Tactical Scheduling for Precision Air Traffic Operations: Past Research and Current Problems

(Preprint. Accepted by the AIAA Journal of Aerospace Information Systems.)

D. R. Isaacson
NASA Ames Research Center, Moffett Field, CA 94035-0001
A. V. Sadovsky†
NASA Ames Research Center, Moffett Field, CA 94035-0001
D. Davis‡
UCLA, Department of Mathematics, Los Angeles, CA 90095-1555

Future air transportation systems stand to benefit significantly in safety and efficiency from the predictable movement of aircraft along precisely defined paths in the airspace. Such aircraft movement, hereafter referred to as Precision Air Traffic Operations (PATO), is not widely used during periods of peak air traffic in today’s system, but is the foundation of high-capacity operations envisioned for the future. Automation and deployment of PATO, being a relatively young field of research, has not had time to establish structured theories and standardized reference literature. As a consequence, researchers interested in entering this field have difficulty applying to it classical techniques of operations research and optimal control theory. The main obstacle to such applications is the lack of access to the finer domain knowledge of air traffic operations (most importantly, knowledge of the operational constraints) needed to formulate research problems that promise deployable automation tools. Such a formulation requires the researcher to characterize and assess the research efforts and tendencies that emerged in the recent decades to address diverse problems, some outwardly similar yet essentially different. The research field of PATO has now matured to a stage where this requirement can start being met. This paper, aimed as a step in this direction, provides (a) a formulation of the general problem of defining conceptually, constructing, and using a schedule for PATO that contains specification of merging sequences and provides aircraft separation continuously in time, (b) the context necessary for understanding the formulation and its limitations, and (c) a review of prior research on future Air Traffic Operations (ATO) and, in particular, on the role of Air Traffic Control (ATC) in these operations.

I. Introduction

The public consumer’s increasing demand for air traffic services over the past decades has led to the development of a plan, called the Next Generation Air Transportation System and abbreviated as NextGen [1], to modernize the U.S. National Airspace System (NAS). Similar efforts are underway elsewhere, notably the Single European Sky ATM Research (SESAR) program [2] and the Seamless Asian Sky (SAS) initiative [3] aimed at harmonizing and modernizing air traffic operations (ATO) in Europe and the Asia/Pacific region respectively. The transition from current ATO to NextGen, SESAR and SAS will rely heavily on procedures based on Global Positioning System (GPS) navigation, shared surveillance and a defined route network. The intent is to increase the predictability of aircraft movements within the airspace.

The current efforts to implement the vision for future ATO are founded on the premise that precision procedures are necessary to keep at or below acceptable threshold levels the risk to safety and the environmental consequences of meeting increased air traffic demand. Realization of the expected benefits depends on conformance with these precision procedures [4, 5]. Future ATO based on this dependence are hereafter referred to as Precision Air Traffic Operations (PATO).

*Aerospace Engineer, Aviation Systems Division, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035-0001. Email: douglas.r.isaacson@nasa.gov
†Aerospace Engineer, Aviation Systems Division, NASA Ames Research Center, Mail Stop 210-6, Moffett Field, CA 94035-0001. AIAA Member.
‡UCLA, Department of Mathematics, Los Angeles, CA 90095-1555
While scheduling problems have long been studied in operations research [6] for various industrial applications, the practice of treating the specific problems in air traffic management (ATM) as problems in optimization is, at the time of this writing, roughly but forty years old [7]. A likely cause for this lag is that application of standard operations research theory (e.g., network flows [8]) specifically to ATM faces a number of obstacles. Among these are the constraints, difficult to capture mathematically, that an aircraft in the air must maintain a minimal speed and a required separation from the other aircraft. Obstacles to using the standard theory of network flows include the following facts:

- Static flow network algorithms (e.g., Ford & Fulkerson’s [9]) cannot be readily applied to PATO because the required separation between two in-trail aircraft depends, in particular, on the order in which they enter an edge. This dependence hinders a definition of edge capacity that is constant in time. One may contemplate meeting separation constraints by using the operation-wide maximum required separation between all pairs of aircraft, but this leads to an underestimate of edge capacity and hence to suboptimal solutions.

- Among the difficulties with applying dynamic network flows are the multicommodity [8] nature of the problems (more than one type of flow carried by the same network) and the assignment of different cost to different orders in which flows enter an edge in the multigraph. In particular, the presence of nonlinearities (e.g., in inequality (7) that models separation constraints) and uncertain perturbations (see the end of section II.C.2) in the ATO severely limit the applicability of linear programs (LP).

Yet another obstacle is that an aircraft’s path is not planned with absolute precision; that is, there is often a lack of intent information for a given aircraft. The currently practiced notion of air traffic scheduling does not always prescribe a time-parameterized curve in 3-D space to be traversed by a flight. Rather, air traffic scheduling relies on air traffic control (ATC) to devise and issue legally binding instructions, called clearances, to each aircraft and to modify the schedule when necessary in order to maintain aircraft separation. Researchers have devised algorithms that produce, at a minimum, a schedule of runway arrival operations (for a detailed review, see Section IV). In the process of computing a schedule, most of these algorithms also end up calculating, in entirety or in part, aircraft trajectories. Currently, however, such automated trajectory synthesis does not meet simultaneously a given schedule and the relevant operational constraints. Furthermore, it is unclear how systems implementing automated trajectory synthesis will account for uncertainties and stochastic processes inherent to ATO; adaptability of human controllers and pilots is vital to the safe operation of current ATO. Consequently, the automated trajectory synthesis envisioned for PATO will raise the questions of how the responsibility for scheduling and for separation of aircraft should be divided between the automation system and the human controller, and how reliability is assured for PATO using automated trajectories. Having this synthesis automated, and aircraft separation handled by the controller, would leave the controller with the full weight of responsibility for separation while considerably restricting their freedom in the choice of aircraft maneuvers. Such a practice is likely to limit the benefits associated with the precision procedures envisioned for PATO. This consideration has led us to attempt developing a rigorous modeling framework for defining an aircraft’s trajectory and to research trajectory-generating algorithms that aim, as a controller does today, to maintain separation along the entire continuous trajectory for each aircraft. This framework has been used in recent literature to treat a number of special cases of PATO and other transportation problems [10, 11, 12, 13, 14].

The main motivation behind this paper is to aid those who wish to enter the research field of ATO but lack access to the required finer domain knowledge of operational constraints that must be included in a realistic model. Hitherto, such knowledge was unavailable in a single, widely accessible source. The organization of the paper is dictated by its central goal: to formulate, mathematically, a general problem of defining conceptually, constructing, and using a schedule for PATO that contains specification of merging sequences and provides aircraft separation continuously in time. Such a formulation is the content of Section III. A context necessary for understanding the formulation and its limitations includes the following three aspects of ATO: Federal Aviation Regulations, current practices, and PATO envisioned for the future. This context is provided in Section II. A detailed classification of the constraints (the key component of the finer domain knowledge) arising in the problem formulation is given in Section IV. Problem formulation precedes the constraint classification. This allows the operational description of each constraint in section IV to be accompanied, for clarity, by a mathematical formulation of the same constraint. Finally, to assess what is novel and significant in Sections III and IV, one needs certain knowledge of prior research on future ATO and, in particular, on the role of ATC in these operations. This research is reviewed in section V. The key ATO terms used throughout this paper are gathered and defined in Table 1.

At least three programs (NextGen, SESAR, and SAS) are currently underway to modernize ATO. While some differences exist in the programs’ implementations, each has the goal of harmonizing globally ATO, and all share the basic principles of PATO. The remainder of this paper provides background and examples specific to the U.S.
NAS, but the authors believe that the problem statement and constraint formulations are applicable for all future PATO implementations.

### Table 1: A glossary of ATO terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
</table>
| 4-D trajectory             | 4DT     | an aircraft trajectory either parameterized in time or with prescribed airspeeds along its route of flight
| adaptation                  |         | a set of data used by an ATC automation system, and specific to modeling ATO, ATC procedures and ATC practice |
| Air Route Traffic Control Center | ARTCC  | an air traffic facility established to provide services to aircraft operating on IFR Flight Plans generally in the en route phase of flight [16] |
| Air Traffic Control         | ATC     | services provided by the appropriate authority to promote the safe, orderly and expeditious flow of air traffic |
| air traffic demand          |         | consumer desire for air traffic services such as air travel or air cargo (i.e., the economic demand, see Section II.C.2) |
| Air Traffic Operations      | ATO     | the collected movements and management of aircraft within the NAS                           |
| aircraft weight class       |         | a classification of aircraft used to prescribe safe following distance for the purposes of wake turbulence hazard avoidance (see Reference [16]) |
| airspace                    | ATC     | a volume above the surface of the earth used for ATO                                         |
| Airport Traffic Control Tower | ATCT   | an air traffic facility established to provide services to aircraft in the vicinity of the airport or on the airport surface |
| altimeter                   |         | instrument used to indicate aircraft altitude to the flight crew                            |
| altimeter setting           |         | the baseline barometric pressure used to compensate for variations in existing atmospheric pressure, or the pressure at sea level for the international standard atmosphere when aircraft are operating above the transition altitude |
| Area Navigation             | RNAV    | a method of navigation which permits aircraft operation on any desired flight path within the coverage volume of ground-based, space-based, or self-contained navigation aids |
| ATC procedure               |         | a preplanned IFR procedure published for pilot use in graphic and/or textual form             |
| clearance                   |         | an authorization for an aircraft to proceed under conditions specified by ATC                |
| delay                       |         | a term used within the literature with varying meaning, typically referring to additional flight time incurred by an aircraft compared to an assumed estimated time of arrival. |
| density altitude            |         | altitude in the international standard atmosphere that corresponds to the measured air density at the aircraft. The pressure altitude corrected for non-standard air temperature. |
| deregulation                |         | refers in ATM to the enactment, in 1978, of the Airline Deregulation Act [17] that removed the requirement for U.S. government approval of airline routes (i.e., service between airport origin-destination pairs), schedules and fares. Deregulation of other markets followed (e.g., the European Union during the 1990’s). |
| en route                    |         | the portion of an aircraft’s flight occurring in between the departure and destination terminal areas |
| flight level                | FL      | aircraft pressure altitude in 100’s of feet, referenced to standard sea level pressure, and restricted to be multiples of 500 feet. Flight levels are only used above the Transition Altitude (18,000 feet in the U.S. and Canada, but varies internationally). (see Transition Altitude) |
| Flight Management System    | FMS     | hardware on board the aircraft used by the pilots of that aircraft to interface with the autopilot and manage the flight according to an objective that includes fuel and labor costs |
| Flight Operations Manual    | FOM     | a manual of procedures and guidance for the conduct of flight for an aircraft, typically containing requirements and conditions of operations set forth for the aircraft as part of its airworthiness certification, as well as preferred practice as prescribed by the manufacturer or airline company |
| flight plan                 |         | information related to the intended flight of an aircraft filed with the ATC authority       |
| flight route                |         | the path which aircraft traverse over the surface of the earth as defined in its flight plan by a sequence of successively traversed waypoints |
| ground speed                |         | the speed of an aircraft relative to the surface of the Earth                                |
| ground track                |         | the projection of an aircraft trajectory, or portion thereof, onto the surface of the Earth  |

Continued on next page

---

\(^a\)This use of the term is inconsistent with the standard terminology of dynamical systems; e.g., see [15].

\(^b\)A typical flight route includes reference to ATC procedures and jet routes that are subsequently defined as a sequence of successively traversed waypoints.
<table>
<thead>
<tr>
<th>Term</th>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>indicated airspeed</td>
<td>IAS</td>
<td>the speed of an aircraft as displayed on the aircraft’s airspeed indicator. The equivalent TAS under standard sea-level conditions of the international standard atmosphere</td>
</tr>
<tr>
<td>indicated altitude</td>
<td></td>
<td>the altitude indicated by an aircraft’s altimeter (with the altimeter setting adjusted as required for the flight operation)</td>
</tr>
<tr>
<td>Instrument Approach Procedure</td>
<td>IAP</td>
<td>an ATC procedure for landing aircraft that prescribes maneuvers for the orderly transfer of an IFR aircraft from a waypoint within terminal airspace (i.e., the initial approach fix) to a landing or to a point from which a landing may be made visually</td>
</tr>
<tr>
<td>Instrument Flight Rules</td>
<td>IFR</td>
<td>rules governing the procedures and conduct of flight under IMC [1]</td>
</tr>
<tr>
<td>Instrument Meteorological Conditions</td>
<td>IMC</td>
<td>meteorological conditions expressed in terms of visibility, aircraft distance from nearest cloud, and cloud ceiling less than the minima specified for VMC</td>
</tr>
<tr>
<td>intent information</td>
<td></td>
<td>knowledge of the intended flight path of an aircraft</td>
</tr>
<tr>
<td>Lateral Navigation</td>
<td>LNAV</td>
<td>a function of RNAV providing electronic guidance to the pilot or autopilot for navigation over a ground track</td>
</tr>
<tr>
<td>National Airspace System</td>
<td>NAS</td>
<td>the collection of U.S. airports, airspace, ATC facilities and services, and supporting infrastructure, technologies and procedural publications</td>
</tr>
<tr>
<td>Pilot in Command</td>
<td>PIC</td>
<td>the person authorized and responsible for the operation of an aircraft</td>
</tr>
<tr>
<td>Precision ATO</td>
<td>PATO</td>
<td>envisioned future ATO which relies on conformance with RNAV and RNP procedures</td>
</tr>
<tr>
<td>pressure altitude</td>
<td></td>
<td>The altitude indicated on an aircraft’s altimeter when the altimeter setting is set to an agreed baseline barometric pressure</td>
</tr>
<tr>
<td>radial</td>
<td></td>
<td>a magnetic bearing extending from a ground-based navigation aid</td>
</tr>
<tr>
<td>requested rate of traffic acceptance</td>
<td></td>
<td>the rate at which requests capable of being fulfilled come in at an air traffic resource such as an airport or TRACON (see Section II.C.2)</td>
</tr>
<tr>
<td>Required Navigation Performance</td>
<td>RNP</td>
<td>a statement of the necessary spatial navigational point performance for operation within a given airspace or for a specific air traffic procedure</td>
</tr>
<tr>
<td>route segment</td>
<td></td>
<td>a consecutive pair of waypoints in the flight route</td>
</tr>
<tr>
<td>sector</td>
<td></td>
<td>a contiguous portion of airspace allocated to and managed by an air traffic controller providing ATC services to those aircraft within its volume</td>
</tr>
<tr>
<td>separation minima</td>
<td></td>
<td>the minimum longitudinal, lateral, or vertical distances by which aircraft are spaced through application of ATC and defined by policy to assure safety</td>
</tr>
<tr>
<td>Standard Instrument Departure</td>
<td>SID</td>
<td>an ATC procedure for departing aircraft that provides for obstacle clearance and defines a transition from the terminal area to the en route airspace</td>
</tr>
<tr>
<td>Standard Terminal Arrival</td>
<td>STAR</td>
<td>an ATC procedure for arriving aircraft that defines a transition from the en route airspace to an IAP or waypoint in the terminal area</td>
</tr>
<tr>
<td>terminal area</td>
<td></td>
<td>airspace in which approach control service or airport traffic control service is available (from [16])</td>
</tr>
<tr>
<td>Terminal Radar Approach Control</td>
<td>TRACON</td>
<td>an air traffic facility established to provide services primarily to arriving and departing aircraft</td>
</tr>
<tr>
<td>trajectory</td>
<td></td>
<td>a curve through space flown by an aircraft along its route of flight</td>
</tr>
<tr>
<td>transition altitude</td>
<td></td>
<td>the pressure altitude above sea-level above which aircraft are required to use flight levels for navigational purposes (18,000 feet in the U.S., but varies internationally)</td>
</tr>
<tr>
<td>true airspeed</td>
<td>TAS</td>
<td>the speed of an aircraft relative to the surrounding air mass</td>
</tr>
<tr>
<td>vector</td>
<td></td>
<td>a heading issued to an aircraft to provide spatial navigational point guidance by radar [16]</td>
</tr>
<tr>
<td>Vertical Navigation</td>
<td>VNAV</td>
<td>a function of RNAV which provides electronic guidance to the pilot or autopilot for navigation to a vertical path or to meet vertical constraints at waypoints along the route of flight</td>
</tr>
<tr>
<td>visual approach</td>
<td></td>
<td>an approach to landing conducted by the PIC using visual separation and requiring that the aircraft remain clear of clouds and that the PIC maintain visual contact with airport or the preceding aircraft on approach</td>
</tr>
<tr>
<td>Visual Flight Rules</td>
<td>VFR</td>
<td>rules governing the procedures and conduct of flight under VMC</td>
</tr>
<tr>
<td>Visual Meteorological Conditions</td>
<td>VMC</td>
<td>meteorological conditions expressed in terms of visibility, aircraft distance from clouds, and cloud ceiling equal to or greater than some specified minima</td>
</tr>
</tbody>
</table>

*In comparison to RNAV, RNP requires the additional on board functions of spatial navigational point performance monitoring and alerting for non-conformance. Existing RNP standards only treat lateral RNP; standards for vertical RNP are under development.*
### Table 1—continued from previous page

<table>
<thead>
<tr>
<th>Term</th>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>visual separation</td>
<td></td>
<td>the practice of spacing aircraft by visual means, either the pilot maneuvering the aircraft to see and avoid other aircraft, or an ATCT controller seeing the aircraft involved and issuing instructions to ensure that the aircraft avoid each other</td>
</tr>
<tr>
<td>waypoint</td>
<td></td>
<td>a geographical location on the surface of the earth that among its ATO uses are spatial navigational point reference and ATC procedure definition</td>
</tr>
</tbody>
</table>

### II. Current and Envisioned Air Traffic Operations: An Overview of Policies and Practice

This section provides PATO scheduling context for readers unfamiliar with current ATO. As a way of an exposition to the basic responsibilities of pilots and air traffic controllers, section II.A provides an introduction to the rules and procedures applicable to current ATO. An overview of current ATO, in section II.B, is aimed to describe and explain the existing system, for which most of the prior research was conducted. An overview of future ATO (section II.D) is presented as a discussion of the transition from current operations to the PATO environment.

#### II.A. An Introduction to the Rules Governing the Operation of Aircraft

This section is a brief summary of the rules governing the operation of aircraft in proximity to other aircraft. A full specification of these rules can be found in Federal Aviation Regulations (Title 14 CFR Part 91: General Operating and Flight Rules [18]) and in FAA Job Order 7110.65U [16], which “prescribes the procedures and phraseology for use by personnel providing air traffic control services.”

Reference [16, Chapter 5] prescribes the procedures for Air Traffic Control (ATC) radar separation of aircraft, hereafter referred to as separation service. An air traffic controller directing aircraft to maintain a safe separation between aircraft of 5 nmi en route is providing a separation service to those aircraft. The pilot in command (PIC) is ultimately responsible for the operation of her or his aircraft (14 CFR Part 91.3) and is responsible for seeing and avoiding other aircraft when weather conditions permit (14 CFR Part 91.113). Document 14 CFR Part 91.221(b) requires that pilots use the Traffic Alert and Collision Avoidance System (TCAS) if it is operable on the aircraft. TCAS is a collision avoidance system that provides aural and visual alerts to pilots when TCAS determines that proximate aircraft present a collision risk. When receiving separation service from the Air Traffic Control (ATC) authority, the pilot must additionally comply with ATC clearances except in the case of emergency (14 CFR Part 91.123). These regulations are mostly expected to apply in their present form to PATO, as there are currently no plans to change them. It is assumed in this paper that all aircraft included in the operations are subject to, and in compliance with, ATC clearances. In other words, self-separating aircraft are not considered in the problem formulation, and ATC may be engaged to instruct aircraft to comply with a schedule of operations. This is not to imply that separation will be maintained across all conditions (even those unanticipated by ATC automation) or that aircraft in PATO will always be operating under IFR. Mechanisms will need to be developed to handle such exceptions to standard procedures. While the robustness of PATO to these exceptions and to failure modes (including failures in ATC and flight deck automation) is an area of research that must be addressed for PATO to be realized, all discussions below are set in the context of the aforementioned IFR operations. (Aircraft operating under Visual Flight Rules (VFR) are not addressed in this paper. Such aircraft are not subject to ATC instruction, and even though mechanisms and procedures must be developed to accommodate operations under VFR in some PATO airspaces, they are considered off-nominal operations and are beyond the scope of this paper.)

#### II.B. Overview of Current Operations and Practices

**II.B.1. Air Traffic Control, flight plans, and route networks**

Air Traffic Control is defined [16] as “a service operated by appropriate authority to promote the safe, orderly and expeditious flow of air traffic.” As part of this service, aircraft are prescribed routes that consist of concatenated route segments defined by: SIDs, jet routes or airways, STARs, and IAPs. The assignment of an aircraft to a specific route of flight is determined prior to departure in current ATO, but may not be fully specified from origin to destination. This assignment is referred to as the aircraft’s flight route and, together with certain other information, constitutes the

---

\[\text{d}\]Requirements for TCAS equipage are based on an aircraft’s: operating authority (civil vs. public), method of propulsion (e.g., turbine powered), maximum number of passengers, maximum takeoff weight, and/or maximum payload capacity. The specific criteria vary by jurisdiction.

\[\text{e}\]Self separation refers to a provision for separation service in which the PIC for an aircraft is responsible for safe separation from other aircraft and may not be subject to the rules of 14 CFR 91.123 (compliance with ATC instructions). Self separation procedures and methods for scheduling self separating aircraft are still subjects of research.

5 of 34
aircraft’s flight plan. Deviations from and amendments to an aircraft’s flight plan may occur as the flight progresses toward destination. The rules governing normal aircraft operations described in Section II.A require that all flight plan deviations and amendments are directed or explicitly approved by ATC. However, such deviations are sometimes the result of human adaptation to abnormal conditions or failure conditions that were not predicted or (in some cases) identified by ATC automation systems. The collection of all routes available for aircraft navigation in a given airspace is hereafter called the route network. Figure 1 shows the high altitude jet route network for the NAS over the Continental United States. Aircraft navigating closer to their origin or destination airport follow arrival, departure, and runway approach procedures. Figure 2 shows the collection of arrival, departure, and approach procedures that comprise the route network for a fictitious, two-runway airport. The route network, the rules governing aircraft deviation from the approved flight plan, and pilot and controller adaptability to unforeseen events and failures, are all critical to safe and orderly provision of air traffic services.

II.B.2. Airspace partitioning into sectors

While the route network provides a framework for organizing air traffic flows in congested airspace, sectorization of airspace (i.e., the practice of partitioning an airspace into subregions called sectors—see Section IV.B.1 for details) is used to manage air traffic controller workload by limiting the number of aircraft that are being serviced and the scope of problem being managed by a single air traffic controller at any given time. Each sector is managed by a single controller, sometimes with support from 1 or 2 planning controller(s). Controllers are responsible only for the aircraft within their sector and have no other part of airspace at their disposal. Sectors are designed to have sufficient volume for vectoring (see section II.C) aircraft to ensure required separation, to accommodate unanticipated increases in the number of aircraft occupying the sector, and to manage off-nominal events or in-flight emergencies. A flight crew will communicate with only one controller within each sector.

En route and terminal airspace is partitioned into sectors that generally follow the major flows of air traffic through the airspace (i.e., the most congested routes in the route network). In en route airspace near major airports, and especially in congested TRACON airspace, air traffic flows often compete for limited airspace. Figure 3 shows regions where the arrival and departure routes of Figure 2 conflict. Sectorization of airspace serves to simplify the air traffic controller’s tasks by procedurally separating these otherwise conflicting air traffic flows. The resulting airspace structure is often complex, with many sectors interlocking in a rarely intuitive fashion. The example airspace shown in Figure 4 is provided to demonstrate the complexity in modeling sectors in congested terminal airspace. Figure 4b
Figure 2. Example TRACON route network

II.C. ATC clearances and intent information

In order to provide required separation between aircraft and to lead the aircraft along the paths prescribed in their flight plans, air traffic controllers issue legally binding instructions, called clearances, to the aircraft currently receiving their separation service.

Clearances issued by a controller to an aircraft flight crew can take a number of forms. Each clearance, however,
can be generally classified as either closed or open. A closed clearance is one that, if it directs an aircraft to deviate from the flight route, brings the aircraft back to its flight route. An open clearance, also called a vector\(^3\), does not return the aircraft to its flight route without additional direction from ATC. In general, it is not known at a given point during the flight when, and whether, ATC will issue further instructions to restore an aircraft to its flight route. Thus, open clearances introduce considerably more uncertainty into aircraft trajectory predictions than do closed clearances. Often, controllers issue open clearances to aircraft in the process of developing a plan for assuring separation, and at such times the controller does not yet have all the clearances planned for the entire duration of the flight, but rather chooses and issues them as the flight progresses. Thus, ATC automation systems lack accurate intent information (see Table 1) for aircraft that have been vectored (issued an open clearance) without a return to flight route.

The example shown in Figure 5 illustrates the difference between open and closed clearances for an aircraft originally navigating to waypoint WP1. Figure 5a shows the expected flight path for an aircraft that has been instructed to “turn right, heading zero-niner-zero, intercept the WP2-one-eight-zero radial, resume course at WP1.” This clearance instructs the PIC to turn the aircraft to the right until reaching a magnetic heading of 90 degrees, thence to follow the 90 degree heading until intercepting a ray emanating from the waypoint WP2 at 180 degrees magnetic heading (due south direction) until reaching the WP1 waypoint, whence navigation of the previously defined route should resume (beginning with WP1 in the original flight route). Figure 5b shows a number of possible flight paths that could be expected following the open clearance “turn right, heading zero-niner-zero.” The range of possible future positions of the aircraft in Figure 5b illustrates the difficulty faced by automation systems in accurately predicting the future positions of aircraft receiving open clearances. For example, one cannot determine the required speed for minimal spacing with a preceding aircraft without a precise definition of path of flight.

When issuing clearances to assure separation, air traffic controllers are required to use the radar separation minima dictated in [16, Chapter 5], or to delegate responsibility for separation from a specific hazard or aircraft temporarily to the flight crew via a visual clearance. The radar separation minima were assessed to include a margin of safety sufficient to mitigate the uncertainties in surveillance, communication, navigation and control of IFR operations receiving ATC separation services\(^1\). When the number of aircraft occupying an airspace approaches the capacity of that airspace, controllers strive, in addition to assuring separation, to separate aircraft as near the required minima as is practical (i.e., to minimize excess separation). This minimization is intended to limit unnecessary delay and fuel burn, to increase runway and airport throughput, and if possible to lower the controllers’ workload by maximizing airspace available for separation of other aircraft. The controller’s task of separating aircraft by at least the required minima without introducing significant excess separation becomes increasingly difficult as the average number of aircraft per unit volume of available airspace increases.

For most arrival and departure operations in congested terminal airspace, a normal separation practice is for controllers to issue open clearances for aircraft. Furthermore, many current terminal procedures (STARs and SIDs \([20]\)) include the phrase “expect vectors” that implicitly requires open clearances. The alternative, i.e., an implementation of closed clearances in a congested terminal environment, would require a substantial increase in the controllers’ workload. A key goal of PATO \([1]\) is to increase the safety and capacity of the NAS by adding considerably to the precision of aircraft navigation and of ATC practices while maintaining safe levels of controller workload.

\section*{II.1. Henceforth, the noun vector will be used only with the mathematical meaning, i.e., an element of a vector space.}

\section*{II.C.1. ATC application of aircraft speed adjustment}

Reference [16, Chapter 5, Section 7] prescribes preferred ATC practice for adjusting aircraft speed. Speed clearances are expressed in terms of knots indicated airspeed (IAS) below FL 240, and may also be expressed in terms of Mach number at or above FL240. Speed adjustments expressed in terms of knots IAS are provided in 10-knot increments and those expressed in terms of Mach Number in 0.01 increments. At high altitudes and high speed, significant error can be introduced to the IAS due to compressibility effects \([21]\). For this reason, speed adjustments for jet aircraft at higher altitudes (i.e., at or above FL 240) are typically expressed in terms of Mach Number.

\footnote{\(^3\)Also used as a verb: to vector.}

\footnote{\(^1\)New technologies supporting aircraft communications, navigation and surveillance (including their associated failure modes) call into question the fitness of existing separation standards for PATO. This fitness (an area of ongoing research) must be considered when evaluating PATO safety and reliability.}
II.C.2. Air traffic demand and requested rate of traffic acceptance

Throughout ATM literature, the term *air traffic demand* appears with at least two different meanings. One is economic: *air traffic demand* refers to a consumer’s desire for certain air travel. In this paper, the term *air traffic demand* will be used only with this, economic, meaning, consistent with common economic terminology.

The other meaning is specific to ATM operations: given two airspace facilities (e.g., sectors), one producing an outflow of air traffic to be accepted by the other, *air traffic demand* refers to the rate of the outflow. In this, latter, use, “demand” is actually the rate at which requests capable of being fulfilled come in at a service station in a queuing system. In what follows, this rate will be referred to as the *requested rate of traffic acceptance*.

Perturbations to ATO (e.g., an unforeseen reduction in airport arrival capacity due to weather forecasting errors) may reduce the capacity of the service station (e.g., an airport) to a level below the requested rate of traffic acceptance. Such a reduction effects a service delay for an aircraft, through two mechanisms, which may be used by the ATC either one at a time or together: a reduction in the speed of aircraft along the route of flight or an increase in the distance to be flown (i.e., vectoring or rerouting). When service delay becomes excessive, ATC typically places aircraft into holding patterns (which can be modeled as *cycles* [23] in the route network) and release aircraft from holding patterns when the situation allows.

II.D. Envisioned Precision Air Traffic Operations: an Overview

Under PA TO, aircraft will navigate procedures called *Area Navigation (RNAV)* and *Required Navigation Performance (RNP)*. RNAV and RNP procedures will provide the precise definition of the expected route of flight needed by ATC automation systems. ATC automation systems will use the intent information provided by RNAV and RNP procedures to predict a trajectory for each aircraft with much greater certainty than is possible in today’s operations.

Accurate aircraft surveillance will be provided by the Automatic Dependent Surveillance-Broadcast (ADS-B) system mandated for most aircraft. The required ADS-B system consists of two components: an IFR-certified GPS receiver and a transmitter to broadcast aircraft position. ATC will then employ, as decision aids, automation systems that build on the improved surveillance and trajectory predictions to manage safely the expected increase in requested rate of traffic acceptance. For example, NASA’s *Terminal Area Precision Spacing and Scheduling System (TAPSS)* is being developed to aid controllers in spacing arrival aircraft on RNAV and RNP arrival procedures [24]. In TAPSS, spacing cues are provided to the controller and are based on the aforementioned improved trajectory predictions for the final 200-400 nmi of flight. Use of RNAV and RNP procedures in PA TO, however, faces a number of obstacles. The three most significant ones are as follows.

- **Use of speed control as primary aircraft separation mechanism**: Since air traffic demand is expected to increase significantly over the next decades, compliance with RNAV and RNP procedures may not be possible with existing ATC separation practice and automation systems. However, with the more accurate intent information of RNAV and RNP procedures, PA TO will allow improved scheduling of aircraft operations and a more strategic treatment of separation assurance. Thus, we expect that separation of aircraft will be based almost exclusively on speed control instructions, with flight path changes during the flight allowed only when special considerations require. Separation service based on speed control requires careful treatment of the problem constraints, as aircraft inertia prevents speed control from being effective in tactically separating aircraft. Constraints typical of PA TO are described in Section IV.

- **ATC promotion of safe and expeditious air traffic flow**: As previously mentioned, air traffic control must promote “safe, orderly and expeditious flow of air traffic.” While Traffic Flow Management (TFM), a tool for the strategic treatment of air traffic flows [25], plays a critical role in ATC services, such treatment itself is beyond the scope of this paper. The focus here is on ATC services for flows of air traffic established after TFM has limited these flows to be within reasonable bounds for airspace and airport capacity. For such air traffic flows, RNAV and RNP procedures and the prescribed airspace structure will serve to provide orderly flow of air traffic, provided mechanism(s) for separating aircraft safely and for serving air traffic flows in an expeditious manner are developed. The ability of ATC to employ speed control for aircraft separation safely, while simultaneously promoting expeditious flow of traffic, is a key challenge to achieving the safety, efficiency, and capacity goals of NextGen and SESAR.

- **The principally unpredictable and little-understood perturbations that jeopardize the reliability of an operation based on RNAV and RNP**: Such perturbations include, but are not limited to: unexpected changes in the weather conditions, errors in the flow of information from ATC to the flight crew, tactical decisions on the part of

---

1 This use is consistent with the definition of *demand* in the classical monograph [22].
the flight crew to deviate from an ATC clearance (e.g., in response to a change in weather or to an in-flight emergency), and errors on the part of the flight crew in the execution of a clearance. While none of these perturbations admits a deterministic predictive model, the extent to which they can be modeled stochastically (i.e., as multiple repetitions of an experiment that obeys a stable probability distribution) is an open question. It is known from ATC and flight crews that they correct tactically for certain types of perturbations to the planned ATO, but the statistical frequencies and patterns of such corrections are unknown. Thus, automation of a substantial part of ATC, though it removes the corresponding sources of human error, carries the unknown risk of losing human adaptivity to the aforementioned perturbations. Once the stochasticity of a given perturbation is established with an acceptable level of confidence, the sensitivity of the system to this perturbation can be studied using probabilistic techniques (see, e.g., [26]).

PAT0 air traffic flows will mostly resemble those of today, but will be characterized by the precisely defined paths of RNAV and RNP procedures. The combination of all procedures in use by a TRA0N and its adjacent Air Route Traffic Control Centers (ARTCC’s) at a given time dictates what air traffic flows occur and which routes of flight are allowed for aircraft operating in those ATC facilities. The examples of this section illustrate a simplified view of PAT0 aircraft flows and terminal airspace structure, focused primarily on air carrier operations originating and terminating at major airports. The reader is referred to Reference [27, Chapter 3] for a comprehensive treatment of U.S. airspace classes and for additional instrument and visual procedures in use today.

III. Problem Definition

Throughout the rest of the paper, the mathematical machinery corresponding to, and immediately following, specific operational considerations is typeset in italics.

III.A. The need for a standard and precise conceptual framework

Many publications state, in some form, PAT0 Scheduling as their central topic. Although pursuing similar research goals, these publications share neither a standard and unambiguous definition of the term “schedule,” nor a precise and universal formulation of a scheduling problem, in the context of PAT0. This lack of standards and structure makes it difficult for researchers to benefit from each other’s work and, as a result, for the research community to organize its efforts.

This section offers such a definition and formulation for the nominal PAT0; i.e., for those where each aircraft is expected to follow precisely the route prescribed by its flight plan. The formulation below is based on a mathematical model of the relevant ATO. Special cases of the model have been used in a number of recent research efforts [14, 13, 12]. The PAT0 concepts involved in the model, their mathematical counterparts, and references defining the latter, are listed in Table 2.

<table>
<thead>
<tr>
<th>PAT0</th>
<th>mathematics</th>
<th>reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>airspace</td>
<td>a region in (or a submanifold of) a Euclidean</td>
<td>[28, 29]</td>
</tr>
<tr>
<td>route network</td>
<td>multigraph</td>
<td>[23]</td>
</tr>
<tr>
<td>waypoints and runway thresholds</td>
<td>vertices in the multigraph</td>
<td>[23]</td>
</tr>
<tr>
<td>route segments</td>
<td>arcs in the multigraph</td>
<td>[23]</td>
</tr>
<tr>
<td>a flight route</td>
<td>a path in the multigraph</td>
<td>[23]</td>
</tr>
</tbody>
</table>

We first formulate the model. Each of the following subsections is devoted to an aspect of the model. A discussion of the aspect starts with the PAT0 considerations and ends with the italicized corresponding mathematical machinery. The model, once formulated, provides the precise terminology for stating and discussing the problem unambiguously and concisely. A short discussion of the inherent uncertainties in PAT0 is included in Appendix A.

\(^{k}\)For example, an isobaric surface in the atmosphere, with non-negligible curvature for long-range flights.
III.B. Dimensionality of the physical airspace

The choice of dimension of the physical airspace to be modeled depends on the airspace domain and on the procedural constraints. In some cases, it may be practical to exclude the vertical dimension without sacrificing operational realism in the solution. Among the special cases in which the vertical dimension may be neglected are the following two:

- Arrival and departure procedures are spatially segregated for aircraft in close proximity to the airport.
- Only airport surface traffic is considered. (Surface ATO is an area of active research in its own right; see, e.g., [30].)

The physical airspace is viewed as a region in a Euclidean vector space, henceforth denoted $E$, of dimension 2 or 3.

III.C. Waypoints and spatial navigational points

A **waypoint** is a specific point on the surface of the earth. In ATO, waypoints are used for aircraft navigation and definition of procedures for arrival, departure, and approach [16]. A **flight route** is a sequence of waypoints; a **route segment**, a consecutive pair of waypoints in a flight route.

Missing from these specifications is that of altitude. In ATO, waypoints are defined in terms of latitude and longitude according to the World Geodetic System (WGS-84) [31], while an aircraft’s altitude is defined by the equivalent pressure altitude (referred to as barometric altitude) and is dictated mostly by controller clearances. The barometric altitude assigned at a given waypoint determines the height above and normal to the surface of the WGS-84 reference ellipsoid (approximation of the Earth’s surface). Since **vertical navigation** (VNAV)—i.e., control of altitude—and **longitudinal navigation** (LNAV) are handled separately from each other, an aircraft’s flight plan, while including a flight route, generally does not specify a curve in 3-D space to be traversed by the aircraft. However, most ATO situations of interest can be modeled with the assumption that each aircraft is to traverse such a path, which is known in advance. Therefore, in the formulation of our modeling framework, we adopt the following assumption (situations where it fails are discussed below, in Section IV.C.3).

**Assumption III.1.** Each flight is to traverse a specified curve, called a path, in 3-D space.

The point of intersection between such a path and the abovementioned ray corresponding to a waypoint will be called a **spatial navigational point**. The mathematical counterpart of this construct is as follows.

The spatial navigational points (for convenience, this term will also include runway thresholds) are viewed as points in $E$. The spatial navigational points, which constitute a finite set, henceforth denoted $V$, are then used as the vertices of a multigraph, henceforth denoted $G$, whose arcs are constructed as follows: for every route segment specified (by providing not just the two waypoints, but also two altitudes) to go from spatial navigational point $p_1$ to spatial navigational point $p_2$, the multigraph $G$ has an arc from $p_1$ to $p_2$.

III.D. Route segments

The actual spatial curve corresponding in air traffic procedures to a route segment is generally curvilinear. Furthermore, since it must be navigable by inertial dynamical systems (aircraft), such a curve is assumed rectifiable and capable of arc length parameterization which is continuous and piecewise continuously differentiable.

Each arc in $G$ from $p_1$ to $p_2$ is identified with a curve in $E$ from $p_1$ to $p_2$ whose arc length parameterization is continuous and piecewise differentiable. In what follows, “arc” will refer to such a curve.

III.E. Movement of aircraft in a route network

Aircraft traverse airspace according to their routes of flight, as specified in each aircraft’s flight plan.

The aircraft modeled are regarded as the elements of a finite set $A = \{1, \ldots, A\}$ of moving agents $\alpha \in A$, each of which is a point moving along an arc of $G$. To each moving agent we associate a path in the multigraph $G$. Such a path, being a concatenation of continuous and piecewise differentiable curves, is itself such a curve. Once an arc length coordinate $s^{(\alpha)}(t)$ is introduced on this curve for moving agent $\alpha$, the value of this coordinate at time $t$ completely specifies the physical position of agent $\alpha$ at time $t$. It follows immediately that the time derivative of $s^{(\alpha)}(t)$ gives the agent’s instantaneous ground speed along its path.

---

1Such a path is not to be confused with the term 3D-trajectory, commonly used in ATO with a context-dependent meaning.
Remark III.1. The aircraft are indexed here using the symbol \( \alpha \) so that the reader, instead of having to memorize an association between an arbitrary symbol (say, i or j) and aircraft, has the mnemonic association with the first letter of the word aircraft. (The Latin letter a as the index symbol would have likely been confused with the indefinite article.)

When the symbol \( \alpha \) appears as a superscript, it is parenthesized, to avoid having it read as an exponent\(^\text{m}\): \( s^{(\alpha)} \).

III.F. Restrictions on the movement of aircraft

Among the numerous operational restrictions (enumerated in Section IV) on how air traffic may move in a route network are the following three: minimal required separation, permissible speed range, and airspace boundaries. Restrictions commonly left out of explicit consideration include the initial positions of the aircraft, as well as any requirements for an aircraft to arrive at a given waypoint within a given time interval. All mathematical models formulated in this paper are based on the following.

Axiom III.1. All restrictions on the movement of air traffic in a route network can be described as constraints on the \( s^{(\alpha)}(t) \)'s and on their time derivatives of order 1 or higher, either for all meaningful continuously varying values of \( t \), or on some finite set of these values (i.e., as boundary or intermediate constraints), as appropriate. (See [32] for a treatment of optimal control problems with intermediate constraints.)

III.G. Problem statement

The flow of air traffic is the movement of the aircraft in \( \mathcal{A} \) along the route network, according to each aircraft’s flight route and subject to a set of constraints specified by the researcher to reflect the realism of operations under analysis. Once a flight route is specified for aircraft \( \alpha \), this route will be assumed parameterized by arc length \( s^{(\alpha)} \).

A specification for each aircraft in \( \mathcal{A} \) of a flight route will be called a routing for \( \mathcal{A} \). An example of a 2-aircraft set \( \mathcal{A} = \{1,2\} \), each aircraft with its assigned route, and the corresponding coordinate space are shown in Figure 6. Accordingly, a general class of PATO problems where the perturbations are stochastic can be formulated as follows (recall the notation described in remark III.1).

Given

- for each \( \alpha \in \mathcal{A} \), a point of origin in \( G \) and a desired destination in \( G \),
- the set of all relevant constraints on the \( s^{(\alpha)}(t) \)'s and on their time derivatives (including all the required times of departure from origin and arrival at destination, if these times are specified as either time instants or time intervals),
- a controlled dynamical system (generally stochastic) for the evolution of the \( s^{(\alpha)}(t) \)'s and, possibly, of their time derivatives (see section IV.A, top paragraph), and
- an objective function of the time-dependent coordinate vector

\[
s(t) = \left( s^{(\alpha)}(t) \right)_{\alpha \in \mathcal{A}} = \left( s^{(1)}(t), s^{(2)}(t), \ldots, s^{(\mathcal{A})}(t) \right)
\]  

(1)

and, possibly, of its time derivatives (see section IV.A, top paragraph),

find all appropriate routings for \( \mathcal{A} \) (in special cases, a routing may be partially or fully specified, e.g., see [14]) and, for each such assignment, a control strategy [33] that results in an \( A \)-dimensional and time-dependent coordinate vector \( (1) \) which gives a feasible \([33, 23]\) (i.e., constraint-compliant) state trajectory and, if needed, minimizes or maximizes the objective function, as appropriate to the operation modeled.

For treatment of certain special cases of this problem, see [14, 13, 12]. Generalizations of this problem are discussed in Appendix A. The choice of the objective function for a PATO problem is dictated by the specific operational needs being addressed by the model (i.e., those qualities of a feasible solution important to the ATC and airspace users in the operational environment). Quantities in a PATO objective function may include:

- flight delay\(^n\),
- fuel burn,
- sensitivity to stochastic perturbations, and
- accommodation of user preferences.

\(^m\)A superscript is sometimes used as an index even without parentheses; see, e.g., [22].

\(^n\)The meaning of the term delay varies widely throughout the ATM literature.
IV. Problem Constraints

To model ATO, a researcher needs to know not only the system constraints, but also, among other things, how these constraints are (or can be) implemented and enforced. In other words, the researcher needs to know the relevant regulations, policies, and practices aimed at safety of flight. Thus, while such safety is the ultimate purpose of each of the constraints described in this section, the aim here is to classify the constraints by the primary considerations behind their implementation. (The aspect of reliability, discussed briefly in section II.D, is not included in this classification because the required context—a functional model of PATO at a sufficiently low level and methodologies of assessing the safety of automated ATC—is yet to be developed, and the regulators’ final approval of such methodologies, yet to be obtained.)

These considerations can be roughly classified into the following types:

• **Physical** factors, i.e., the physics of flight and aircraft control,

• **Policy** factors, i.e., the official policies imposed on ATO by the FAA and other authorities (e.g., an airport authority imposing a curfew),

• **Procedural** requirements, i.e., restrictions imposed on the aircraft by the procedure (i.e., STAR, SID or IAP) that the aircraft has been cleared to execute, or by procedures defined in an aircraft’s flight operations manual, and

• **Operational** considerations, which do not come directly from physics, policy, or procedures, yet are supported by practical considerations and are a regularly practiced course of action in the execution of some ATO operation.

These four types are not mutually exclusive: a constraint can arise from any subset of them. Nevertheless, we now attempt to classify them by the single most relevant of the above four factors. The other factors, if any, involved in a constraint are also listed.

IV.A. Physical Constraints

A controlled dynamical system developed for each aircraft must be consistent with the performance limitations for the aircraft type and its operational characteristics (e.g., aircraft weight). (Those limitations concerning structural integrity and controlled flight are generally not directly encountered in ATO because other constraints (enumerated below) keep the system from reaching these limits. For a thorough treatment of the limitations on aircraft loads and controlled flight, the reader is referred to [21, section 12.7].) Many dynamic models for aircraft have been developed in the literature, and the suitability of a model is highly sensitive to the specific context. For example, kinematic aircraft dynamic models may be sufficient for modeling aircraft movement en route, but kinetic models are preferred for modeling operations near the airport, because such models are better at capturing multiple, sustained accelerations, typical in the terminal airspace. Two types of systems of state equations [28, section 11.8-1] in the modeling framework of Section III that have been used in the literature [13, 14, 12] to describe the time evolution of an aircraft’s arc length coordinate \( s^{(\alpha)} \) are the inertia-free system

\[
\frac{ds^{(\alpha)}}{dt} = v_{\text{ground}}^{(\alpha)}, \quad \alpha \in \mathcal{A},
\]  

(2)

with the ground speeds \( v_{\text{ground}}^{(\alpha)} \) (of the aircraft’s movement along the assigned path) the control variables, and the inertial system

\[
\frac{ds^{(\alpha)}}{dt} = v_{\text{ground}}^{(\alpha)}, \quad \frac{dv_{\text{ground}}^{(\alpha)}}{dt} = a_{\text{ground}}^{(\alpha)}, \quad \alpha \in \mathcal{A},
\]  

(3)

where the state variables are the \( s^{(\alpha)} \)'s and the \( v_{\text{ground}}^{(\alpha)} \)'s, and the control variables are the accelerations \( a^{(\alpha)} \). The decision whether to neglect or include inertia in the state equations should be made for each specific modeling effort, as appropriate. Given some acceptable tolerance on the physical execution for a given model, a suggested criterion for such a decision is whether the physically realistic execution of a computed control strategy is sufficiently close to that strategy.

If \( s^{(\alpha)} \) is the arc length coordinate for aircraft \( \alpha \), let

\[
X^{(\alpha)}(s^{(\alpha)}) = \left( x^{(\alpha)}(s^{(\alpha)}), y^{(\alpha)}(s^{(\alpha)}), \nu_{\text{ind}}^{(\alpha)}(s^{(\alpha)}) \right).
\]  

(4)

\(^{o}\)While ATO policies and clearances are typically stated in terms of indicated airspeed (see Table 1), state equations in the modeling framework proposed rely on ground speeds. The researcher must mind this in the process of modeling.
(here the coordinates $x^{(\alpha)}, y^{(\alpha)}$ are in a “horizontal plane,” and the coordinate $h_{\text{ind}}^{(\alpha)}$ denotes the indicated altitude) denote the position of the aircraft (at point $s^{(\alpha)}$ on its path) in the physical space (of dimension 3 or, in some simplified settings, 2).

Regard the physical airspace as a Euclidean space [28, 5.1-1] with a dot product [28, 5.2-6] of vectors $a, b$ denoted by $a \cdot b$. With this dot product, the norm [28] of a vector $a$ is defined to be

$$||a|| = \sqrt{a \cdot a}.$$  

The vector

$$d^{(\alpha)} \left( s^{(\alpha)} \right) = \left. \frac{d\xi}{ds} \right|_{\xi = s^{(\alpha)}}$$

has norm 1, is tangent to the path of aircraft $\alpha$ at the point $s^{(\alpha)}$, and points in the direction in which $s^{(\alpha)}$ increases (hence, by the convention chosen in our models, in the instantaneous direction of the aircraft’s motion).

If the wind velocity is known at every position $x$ in the physical airspace (at least, at every point on aircraft’s path) to be $w(x)$, and if $s^{(\alpha)}(t)$ is the motion of the aircraft along the path, then the (scalar) ground speed of the aircraft at time $t$ is

$$v_{\text{ground}}^{(\alpha)}(t) = \left. \frac{ds^{(\alpha)}(\tau)}{d\tau} \right|_{\tau = t},$$

and the (scalar) airspeed of the aircraft is obtained by adjusting the ground speed by the parallel (to it) component of the wind:

$$v_{\text{air}}^{(\alpha)}(t) = v_{\text{ground}}^{(\alpha)}(t) - w\left(x(s^{(\alpha)}(t))\right) \cdot d^{(\alpha)} \left( s^{(\alpha)}(t) \right). \quad (5)$$

Physical constraints of the aircraft dynamics and control of particular interest to PATO Scheduling include a number of parameters, described in the next several sections.

\textbf{IVA.1. Feasible Speed Range}

The feasible speed range depends on the aircraft configuration (e.g., “clean/flaps-retracted”, “landing/flaps-deployed”, or “one engine inoperative”), and must consider winds aloft (i.e., the winds exerting forces on the flying aircraft) when determining the minimal and maximal speeds at which aircraft progress along route segments.

The feasible speed range for an aircraft in a given flight configuration is specified as a minimal and a maximal airspeed, which we denote, respectively, $V^{\text{min; } \alpha}$ and $V^{\text{max; } \alpha}$. From the minimal and maximal airspeed and a knowledge of winds aloft, one can determine the minimal and maximal ground speeds (from [Eqn. (5)]) for an aircraft at a specified distance along the path, $V_{\text{ground}}^{\text{min; } \alpha}$ and $V_{\text{ground}}^{\text{max; } \alpha}$, and impose the constraint in the form

$$V_{\text{ground}}^{\text{min; } \alpha} \leq \frac{ds^{(\alpha)}}{dt} \leq V_{\text{ground}}^{\text{max; } \alpha}, \quad \alpha \in A. \quad (6)$$

\textbf{IVA.2. Feasible Range of Longitudinal Acceleration and Deceleration}

While the longitudinal acceleration (i.e., acceleration in the direction of the route of flight) is often ignored (e.g., in inertia-free models), the impact of this rate on predicted separation between aircraft can become significant when large speed changes are commanded by the pilot or through the flight deck automation. Winds aloft must be considered when determining the constraints on longitudinal acceleration along a route segment.

The permissible acceleration range for an aircraft $\alpha$ in a given flight configuration is specified by a minimum $A_{\text{g}}^{\text{min; } \alpha}$ and a maximum $A_{\text{g}}^{\text{max; } \alpha}$ allowed. From a knowledge of winds aloft, one can determine the corresponding ground-referenced minimal and maximal accelerations, $A_{\text{g}}^{\text{min; } \alpha}$ and $A_{\text{g}}^{\text{max; } \alpha}$, and impose the constraint in the form:

$$A_{\text{g}}^{\text{min; } \alpha} \leq \frac{d^2 s^{(\alpha)}}{dt^2} \leq A_{\text{g}}^{\text{max; } \alpha}, \quad \alpha \in A.$$  

The values $A_{\text{g}}^{\text{min; } \alpha}, A_{\text{g}}^{\text{max; } \alpha}$ depend on the aircraft configuration and on its rate of climb or descent.

Feasible values of longitudinal acceleration vary from one aircraft type to another, but in most of the published research have for simplicity been assumed identical (e.g., 1 knot IAS per second deceleration for all descending aircraft).
IVA.3. Feasible Range of Vertical Speed

The range of feasible vertical speeds for an aircraft follows indirectly from the specification of maximal rates of climb and descent. This range depends on a number of factors, including: aircraft weight, altitude, air temperature, relative humidity, and aircraft configuration (e.g., flap setting).

The mathematical formulation of this constraint is analogous to that in section IVA.1.

IVA.4. Limitations imposed by the amount of fuel

The amount of fuel on board an aircraft determines the range and endurance of an aircraft and may, in rare cases, limit the amount of time an aircraft is able to remain in flight, the number of maneuvers that an aircraft may be able to execute within a given schedule, or even the ability of an aircraft to land safely. An aircraft that has declared a low-fuel emergency is given priority in the schedule over all other aircraft. Because the remaining amount of fuel on board an aircraft is unknown to ATC in current ATO, this constraint is typically ignored in the modeling, but operationally a mechanism must be provided to accommodate aircraft that have declared emergency within a schedule (e.g., see Section IV.D.2). The authors believe that this constraint (if fuel data become available to ATC) should be modeled in PATO to avoid managing low-fuel aircraft as exceptions; giving schedule priority to aircraft low on fuel may increase the safety of PATO and reduce pilot and controller workload.

If $B^{(\alpha)}(s^{(\alpha)}, \frac{ds^{(\alpha)}}{dt}, \ldots)$ is the instantaneous rate of fuel consumption and depends on finitely many of the values

$$\frac{d^k s^{(\alpha)}}{dt^k}, \quad k = 0, 1, 2, \ldots,$$

then an upper bound $F^{(\text{max};\alpha)}$ on the total amount of fuel to be consumed by aircraft $\alpha$ over a time interval $[t_1, t_2]$ can be imposed in the form

$$\int_{t_1}^{t_2} B^{(\alpha)}(s^{(\alpha)}(t), \frac{ds^{(\alpha)}}{dt}(t), \ldots) \, dt \leq F^{(\text{max};\alpha)}$$

IVA.5. Control Latency

Control latency is the amount of time needed (and constrained by the physics of flight) to execute a maneuver instructed by the ATC authority (hence dictated by policy). The control latency includes the time required for communication of the instruction, by voice or data, as well as the time necessary to input the commanded maneuver to the flight controls to initiate the maneuver (i.e., to effect an aircraft control surface actuation, but not to complete the commanded maneuver). Control latency is a function of the communication mechanism (physical) and protocol (policy), ATO practice (operational), flight crew procedures (procedural), flight crew workload, and flight crew proficiency. While an accurate model of control latency would be inevitably complex, it is usually either neglected or approximated as a static value. To ensure reliability of PATO solutions in the presence of varying control latency, new models of communication between ATC and flight crews and of command execution timing variance will be needed.

A model of control latency with a constant time lag of $\tau^{(\alpha)}$ specific to aircraft $\alpha$ can be included in the control dynamical law (2) by writing

$$\frac{ds^{(\alpha)}}{dt}(t + \tau^{(\alpha)}) = v^{(\alpha)}_{\text{ground}}(t), \quad \alpha \in \mathcal{A},$$

and in the control dynamical law (3) by writing

$$\frac{d^2 s^{(\alpha)}}{dt^2}(t + \tau^{(\alpha)}) = a^{(\alpha)}_{\text{ground}}(t), \quad \alpha \in \mathcal{A}$$

IVA. Policy Constraints

IVA.1. Airspace Boundaries

As discussed in Section II.B, airspace is partitioned into bounded 3-D regions (henceforth called sectors), each managed by an air traffic controller. Airspace boundaries are defined by the FAA and updated on a 56-day cycle. However, preferred ATO practice often combines sectors when the requested rate of traffic acceptance is low, allowing a single controller to provide services for a larger airspace volume.
Geometrically, a sector is a bounded, connected region in 3-D which is a union of cylinders (not necessarily circular), each cylinder satisfying the following conditions:

1. The cylinder’s base lies in the “horizontal plane” that approximates the portion of the earth’s surface “directly below” the aircraft.
2. The cylinder’s base is bounded by a closed curve which is simple (devoid of self-intersections), continuous, and piecewise differentiable.
3. The generators of the cylinder are strictly vertical (i.e., orthogonal to the horizontal plane).
4. The height of the cylinder is finite.

By intersecting the lateral surface of the cylinder with a plane parallel to the abovementioned horizontal plane, one obtains a closed curve called the horizontal boundary of the sector for the altitude (height) of the intersecting plane.

Unless transitioning from one sector to another, an aircraft must maintain a specified horizontal distance from the sector’s horizontal boundary at the aircraft’s flight altitude. The ceiling and floor of the sector restrict the aircraft’s altitude, but unlike with horizontal boundaries, no cushion space is required: the aircraft is allowed to be at the floor or ceiling of the sector.

IV.B.2. Maximal allowed altitude

An aircraft’s maximal allowed density altitude (operational ceiling) defines the maximal allowed pressure altitude or FL. The ceiling is dictated by the requirement for a minimal climb rate, and thus depends on the same factors as does feasible vertical speed (section IV.A.3). Since aircraft weight is one of those factors and changes during the flight, so does an aircraft’s maximal allowed altitude.

IV.B.3. Separation Minima

In the U.S., FAA Order JO 7110.65U [16] dictates the required separation minima for aircraft operations under ATC separation provision, although these minima may change in the future (see footnote II.C). This paper includes a partial treatment of the separation minima, tailored for the purpose of scheduling aircraft operations in dense airspace surrounding major airports. (For a full treatment of aircraft separation requirements, the reader is referred to [16].)

Two types of separation arise in the context of scheduling for PATO: Radar Separation and Visual Separation.

**Radar Separation Minima** Reference [16, Chapter 5] specifies the vertical and horizontal separation minima that ATC must provide between aircraft under radar separation procedures. While the minima include many exceptions and context-dependent variations, only the most commonly applicable minima are described herein. En route aircraft operations (i.e., those conducted in ARTCC airspace) typically require separation by at least 5 nmi horizontally or 1,000 feet vertically; i.e., a vertical separation of 1,000 feet or more absolves the aircraft pair of any requirement for horizontal separation (see inequality (7)). Terminal operations (i.e., those conducted in TRACON or ATCT airspace) are typically required a horizontal separation of at least 3 nmi horizontally or a vertical separation of 1,000 ft. The most common exception to the standard terminal and en route separation minima is the application of wake vortex separation minima intended to provide safe operation for in-trail aircraft (i.e., for those aircraft operating directly behind another), typically on the same arrival, departure or approach procedure. If an aircraft is in-trail*, the required horizontal separation minima depend on the weight category of each aircraft in the pair, as specified in Table 3.

Other factors affecting radar separation minima specific to arrival, departure, and approach include: the runway layout of the airport, the approach and departure procedures associated with the runways, and the staffing and equipment of the ATC facility. Finally, separation minima at the airport are dictated by the regulations stated in Reference [16, Sections 9,10, Chapter 3]. These minima are defined spatially in some cases and temporally in others.

Separation minima apply to each aircraft, at all times along their paths of flight, and on the ground while conducting arrivals and departures (i.e., while on the runway). While the separation requirements are continuous in time, the separation constraint may be met by imposing separation minima at discrete locations and placing additional restrictions on the movement of aircraft along their routes (between the locations where separation minima are imposed). A discussion of research efforts to model separation using such discretization is included in Section V.C, below.

---

*P.i.e., directly behind another, or directly behind and below by at most 1,000 feet; see [16].
### Table 3. Separation minima in nautical miles for in-trail aircraft with wake turbulence application (legend: * may be reduced to 2.5 nmi in some conditions; † reduced by 1 nmi away from final approach course)

<table>
<thead>
<tr>
<th>Lead Aircraft Weight Class</th>
<th>Trail Aircraft Weight Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A380</td>
</tr>
<tr>
<td>A380</td>
<td>3</td>
</tr>
<tr>
<td>Heavy</td>
<td>3</td>
</tr>
<tr>
<td>B757</td>
<td>3</td>
</tr>
<tr>
<td>Large</td>
<td>3*</td>
</tr>
<tr>
<td>Small</td>
<td>3*</td>
</tr>
</tbody>
</table>

**Visual Approach**

A visual approach [16, Chapter 7, Section 4] is an ATC authorization for an aircraft on an IFR flight plan to proceed visually to the airport of intended landing. A pilot who has accepted this authorization bears the sole responsibility for separation and is no longer bound by explicit separation minima with preceding aircraft on approach. The aircraft on visual approach, however, will not be cleared to land before the landing aircraft directly in front (the leading aircraft) has exited the runway. This restriction dictates a practical lower bound on the separation not prescribed by any formal policy or procedure. Thus, rather than place an additional constraint on the PATO scheduling problem, visual approaches allow that the radar separation minima constraint (Section IV.B.3, Radar Separation Minima) be removed for a part or all of the final 10 n mi of the flight. This removal, regularly practiced in VMC, effectively allows an increase in runway capacity: pilots using visual separation follow the preceding aircraft more closely than is permitted by the radar separation minima. It follows that problem formulations intended for implementation in an operational system need to consider visual approaches to avoid wasting VMC runway capacity. Treatment of visual approaches as a constraint in the problem formulation requires the following determinations:

- **The point at which an aircraft conducting an instrument approach can be cleared for a visual approach** depends on a number of factors, including: visibility, the ability of the PIC to visually acquire the lead aircraft (and/or the landing runway), and the workload of the approach controller.

- **The minimum interval between the STAs of the leading and trailing aircraft conducting a visual approach is dictated by the following factors:** the time it takes for the lead aircraft to exit the runway upon landing⁸, the subjective assessment (by the PIC of the trailing aircraft) of a safe following distance from the lead aircraft (including consideration of wake hazard avoidance), and the landing speeds of the lead and trailing aircraft according to their respective FOMs.

The authors are unaware of any studies aimed to determine the above constraints comprehensively for the purpose of predicting inter-arrival spacing of aircraft conducting visual approaches. Such studies would significantly benefit PATO by ensuring maximum runway utilization during VMC.

Separation constraints can be formulated in the above model as follows. Since the flight path of aircraft $\alpha$ is known, the arc length coordinate $s^{(\alpha)}$ of the aircraft completely determines its physical (3-D) position whose coordinate vector, in some coordinate system global for all aircraft, can be written in the form (4).

If two aircraft $\alpha_1$ and $\alpha_2$ are required to be “horizontally separated” by distance $r$, this requirement can be written

$$ (x^{(\alpha_1)} - x^{(\alpha_2)})^2 + (y^{(\alpha_1)} - y^{(\alpha_2)})^2 \geq r^2 $$

or, more conveniently in what follows,

$$ \frac{1}{r^2} \left[ (x^{(\alpha_1)} - x^{(\alpha_2)})^2 + (y^{(\alpha_1)} - y^{(\alpha_2)})^2 \right] \geq 1. $$

To impose the additional condition that horizontal separation is not required if the two aircraft’s altitudes differ at least by quantity $q$, the entire constraint can be written

$$ \max \left\{ \frac{1}{r^2} \left[ (x^{(\alpha_1)} - x^{(\alpha_2)})^2 + (y^{(\alpha_1)} - y^{(\alpha_2)})^2 \right], \frac{1}{q^2} (h^{(\alpha_1)}_{ind} - h^{(\alpha_2)}_{ind})^2 \right\} \geq 1. $$

---

⁸Allowances are made for smaller aircraft to land while the lead aircraft is a prescribed minimum distance down the runway (see [16]).
IV.C. Procedural Constraints

IV.C.1. Airspeed Restrictions

Restrictions on the airspeed of flight may be directed by policy, required by procedure, or preferred in practice. Speed restrictions are common for A TO in terminal airspace. For example, a policy-directed restriction for most aircraft is that the maximal speed allowed at altitudes below 10,000 feet is 250 kts. Future operations may require minimal speeds in some airspace. In today’s A TO, a preferred practice is to keep the minimal speeds recommended in Reference [16, Section 5-7] for different aircraft types (e.g., turbojet, turboprop, or reciprocating engine aircraft). The landing speed is currently determined by the flight crew in accordance with the landing procedure dictated in the flight operations manual. Procedural restrictions on airspeed are usually captured by mathematical conditions of the following three types:

\[ v \leq V^{\text{max}}, \quad v = V^{\text{required}}, \quad v \geq V^{\text{min}}, \]

where \( v \) is the airspeed of the aircraft. The specific constraints are determined by the procedure or policy; e.g.

\[ v \leq V^{\text{max}} = 250 \text{kts at altitudes below 10,000 ft}, \]

or

\[ v = 280 \text{kts at a specified waypoint}. \]

Again, to impose the corresponding constraints on ground-referenced speeds, one must convert airspeed to ground speed.

IV.C.2. Minimum Climb Gradient

This constraint arises when obstacle clearance is a concern during the departure phase of flight, and is dictated by the SID procedure. The maximal physically feasible value of the climb gradient is affected by the same factors as is the feasible vertical speed (section IV.A.3) and, also, by winds aloft. U.S. terminal procedures specify a minimum climb gradient on departure, in feet/nmi, that must be kept until reaching a specified altitude (e.g., 560 feet per nmi to 2,000 feet).

IV.C.3. The Prescribed Route of Flight

Aircraft in PATO navigate a route of flight as specified in the aircraft’s flight plan (typically including portions of the route of flight dictated by SID, STAR and IAP procedures). As pointed out in Section II.B, deviations from the prescribed route of flight can be at the instruction of ATC or in response to an emergency condition (declared according to criteria defined by preferred practice). It is assumed in the PATO formulation that any such deviation will result in an amended flight plan that results in a new route of flight (see Section II.B).

Mathematically, this requirement consists in parameterizing the flight path of aircraft \( \alpha \) by arc length \( s^{(\alpha)} \). A motion of the aircraft along the path is then a time parameterization \( s^{(\alpha)}(t) \) of the aircraft’s position on the path. In the notation of (4), the instantaneous ground-referenced velocity of the aircraft at time \( t = t_0 \) is then given by

\[
\frac{d \mathbf{x}^{(\alpha)}(s^{(\alpha)}(t))}{dt} = \left\{ \frac{d \mathbf{x}^{(\alpha)}(s^{(\alpha)})}{ds^{(\alpha)}} \right\}_{s^{(\alpha)}(t_0)} \left( \frac{ds^{(\alpha)}(t)}{dt} \right)_{t=t_0}
\]

Since \( s^{(\alpha)} \) is the arc length coordinate, the vector in \{ \}’s is the unit vector tangential to the path of flight and points in the instantaneous direction of the flight.

IV.C.4. Altitude Restrictions

Altitude restrictions are commonly imposed by terminal procedures (e.g., STARs and SIDs) and are typically associated with crossing a waypoint. Altitude restrictions may also be imposed as a preferred ATO practice, generally governed by Letters of Agreement (LOAs) between different ATC facilities and between sectors within an ATC facility. The types of altitude restrictions that may be imposed by a procedure include:

1. Cross a specified waypoint at a specified altitude.
2. Cross a specified waypoint at or above a specified altitude.
3. Cross a specified waypoint at or below a specified altitude.

4. Cross a specified waypoint with an altitude in a specified range (a simultaneous imposition of restriction types 2 and 3).

In the mathematical formulation, this constraint is imposed by specifying a permissible collection of routes of flight for a given aircraft.

IV.D. Operational Constraints

By operational constraints we mean constraints that cannot be categorically classified as arising directly from the physics of flight, or from official policy, or from procedural requirements. Dictated primarily by the pragmatic side of operations, the constraints described in this section may nevertheless also involve one or more of these aspects, as well as other considerations of practicality and feasibility not directly implied by physics or explicitly formulated in legal documents.

IV.D.1. Availability of a runway for landing or departure

A runway may not be all the time available for both landings and departures. In a given time period, some runways may be reserved for only landings, others for only departures, and still others open for both. This constraint arises in a number of situations, including: when a change of landing or departure runways assigned to an aircraft is considered, when the specification for one or more runways (whether it is open for departures, landings, both, or neither) changes unexpectedly, or when the expected landing or departure time for an aircraft precludes the use of the runway assigned for that procedure in the flight plan (e.g., if excessive delay, unforeseen at initial flight planning, is incurred). The available runways are dictated by preferred practice and sometimes by policy. For example, noise impact on surrounding communities may lead to a policy that forbids the use of some runways during night time hours.

In the framework of Section III, this constraint, if imposed, restricts the choice of a path (Section III.E) that can be assigned to a given aircraft.

IV.D.2. Traffic Flow Management Restrictions

TFM restricts aircraft flows in order to manage requested rates of traffic acceptance when capacity is limited. The many forms of TFM initiatives are beyond the scope of this paper. We confine ourselves to listing the few TFM methods that are commonly used in today’s system, affect current scheduling terminal operations, and are likely to affect PATO in the future.

REQUIRED TIME OF ARRIVAL. In PATO, an aircraft may be subject to a Required Time of Arrival (RTA), imposed to coordinate aircraft movements or to promote efficient operations. RTAs may be specified at any waypoint along an aircraft’s flight route, and may in PATO be specified for more than one such point. The PIC enters an assigned RTA into the aircraft FMS. Future arrival procedures may include RTAs to assist the air traffic authority in spacing aircraft efficiently. One potential use of an RTA is to impose it at a point at which the aircraft is to start a continuous descent.

If an aircraft \( \alpha \) is assigned a path with arc length \( s^{(\alpha)} \) and is required to arrive at destination \( s^{(\alpha)}_{\text{DEST}} \) on that path at time \( t^{(\alpha)}_{\text{RTA}} \), this constraint can be written

\[ s^{(\alpha)} \left( t^{(\alpha)}_{\text{RTA}} \right) = s^{(\alpha)}_{\text{DEST}}. \]

If the RTA is the same for all aircraft, the latter equation is a boundary (more specifically, end-time) constraint. Otherwise (see, for example, [13]), it is an intermediate constraint.

MILES-IN-TRAIL RESTRICTIONS. Miles-in-trail (i.e., the distance between an aircraft and the one directly in front) is sometimes subject to an increased lower bound, which is above the required separation minima IV.B.3. The intent is to adjust artificially the sparsity of traffic to the traffic-carrying capacity of an airspace region, such as a sector or a TRACON, or to allow for insertion of additional aircraft into an established flow. (An electrical circuit analogy is a voltage source in a loop with a resistance. The voltage potential corresponds to the requested rate of traffic acceptance. An increase in the resistance—imposition of larger minimal separation—decreases the current.) For example, if weather conditions at Chicago’s O’Hare airport (ORD) significantly reduce the airport’s arrival capacity, miles-in-trail may be restricted for all aircraft bound for ORD well in excess of the separation minima. Under these circumstances, aircraft departing the New York area and bound for ORD may be subject to a restriction of 30 (nautical) miles-in-trail.
The restriction would be imposed on the NY-ORD flow of aircraft and might apply as well to aircraft pairs that are not successive along a route segment. More common to terminal scheduled operations are miles-in-trail restrictions on those runways used for both landings and departures. Excess separation between arrivals may be warranted to interleave departures with landings on the same runway.

This constraint is a special case of that described at the end of section IV.B.3.

**AIRCRAFT ACCEPTANCE RATE** An aircraft acceptance rate puts a limit on the number of aircraft per hour that an airport or air traffic facility will accept into their airspace or (in the case of an airport) will clear for landing or departure on the airport’s runways. *Airport Arrival Rate* (AAR) and *Airport Departure Rate* (ADR) are typically enforced on a smaller time interval (e.g., 15 minutes) to ensure short-term requested rate of traffic acceptance does not significantly exceed the AAR or ADR. AAR and ADR are limited either by available runway, or by airspace capacity, or by both, according to the required separation minima (section IV.B.3), and are generally further restricted by considerations of controller workload (section IV.D.3).

The form in which this constraint would enter a mathematical model depends on the boundary conditions imposed. If, for example, each aircraft is subject to a required time of arrival at the destination, those of the arrival times at the same destination should be checked for the desired sparsity. If, on the other hand, the arrival times $t_{RTA}^{(\alpha)}$ of the aircraft are unknown, while the destinations $s_{DEST}^{(\alpha)}$ are known, then the constraint can be enforced as follows. From among the aircraft in $A$, pick all those, denoted here by $\alpha_1, \alpha_2, \ldots, \alpha_N$, headed for the same destination. One way to guarantee a lower bound on the aircraft acceptance rate at that destination is to restrict the inter-arrival times to be at least a suitable number $\Delta t$ of minutes apart:

$$|t_{RTA}^{(\alpha_{k_1})} - t_{RTA}^{(\alpha_{k_2})}| \geq \Delta t, \quad 1 \leq k_1 < k_2 \leq N$$

Computationally, however, the latter condition may be prohibitive due to its combinatorial nature.

**BLOCKED SLOT** A blocked slot is a period of time reserved—in a schedule of operations for a runway—for use other than for scheduled ATO. Blocked slots provide flexibility to controllers and air traffic managers, allowing for accommodation of unforeseen events such as declared emergencies and minor runway maintenance (e.g., clearing debris or replacing a light) with minimal interruption to scheduled operations.

If $[T_1, T_2]$ is a time period corresponding to the blocked slot, then an aircraft flying or taxiing to reach the blocked location at an unknown arrival time $t_{RTA}$ can be subject to the constraint

$$t_{RTA} \notin [T_1, T_2].$$

**IV.D.3. Controller Workload**

Air traffic controller workload can affect air traffic operations in the following ways: 1) by imposing cautionary measures on air traffic to limit congestion, and 2) by reducing the controllers’ ability to provide expeditious service when they are consumed by tactical aircraft movements for safe separation. Hence, controller workload must be considered when developing new procedures or automation systems. While such workload is not easily measured, the following constraints have been claimed to account, in part, for the impact of controller workload in today’s ATC system.

**PRECEDENCE** A precedence constraint, also known as the no-passing constraint, forbids that one aircraft overtake another on the same route segment. While controllers, in principle, retain the authority to issue an open clearance to an aircraft to overtake another (slower) aircraft, controllers tend to refrain from such maneuvers. In practice, such maneuvers are usually reserved for aircraft on route segments that differ only in altitude, since such aircraft are typically of similar performance capabilities and, if in trail, would require a long time to conduct an overtake.

Such passing is prevented by enforcing, for the two aircraft flying the same route, the separation constraints discussed in section IV.B.3.
**Constrained Position Shift** The Constrained Position Shift (CPS) (see Section V.A) is another constraint aimed to address controller workload by limiting the allowed deviation from a first-come-first-served (FCFS) sequence of operations (i.e., runway arrivals or waypoint crossings). Prior research [34] demonstrated that most of the sequence optimization benefit that can be achieved with unconstrained position shifting, can also be achieved with position shifts constrained to 2 or 3 positions from the FCFS sequence (the two heuristics abbreviated CPS-2 and CPS-3, respectively). This heuristic sequence optimization is achieved by grouping aircraft of the same weight class to avoid inefficient sequences that would otherwise result from enforcing the wake vortex separation minima IV.B.3 between aircraft in a FCFS sequence. CPS is also used to attain perceived fairness in a schedule (with large deviations from FCFS perceived as unfair) by limiting the penalty and advantage that accrues to each aircraft on deviation from FCFS.

From among the aircraft in \( A \), pick all those, denoted here by
\[
\alpha_1, \alpha_2, \ldots, \alpha_N,
\]
headed for the same runway, and indexed in the FCFS sequence of arrival (i.e., first arrives \( \alpha_1 \), then \( \alpha_2 \), etc.). The permissible permutations \( \sigma \) of the indices \( \{1, 2, \ldots, N\} \) are specified by the constraint
\[
|\sigma(k) - k| \leq 2, \quad k = 1, \ldots, N
\]
for CPS-2, and by the constraint
\[
|\sigma(k) - k| \leq 3, \quad k = 1, \ldots, N
\]
for CPS-3.

**Maximum Aircraft Count** Current ATO manage the workload of controllers in charge of en route sectors by placing an upper bound on the number of aircraft predicted to occupy a sector during a specified timespan [35]. The Monitor Alert Parameter (MAP) value for a sector prescribes the upper bound on predicted aircraft count in a sector and is monitored by the Traffic Management Unit (TMU) more than one hour in advance of predicted aircraft entry into a sector. If the predicted aircraft count exceeds the MAP value, the TMU receives an alert, and typically places Miles-in-Trail restrictions (see Section IV.D.2) on aircraft routed through the sector in question. MAP values depend on average flight occupancy time within a sector and can be adjusted based on input from the sector controller and on a variety of factors (e.g., high winds or specific operations such as refueling carried on in the sector).

### IV.D.1. Computational Lag

Flight crew workload and passenger comfort can affect the economic performance of an airline company, hence will require attention in scheduling PATO. Complex procedures and highly repetitive tasks can increase workload for the flight crew and for the air traffic controller. One way to avoid this increase is to limit repetitive control actions by restricting the total permitted number of commanded maneuvers. Passenger discomfort associated with vertical, lateral, and longitudinal aircraft accelerations can also be mitigated by limiting the number and severity (assessed in terms of passenger motion sickness) of aircraft maneuvers. An upper bound on the number of maneuvers allowed is sometimes used for limiting flight crew workload and preserving passenger comfort. The researcher will do well to note that most of the simple formulations of this constraint (including the following) ignore the fact that different maneuvers have different impacts on the flight crew and ATC. For example, an ATC clearance instructing an aircraft to conduct a specified arrival procedure may be simple in cruise, but involve considerable flight crew workload in descent. Enhanced formulations may include weighting of various maneuvers according to their impact on those tasked with implementing and executing them (i.e., on the ATC and flight crew).

A mathematical formulation of such constraints requires a precise definition of the concepts of "maneuver" and "maneuver severity." In current ATO, however, these terms have not acquired an interpretation sufficiently precise. To indicate an example of a mathematical condition partly successful at capturing this constraint in its "commonly accepted" intuitive sense, we resort to the notation of (4) and state that, over a time period \([t_1, t_2]\), the total "severity density" of maneuvers by aircraft \( \alpha \) can be restricted in the form
\[
\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left( \frac{d^2 x^{(\alpha)}}{dt^2} \right)^2 \, dt \leq (\text{a suitable constant}),
\]
while the severity of each individual maneuver over the same time period can be restricted in the form
\[
\max_{t_1 \leq t \leq t_2} \left( \frac{d^2 x^{(\alpha)}}{dt^2} \right)^2 \leq (\text{a suitable constant}).
\]
IV.D.5. Maximal Permitted Aircraft Bank Angle

The maximal aircraft bank angle constraint arises when route modification is considered. Such modification requires turns. The maximal bank angle determines the minimum turn radius for a given airspeed. Aircraft bank angle is commonly limited by procedures in the aircraft’s Flight Operations Manual [36]. This constraint needs no explicit consideration when aircraft are navigating RNP procedures, since aircraft bank angle limits are built into the design of these procedures.

In cases when this constraint must be included in the model formulation, it can be captured in a number of ways. One way is an inequality that bounds the speed from above by a quantity that depends on the turn radius \( r \) and on the maximal bank angle \( \phi_{\text{max}} \):

\[
\frac{ds(\alpha)}{dt} \leq \sqrt{r \ g \ \tan(\phi_{\text{max}})}
\]

V. Review of Prior Research in ATM

Prior to 1970, ATO were concerned mainly with separation of aircraft. A series of mid-air collisions necessitated ATO modernization. Efforts to meet this necessity included widespread radar surveillance, automated flight tracking, and display of flight information furnished for each aircraft tracked on controller displays [37]. Before radar surveillance came into widespread use, the distance required to separate aircraft safely increased as aircraft performance (in particular, maximal speed) improved (i.e., with the introduction of turbojet aircraft operating at higher altitudes). This practice led to separation minima of 100 nmi in some cases and posed a serious threat to the capacity of the route system. As a result, models of and research into ATO during this time focused on sector capacity and airport planning (one research direction being evaluation of the need for new airports and runways) [38]. As the public demand for air travel increased, leading up to, and accelerated by, deregulation of the U.S. airline industry in 1978, research focus shifted to managing air traffic flow at and near congested airports. Two threads of ATO research dominate the literature during this period: air traffic flow management (ATFM) and schedule optimization. ATFM researchers aimed to develop techniques to prevent clustering of flights and to balance flow to ATO resources (both terminal and en route) with predicted available capacity. Research in ATFM has been mostly independent of that in schedule optimization. ATFM techniques address the balance of flow with capacity without regard for accurate modeling of the movement of individual aircraft. By contrast, schedule optimization requires accurate information about individual flights for compliance with (sequence-dependent) separation requirements in terminal airspace (see Section IV.B.3).

The evolution of technologies of communication, computation, and navigation has brought new systems and capabilities: Flight Management System (FMS), Area Navigation (RNAV), and eventually the Global Position System (GPS). These new systems enabled higher precision in the management of flights and showed promise to improve the efficiency and safety of ATO. ATM research saw a new trend, that of formulating and solving an optimization problem whose solution not only observes the required separation, but also minimizes an objective function (see, for example, the work of Dear [7] in 1976). A broad class of such problems is aimed at scheduling air traffic, either to arrive or to depart from a given airspace resource (e.g., runway or arrival/departure fix).

Deregulation further increased the role of this new trend in ATO schedule optimization as airlines responded to the increasingly competitive marketplace. Before the deregulation, airlines operated most flights directly between city pairs as regulated by the Civil Aeronautics Board (CAB). Following deregulation, airlines switched to the so-called hub-and-spoke operations, which have most flights go through a few major airports (hub airports) with connecting flights to many smaller airports. This new airline practice created spikes in arrival and departure flow at hub airports as large numbers of aircraft arrived or departed in short periods of time for connecting flights. Hub-and-spoke operations were dominant through most of the 1980’s and into the 1990’s, and remain widely used today, albeit with a re-emergence of point-to-point flights as well [37]. Similar deregulation of the European market occurred during the mid-1990’s, but with one significant difference. Airport capacity is more tightly controlled in Europe by allocating landing slots to airlines so that the requested rate of traffic acceptance rarely exceeds the airport capacity. In the U.S., airline schedules are generally based on peak airport capacity, which leads to periods when the requested rate of traffic acceptance significantly exceeds capacity. The less restrictive U.S. approach may allow for more scheduled flights to and from an airport, albeit at the expense of robustness (e.g., with increased delays and flight cancellations when adverse weather reduces airport capacity; see section I.C).

Future ATO are expected to be characterized, as are current ATO, by a small number of large airports in the NAS operating at or near capacity for significant contiguous periods of time [39]. Over the next decades, increased air traffic flows are expected to add to the environmental concerns (emissions and noise impact) and to fuel costs. This
leads to the needs both for systems that address immediate and near-term impacts at currently congested airports and for research into longer-term solutions to projected capacity shortfalls and environmental impacts [1]. This dichotomy of research needs for ATO, near-term congestion relief vs. long-term solutions, has existed since the 1960’s [37], if not earlier.

The discussion of literature throughout the remainder of this section is focused specifically on tactical scheduling of PATO. The topic of this discussion is the division of the research community--into two camps, the “theoretical” and the “operational”--that has developed over the past decades. The exposition is limited here to the emergence of those trends that originated during that period and were novel in including PATO constraints in their formulation of a tactical ATO scheduling problem. The more recent efforts following these trends are not reviewed; the reader is referred to [40] for a thorough survey of those.

Most research efforts in ATM scheduling can be intuitively placed somewhere on the continuous spectrum between two approaches to research, which we will call here operational and theoretical (described below), usually near one end of the spectrum. The operational approach is one that tends to assume practices current or anticipated short-term, and to impose first and foremost the requirement that a result must ultimately become implemented “in the field” to address pressing needs. The roots of this approach appear to lie mainly in the course of development taken over the last 80 years by technology research, airline industry, and their interaction with government regulations. The theoretical approach aims not only to analyze and optimize current operational practices, but also to suggest models for future (including long-term) operations, envisioned on the basis of the observed and forecast growth of the number of air traffic operations. On the other hand, the theoretical approach is usually not suited for immediate field implementation because it does not address operational constraints to a sufficient extent. These approximate characterizations of the two methods can be found in Table 4.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Relative priority ascribed to: improving current or short-term ATO, developing new models, also for future ATO, producing a tool ready to be deployed “in the field”</th>
</tr>
</thead>
<tbody>
<tr>
<td>operational</td>
<td>high, low, high</td>
</tr>
<tr>
<td>theoretical</td>
<td>medium, high, low</td>
</tr>
</tbody>
</table>

Table 4. A rough comparative characterization of the operational and theoretical approaches to research in ATM

As the research field of ATM scheduling evolved, the divide that always existed between the two approaches further widened. The requirements for a scheduling algorithm to support PATO will likely include the merits of both the operational and theoretical approaches to scheduling. We enumerate the key merits and shortcomings of each in Table 5.

Another shortcoming, currently present in both the operational and theoretical settings, is the absence of a universal terminology or a mathematical framework for modeling air traffic. This shortcoming hinders communication between researchers. While the terms “throughput” and “demand”, for example, have a precisely defined meaning in dynamic network flows [8], their usage in aerospace literature frequently appears without reference to a framework that gives them precise meaning (see Section II.C.2).

V.A. Research trend 1: Optimization of various dynamic network flow formulations

Since 1970, many authors have formulated the air traffic scheduling problem as one on a dynamic network flow and have devised algorithms to solve it optimally for some objective function, subject to a set of operational constraints. Among the constraints modeled in the literature are:

- Precedence (section IV.D.3): Relative order between aircraft is maintained as aircraft traverse the same route segment.
- Runway separation (the special case of that described section IV.B.3 that is imposed only at the point of aircraft crossing the runway threshold): A lower bound on the separation between successive operations on a runway (or on interdependent runways) is imposed. The required separation between successive arrivals and successive departures is sequence-dependent [16, Chapter 5], but is not always so modeled.
### Operational Approach to Scheduling

**Merits**

*High-fidelity operational constraint modeling* realistically includes aircraft performance limitations, procedural models, and TFM initiatives.

*Consideration of scheduling constraints and controller preferences* allows the input from air traffic personnel to define scheduling constraints and preferences that may otherwise be at odds with the result of the scheduling algorithm.

### Theoretical Approach to Scheduling

**Theoretical foundation** provides rigorous formulations based on proven methods, which can potentially serve as a basis for safety assessment methodologies.

### Shortcomings

**Little regard for optimality** precludes substantial savings of resources. To date, development of operational systems has focused on helping the air traffic personnel perform effectively. Since controllers have found it difficult to deviate from FCFS, operational schedulers are currently focused on providing a workable schedule, which is necessarily similar to the (suboptimal) FCFS sequence.

**Lack of theoretical foundation** can substantially hinder systematic analysis and validation of operational systems for conceptual integrity, reliability, safety, and optimality. Such analysis is required, for example, for safety-critical automation.

### Table 5. A comparison table between the operational and theoretical approaches to air traffic scheduling

<table>
<thead>
<tr>
<th>Operational Approach to Scheduling</th>
<th>Theoretical Approach to Scheduling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Merits</strong></td>
<td></td>
</tr>
<tr>
<td><em>High-fidelity operational constraint modeling</em></td>
<td><em>Theoretical foundation</em> provides rigorous formulations based on proven methods, which can potentially serve as a basis for safety assessment methodologies.</td>
</tr>
<tr>
<td><em>Consideration of scheduling constraints and controller preferences</em></td>
<td></td>
</tr>
<tr>
<td><strong>Shortcomings</strong></td>
<td></td>
</tr>
<tr>
<td><em>Little regard for optimality</em></td>
<td><em>Limited fidelity in physical and operational constraints</em> results from the challenge of capturing key aspects of ATO realistically and computing solutions suitable for real-time operation. Even with aircraft separation being solely the controller’s responsibility, assessment of feasibility of a schedule solution will become critical as the requested rate of traffic acceptance increases and control flexibility decreases.</td>
</tr>
<tr>
<td><em>Lack of theoretical foundation</em></td>
<td></td>
</tr>
</tbody>
</table>

- **Arrival window:** Scheduled time of arrival and departure is restricted to an acceptable or feasible range. This range is determined by the aircraft’s feasible speed range (section IV.A.1) and by the prescribed route of flight (section IV.C.3).

- **Constrained Position Shift (CPS, section IV.D.3):** An aircraft’s place in the landing or departing sequence is restricted to be within a pre-determined number of position shifts from that in a nominal (often First-Come-First-Serve) sequence.

CPS [7] was first proposed as a constraint in air traffic scheduling algorithms that was called upon to achieve a satisfactory compromise between solution optimality, computational tractability, and perceived fairness. Explicit enumeration of possible sequences was found computationally costly for large numbers of aircraft and for large Maximum Position Shift (MPS). (The cause of this high cost can be explained by applying an elementary fact from combinatorics to the extreme example of unconstrained position-shifting: a set of \( N \) aircraft can be arranged in \( N! \) different sequences.) Much of the air traffic scheduling research subsequent to [7] was aimed at formulations and solution techniques that improved on the computational efficiency and (with a lower priority) on the operational fidelity of Dear’s work [7].

Various air traffic scheduling problems (e.g., runway arrival scheduling, arrival metering, runway departure scheduling), with different subsets of the constraints listed in Section IV, have been formulated and solved computationally as Nonlinear Programs, Mixed Integer Linear Programs (MILP), Dynamic Programming recursions, and other problems in Combinatorial Optimization (which we classify simply as Combinatorial Optimization). We now briefly review this work.

**Nonlinear Programs and Mixed Integer (Non)Linear Programs:** Bianco et. al. [41] formulated the runway scheduling problem as a special case of the job-shop scheduling problem (a subclass of the MILP) and computed an optimal solution using a tree-based search. Beasley [42] extended the MILP formulation of [41] and the
genetic algorithm approach proposed in [43] to accommodate multiple runways and to include the constraints on runway separation, arrival window, position shift, and precedence.

**Dynamic Programming:** Psarafitis [44] formulated the arrival sequencing problem as a special case of the job-shop problem and computed solutions in polynomial time using dynamic programming recursion. He imposed the runway separation and position shift constraints. Psarafitis’s formulation was modified by Venkatakrisman et. al. [45] to include the arrival window constraint. Balakrishnan and Chandran [46] formulated the runway scheduling problem as a modified shortest path problem [9] and used a dynamic programming algorithm with computation time linear in the number of the aircraft. The formulation in [46] included the constraints on runway separation, arrival window, position shift, and precedence.

**Combinatorial Optimization:** Dear and Sherif [47] proposed a heuristic algorithm for solving the scheduling problem with the runway separation and CPS constraints. Neuman and Erzberger [34] extended the model in [47] by including the arrival window and precedence constraints in their formulation. For this new formulation, they developed a new heuristic algorithm and compared it to a modified FCFS formulation, and to two formulations which allowed Time Advance, i.e., increasing throughput by accelerating selecting aircraft to eliminate gaps resulting from fluctuations in the incoming air traffic flow. Time Advance was found particularly effective in reducing average delay with the precedence constraint imposed. Besides the latter heuristic algorithm, the work [34] also offered a branch-and-bound [23] solution technique that used a simplex-bounding algorithm. The formulation divided the scheduling problem into devising algorithms for aircraft sequence and for Scheduled Time of Arrival (STA). This approach showed high efficiency for large numbers of aircraft. The Implicit Enumeration (IE) algorithm proposed in [48] used a branch-and-bound technique to find an optimal arrival schedule with constraints on runway separation, arrival window, constrained position shift, and precedence.

**V.B. Research trend 2: Constraint-based scheduling**

Following airline deregulation, and concurrently with the development of air traffic scheduling formulations and techniques of optimization, actual air traffic demand (and hence also the requested rates of traffic acceptance) and delay grew to levels that stressed the existing ATO system. This stress necessitated near-term improvements to the procedures of sequencing and scheduling used at the time. Time-based metering was proposed [49] as a measure to address the increasing delays and congestion. Under time-based metering, controllers are given STAs and additional information to help them meet the STAs and prescribed rates of flow into congested airspace. The earliest automation tools for time-based metering were the En Route Metering (ERM) program and Arrival Sequencing Program (ASP) [50]. We now review one such tool that was developed subsequently and is currently used in operations.

The Traffic Management Advisor (TMA) was developed as a replacement for the ASP. The purpose of the TMA was to supply a sequenced list of arrivals and corresponding STAs to the arrival sector controllers in the ARTCC. The controllers were to use the list to meter aircraft within their airspace according to the TMA-prescribed delay. Development and execution of the control strategy to meet the prescribed STA for each aircraft remained the responsibility of the controller. Early research prototypes of the TMA relied on Brinton’s IE algorithm [48]. When tested in simulations, however, these prototypes were found unacceptable by the participating controller subjects, who reported that high workload levels were required to implement the advised schedules. Efforts to refine the IE algorithm continued independently of TMA development, but none brought results that would address the urgent need to supply the TMA with a scheduling algorithm needed for concept assessment and operational evaluations toward rapid deployment in the field. To meet this need, a constraint-based scheduler, called the Dynamic Planner (DP) [51], was developed. The DP used a modified FCFS heuristic algorithm that imposed constraints on runway and terminal entrance separation (section IV.B.3), arrival windows (sections IV.A.1 and IV.C.3), and precedence (section IV.D.3). In addition to satisfying these constraints, the DP adhered to prescribed airport acceptance rates (section IV.D.2) and allowed for controller-assigned sequences (a variation of RTA, Section IV.D.2) and for blocked slots (section IV.D.2). Research for the TMA has focused on an interface used by the radar controller, with acceptability to the controller being the primary requirement on the scheduler. Developed mainly to supply the controller with a satisfactory schedule, the DP gave but limited consideration to sequence efficiency. The TMA was successfully deployed into 20 ARTCCs and constitutes the basis for Time-Based Flow Management in the current U.S. air traffic system.

---

1Here delay refers to the sum of the individual delays incurred by a collection of flights.
V.C. Research trend 3: Combine scheduling with separation considerations

In most dynamic network flow formulations of air traffic scheduling problems, separation (section IV.B.3) at the runway threshold and precedence (section IV.D.3) are the only constraints that address aircraft separation. Instead of constraining pairwise distances between aircraft throughout the airspace, as is required of actual ATO, the common network flow approach is to constrain only the spacing of consecutive arrivals (departures) at the runway and to forbid overtakes on common routes. In real operations, therefore, such models cannot guarantee separation as required by the FAA. Thus, they lack self-sufficiency: a schedule they produce must still be thoroughly tested and, if possible, mathematically verified for feasibility, with proper consideration of required aircraft separation, aircraft performance limitations, procedural restrictions, and usable airspace. Among the scheduling formulations that appear in the literature, the following three have been used to address required separation between aircraft along their entire flight paths.

Graph-theoretic approach to air traffic scheduling: Adacher et al. [52] proposed a graph-theoretic formulation for the problem of sequencing aircraft. This formulation led to a linear program that imposed Boolean (binary) constraints (hence is a MILP). In addition to the constraints imposed by other methods, the formulation in [52] imposes the following ones:

- Speed: The feasible speed range (section IV.A.1) for an aircraft is included by setting upper and lower bounds on the time it takes an aircraft to traverse a route segment.

- Separation Time Interval (STI): A minimal time interval between the times of two consecutive aircraft entrances a route segment. The STI is consistent with the required separation between consecutively entering aircraft and is modeled as sequence-dependent (section IV.B.3). In real ATO, the STI depends on the speed of the aircraft (required separation is specified in spatial rather than temporal terms). The proposed formulation, by contrast, assumes an STI based on a nominal, fixed speed.

The STIs and the constraints on speed are intended to meet the operational constraint that consecutive aircraft traversing a route segment must maintain at least a minimal required separation. However, there is no explicit treatment of separation between aircraft during the traversal, and the possibility for loss of separation remains, unless further constraints are introduced.

Concurrent scheduling and deconfliction (the active Final Approach Spacing Tool): Robinson and Isaacson [53, 54] developed an algorithm that simulated and evaluated instructions issued by a skilled controller for separation, with each relative sequence considered in the construction of the schedule.

The active Final Approach Spacing Tool (aFAST) that used this algorithm relied upon a voluminous adaptation to model controller technique in sequencing and spacing arrival aircraft. In addition to the constraints on runway separation, arrival window, and precedence, the aFAST algorithm included these:

- Along-Path Separation (section IV.B.3): A required spatial separation between aircraft along their path of flight.

- Aircraft Performance (section IV.A): Aircraft movements subject to realistic models of aircraft dynamics. This constraint “inherits” the errors in the models, both dynamic and procedural, and in the weather forecasts.

- Procedural/Airspace: Aircraft states (spatial position, speed) restricted to those allowed by procedures and maintained within the airspace allotted for those procedures, even when control actions are considered. The aFAST imposed the following procedural and operational constraints: Speed Restrictions (section IV.C.1), Route of Flight (section IV.C.3), Altitude Restrictions (section IV.C.4), Permissible Runways (section IV.D.1), Airspace Boundaries (section IV.B.1), Blocked Slot (section IV.D.2), Precedence (section IV.D.3), and (implicitly) Maximum Number of Maneuvers (section IV.D.4).

An output from the aFAST scheduling algorithm is a set of trajectories that meets the prescribed schedule and all constraints. These trajectories were the basis for speed- and heading control advisories provided to the air traffic controller to meet the schedule. Thus, the aFAST scheduling algorithm provided not only a sequence of arrivals with STAs, but also a control strategy for each aircraft. Robinson’s formulation [53, 54] explicitly included aircraft separation (section IV.B.3) and synthesis of a control strategy as a result of the scheduling process, but did not achieve optimality in terms of the objective functions typical of the dynamic network flow formulation (e.g., delay, makespan, fuel burn). More importantly from the operational perspective, the formulation required a substantial and complicated adaptation. A system relying on such an adaptation may be difficult to verify for safety, a likely requirement in the future.
MILP FORMULATION FOR SECTOR TRAJECTORY PLANNING: Huang and Tomlin [55] presented a methodology for generating conflict-free trajectories for aircraft traversing a dynamic network flow model of an en route sector. They formulate the problem as a Mixed Integer Linear Program with the following constraints:

- Precedence (section IV.D.3): Two aircraft in-trail on the same path cannot trade leadership (i.e., no passing).
- Permissible Speed Range (section IV.A.1): Minimal and maximal speeds are imposed on each aircraft for each route segment traversed. While the formulation uses a static value for each speed bound, the methodology appears to allow for speed bounds for each route segment (i.e., such as would be required to model the effects of wind on the ground speed of the aircraft).
- Separation (a limited version of the treatment in Section IV.B.3): The separation minima are imposed along the entire flight path of each aircraft, but come with additional, artificial, restrictions absent from ATO. First, the separation minima are assumed symmetric (i.e., independent of order of arrival at an intersection or merge), an assumption in violation of the separation minima constraint described in Section IV.B.3 and specified in Table 3. Furthermore, the formulation requires that aircraft speed on a route segment be constant (i.e., the speed can change instantaneously as an aircraft goes from one segment to another). The treatment of the aircraft as an inertia-free system, while common in the literature, neglects the feasible acceleration range constraint (section IV.A.2).

While the computational cost of the algorithm is exponential in the number of routes in the sector, results with representative traffic samples demonstrated reasonable physical computation times for the modeled sector. The computational efficiency of the algorithm has not been evaluated in the context of operations similar to terminal PATO (i.e., for operations with many merging routes and with desired spacing near the separation minima). Further, the artificial restriction of constant aircraft speed on each arc has not been assessed for impact on optimality compared to a formulation that places no such restriction on aircraft speeds.

VI. Summary of Purpose and Recommendations for Future Research

The purpose of this paper is to define concisely and unambiguously the problem of PATO so that researchers, with expertise in other fields but new to ATM, gain access to the finer domain knowledge often implicitly assumed, but mostly unavailable, in the ATM literature. To that end, this paper offered an overview of current and future ATO with a focus on congested air traffic operations near major airports. The problem of scheduling and determining aircraft trajectories for PATO was formulated with uncommonly close, but absolutely necessary, attention paid to the separation requirement (Section IV.B.3). Section IV was provided as a reference for the many constraints arising in PATO, some of which are frequently ignored in the literature, which precludes operational application. Finally, a review (Section V) of prior work in the ATM field relevant to PATO provided the background necessary for researchers to characterize and assess the research efforts and tendencies that emerged in the recent decades. We believe that this paper provides the necessary foundation for applying to PATO the modeling techniques from optimization, operations research, and computer science in a way that meets the essential and oft-ignored operational constraints. Researchers should be aware, however, that models with adequate constraints, while necessary, are alone insufficient for operational application. Such application will require also further research into a number of topics, including: path control (Appendix I.A), uncertainty in control strategy execution (Appendix I.B), robustness of PATO (Appendix I.C), and operational reliability (Appendix I.D).

A. Possible generalizations of the problem

I.A. Path control

An assumption fundamental to the model formulated in this section is that the aircraft is constrained to move along a fixed, well-defined path (Assumption III.1 in Section III.C). In some operational situations, however, this assumption fails to hold. An aircraft may be allowed certain freedom in the choice of altitude (see, for example, Section IV.C.4), or directed by ATC to modify the originally planned path (see Section II.A). A formulation modeling the physical airspace in 2 dimensions may be unaffected by the freedom in the choice of altitude, but modifications to an aircraft’s route of flight would require reconstructing the multigraph for both 2-D and 3-D airspace model formulations.

I.B. Uncertainty in the execution of a control strategy

Even if a required control strategy (section III.G) prescribing ground speeds has been computed, the precision with which it can be executed is limited. The error in execution has in current ATO six main sources:
1. Aircraft spatial navigational point error.

2. Aircraft intent modeling error (i.e., an open clearance providing little information about the future position of the aircraft).

3. Human factors in the execution of ATC services and pilot duties.

4. Format limitations and lag in ATC-pilot communication.

5. Errors in the forecast of wind.

6. Errors in the forecast of other adverse weather effects (i.e., a storm that closes a region of the airspace to aircraft passage).

The first three sources are expected to be mitigated in PATO mostly by the mass use of RNAV and RNP ATC procedures. Introduction of data communications between the aircraft flight crew and ATC will minimize the impact of the fourth source. The ability to mitigate the last two sources strategically is limited by the stochastic and complex nature of the nonlinear models of fluid motion derived from first principles (e.g., the Navier-Stokes equations). A number of approximations and empirical studies have been carried out (see, e.g., [56, 57, 58, 59, 60]), but all compensate the lack of sufficient data and analytical solutions by assumptions about the probability densities governing the stochastic field that can describe wind accurately.

I.C. PATO Robustness

ATO are subject to various types of perturbations that cannot be predicted with certainty. Such perturbations include, but are not limited to: errors in weather forecasting, errors in control execution (Appendix I.B), malfunctions in the air traffic facility (e.g., the necessity to close a runway to clear it of debris or to replace a light). By PATO robustness we mean here the ability of the system to keep operating without radical changes to the flight schedules despite the perturbations that have occurred or are occurring.

While the aforementioned perturbations cannot, as we noted, be predicted with certainty, it is not clear how successfully they can be modeled probabilistically. Such modeling is hindered by the rarity of failures and of other events of interest (e.g., large navigational errors, ATC service disruptions, and in-flight emergencies) and by a principal absence of multiple trials of a relevant experiment (such as having a set of aircraft with prescribed initial conditions move through a route network) under the same operational conditions. Attempts to test for the presence of stable statistical properties using data that were simulated or recorded live must be made with care to understand whether the data can be interpreted as realizations of “the same” probabilistic experiment. For example, a human air traffic controller repeatedly presented with the same traffic conditions will likely respond with different decisions from time to time, depending on the controller’s physical, mental, and emotional state. The variation may become even greater from one controller to another. Other sources of variation include changes (over time that can be as short as on the order of weeks) in the following:

- ATC procedures,
- procedures documented in Flight Operations Manuals,
- Communication, Navigation, and Surveillance (CNS) capabilities, and
- available air traffic service facilities (e.g., airports, runways, taxiways, and gates).

I.D. PATO Reliability

By PATO reliability we mean here the ability of PATO to function without any of its components causing the entire system to fail. Reliability is distinct from robustness: a component or the entire system may fail even in the absence of an external perturbation. The following is a schematic description of the functional components of ATO.

Functionally, all ATO (whether current or envisioned) can be viewed as the interaction between three major conceptual components; for the moment, we call them, (a) the current state of air traffic, (b) a scheduler, and (c) a mechanism of schedule issuance (MSI). The overall intuition intended by these names is the following exchange of signals. Component (a) comes with some description of the service needed by the air traffic. (This service consists of a collection of instructions how the aircraft are to proceed safely toward the scheduled destinations.) This description is fed, as a signal, into component (b), which comes up with the required instructions, and feeds them into component (c). The latter is responsible for delivering the instructions to the parties controlling the constituent aircraft of the air
traffic. A summary of this interaction is shown diagrammatically in Figure 7. The actions of the flight crew (e.g., to update the current state of traffic by responding to the issued schedule) are not included in the diagram.

In the current ATO, the functions of components (b) and (c) are carried out mostly by the human ATC. Speed clearances and other instructions to be issued to the aircraft crews are both determined and communicated to the aircraft crews (by voice over the radio) by the ATC.

References


Figure 3. Example TRACON conflicting routes
Figure 4. TRACON airspace position: a) symbology, b) arrival feeder East position and c) departure North/East position

Figure 5. Examples of (a) closed and (b) open ATC clearances
Figure 6. A notional example with two aircraft, each assigned to a route. (A) Two aircraft (shown schematically as solid black triangles), each on its route. The route of each aircraft is shown as a curve, solid for $\alpha = 1$ and dashed for $\alpha = 2$. (B) The corresponding $s^{(1)}, s^{(2)}$ coordinate space and the coordinate vector $s(t) = (s^{(1)}(t), s^{(2)}(t))$ at time $t$ of the 2-aircraft set shown in panel (A).

Figure 7. The three major components of ATO and their interaction (arrows with captions).