders should not deviate significantly from the FCFS order.

IV. A HIERARCHICAL FRAMEWORK FOR ATM

The decentralized flight planning algorithm developed in the last section leads to a hierarchical framework for the entire route air traffic management system as illustrated in Figure 7. The framework contains three interacting layers consisting of the air traffic management (ATM), air traffic users (individual flights), and the flight management system (FMS). The role of each layer is described below.

According to our hierarchical framework, the role of ATM is to, in real time, gather various measurements, update detailed weather and traffic forecasts accordingly, and send this information to the users (individual flights) upon request. At the beginning of each time period, the ATM receives new weather forecast data and new filing requests of flight plans. For each proposed flight plan, the ATM will check whether it satisfies all the weather and traffic constraints, namely, whether it passes through certain weather forbidden zone or congested sectors. If the constraints are all satisfied, then the plan will be accepted and the traffic regulation function \( \lambda \) will be updated according to Algorithm 1. These updated weather and traffic predictions \((\lambda, \mathcal{W})\) are made accessible to individual flights in the second layer.

The main goal of the second layer is to optimize performance. At this layer, each flight, before taking off, receives the weather and traffic information from the ATM layer, and regards these information as traffic rules. Subject to these rules, the user optimizes its path plan according to its own cost metric using the algorithm described in Section III-B. The resulting plan will meet the regulation rules and will be accepted by the FAA in the ATM layer.

The bottom layer can be viewed as a physical layer, where the on-board FMS receives the high-level (waypoint-based) path and generates low-level control signals to control the aircraft according to the high-level path. Since our planning algorithm respects the underlying aircraft dynamics, the 4D paths generated by the algorithm can indeed be carried out by the aircraft.

A. Improvements over the Current System

An important distinction of the proposed framework compared with the current ATM system is the provision of the traffic information to the end users. In the current system, the detailed traffic information, based upon which FAA accepts/rejects flight plans, is not available to the users during their flight planning processes. Due to the lack of a common situation awareness, the users may propose paths that violate traffic restrictions, resulting in departure delays that are often unnecessary. For example, Figure 8 illustrates a case where one of the sectors along the nominal path of a flight is blocked by weather or traffic congestion, which will cause a departure

\[ Z_+ \text{ and a predefined ordering of all the flights } \{i_1, \ldots, i_N\}. \]

The function \( \Gamma \) keeps track of the usage of all the sectors and is updated after each flight files its plan. Once \( \Gamma(m, t) = c_m \), the corresponding element in the traffic regulation function \( \lambda \) will be set to infinity, indicating that the system will no longer accept any new requests of using sector \( m \) during the time period \([t, t+1]\). The detailed procedure of this approach is summarized in Algorithm 1. It can be easily verified that constraint (3) is met at any stage of the algorithm, which guarantees the safety for all nominal paths generated by the algorithm.

The flight ordering \( \{i_1, \ldots, i_N\} \) required by Algorithm 1 captures the priorities of different flights in using congested sectors. If the order is assigned according to flight departure times, then it corresponds to the First-Come-First-Served (FCFS) policy used in the current ATM system. Even though the FCFS ordering may not result in a fair allocation of airspace resources, it is advantageous in practice because it allows Algorithm 1 to be carried out “on-the-fly”, where each flight solves Problem 3 shortly before departure with updated traffic and weather information, and more importantly, it need not modify its path once airborne. The fairness concerns can be addressed, to some extent, by alternating among multiple
delay in the current system, although a small modification of the nominal path can avoid such a delay. Sharing the traffic regulation information with the users as suggested by our framework, allows the users to see which parts of the airspace are still open, and thus make the most efficient use of the airspace under the current traffic conditions.

The weather forecast already plays a crucial role in the current ATM system, however, it has not been used in a systematic way. The forecast is often sent to pilots through weather briefings or even through verbal communications. The weather avoidance is usually achieved by manually modifying the waypoints in an ad-hoc manner, which could be rather conservative. Under our framework, the detailed weather information is sent to each flight and can be used to automatically generate the best way to avoid weather while respecting the traffic constraints. In addition, the consideration of the stochastic nature of the weather forecasts through inclusion of the expected weather risk term \( \lambda \) in the cost function, has a great potential to improve the average efficiency of the current ATM system that is operated mostly based on the worst-case scenario. The absolute safety can still be guaranteed through online planning when an unpredicted rare weather event occurs.

Lastly, the passing of information from the higher level to individual users allows for decentralized implementation of the centralized optimization without loss of safety of the overall system. The complexity of the decentralized planning algorithm does not depend on the number of aircraft in the system. Thus, the proposed framework can handle an increasing air traffic volume as predicted by the FAA. In addition, since this planning can be done in an automatic way, it can reduce the workload of air traffic controllers and potentially reduce the risks due to human errors.

**B. Integration with the Current System**

Information exchanges between the ATM layer and the users have already been utilized for path planning in the current air transportation system. Therefore, the implementation of the new framework does not require substantial infrastructure modifications, but instead would only require (i) modifying the information sent to the users from simple weather briefings to a better structured data set \( (W, \lambda) \), and (ii) encouraging the users to utilize the detailed information in a more systematical way as illustrated in Figure 6. Both of these two changes are implementable in the near future.
In addition, due to the hierarchical decomposition, the planning decision of any user depends only on the regulation signal \((\mathcal{W}, \lambda)\) received from the ATM layer, but not on the planning decisions of other users. Hence, the proposed framework will not be affected by the existence of non-participating users that do not use the proposed planning algorithm. This property is crucial because it allows the current ATM system to be gradually transformed into the new framework.

Although the hierarchical framework is described mainly in the context of pre-departure path planning, it can also be used to reroute flights to deal with unexpected traffic and weather conditions. In this case, the flight needs to first obtain the latest weather and traffic information from the ATM layer, then recompute a new path with initial location updated to the current aircraft position, and finally file the modified path plan with the ATM layer. Clearly, in the new ATM framework, the physical operations of flights depend more crucially on the communications between the flights and the ATM layer. This stronger integration of the cyber and physical components should be supported by more dedicated security solutions for data communications.

V. SIMULATION RESULTS

We now illustrate the use of the proposed hierarchical framework through a large-scale simulation motivated by real traffic data in the contiguous United States. To this end, the origin-destination pairs and the departure times of the flights that travel among the 34 continental airports in the FAA’s Operational Evolution Plan (OEP)\(^3\) are extracted from the Enhanced Traffic Management System (ETMS) data of Aug 24th, 2005. The departure time distribution of these flights is shown in Figure 9. We consider all the flights whose departure times are between 12 p.m. GMT (7 a.m. EST) and 10 p.m. GMT (5 p.m. EST) and use them as our test data set, which has 5419 flights.

In this simulation, we compute the en-route path plans characterized by planar waypoints (with no altitude information) and the corresponding time stamps. In addition, we assume the

\(^3\)There are 35 airports in the OEP plan, which account for about 69% of total operations in the NAS [39]. Our simulation is based on these airports, excluding HNL (Honolulu International Airport).
airway graph consists of a uniform grid with grid size equal to 1 nautical mile. The discrete time step is set to be $T_s = 1 \text{ min}$ and the running cost function is chosen to be

$$L^i(t, v, u) = \begin{cases} 0 & \text{if } v = x^i_f \\ \|f^i(v, u) - v\| + c & \text{otherwise,} \end{cases}$$

for all $i \in I_F$, which penalizes a weighted sum of the traveling distance and traveling time. The constant $c$ is chosen to be $6T_s$. The particular values of the cost parameters and the grid point locations are not critical for the proposed algorithm, and they are adopted to simplify the simulation and presentation.

With the chosen parameters, we first compute the optimal path for each of the flights in the data set without considering any traffic constraints. In this case, all the flights can fly their optimal paths. Based on these unconstrained paths, the average sector count over the time window between 12 p.m. EST and 5 p.m. EST is calculated for each of the 284 high-altitude sectors in the continental US, and their values, in normalized scale, are shown in Figure 11. It can be seen that without any regulation, the traffic tends to concentrate on a few sectors and the majority of the rest of the airspace remains under-utilized.

To demonstrate the proposed ATM strategy, all the sector capacities are set to 8 and the flight plans are recomputed using Algorithm 1 subject to these constraints with priorities assigned according to their departure times. Normal capacities typically range from 10 to 20. The obtained paths satisfy the capacity constraints at all times, while the previous unconstrained paths result in 40 sectors exceeding the capacity over some time period in the planning interval. An example of aircraft-count improvement is illustrated in Figure 10 for Sector ZTL15. Figure 12 shows the average sector counts based on the constrained flight plans under the same condition as described in the last paragraph. In comparison to Figure 11, it is clear that the proposed planning strategy yields a better utilization of the airspace over time with the traffic density in the congested sectors properly diffused into their neighbors.

The total costs corresponding to the unconstrained planning and our hierarchical planning strategy, denoted by $J^*_{uc}$ and $\tilde{J}^*$, respectively, are also computed. It is observed that imposing the capacity constraints only incurs a 0.71% increase of the total cost under our hierarchical planning strategy, i.e., $\tilde{J}^* - J^*_{uc} = 0.71\%,\tilde{J}^*$. In addition, it is clear that $J^*_{uc} \leq J^* \leq J^*$, where $J^*$ denotes the minimum cost corresponding to the centralized optimal solution of Problem 1. Therefore, the performance loss due to the hierarchical decomposition, namely, $J^* - J^*$ must be less than 0.71% for this particular example.

To illustrate the weather avoidance feature of the proposed planning algorithm, we assume that from 11 a.m. EST to 1 p.m. EST, there are two severe convective weather storms: one of them completely blocks the 5 shaded sectors around the center of Figure 13-(a), while the other one partially covers another 5 outlined sectors on the left half of Figure 13-(a) and reduces their capacities to 4. The unconstrained path plans indicate that 116 (resp. 128) flights are scheduled to travel through the first (resp. second) weather region over that time window. According to the current ATM strategy, some of these flights may be delayed substantially. However, the path plans generated by our algorithm under the weather-induced sector-capacity constraints, indicate that all of these affected flights can leave on time without violating the constraints, and the average traveling distance of these flights increases by only 0.52%. Figures 13-(a) and 13-(b) show all the aircraft locations at 12 p.m. EST according to the paths generated by our algorithm with and without the weather constraints, respectively. The effectiveness of our algorithm in handling this situation is clear from the figure. Further testing of our algorithm based on real forecast weather data will be conducted in our future work.
VI. Conclusion

The current ATM system is responsible for both safety regulation and performance optimization. This paper presents a hierarchical decision architecture for the future ATM system that distributes the optimization task to the users. With this architecture, the ATM serves as a service provider that provides regulation information based on traffic and weather conditions for the users to optimize their own cost functions and intervene in decisions only when they violate the regulation rules. A general approach to solve individual flight planning problems is also proposed, which considers both traffic constraints and weather risks. Simulation results based on real air traffic data have shown the effectiveness of the proposed framework in congestion control and weather avoidance.

As for future research, an immediate step is to study the potential benefits of our framework in terms of energy savings through using realistic fuel consumption and wind forecast models. We also plan to extend this framework to explicitly consider airport arrival and departure capacity constraints as well as to explore various ordering schemes for flight planning to improve fairness in airspace resource allocation. In addition, although the proposed framework adopts the most general probabilistic model to describe weather uncertainties, a unified way to extract and quantify hazard probabilities from various existing weather products still require substantial engineering and research efforts.

Finally, we envision that the hierarchical ATM framework developed in this paper also provide useful insights for the modernization of other infrastructure systems, such as power systems and ground transportation systems. In particular, both the concept of decomposing performance optimization from reliability/safety regulations, and the concept of distributing performance optimization tasks to individual users while maintaining centralized authority for determining safety rules and policies, are effective methodologies for continuously leveraging new technologies for legacy infrastructure systems without compromising safety standards.

Wei Zhang (S’05-M’10) received a B.E. degree in Automatic Control from the University of Science and Technology of China, Hefei, China, in 2003, an M.E. degree in Electrical Engineering from the University of Kentucky, Lexington, KY, in 2005, and a PhD degree in Electrical Engineering from Purdue University, West Lafayette, IN, in 2009. Between January 2010 and August 2011, he was a postdoctoral researcher in the Department of Electrical and Computer Engineering, University of California, Berkeley. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH. His research interests include control and estimation of hybrid and stochastic dynamical systems, and their applications in various engineering fields, especially air transportation systems, robotics, and energy systems.

Claire J. Tomlin (S’94-M’98-SM’06-F’11) is a Professor in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, and holds a part-time Research Professor position in the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA. She has held visiting researcher positions at NASA Ames and Honeywell. Her research is in the area of hybrid control systems, with applications to air traffic systems, unmanned aerial vehicles, and systems biology.

Dr. Tomlin received the MacArthur Fellowship (2006), the Okawa Foundation Research Grant (2006), and the Eckman Award from the AACC (2003).
REFERENCES


