Two Perspectives on Graph-based Traffic Flow Management

Aude Marzuoli, Vlad Popescu, Eric Feron
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0150, USA
{aude.marzuoli, vlad.popescu, feron}@gatech.edu

Abstract—New en-route Traffic Flow Management (TFM) tools are needed to handle the predicted air traffic growth by mitigating congestion, reducing delays and maintaining high safety levels. A large-scale graph abstraction of traffic through an airspace center is extracted from historical navigation data. The network graph model represents aggregated air traffic flows through the airspace. Although exemplified for a typical center, the methodology is generalizable to any airspace. The network is used to support two perspectives on TFM: a centralized approach based on network flow optimization with workload-derived sector constraints, and a decentralized approach based on mean field games.

I. INTRODUCTION

With the forecast increase in air traffic demand over the next decades, achieving an optimal functioning of the network system is essential. The network system is comprised of the aircraft operators, airports and Air Traffic Management (ATM). In the coming years, the current airspace capacity limits, i.e. the maximum number of aircraft allowed in a given airspace, are expected to be exceeded. Delays caused by congestion or weather perturbations are becoming more acute at many airports and in many airspaces.

It has been observed that traffic demands are spatially and temporally heterogeneous, at times leaving substantial regions of the National Airspace System (NAS) with underutilized resources while capacities become saturated in other regions [1]. Robust tools are needed to better manage congestion and delays [2] by air traffic flow managers. These tools will come into effect under the ‘paradigm shift’, supported by innovative technologies [3].

The focus of this paper is on strategic medium- and long-term planning. This paper presents a methodology for constructing a graph abstraction of the airspace which may be used as a Traffic Flow Management (TFM) framework. Two perspectives, one using centralized control and the other using decentralized control are presented in this paper. Both approaches use the TFM framework to optimize aircraft routing in order to reduce congestion and delays, while allowing more aircraft in an airspace and ensuring high safety levels.

The TFM framework is based on data mining and modeling of the airspace. Many tools in TFM require knowledge of the initial aircraft positions and trajectories (with or without uncertainty). Such precise knowledge is useful for ATM activities such as planning a safe trajectory for an intruder aircraft through the airspace [4]. Conversely, strategic planning for mid-term or long-term ATM with time horizons greater than 30 minutes requires knowledge of air traffic patterns and flows characteristics. Among the relevant flow characteristics are the geometrical configuration, flight plans and average distances between two consecutive aircraft. The current position and intent information of individual aircraft are irrelevant at that aggregated scale.

Representing air traffic as flows enables the computation of more predictable macroscopic estimates than following individual aircraft. This approach provides a higher-level view of air traffic to traffic flow managers [5]. Furthermore, the representation according to dominant traffic flow patterns and traffic flow interactions is consistent with the mental models used by air traffic controllers in their abstraction of sectors [6], [7]. The mental models that controllers develop to aggregate traffic flows rely on features such as the number of flows, the major flow and its size and the number of crossing flows [5]. These criteria can further be used to estimate and predict sector capacity as a function of the traffic flow pattern and to study the impact of severe weather [8].

The graph abstraction of traffic flows is the basis for two perspectives on flow management. A data-mining approach is adopted to build flows based on data [9], and then to create a network from the previous flows. This methodology can be adapted to any airspace to obtain the graph abstraction. The network supports a centralized perspective that intends to simulate realistic traffic traveling through the given airspace. The Air Traffic Flow Management Problem was examined in 1994 by Odoni [10] and aims at solving complex situations in Air Traffic. A new linear formulation of the TFM problem is developed, using estimates of controller workload based on flow geometry [11].

Conversely, the decentralized perspective is aimed at identifying and forecasting the emergence of systemic
congestion and delays as a result of local interactions and strategies of aircraft. Previous research on decentralized ATM has elaborated control laws for Eulerian flows [12]. The perspective introduced in our research is based on the recent theory of mean field games [13], [14]. The novelty of this approach resides in the coupling of two equations, where microscopic agents with limited local influence optimize their strategy based on a rational anticipation of macroscopic population dynamics. The coupling provides a robust methodology for identifying and forecasting various emerging phenomena. In the case of air traffic, the aircraft population evolves in a state space constrained by the flow graph abstraction [15].

This paper is organized as follows. Section II develops the methodology from building the en-route TFM framework. Section III focuses on implementing a centralized approach based on the previous framework. Section IV explains the decentralized mean field games approach. Finally, Section V provides concluding remarks and the efforts to be pursued.

II. GRAPH ABSTRACTION OF THE AIRSPACE

A. Clustering of Trajectories into Flows

This subsection introduces a methodology to develop a mode of en-route aircraft operations through the extraction of air traffic flows within an air space from archived flight data.

The data used to construct the airspace model is taken from Enhanced Traffic Management System (ETMS). Cleveland center is selected because of its significance to the NAS. The data includes aircraft trajectories, spatially sampled (longitude, latitude, altitude) every minute. During the 123 days (May to August 2005) covered by the data, all 526,840 aircraft trajectories with at least one point over FL250 are considered, which gives us a subset of data covering the majority of en-route aircraft. After filtering inconsistencies in altitudes, a ‘clean’ data-set of 338,060 trajectories remains. Figure 1 represents the air routes and jet routes in the Cleveland center overlayed with a density plot depicting the spatial distribution for a day of traffic. It demonstrates traffic is far more diverse than what the routes alone suggest.

The trajectory Clustering Algorithm is defined by the following steps, according to [9]:
1. Clean and format the trajectories.
2. Augment dimensionality of the data by adding features such as heading, polar coordinates, etc.
3. Apply hierarchical clustering. Organize and divide the trajectories by altitude and attitude to create separate data sets.
4. Normalize each feature and concatenate the data into a single row vector for each flight. Each column corresponds to a feature.
5. Apply a Principal Component Analysis (PCA) on the matrices, and reduce the dimensionality of the data by keeping only some principal components.
6. Cluster the values of the projections using a density-based clustering algorithm (DBSCAN).
7. Obtain clusters of trajectories and outliers for each altitude and attitude category. Figure 2 shows an ascending flow obtained by the above clustering algorithm.

Following categorization and clustering, about 80% of the trajectories are grouped into 690 clusters, or flows, and the remaining 20% are outliers (modeled separately). Figure 3 presents a 2D and a 3D view of the centroids of all clusters. Blue lines represent westbound traffic, yellow eastbound, green descending, red ascending. The major airports of Cleveland Center - Cleveland Hopkins airport (CLE), Detroit Metropolitan Wayne County Airport (DTW) and Pittsburgh International Airport (PIT) - are clearly identifiable by the clusters corresponding to ascending and descending traffic. The fraction of outliers is relatively constant throughout the day. The results obtained demonstrate that the clustering remains consistent over a broad range of parametrizations (time of day, altitude/attitude). Therefore, clustering yields a model that can be utilized for subsequent complexity analysis.
B. Network Flow Model of the Airspace

A network is a system of nodes, with directed edges linking them. The edges represent portions of flow corridors in which aircraft fly, whereas the nodes correspond to spatial areas where aircraft may enter, change, or leave a flow corridor. To generate the network, the following steps were carried out.

The first step is to locate the regions where aircraft can leave a flow and join another. The areas where flows spatially interact correspond to areas that engender a high probability of conflict. Such areas include intersections of flows and flow merging. Some of these areas are such that an aircraft may leave the flow it was traveling on and join another. If these conflict areas have specific geometric features, they become a part of the network allowing for rerouting options of aircraft. There are 1198 nodes for rerouting for the network representing the Cleveland center (ZOB). If a conflict area does not satisfy the geometric properties above, it is called a crossing, and is not used to build the network, but is considered in workload calculations. There are 14,953 crossings and they intervene in the complexity of the airspace as spatial areas with higher probabilities of conflicts.

The second step is to define the nodes corresponding to spatial areas where aircraft enter or leave the airspace, whether on the boundaries of the center, or at airports located in the center. All flows have an entry and an exit. For en-route flows, the entrances and exits are located at the boundary of the center. For arriving or departing flows, the entrances and exits are at the ground level within the center. Observing the distribution of the entrances and exits of the flows in 3D enables grouping them into shared entries and exits. This is done by applying a calibrated k-means clustering algorithm. The resulting centroids for each entry cluster or exit cluster are defined as the entry nodes and exit nodes of the network. We obtain 40 entry clusters and 50 exit clusters. Along with the nodes for rerouting, there are thus 1288 nodes in the network.

The third step is to create the edges that link the nodes of the network, to re-create the possible flow routes an aircraft can travel on. On each flow, an edge is defined between all consecutive nodes (whether entry, intersection enabling rerouting, exit) along the flow. Any redundant edges, i.e. those edges corresponding to two flows, but linking the same nodes, are removed. A 2D view of the resulting network is shown in Figure 4.

Our interest lies in simulating traffic as realistically as possible, that is simulating traffic flying from its origin in the center to its destination, as shown by historical data. In order to do this, the origin-destination pair for each flow is stored, using the entry and exit data gathered by the k-means clustering in the previous step. Hence the 218 origin-destination node pairs of the present network are obtained. A commodity is defined as all aircraft travelling on the same origin-destination nodes pair. The number of trajectories clustered in each flow provides the relative importance of each flow with regard to the total traffic.

This process is extended to the origin-destination pairs, or commodities, to determine the relative importance of each commodity in regard to the total traffic. The fraction of the total traffic historically associated with each commodity is denoted as \( f_k \), for \( k \) between 1 and 218. Thus the main routes traveled in the center are identified, as illustrated in Figure 5. For instance, 50% of the traffic is historically associated with 8% only of the commodities, while 90% of traffic is travelling on 40% of the commodities.
III. TRAFFIC FLOW MANAGEMENT ON THE NETWORK MODEL

A. Formulation of the en-route Traffic Flow Management problem based on flow geometry

The network model is intended to provide support for further analysis of the airspace. In this section, different means of addressing en-route traffic flow management optimization problems are discussed, using the previous network and linear formulations. A set of common constraints for various en-route Traffic Flow Management problems are defined, thereby providing a framework to modify the objective function and provide additional constraints as necessary. Next, a non-standard set of constraints is added, to account for sector capacity as a result of controller taskload. The flow and sector constraints are described in the next paragraphs. The general problem is of the following form:

$$\max \text{ objective}$$
$$st: \left\{ \begin{array}{l}
\text{flow constraints} \\
\text{sector constraints}
\end{array} \right. \quad (1)$$

The following flow constraints are enforced, resulting from the network formulation. The flow rate is the upper bound on each edge, to ensure safe separation distance between aircraft according to their average speed. At each node corresponding to an intersection, flow conservation is required. Throughput conservation is required, meaning that the number of aircraft entering the center is to be equal to the number of aircraft leaving the center. To keep track of which edges aircraft of a given commodity travel on, the total flow rate on an edge is defined as the sum of the flow rates of all commodities on the edge. The demand of a commodity is the flow rate of this commodity on all edges entering through the associated entry node. Thus the throughput of the center is the sum of all the demands. This formulation results in an unsimplified linear program of approximately 273,000 lines.

However, the flow constraints do little to impede the traffic in the airspace. In reality, traffic throughput is bounded, typically as a result of weather or controller taskload constraints engendered by the existence of a human-in-the-loop control system. To account for this bound, sector constraints based on a taskload and communication interpretation of aircraft management are introduced, supported by prior research on dynamic density and analysis of controller and pilot communication times. Historically, a sector capacity is the number of aircraft present in a sector, established by the Monitor Alert Parameter (MAP) and is appropriate for considering nominal traffic patterns. Yet, when dynamics are present (e.g. weather and changing traffic patterns), MAP values no longer accurately represent sector capacity - and often times lead to congestion, or conversely, under-utilization of the airspace. To allow for the ability to consider off-nominal conditions, a new measure is proposed.

A simple taskload model is chosen to approximate constraints on the expected subjective taskload a controller should be exposed to while taking into account some factors introduced by dynamic density and controller-pilot communication times [5]–[8]. The model includes the following common tasks, and estimates the amount of time-effort the controller must spend on each:
- aircraft acknowledgments,
- altitude clearances,
- hand-offs,
- monitoring turning aircraft,
- resolving potential conflicts.

A running cost based on a weighted sum of the number of events associated with each task and associated with airspace management tasks is calculated. The proposed taskload measure can be computed with the flow rates throughout the center. We believe that a model that considers event rates across each sector, and limits an expected taskload estimate, is potentially more relevant than constraining aircraft counts. The re-routing of traffic produces spatial aircraft distributions that can either simplify or complicate traffic management. By accounting for potential conflicts a more meaningful measure is introduced.

For an arbitrary sector $S$, the taskload constraint is given by summing the weighted effort required for the five tasks listed above (acknowledgments, clearances, turning aircraft, etc.). The associated equation is

$$\sum_{i=1}^{5} C_i^S R_i^S \leq W^S,$$

where $W^S$ is a measure of the maximum allowed taskload. In Equation 2, the value $R_i^S$ represents the expected rate associated with the $i^{th}$ task inside of sector $S$. The weighting $C_i^S$ is the average amount of time spent on the corresponding task. Accordingly, the total sum is the total amount of time effort expected by the air traffic controller for a given period of time. To maintain reasonable taskload, the bound on the constraint, i.e. $W^S$, should be selected carefully. Normalizing all the weightings and rates, the value $W^S$ represents the
In order to estimate the demand patterns for suitable time experiences peak-demand hours and lower-demand times. Demand patterns vary throughout the day, and the airspace the network and the bounds accordingly modified. Moreover, the state of the airspace, because edges can be removed from lies in the fact that the sector constraints can be adapted to is safe to travel. The strength of the above TFM formulation travel on a sub-network, comprised only of the edges that it is therefore the entire network can be opened, or the airspace traffic, the state of the network has to be defined. Indeed, explained in Section II.B. Y et, in order to realistically simulate for the center. On average, this can simply be obtained as are travelling in Equation 4, which is very unlikely. Never serviced. For instance, only 40 commodities out of 218 the more traffic prevail, and some origin-destination pairs are the more traffic that can be expressed as the sum of all demands. The first problem, expressed in Equation 3, only takes into account the flow constraints, whereas the second problem, expressed in Equation 4, is also subject to the sector constraints.

\[
\begin{align*}
\text{max} & \quad \text{Throughput} \\
\text{st} : \quad & \text{flow constraints} \\
\end{align*}
\]

\[
\begin{align*}
\text{max} & \quad \text{Throughput} \\
\text{st} : \quad & \text{flow constraints} \\
& \text{sector constraints} \\
\end{align*}
\]

The results show that the center throughput obtained by Equation 4 is only 14% of the throughput obtained by Equation 3. Besides, in case Equation 3, 3020 edges out of 3085 are occupied, while 570 edges are in Equation 4. The network being almost fully occupied at any interval of time is unrealistic. This demonstrates the importance of adding sector constraints in the TFM formulation. Nevertheless, to simulate traffic through an airspace, a demand pattern needs to be fixed, otherwise the routes capable of accommodating the more traffic prevail, and some origin-destination pairs are never serviced. For instance, only 40 commodities out of 218 are travelling in Equation 4, which is very unlikely.

Another interesting question is to define the demand pattern for the center. On average, this can simply be obtained as explained in Section II.B. Yet, in order to realistically simulate traffic, the state of the network has to be defined. Indeed, the airspace can be working under nominal conditions, and therefore the entire network can be opened, or the airspace can be experiencing weather perturbations, and aircraft may travel on a sub-network, comprised only of the edges that it is safe to travel. The strength of the above TFM formulation lies in the fact that the sector constraints can be adapted to the state of the airspace, because edges can be removed from the network and the bounds accordingly modified. Moreover, demand patterns vary throughout the day, and the airspace experiences peak-demand hours and lower-demand times. In order to estimate the demand patterns for suitable time intervals, that is of about 15 minutes, through the day, the ETMS data used to construct the flows proves to be useful again. We are currently data mining, in order to identify days under nominal conditions, days under perturbations, what type of perturbations, and define suitable demand patterns. The goal is to use the information gathered on which routes were historically traveled by aircraft through any day, under which conditions, in order to refine the TFM formulation.

The model presented enables us to determine whether a given demand pattern can be accommodated or not, depending on the controller’s capacity to handle traffic. If the model shows that the demand pattern can be accommodated, it also provides the best routes for aircraft to follow. The best routes balance the preferences of aircraft to travel the shortest routes and the need to maintain acceptable taskload levels for the controller. Besides, it gives a clear view of which routes could accommodate more traffic if the demand were to change, i.e. the remaining capacity. If the model shows that the demand pattern cannot be accommodated, it points out where the congestion would occur and what share of the demand could travel while ensuring safety.

### IV. Mean Field Games Approach

The decentralized approach uses the abstracted flow graph presented in Section II according to the theory of mean field games.

#### A. A Brief Introduction to Mean Field Games

Mean Field Games (MFG) theory considers a class of games with large numbers of players, and establishes equations for calculating the Nash equilibria of long-term stochastic problems. Deterministic limits can be identified for stationary systems, but the theoretical framework is also well suited to an interpretation in terms of optimal control with finite horizon.

The mean field approach brings together two seemingly opposed methodologies. One is the atomistic approach where the microscopic dynamics of the agents and their interactions are precisely described. In the case of air traffic the continuous approach has conventionally been used with agent-based simulations that describe conflicts and controller taskload [16]. Second is the continuum approach traditionally used in stochastic control and statistical physics approximations, where the macroscopic system properties and evolution are considered in terms of partial differential equations. In the case of air traffic this was used with Eulerian flow models [17].

In order to match the atomistic and continuum scales, the theory of MFG considers a feedback coupling (see Figure 6). Microscopic rational agents whose dynamics are controlled by a stochastic differential equation seek to minimize their costs by relying on their rational anticipation of the macroscopic system dynamics and their value function. Each agent’s
optimal strategy to reach a future state is associated to a value function which solves a Hamilton-Jacobi-Bellman (HJB) partial differential equation. The rational anticipation by the agents yields, under certain conditions, convergence to a Nash equilibrium. The macroscopic system dynamics are piloted by a transport equation of the Kolmogorov-forward type (also known as a Fokker-Planck equation). The overall dynamics along with the micro-macro equilibria thus emerge from the feedback coupling of the HJB and the Kolmogorov equations.

B. Application of MFG to Air Traffic

MFG are well suited to describe interactions between numerous agents with marginal local influence, which is applicable to air traffic. Instead of considering individual aircraft, MFG are based on continuous density measures to approximate such large numbers of similar agents. As illustrated by Figure 7, mean field games convert large numbers of aircraft navigating in the airspace to a scalar measure of traffic density along respective air routes. By characterizing the density of traffic along flows rather than individual aircraft, the mean field games approach fits well into the flow network paradigm introduced in Section II.

Aircraft must traverse the airspace by passing as close as possible to certain waypoints and following the shortest available routes between the waypoints. The waypoints and routes are defined by graphs (or networks) such as the ones presented in Section II. Figure 8 represents a reduced version of the network for ZOB (Cleveland) Air Route Traffic Control Centers (ARTCC).

The network is introduced into the MFG formulation as a constraint on the available state space. The subjacent cost used in the optimization penalizes both the expected congestion on the possible branches of the network and also the total length of the chosen path. The mean field dynamics are defined by the coupling of the Fokker-Planck and Hamilton-Jacobi-Bellman equations. Fokker-Planck propagates the population density forward in time, while the HJB equation propagates the value function (the optimal cost) backward in time, based on agents’ expectation of future evolution (see Figure 6). The optimal solution to the coupled equations is a decentralized control that regulates congestion and minimizes deviations and delays.

The main interest of the MFG approach lies in its ability to incorporate several distinct populations of aircraft cohabiting in a shared airspace with varying equipage or autonomy levels and possibly diverging objectives. Congestion and delays...
are treated as intrinsic emerging factors. The optimality of aircraft trajectories thus contributes to self-regulation of the system. At the macroscopic scale, the existence of congestion can be used as an indication of when and where additional directed control is required. The MFG perspective is therefore suited to assessing the impact of ATC paradigms where the locus and amount of control varies on a scale ranging from centralized to decentralized/autonomous.

A precise analytical formulation and numerical discretization scheme for using MFG in the context of ATM congestion forecasting and control has been achieved [15]. The flow graph plays a role in the cost function by penalizing deviations away from its edges. It is important to note that this graph representation implicitly uses agent preferences (origin, destination, ideal routing). The full cost is modeled in Equation (5) and incorporates congestion, distance and discount terms.

In order to represent aversion to congestion, high density regions are penalized by increasing the cost of traveling through them. The $\frac{\alpha^2}{2} m^2$ term implies control $\alpha$ through high density (higher $m$) regions will be more expensive than low density regions. $\beta > 0$ is a parameter that represents the magnitude of aversion (how fast the cost will increase as $0 < m < 1$ increases).

To the quadratic congestion term we then add the distance term $\lambda d^2(x, x_{gc})$ to penalize deviations away from great circle routes (flow graph edges) connecting waypoints (flow graph nodes) - see Figure 8. The parameter $\lambda$ represents the tolerance to deviating against traveling through congested regions.

Finally, a discount factor $e^{-rs}$ represents the urgency (the relative balance of short-term against long-term costs). The parameter $r$ models the flexibility in trading-off the present against the future.

$$J = \int_{t_0}^{T} \left[ \int_{\Omega} \left( \frac{\alpha^2}{2} m^2 \beta + \lambda d^2(x, x_{gc}) \right) \, dm \right] \, e^{-rs} \, ds \tag{5}$$

The coupled forward (Fokker-Planck) and backward (HJB) optimality equations in the time-varying case are shown in (6). A more detailed derivation and discussion may be found in [15].

$$\partial_t m - \frac{\alpha^2}{2} \Delta m + \nabla \cdot \left[ \frac{\nabla v}{m^2} \right] = 0$$

$$m|_{t=0} = M_0$$

$$\partial_t v + \frac{\alpha^2}{2} \Delta v + \frac{|\nabla v|^2}{2m^2} - rv = -\lambda d^2(x, x_{gc})$$

$$v|_{t=T} = V_T \tag{6}$$

We are currently developing numerical resolutions methods apt at solving the forward-backward coupling.

V. FUTURE WORK AND CONCLUSION

A data-based methodology for modeling an airspace as a large-scale 3D graph has been presented. This abstraction of a complex system provides aggregated information about the airspace, its complexity, and the location of areas with stronger probability of conflict.

Two perspectives on this framework for en-route Traffic Flow Management have been developed. The centralized approach offers a formulation of the en-route TFM problem in order to simulate traffic through the airspace. The formulation is adapted to the network flow model by incorporating results from complexity metrics which account for controller workload and ensure safety requirements. The decentralized approach applies mean field games theory to identify, forecast, and control the occurrence of systemic congestion and delays. The aircraft population lives in a state space constrained by the graph and evolves according to optimization criteria which account for expected congestion and rerouting costs.

The work presented here is undergoing and will offer several applications for Traffic Flow Management. The ultimate goals are to be able to predict congestion and mitigate its effects, both under nominal and perturbed conditions, while accommodating larger shares of traffic.

The next step toward accomplishing these goals is to add stochastic weather models to the TFM framework. Weather perturbations will be incorporated in the network formulation through morphisms acting on the graph according to relevant spatial and temporal correlations. The modified network structure will then be used by both the centralized and decentralized approaches, which will result in stochastic optimization problems.

Moreover, comparing the results of both approaches may lead to the implementation of a mixed strategy. Indeed, the centralized approach is likely to violate individual aircraft preferences, whereas the decentralized strategy does not take into account bounds on controller workload. Therefore, a mixed strategy that incorporates elements from both perspectives is likely to produce results that better balance aircraft and controller objectives.

ACKNOWLEDGEMENTS

The research was supported by NASA grant NNX08AY52A (Influence of Degraded Environment on Airspace Safety) and Air Force grant FA9550-08-1-0375 (Coordinated Multi-Disciplinary Design of Complex Human Machine Systems).

REFERENCES


