Towards a Functionally-Formed
Air Traffic System-of-Systems

Sheila R. Conway
Old Dominion University
Norfolk, Virginia, USA
NASA Langley Research Center
Hampton, Virginia, USA
Sheila.R.Conway@NASA.gov

Maria C. Consiglio
NASA Langley Research Center
Hampton, Virginia, USA
Maria.C.Consiglio@NASA.gov

Abstract - *Incremental improvements to the national aviation infrastructure have not resulted in sufficient increases in capacity and flexibility to meet emerging demand. Unfortunately, revolutionary changes capable of substantial and rapid increases in capacity have proven elusive. Moreover, significant changes have been difficult to implement, and the operational consequences of such change, difficult to predict due to the system's complexity.*

Some research suggests redistributing air traffic control functions through the system, but this work has largely been dismissed out of hand, accused of being impractical. However, the case for functionally-based reorganization of form can be made from a theoretical, systems perspective.

This paper investigates Air Traffic Management functions and their intrinsic biases towards centralized/distributed operations, grounded in systems engineering and information technology theories. Application of these concepts to a small airport operations design is discussed. From this groundwork, a robust, scalable system trans-formation plan may be made in light of uncertain demand.

Keywords: Air Traffic Management (ATM), distributed vs. centralized control, functionally driven, Air Traffic Control (ATC).

1   Motivation

Changes in the demand for air transportation are inevitable, and indeed seem to be upon us.[1] The National Airspace System (NAS) improvement initiatives are largely focused on incremental improvements in today's Air Transportation System (ATS), but are unlikely to satisfy future demand. The Federal Aviation Administration (FAA) admits that their own plans to increase system capacity by 30% or more will be insufficient to meet their own projected demand growth stemming from expansion of existing airline and general aviation operations. Cost, maintenance, and integration complexity burdens continue to rise as new systems and technologies are added to the NAS, creating not a "system-of-systems" as commonly defined in the literature,[2] but rather a collection of poorly integrated legacy elements.

The uncertainty of future demand is reflected in comments such as those of Secretary of Transportation Norman Mineta, who recently called for tripling the air traffic capacity of the United States in the next 15 to 20 years. He and others are projecting a substantial impact of new transportation modes such as jet taxies (e.g. Dayjet, publicly launched April 2005) and unmanned aerial vehicles on the character and volume of future traffic. He stated, "The changes that are coming are too big, too fundamental for incremental adaptation…We need to modernize and transform our global transportation system, starting right now." [3]

2   ATS and ATM: The Status Quo

The operational ATS element that involves the safe and efficient management of flights is commonly referred to as Air Traffic Management (ATM). ATM is accomplished by the combined effort of pilots, air traffic controllers, airline dispatchers, and flow managers, all of whom adhere to procedures and regulations that limit the possibility of traffic conflicts. The air traffic control system can be thought of as a heterogeneous distributed system composed of multiple, highly interconnected subsystems that interact and share data and resources. Air Traffic Control (ATC) is also often considered to be highly centralized, as all local control of aircraft stems from a central ATC entity (itself a system of entities). It seems then that the amount of centralization is a matter of degree, and its determination highly dependent on the observer.

ATS is nominally considered to encompass commercial air carriers, general aviation operations, the passengers they serve, as well as the infrastructure within which they operate. All of these entities operate for their own purposes, optimizing or "gaming the system" to benefit their own set of criteria which may or may not overlap with those of others. Regardless of the complement of system elements one might choose to include, or the labels that are applied, there is an overriding necessity for ATM.

The FAA states that the fundamental purpose of ATM is to ensure for a safe and efficient system which means accident prevention and management of traffic. Beer's

systems science work suggests that for continued success, the system must accommodate both of these functions independently while also mediating their interaction.[4] Additionally, the balance of these activities must be continuously revisited and adjusted to meet present and expected future objectives. These functions could be captured by an enumeration of operational tasks: arbitration of resource conflicts, optimization of limited resources, ensuring security, data collection and dissemination, traffic conflict detection and resolution, and demand limitation (the last line of defense against system overload). Itself a system-of-systems, ATM provides operational control for the larger ATS.

2.1   Today's ATM Implementation

ATM is often described by its constituent actors (e.g. controllers, pilots, etc.) or entities (e.g. sector control, approach control, etc.). Though there are some functional ATM descriptions, they often focus narrowly on one or two tasks (e.g. traffic separation) while not addressing others. It is likely that any future ATM system, regardless of its form, will have to provide the same functions as the present system. In this light, the nature of these functions and the core activities in the present-day ATS that impart them warrant further inspection. The following rendering is a more elemental, functional image, ordered roughly by operational and temporal scope:

2.1.1   Regulating scheduled over-demand

Control of demand is the most extreme regulatory tool available in the ATS. This function can be invoked when the forecasted demand exceeds system capacity and threatens viability. Operations can be disallowed from the system, grounding aircraft or delaying flights. Alternately, this function can be used as an optimization tool, improving average operating efficiency for aircraft in use by delaying excess flight operations until such time they could be expected to be serviced.

The expression of this function in today's ATS is flow management, a toolset including ground holds which limit demand for a critical resource such as an approach.

2.1.2   Regulating dynamic over-demand

A less aggressive regulation function reallocates flight operations away from system bottlenecks into areas of excess capacity.

This function, akin to Beer's autonomic function, is exemplified by rerouting flights from busy enroute sectors.

2.1.3   Route Planning and hazard avoidance

Flight planning and real-time re-planning functions must also be supported. Unlike fixed road infrastructure, aviation routing is more adaptable, being responsive to city pairings, weather, prevailing winds, and other dynamic factors. Not only does ATM have to accommodate various routes, it plays an active role by creating and disseminating system information and delineating hazards.

In today's system, airline pilots and dispatchers perform the lion's share of this function, though the FAA has increased its role with the relatively recent implementation of the national route program. FAA participation entails their acceptance of user-preferred routes (assumedly for flight optimization or area hazard avoidance) implicitly agreeing to provide continued safe traffic management for the new route.

2.1.4   Sequencing

ATM operations often create situations where resources must be shared, such as multiple aircraft operating at a single airport or runway. Sequencing is the mediation of a shared resource by forming an ordered queue for resource allocation, thus eliminating ambiguity or "ties" in resource requests. Sequencing can be granted first-come, first-served, or can be determined with other schema such as system utilization maxima.

Functionally speaking, sequencing is a vital portion of many ATC instructions, and is implicit in all ATC-defined routing (vectoring).

2.1.5   Separation

Primarily a safety constraint, the separation function affords a physical interval between aircraft, intended to eliminate collision risk. The interval is operation-specific and dependent on conditions and operational uncertainty.

Typically, today's operations using secondary radar data strive to maintain a minimum distance between aircraft in flight. ATC clearances (procedure initiation, direction, altitude and speed of flight, etc.) are largely predicated on providing separation. Special considerations are given to operations constrained to precision guidance approaches and other guidance-based procedures that reduce allowable separation minima.

2.1.6   Spacing

Spacing requires one aircraft to control its position relative to another, e.g. x miles in-trail (a fixed distance along a common guidance path). Because both an order and an interval are assigned, a spacing operation supplants both the safety function of separation and the mediation function of sequencing.

Because aircraft in today's system have little data regarding other aircraft, responsibility for this function's performance lies largely with ATC.

2.1.7   Collision Avoidance

Though the collision avoidance function is also safety related, it is distinct from separation in a number of ways: While the primary function of separation achieves safety by minimizing opportunity for collision, it offers no strategy to resolve conflicts. However, if a traffic conflict persists, or evolves from an abrupt maneuver or breech of procedure to create a very near-term collision risk, a collision avoidance function can offer conflict resolution advisories, providing either or both aircraft an escape maneuver to avoid the collision. Thus both the availability of immediate resolutions and the short time horizon of this

function make it distinct from other system functions.

TCAS, the Traffic Collision Avoidance System, is an airborne system that fulfills this function. The FAA's role is limited to mandating the necessary equipment, enforcing compliance to TCAS advisories, and employing ATC procedures for coping with aircraft responses to advisories.

2.2 Improvement Programs

Recent attempts to modernize the ATS have mainly focused on the technological infrastructure legacy, commonly known as the National Airspace System (NAS), and bringing its components up to date. But these technological improvements have largely proven operationally fruitless. A case in point is the enroute controller station replacement program. After many years of hard work getting modernized stations installed in ATC facilities, there has been little difference in the methodology or effectiveness of the controllers using the new equipment.[5] Another case is the FAA's CAPSTONE program that successfully deployed (in a limited region) an entirely new surveillance system capable of both conventional air-to-ground and innovative air-to-air data dissemination. Unfortunately, the program is only exploring traffic management functions that use these data identically to those in a traditional radar environment.

Systemic change driven by such "technology-push" has not been successful outside the NAS either. Airframe manufacturers and airlines have been investing in improved airborne avionics, but have not strongly advocated changes in the ATC system that would allow the capabilities of the new equipment to be realized. Many aircraft carry systems capable of direct point-to-point routing and more efficient altitude profiles, but they are unable to utilize these functions in ATC operations that rely on fixed routes and altitude stratification. Slowly, there has been some lifting of constraints such that users can, for example, request an altitude that may be more fuel and time efficient, but these opportunities are still the exception rather than the rule.

The common thread in all of these failures seems to be allowing the demands and constraints of technology (particularly the legacy systems of the ATS), rather than the functional demands of the new air traffic environment to drive the design process. System design based on functionality rather than technology seems obvious, yet design always involves tradeoffs between implementation (cost, complexity, maintenance, etc) and functionality. In the case of ATS, the functional requirements of the system have been overwhelmed by these constraints.

Some researchers have indeed taken a broader look, suggesting redistribution of ATC function through the system [6], though their work is often regarded as unrealistic, or at the very least, impractical. Naysayers raise issues of predictability, certification, and acceptance of new roles and responsibilities. But the case for such functionally-based redistribution of form from a theoretical, system-of-systems perspective is strong.

3 Form and Function

There is a growing gap between conventional ATM methods and the nature of the demand for ATM services. The continued safety and efficient management of flights begs for a rethinking of the way the ATS provides ATM functionality: A *systemic,* flexible, and purposeful design is required. The ATS can no longer afford to be constrained to legacy equipment or operations. Granted, a transformed system will provide much of the same functionality as the current ATS, but perhaps through different means. The functions by themselves do not define the design, but may shape a system's architecture or *form.*

"Form follows function" is a concept stemming from the architectural realm that has been adopted by designers in many other areas, from biology to manufacturing. It implies that the inherent nature of things, including both their purpose and the constraints that may limit their manifestation, favor particular designs. This is not to suggest that function is the only design influence, but that it should be prominent, well considered, and most of all, satisfied. This notion is well accepted in information technology (IT), where there is an expectation that system architectures are compatible with software function.[7]

3.1 System Form: Centralized/Distributed

In IT as well as many other fields, two major classifications of designed system forms or architectures are generally recognized: centralized and distributed. Centralized forms generally consist of a single, large warehouse or processing center for data, control, etc, and single-point collection and redistribution infrastructure to amass and dispense information throughout the system respectively, akin to an airline hub-and-spoke routing design. Distributed forms have data, processing capability, etc. scattered amongst system constituents, and infrastructures that are more loosely coupled. These connections can be regular and follow rigid rules, or can be determined by function, as scale-free networks (often exemplified by Southwest Airlines' routes [8]) appear to be.

The theory of distributed systems that unfolded in the domain of IT to describe the behavior and address the challenges presented by the rapid development of computer networks can also be used to understand other distributed systems. Analytical methods and algorithms developed for distributed computer systems can be applied to analyze other systems such as the ATS.

A distributed system can be defined as a collection of independent subsystems that cooperate to solve a single problem. To its users, it appears to be a single coherent system [9]. According to Sycara [10], a distributed system is one that is comprised of asynchronous subsystems, has no global control, and utilizes decentralized data. However, communication and synchronization among subsystems is required for a distributed system to operate correctly. Different implementation philosophies are often proposed

to address these system properties, ranging from highly centralized to fully distributed solutions. Parallels can be drawn to show how the ATS can satisfy these definitions:

- Systems are Asynchronous: Global time is not known by all subsystems, and actions performed by the individual subsystems must be coordinated to ensure proper ordering. In Air Traffic, while many operations rely on planned time of occurrence, synchronization is loose at best, and is difficult to enforce due to the flight environment's variability. To cope with this issue, operations afford flexibility around timed events (e.g. a take-off window rather than a single departure time) and mechanisms to enforce control in real time rather than to a precise schedule. However, a result of "padding" the event schedule is reduced system-wide capacity.

- Decentralized Control: There is no global system control. No single subsystem can arbitrate over other subsystems without communication and synchronization. Each subsystem has incomplete information or capabilities for solving problems and, thus, has a limited viewpoint. Both supporting and counter examples are seen in the ATS: Each sector controller is responsible for traffic management within their own sector, but there is a systemic control function known as flow control which limits flight volumes in each sector to manageable levels.

- Decentralized Data: No single subsystem has complete knowledge about the system at any given time and subsystems frequently depend on each other to communicate data required to perform their functions. Different models of communications are applied depending on the specific user and data requirements. In airline operations, for example, while controllers and pilots negotiate flight routing in real time, pilots and airline dispatchers do too, as the dispatcher is privy to company information such as gate availability at the destination that the other actors are not.

One of the fundamental problems in all systems, regardless of their form, is shared resource management. Some resources can be shared by many users at a time (i.e. weather reports), while others (e.g. runway, taxi ways, etc) need to be shared in sequential order, by one user at a time. Thus, a safety critical system-of-systems must guarantee *mutual exclusion*. Coordination among users must be enforced through a communication and synchronization mechanisms to ensure the proper and correct use of the resource. Both centralized and distributed approaches are possible, but they each have advantages or disadvantages. How and by whom the resource is controlled depends on implementation.

3.1.1   Centralized Solutions to Mutual Exclusion

Centralized control is conceptually the simplest way to achieve mutual exclusion: A single entity is in charge of arbitrating resources, and is the only entity that has (or needs) knowledge of the resources' state. All users request and release access through an arbiter who grants/denies

access. Communication requirements are proportional to the number of users requesting access to the resource. Also, users may join the user group at any time, since the arbiter keeps track of membership. The main problem with this approach is that the arbiter represents a single point of failure, and can become a bottleneck at times of high demand. Also, the cost of a dedicated subsystem to be the arbiter needs to be considered. All communications must be reliable, i.e. all messages must be acknowledged.

For example, an airport tower providing approach and departure sequencing instructions to pilots represents a centralized solution to the mutual exclusion problem. It is simple to understand and implement, the number of message exchanges is low (proportional to the number of users), and it allows users to join the system at any time. The obvious drawback is that the tower represents a single point of failure, making airport sequencing and separation services solely reliant on this resource.

3.1.2   Distributed Solutions to Mutual Exclusion

A distributed algorithm for guaranteeing mutual exclusion is generally complex, and some times not possible. A fully distributed solution requires a known and stable user group to negotiate the access order and to coordinate the resource use. During a round of negotiations, new users cannot join the group and must wait. In the airport tower example, this would mean that the airport would be closed to new operations during such periods.

Knowledge about the state of the shared resource is distributed among all the users and needs to be communicated to all before any action can be taken. In a distributed architecture, all users send time stamped requests to everybody else in the group using a special time-stamping method that guaranties message ordering [11]. Receivers respond by granting or denying access depending on their state and the message timestamps. Since access to a shared resource requires the permission of all participants in the group, all users are potential bottlenecks. The number of messages exchanged by this approach is on the order of the square of the number of users. Every resource request requires each user to exchange messages with every other user. This algorithm can have multiple failure points and does not scale very well given the large number of messages that need to be exchanged to achieve consensus. As with centralized control, all communications must be reliable.

3.1.3   Tradeoffs: Centralized/Distributed Designs

Even assuming performance could be met by either a distributed or centralized design, other considerations must be weighed. Hildebrand [12] and Harbitter [13] elaborate some general trade-offs between the approaches:

- Bandwidth – Costly bandwidth favors centralized functions since data consolidation and transmission problems have less influence on system performance.

- Cost – Centralization can minimize redundancy, equating to cost savings in staffing and technology.

- Troubleshooting – Hildebrand claims that "there is no doubt that finding and eliminating problems is simpler in a centralized environment".

- Backup and disaster recovery: a double-edged sword: It is easier to backup and recreate a single entity, but they are more susceptible to site-induced problems such as power outage, etc, and offer no geographic protection.

- Investment – Distribution affords cost sharing and incremental growth, while centralization requires a larger initial investment to establish central resource.

- Security – Centralized approach affords access control and unified guidelines for system participation. Though security exposure is extreme at the central location, it is easier to protect this single asset. Both approaches require "electronic trust" or authentication.

- Reliability and accuracy – Distributed point-source users are motivated by their own discipline, policies, and requirements to keep information current.

- Scalability – Harbitter states, "A distributed strategy has a clear advantage in the area of scalability".

3.2 Form vs. Function: Scope Matters

The centralization vs. distribution of an approach is in the eye of the beholder. Take for example a simple re-routing of two aircraft to avoid a traffic conflict. From a pilot's perspective, the system is highly centralized. The local actors, the pilots, have no direct control of the aircrafts' closure geometry, and rely on a central, common resource (a controller) to resolve the problem. However, looking at the bigger picture of resolving traffic conflicts in the NAS, the problem is distributed amongst many localized control facilities and controllers, with little coordination related to this function.

As described above, centralized technology does not necessarily imply the same centrality in function. Discrimination between the nature of a function and its implementation is important. A function can be a local, distributed one such as traffic light switching at a busy intersection that is also implemented locally, e.g. with sensors and relay switching at the intersection. However, one could also consolidate local data to a central repository and analysis function that determines an appropriate response to the sensor data and returns an action to the localized system. Continuing the example, if the system at hand was not considered to be an intersection, but rather light control along a busy stretch of road, some level of synchronization and information sharing between the local intersections would be required. Synchronization alone does not necessitate a centralized approach, however, as Strogatz and Stewart demonstrated with their study of self-organizing oscillators [14]. Yet the net effect can be considered to be a global rather than local functionality.

4   Functional System Design

The concept of form following function does not by itself constitute a comprehensive design guideline. Granted, the form must complement the function, but what *are* the necessary functions for a complex, system-of-systems in general and how are these functions manifest in the ATS? A first place to look is the literature, where we find that Beer spent much effort struggling with this question. Beer concluded that irrespective of a complex system's form, there are indeed a set of necessary and sufficient *functions* to ensure its viability. He concluded that viability was maintained by engaging in different activities, keeping them from interfering with each other, managing them together, and providing for review of the former in light of the system's future interests.

Once the system-of-systems' functions are established, an idealized system architecture or form is developed, necessitating a high-level operational concept and goal-set. Subsequently, operational, practical, and pragmatic constraints are heeded, morphing the idealized system to a down-to-earth detailed design which outlines the interfaces between member systems, legacy systems to remain, etc.

The result is a system-of-systems design, potentially implemented both centrally and distributed as required within technical and political realities, rather than letting the realities over-constrain the solution. This becomes a target end state model. Then the work of transformation begins: actions that react to and motivate the changing of constraints to move the system toward the target model.

5   A Design Example

An example of an ATM operation stemming from functional system design is the Small Aircraft Transportation System Higher Volume Operations (SATS HVO) concept which was developed to increase the utilization of non-towered, non radar airports. At the heart of SATS HVO are sequencing and separation functions that were implemented in an innovative, efficient way given the objectives and constraints of the system

The SATS HVO concept relies on a volume of airspace around an airport where pilots assume responsibility for self-separation using onboard equipment and datalink communications together with published procedures [15].

Pilots approaching the airspace are given sequencing information by an automated function at the airport that represents a (locally) centralized form of control. It implements the critical function of mutual exclusion by informing pilots of their relative landing order. Airborne automation providing separation, as described in the concept, represents a distributed form of control otherwise centralized in the ATS. These design choices were derived directly from the application of a functionally-formed air traffic system-of-systems.

Research suggests that SATS HVO can safely provide a 4 to 5X increase in operational capacity relative to today's

operations. While these results may not be typical of potential benefits across the ATS, they are an indication that further application of these functional system-of-systems approaches are worthy of pursuit for ATM design.

6   Conclusions

Some ATM researchers call for a largely centralized solution [16] while others have argued for one that is completely distributed [6]. Not surprisingly, a future system at either extreme has been a hard sell, considering the limitations of scale, safety, security and performance concerns that they both can have. Hybrid solutions have slowly been winning converts in the mainstream ATS research community. There are attributes of both approaches that influence functional performance which may cause one to be preferred over the other for implementations of specific functions within the system

Regardless of these caveats, there are some generalizations concerning centralized vs. distributed approaches to ATM functions that may be illuminating:

- Mediation of a shared, high demand resource favors a centralized approach. Regulating a scheduled over-demand in this manner minimizes opportunities for deadlocks or inaction.

- Optimization in situations with sizable resources favors a distributed, localized approach.

- Optimization of limited local resources demands a "local" centralization.

- Very short-period functions e.g. collision avoidance, may require autonomic-like, distributed approaches.

- Many functions are best acted upon locally in a distributed fashion with centralized oversight.

This latter point is supported by the direction that flow management is going. Spacing also may be seen to fall into this category. It is likely that distributed implementations of spacing tools amongst aircraft will be affordable augmentations to already sophisticated on-board automation in the near future. While spacing is currently accomplished via a combination of ground-generated speed targets and airborne execution, the inner-closed-loop control dynamic beg for local management while outer loop or higher order functions, such as providing interval targets based on traffic flow, could be centralized.

Though it is important to update the system to enable its continued, day-to-day maintenance, there is also a need to ensure the system will meet its goals under future demand. Changes aimed at this latter effort should not be entirely limited by today's operations. Granted, an operationally critical system-of-systems such as air transport cannot be shut down and rebuilt from the ground up. However, changing the ATS substantively may require revisiting current policy and future goals, and focusing transition to a system more aligned with its objectives within realistic constraints. This requires revisiting the basic functions on which the current system operates and the form of its control. This in turn will require a systematic, functionally-based approach to future ATS design. A functionally-formed system may afford scalability and minimal constraint obliged by system science.

If we don't take a functional approach to design and let ourselves continue to be lead by "technology push" and political pull, we will not realize the full potential of the ATS. As air traffic demand grows and changes in nature, the ATS may quickly run into gridlock, or worse, into a scenario where its stellar safety record is compromised.

7   References

[1] Cistone, "Next Century Aerospace Traffic Mgment: The Sky is no Longer the Limit" *J. of Aircft*, v.41 n.1 2004

[2] Keating et al, "Systems of Systems Engineering", *Engineering Management Journal*, v.15, n.3, 2003

[3] Transportation Secretary Mineta, speech to the Aero Club of Wash., US DOT , Jan 27, 2004 www.dot.gov

[4] Leonard, "A Viable System Model: Consideration of Knowledge Management", *Journal of Knowledge Management Practice*, v. 1, August 1999

[5] Shea, "Standard Terminal Automation Replacement System" NATCA testimony to US Congress, March 2001

[6]Hoekstra et al, "Free flight in a Crowded Airspace" 3rd US/Europe ATM R&D Seminar, Napoli, Italy, June 2000

[7] May & Beck, "Architecture for tomorrow's administrative systems", *PennPrintout*, U. of PA, 9:6, 1993

[8] Watts, *Small Worlds*, Princeton Univ. Press, NJ 1999

[9] Tanembaum and van Steen, *Distributed Systems: Principles and Paradigms,*Prentice Hall, 2002

[10] Sycara, K. "Multiagent Systems", *AI Magazine*, American Assoc..for Artificial Intelligence, p79-92, 1998

[11] Lamport: "Time, Clocks and the Ordering of Events in a Distributed System". *Communications of the ACM*, vol 21, no. 7, pp558-565, July 1978

[12] Hildebrand , "Weighing centralized, distributed data centers", *Storage*, Feb 2003

[13] Harbitter, "A critical look at centralized & distributed strategies for large-scale justice information sharing applications" *Intg'd. Justice Info Sys Monograph*, 2004

[14] Strogatz and Stewart "Coupled oscillators and biological synchronization" *Scientific American* 269 (6):102-9, 1993

[15] Williams et al, "Preliminary Validation of the SATS HVO Concept", 24th ICAS. Yokohama, Japan, 2004

[16]Erzberger & Nedell, "Design of Automated System for Management of Arrival Traffic", NASA TM-102201,1989