THE EVOLUTION OF HIGH-RELIABILITY COORDINATION MECHANISMS FOR COLLISION AVOIDANCE

KEVIN CROWSTON, Syracuse University

School of Information Studies, 4-206 Centre for Science and Technology, Syracuse, NY 13244–4100, Tel: +1 (315) 443–1676, Fax: +1 (866) 265-7407, Email: crowston@syr.edu

ABSTRACT

Resource sharing and allocation are important coordination problems in most processes and organizations. They are especially critical in transportation systems, where the resource to be shared and allocated is the space through which various vehicles move and the problem is ensuring that vehicles do not conflict in their use of the space—that is, that they do not collide. Transportation systems are interesting because they accomplish this resource allocation in a highly reliable and often highly distributed fashion.

In this paper, we apply coordination theory to analyze collision avoidance as a coordination problem. Coordination theory suggests that coordination problems are created by dependencies among activities and resources that constrain how the activities can be performed. To avoid or overcome these constraints, additional work must be performed in the form of a coordination mechanism that manages the dependency. From this perspective, transportation systems can be viewed as collections of mechanisms for allocating a scarce resource, namely the space through which vehicles move. The claim of coordination theory is that having identified the type of dependency involved in transportation systems, we can consider alternative coordination mechanisms and more importantly, the tradeoffs between them. More interestingly, we can analyze how the use of information technology differentially affects the costs of different mechanisms thus shifting the tradeoff. As well, the range of coordination mechanisms identified may have implications for resource allocation in other kinds of organizations.

Resource sharing and allocation are important coordination problems in most processes and organizations. For example, most organizations must assign employees to tasks, schedule limited equipment or allocate raw materials to various products. These problems are solved using numerous coordination mechanisms, such as first-come-first-served, managerial fiat, or market prices. Resource allocation is also critical in transportation systems, where the resource to be allocated is the space through which various vehicles move. In this case, the resource allocation problem is ensuring that vehicles do not conflict in their use of the space—that is, that they do not collide—and again, numerous coordination mechanisms have been developed.

In this paper, we will apply coordination theory to analyze alternative approaches to collision avoidance. The contribution of the paper is to identify the coordination mechanisms used for space allocation in transportation systems to provide insight for resource allocation in other settings. As a specific example, we will use the principles of coordination theory and the mechanisms derived from our analysis of transportation systems to suggest several parallel approaches for the allocation of the time of computer-support specialists. A secondary purpose is to identify coordination mechanisms from other processes that might be useful in transportation systems. Finally, we show how the specific characteristics of the systems, such as the use of technology, affect the relative desirability of different mechanisms. This analysis may suggest a comparable technology-related evolution of coordination mechanisms in other settings.

Transportation as a coordination problem

Transportation systems must allocate many scarce resources, such as vehicles with different capabilities, drivers/operators and space of different kinds: runways, parking spaces, roadway capacity and so on. In this paper, we will focus at the finest granularity and consider the second-to-second decisions needed to allocate space to avoid collisions.

Strictly speaking, collision avoidance simply requires that the vehicles not actually make contact—as the cliché states, a miss is as good as a mile. However, because of the degree of imprecision in operations, the goal of these systems is not just avoidance of collisions, but rather maintenance of a more generous degree of separation between vehicles. The size of this space depends on the nature of the vehicles and on the level of safety desired. For example, controlled aircraft are separated by 1000 feet vertically and 5 nautical miles laterally. For automobiles, a commonly taught rule of thumb suggests allowing a car length between cars for every 10 mph of speed, although most drivers seem satisfied with less. Therefore, it is common to speak of collision avoidance schemes as providing "separation" rather than collision avoidance. Essentially, each vehicle defines a chunk of space, with itself at the middle. Ensuring that each vehicle has exclusive use of the space around it at all times is the separation problem. For example, Figure 1a shows two aircraft on crossing courses. Because they occupy different points in space at all times, they do not collide. However, Figure 1b shows that the aircraft still conflicted in their use of space because the extended area around the vehicles did intersect (a loss of separation).

CONTRIBUTION

This paper makes several contributions. First, the paper makes a methodological contribution by demonstrating the use of coordination theory to analyze a system. Second, the use of coordination theory puts the various mechanisms into a common framework, highlighting similarities and differences between the various systems. This analysis can also be extended to show how coordination mechanisms from other processes might be useful in transportation systems (or vice versa). Finally, we discuss how the specific characteristics of the systems, such as the use of technology, affect the relative desirability of different mechanisms. This analysis suggests how coordination mechanisms might evolve with the increasing use of technology.
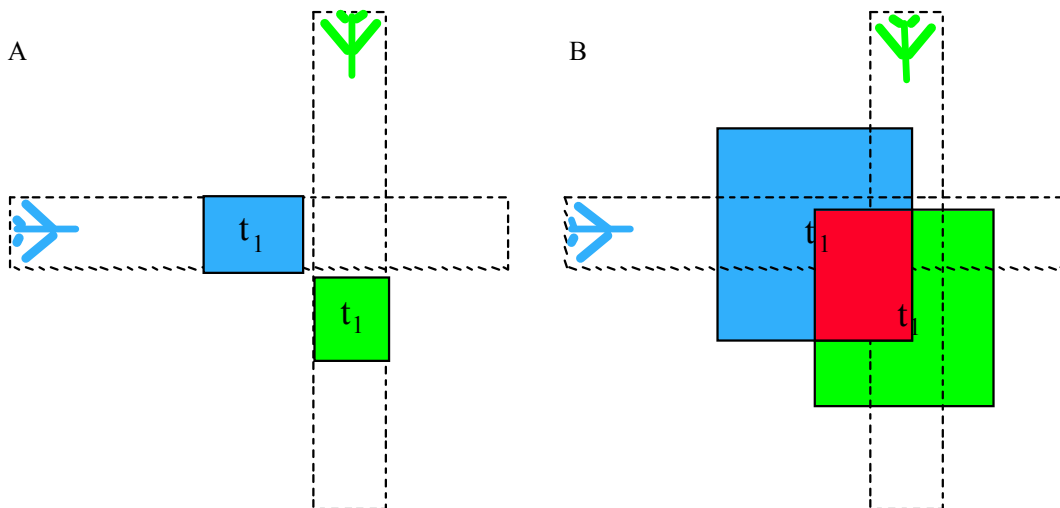
Figure 1. Collision avoidance versus loss of separation. The two aircraft do not collide because they occupy distinct locations at all times, as shown in A. However, because the extended areas around the aircraft intersect, as shown in B, there is a loss of separation.

Why study collision avoidance?

Transportation systems are interesting for several reasons.

- First, transportation is a system, including distributed vehicles and sometimes central control. Importantly, these systems often work in a decentralized fashion with only limited communication between vehicles. Mechanisms from these settings may be particularly applicable to distributed groups.

- Second, the extremely high cost of doubly allocating space—namely a collision—requires a highly reliable resource allocation mechanism. Study of transportation systems may therefore have implications for the design of other high-reliability systems. Particularly interesting is the recent development of meta-control mechanisms to avoid overloading a simpler but less robust allocation mechanism.

- Third, space is a continuously divisible resource, so mechanisms have to define the resource as well as allocate it. As well, vehicles occupy a path through space, so decisions about allocations have to be linked over time.

- Finally, and of particular interest to research in information systems, advances in technology and demand have historically led to new ways of managing resources, as particular functions are automated and increased information provided to vehicles. A principled analysis may suggest new approaches based on other resource-allocation mechanisms.

Overview of the paper

In the remainder of this paper, we will first introduce coordination theory and discuss how it can be applied to transportation systems. We will then consider space allocation in four settings: automobiles, trains, ships and planes. We conclude by discussing how these findings can be extended to resources more common in organizations.

COORDINATION THEORY

To analyze transportation systems, we apply the analytic lens of coordination theory (Malone and Crowston 1994). Coordination theory suggests that *dependencies* among activities and resources create *coordination problems* that constrain how the activities can be performed. To avoid or overcome these constraints, additional work must be performed in the form of *coordination mechanisms* that manages the dependencies. The further development of coordination theory requires 1) cataloging possible dependencies, 2) identifying alternative

coordination mechanisms that can be used to manage each dependency and 3) describing the tradeoffs among these mechanisms. The claim of coordination theory is that having identified the type of dependency involved in a process, we can create new processes by considering alternative coordination mechanisms and the tradeoffs between mechanisms. In particular, we can look for alternatives that are enabled or improved by the use of information technology (IT).

Shared resource dependencies in transportation

Cataloging dependencies is an active research area to which this paper contributes. Malone and Crowston (1994) offered a preliminary list that was later extended by Crowston (2003). In their terminology, the dependency in the case of transportation is a shared-resource dependency between the motion of two vehicles. In order for either vehicle to proceed, it must have exclusive use of necessary resource, namely the space, for a given period of time. To ensure this exclusive use, additional work (i.e., a coordination mechanism) is necessary[1].

Space as a non-shareable non-consumable resource. The choice of mechanisms to manage shared-resource dependencies depends in part on the type and nature of the resource to be shared. Runways, roadways and other kinds of space are non-shareable and non-consumable (NSNC) resources, meaning that they can only be assigned to a single activity at a time, but can later be reused by other activities (at least until they wear out or need maintenance; such factors are outside the time-frame we are considering here). Other examples of such resources include meeting rooms, tools, time on computer networks and the time of human experts. By contrast, information resources can be easily shared among several activities while raw materials and money are typically consumed by the activities that use them. However, investors might consider money as a NSNC resource, since they expect to get the invested money back after some time.

Because NSNC resources are allocated to a single task for a period of time, the allocated resource is specified by a combination of time and resource. The resource allocation problem can therefore be viewed from either perspective. In other words, while we have talked about the need for a buffer in space around each vehicle at every point in time, it would be equally correct to talk about a buffer in time around each vehicle at every point in space. In most of the following discussion, we will take the first perspective, though the principles apply as well to the second.

Coordination mechanisms for resource allocation. Malone and Crowston (1994) suggest three general approaches to managing the allocation of NSNC resources: elimination of the dependency, conflict-detection and pre-allocation. In the first approach, obtaining additional resources eliminates the dependency, thus eliminating the need for a coordination mechanism. For example, if every group has its own dedicated conference room, then they never need to check if the room is available for their meetings. This approach is appropriate for low-cost and high use items, such as staplers, desks and even computers.

In the second, actors simply take the resources they need and resolve any conflicts as and if they arise. For example, a group might simply occupy a meeting room if it is not already in use or look for an alternative meeting room if it is. This basic approach is used for automobile traffic, as drivers simply use the roadway in front of them unless someone else occupies it. More specifically, someone must:

1. determine what resources the activities need;

2. identify a possible resource to use;

3. check if that resource is already in use;

---

[1] Note that this analysis does not address the related problem of navigation, which might be defined as finding the way from one place to another while avoiding stationary obstacles. Rather, we take different degrees of navigational ability as an important characteristic of these systems and note that changes in this ability may require changes to the collision avoidance system.

a) if the resource isn't in use, then use it;

b) if the resource is in use, then determine which activity has priority and
repeat the process to assign different resources to the other.

Of course, these steps are performed with many variations. In step 1, the needed resources may be obvious, or may require significant work to determine. Indeed, in a group, coming to shared agreement about the tasks and resources may require a collective act of sense-making (Crowston and Kammerer 1998). Step 2 might turn up a large number of resources, or just one, known to be available. Step 3b, checking for priority, is optional: the resource might always go to the current user.

In the third approach, the same basic steps are required as in the second, but they are performed ahead of time and step 3 is modified to check for or make reservations. For example, meeting rooms might be reserved in advance or a train might be scheduled to use a particular stretch of track at a particular time. Of course, to avoid conflicts at the time the action is performed, it is necessary to check that the reservation has been honoured and the resource is actually available.

Coordination mechanisms and information processing. Coordination mechanisms differ in the information they need and how it is processed, making them a particularly interesting topic for information systems researchers. Therefore, it is important to note which individuals have the necessary information and to consider how it will be communicated to those who need it. For resource allocation, necessary information includes task needs (step 1), resource availability, current or future (steps 2 and 3) and allocation decisions (step 3). For example, in the case of a meeting room, the status of a room might be indicated by a "Meeting in progress sign" or by the simple presence of a group (seen through a window or by opening the door). Information about reservations might be managed by one person or by keeping a list where all users can find it. Making conflicts visible (required in step 3) is a key problem in sharing data in database systems. Similarly, the ability to see conflicting traffic is important in choosing between different collision avoidance schemes. As well, there must be some basis for choosing between conflicting uses (in step 3b). A meeting room might be allocated first-come-first-served, by the decision of a manager (e.g., based on the perceived importance of the meeting) or even by bidding, as in a market. Most transportation systems are based on first-come-first-served priority, but other approaches might be useful, as will be discussed below.

Tradeoffs among mechanisms. Different coordination mechanisms impose different costs, so typically there is a tradeoff to be made in selecting a mechanism. Obtaining additional resources eliminates the need to coordinate between conflicting users, but at the cost of the additional resources. In the case of transportation, such an approach may be infeasible due to the limited amount of space available and the impossibility of making more. Conflict-detection trades the fixed cost of making a reservation with the (usually lower) cost of checking first that the resource is free, but for some activities adds the cost of hunting around for available resources. Any of these mechanisms might be better than the others, depending on the circumstances. More interestingly, the use of technology will differentially affect the costs of possible mechanisms, again making them interesting for information systems researchers. For example, a computer system might make it possible for all potential users to cheaply make room reservations or find available rooms. Similarly, advances in communications and especially navigation will change the desirability of different collision avoidance systems.

COLLISION AVOIDANCE IN TRANSPORTATION SYSTEMS

In this section we present mini-case studies of collision avoidance in four modes of transportation: automobiles, trains, ships and aircraft. There are many similarities among these modes. For example, each is a system composed of separate vehicles controlled in a distributed fashion by drivers, engineers, masters and pilots, respectively. However, the resource allocation mechanisms differ,

affected by factors such as the different capabilities of vehicles and their freedom of motion—one-dimensional for trains, essentially one-dimensional for cars (which are restricted to the road, even though they can switch lanes), two-dimensional for ships and three-dimensional for planes. Therefore, the comparison among these modes can illuminate factors influencing the choice of mechanisms.

Each mini-case is based on analysis of secondary documents, such as descriptions of:

- the operation of each system, such as manuals for pilots, drivers, ship's officers, etc.;

- the history and evolution of these systems and explanations of their current structure;

- problems with each system and accident reports; and

- discussions of potential new technologies in the literature.

We have also drawn in part on personal experience as a car driver and licensed airplane pilot and on consultations with a licensed marine navigation officer. In each case we briefly discuss how collision avoidance is effected and reduce these mechanisms to basic principles in order to make the comparison between them clearer.

Automobiles

We will start our discussion of specific resource allocation mechanisms with automobiles, probably the most familiar form of transportation to most readers. The question we address is how collisions are avoided in the automobile transportation system, that is, how the scarce resource of road space is allocated so as to avoid conflicting uses. The main principle for conflict avoidance in automobile traffic is "see-and-avoid", meaning that drivers are expected to see possible conflicts and

avoid them, as shown in Figure 2. In some cases, a manager (human or automated) explicitly assigns the resource is to one vehicle or another. A subsidiary principle is to dedicate resources to classes of users to minimize conflicts. These principles will be discussed in turn.

See-and-avoid

In automobile traffic, the primary collision avoidance mechanism is "see-and-avoid". In the terms introduced above, this is a conflict-detection resource allocation mechanism, where step 1 and 2 are determining the future course of the vehicle and therefore what space is needed, and step 3, identifying possible conflicts from other vehicles. A summary of this mechanism is shown in Table 1. If a conflict exists, then in step 3b, drivers determine who has priority and who should alter course. In head-on conflicts both vehicles alter direction to their right (or left, in the UK, Japan and many other countries); in other cases, priority is determined using simple heuristics, such as first-come-first-served or vehicle to the right has priority. The other driver takes action to avoid a conflict, typically by slowing or stopping. These heuristics are taught as part of drivers' education and enforced by law.

This mechanism has two advantages. First, see-and-avoid is a decentralized process, requiring no central control in real-time. In other words, the actors driving the vehicles also perform all of the coordination mechanism. As a result, see-and-avoid is relatively inexpensive to implement. Second, no information needs to be explicitly exchanged by drivers; instead, each looks out for developing conflicts and decides independently how to resolve them. In other words, drivers independently gather and

Figure 2. See-and-avoid for collision avoidance in automobile traffic. Each driver looks for conflicting traffic and diverts to their right (or left, depending on the country) as necessary to avoid a collision.

Table 1. Summary of collision avoidance mechanisms for automobile traffic.

Step	Automobile traffic	Traffic in lanes	Shifting lanes
Step 1—Resource needed	Space ahead of automobile	Space in lane ahead of automobile	Space in lane adjacent and ahead
Step 2—Resource available	Space with maneuvering limits of automobile	Space with maneuvering limits of automobile	Space with maneuvering limits of automobile
Step 3—Possible conflicts	Other automobile on collision course	Automobile in lane ahead or behind	Automobile in adjacent lane
Step 3a—Priority	Unclear—often vehicle to the right or no one if dead ahead	Vehicle ahead has priority	Vehicle in lane has priority (usually)
Step 3b—Alternative resources	Divert to the right, slow down or stop	Slow down or stop	Stay in lane, slow down or stop

process the information they need. To do so, drivers make assumptions about what others will do and plan accordingly. In practice, information exchange is restricted to hints about intended actions, such as signalling lane or direction changes.

Seeing. Reliance on see-and-avoid has implications for the design of the rest of the transportation system. First, see-and-avoid requires that drivers be able to see conflicts— other vehicles—in order to avoid them. Highways, for example, must be engineered with bends shallow enough that drivers can see enough ahead or conversely, the speed limit must be lowered in areas with sharp curves (OwenBowen and the Editors of Life 1967, p. 101). Design rules must ensure visibility at intersections, e.g., by removing hedges or other obstruction or by adding mirrors. Nevertheless, a study has shown that 23% of accidents have as a causal factor improper lookout (drivers fail to check for conflicts), 15%, inattention (drivers fail to notice a conflict that requires them to slow down or stop) and 8%, false assumption (drivers guess wrong about what others will do and miss a conflict) (Treat et al. 1979). View obstruction contributed to a further 12% of accidents (Treat et al. 1979).

Avoiding. Second, having seen, drivers must maneuver so as to avoid conflicts. Improper evasive action contributed to 13% of accidents (Treat et al. 1979). The combination of loss of view and inability to avoid is particularly deadly. For example, accidents frequently happen in fog banks where drivers can no longer see. Even worse, drivers in fog do not behave predictably; some maintain speed, while others slow down or even stop, making it impossible for drivers to predict what others will do. One approach to reducing the toll of fog-related accidents is driver education about the appropriate action to take.

Dedicated resources to reduce conflicts

As discussed above, the one approach to coordination is to dedicate resources to eliminate rather than manage the dependency. This coordination mechanism is also summarized in Table 1. In our framework, these approaches restrict the resources considered in step 2. In automobile traffic, lanes are dedicated to traffic moving in the same direction, as shown in Figure 3, thus reducing collision avoidance in most case to not running into the car ahead (or stopping too suddenly for the car behind). It is much easier for a driver to concentrate on the road ahead rather than worrying about possible conflicts from all directions. Note that many accidents occur in parking lots and driveways, where traffic comes from multiple directions, making seeing harder, and where the rules for priority are less clear, making avoidance harder.

The obvious problem with lanes is where they cross, since streams of traffic use the intersection in both directions, as shown in Figure 4. In a sense, lanes reduce potential conflicting traffic by concentrating it at

intersections. Again, one solution to this shared resource dependency is to provide duplicate resources to eliminate it, i.e., by building overpasses and exchanges, as shown in Figure 4D. Limited entrances and exits also limit possible conflicts and as a result, accidents are rarer on expressways than city streets.

Explicit management of resource allocation

If the dependency cannot be eliminated, it must be managed. As discussed above, two approaches can be used: conflict-detection or pre-allocation. Right of way indications, stop signs and rotaries (also known as round-abouts or traffic circles), shown in Figure 4A and B, establish the priority when there are conflicting uses. As with traffic along the road, drivers arriving at the intersection look for conflicts and proceed if there are none, as shown in Table 2. However, if two drivers arrive at the same time, one defers to the other, priority being determined based on conventional rules, such as the driver on the main road, without a stop sign, or to the right (or left) has priority.
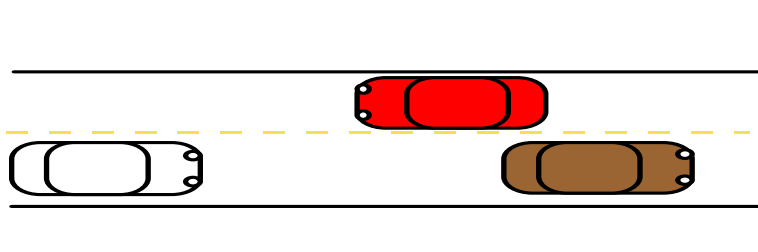

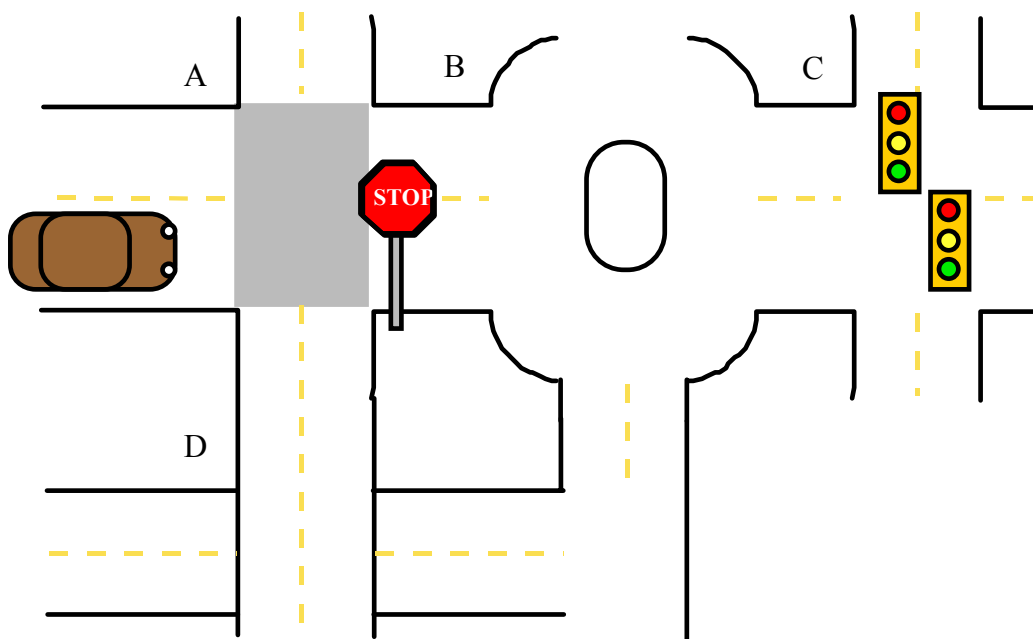**Figure 3. Dedicated resources (lanes) reduce conflicts in automobile traffic.**

Figure 4. Intersections create a shared resource that must be allocated. The intersection may be managed by drivers at a stop sign (A) or rotary (B), by a traffic light (C) or the dependency eliminated by an overpass (D).

Table 2. Summary of collision avoidance mechanisms for automobile traffic at intersections.

Step	Automobile traffic at intersections
Step 1—Resource needed	Space in intersection
Step 2—Resource available	Space with maneuvering limits of automobile
Step 3—Possible conflicts	Other automobile on crossing street on collision course
Step 3a—Priority	Automobile without stop sign or with green light or on main road has priority
Step 3b—Alternative resources	Slow down or stop and wait for road to be clear

Stop signs work poorly in heavy traffic because on each turn only a single car can progress, thus imposing a high overhead for this resource allocation mechanism. On the other hand, without a stop sign, traffic on a side road may have to wait for a long time for a break in traffic on the main road. Rotaries have the advantage that in the absence of conflicts, drivers can proceed immediately, but they too fail under too heavy a load. When the overhead of a distributed system would be too high, a pre-allocation mechanism can be used instead. For example, a traffic light (Figure 4C) performs an explicit resource allocation, giving the intersection to one traffic stream or another for a period of time. This approach has the advantage of distributing the switching time over multiple cars, at the cost of building a signal and possibly requiring traffic to wait even in the absence of conflicting uses. If the traffic increases further, time must be reserved as well to facilitate traffic crossing traffic by turning left (or right).

Changes enabled by new technology

A number of new technologies are being considered to increase the traffic-carrying capacity of highways. For example, researchers are developing an active cruise control that would use radar to maintain a fixed distance from the car ahead. Notice, however, that the use of such a system does not change the resource allocation mechanism. Instead, it substitutes the system's seeing for the driver's, although only for traffic ahead in the lane.

Summary

In summary, the primary collision avoidance system used for automotive traffic is see-and-avoid, a distributed resource allocation system based on ability of all users to see-and-avoid potential conflicts. This mechanism is augmented with rules to set priorities where there are potential conflicts, such as rules for who goes first at a stop sign or to give priority to a main road. As well, as traffic increases, resources are dedicated to handle particular flows, either time-shared (as at intersections) or permanently (as with bridges). The static allocation of resources, such as lanes, overpasses, etc. and the careful design of roads and regulations ensure see-and-avoid is feasible. Even so, many accidents are attributable to failures of this mechanism, which is the tradeoff for the reduced cost of the mechanisms.

Trains

Next we will consider train traffic. Trains are superficially similar to automobiles, moving as they do along a single-dimensional track, also known as a road. However, the increased speed and weight of trains means that their operators (engineers) cannot possibly see far enough ahead to be able to avoid conflicts, especially traffic coming in the opposite direction[2]. Furthermore, train engineers cannot easily avoid conflicts, because a train cannot stop quickly nor move

[2] Because of the cost of laying double tracks, many lines are single track used in both directions.

off the line except at certain limited points (i.e., at sidings). Therefore, the collision avoidance system for trains has evolved quite differently. In the remainder of this section, we will consider the mechanisms used to control train traffic.

Pre-allocation of resources

Because of the limitations discussed above, train traffic is carefully pre-planned using a timetable (Blythe 1951, p. 28) that lays out when trains should run and exactly where they will pass each other. For example, the schedule will indicate when a low-priority train should pull on to a siding or wait in a station to allow an express train to pass. In other words, the limited resource, a track, is explicitly pre-allocated to particular uses by a scheduler using a schedule to check the availability of the track. This mechanism is summarized in Table 3. Figure 5, based on Tufte (1983, p. 31) shows a graphical approach to scheduling. The figure represents a train schedule, with time across the page and distance down. Individual trains are plotted as diagonal lines, allowing crossings to be worked out in advance. More recently, computer systems have been used to develop and check schedules (e.g., Zwaneveld 1997). A key constraint included in such systems ensures headway (i.e., spacing) between trains

on the tracks (p. 30) and in stations (pp. 35–39). The later task is more complicated because of the large number of possible routes through a station. However, an entirely prescheduled system is inflexible; accidents can easily occur when a train breaks down on the line, thus occupying the resource past its scheduled time, and another comes along. As well, adding a "special" train (one not in the regular schedule) requires particular care, as workers tend to assume the track is available if ordinarily unscheduled. For example, it might ordinarily be harmless if a train leaves a station a few minutes off schedule, but this behaviour can lead to an accident if a special train is scheduled to pass at the station.

Facilitated see-and-avoid

To avoid the limitations discussed above as traffic grew, train operators attempted to explicitly control the space occupied by a train in real-time, i.e., augmenting the pre-allocation mechanism with a conflict-detection mechanism. Since trains are limited to their tracks, in general, there is no question of priority, but rather of stopping before a collision. Since engineers cannot see conflicts far enough ahead to stop (for step 3), they relied instead on signals, what might be called facilitated see-and-avoid.

Table 3. Summary of collision avoidance mechanisms for train traffic.

| Step	Train traffic, pre-scheduled	Train traffic with signalmen	Train traffic with block working
Step 1—Resource needed	Track from origin to terminus of train	Track ahead of train	Track ahead of train
Step 2—Resource available	Track—trains are difficult to stop and can't move off track	Track ahead of train	Next block of track ahead of train
Step 3—Possible conflicts	Other train running on same track	Train on track ahead; recent passage signalled by signalman	Train on track ahead; presence communicated by next signalman, or directly sensed
Step 3a—Priority	Determined by company policies	Train ahead has priority	Train ahead has priority
Step 3b—Alternative resources	Schedule trains to pass at siding or station	Slow down or stop	Slow down or stop
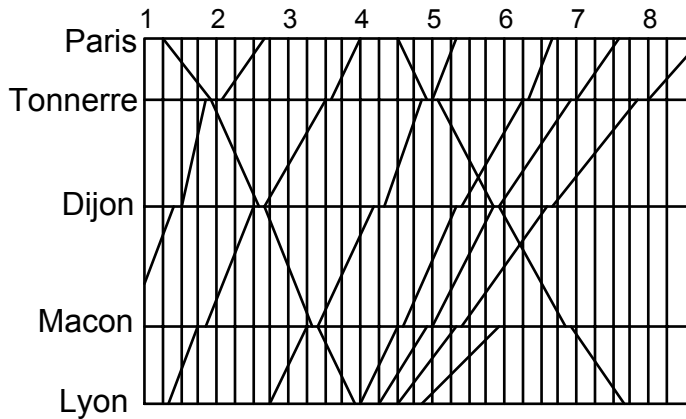
Figure 5. Graphical representation of a train schedule (after Tufte, 1983, p. 31). Time is across and distance down. Stations are indicated by horizontal lines. The paths of individual trains are shown by the diagonal lines; passing situations are shown by intersecting lines.

Originally signalling was done by hand. A railroad maintained stations every mile or so, manned by flagmen who signalled approaching trains that the line ahead was clear, as shown in Figure 6 (Blythe 1951, p. 27). This mechanism allocates track space based on time-interval separation. Flag men simply waited a safe time after a train passed before signalling that line was clear, thus providing the necessary information for step 3, as shown in Table 3. Alternately, they might show a caution flag for some time after a train, allowing others to pass, but requiring they be prepared to stop. This mechanism has the major advantage of requiring no communication between stations. Indeed, each station acts like a memory, simply noting the fact that a train has recently passed.

However, clearance from a flagman cannot prevent a collision with a train broken down on the tracks or approaching from the opposite direction (i.e., the flagmen can only "see" a subset of possible conflicts). To handle the first case, a broken-down train, a flagman would be sent a mile or so from the casualty to stop any on-coming trains or to place detonators (small explosives) on the track to signal other trains to stop.

The second problem could only be handled by pre-allocating the track for one direction or another, and not allowing a train to depart until the one expected from the opposite direction had arrived. Again, such a mechanism is inflexible; special trains are at risk if unexpected. Another approach to the problem of opposite direction traffic is to dedicate resources to eliminate the dependency. Busy railroads often have tracks used only in one direction (e.g., an up and a down track) or for express vs. local traffic, as shown in Figure 7. Having separate lines doubles the cost of the tracks and maintenance but eliminates the need to communicate between ends of the block.
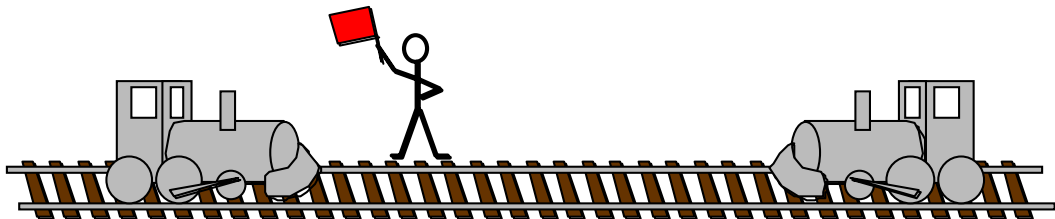

Figure 6. Allocation of track by a signalman. The signalman indicates that another train has passed, thus augmenting the driver's ability to see ahead.
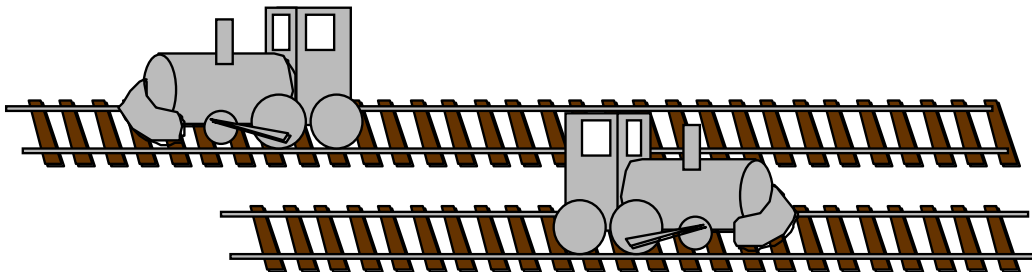
Figure 7. Dedicated resources (one-way tracks) reduce conflicts in train traffic.

Finally, the dependency can be explicitly managed. A key problem in allocating a railway is determining that a piece of track is occupied (step 3 of the mechanism). Various systems have been invented to manage the allocation by representing a line with another resource for which it was easy to determine use. For example, one mechanism reifies possession of a section of track in the form of a baton; before proceeding on to the line, the engineer must be in possession of the baton, picked up at the entrance (Blythe 1951, p. 86). A problem with this mechanism is that trains have to go in each direction alternately in order to move the baton back-and-forth or a runner must be sent to pickup the baton (Blythe 1951, p. 87). Variants allow for a baton with multiple sections, to allow multiple trains in the same direction, a written pass, issued after showing the engineer the baton that proves ownership of the line, or various electrical systems that provide the same function (Blythe 1951, p. 87).

Block working, enabled by new technologies

The next major development in train control was block working. The continuous track is broken up into sections, called blocks, which are allocated by the signals placed at the entry to each block, as shown in Figure 8. No train is allowed into a block until the previous train has cleared it. In other words, the train driver's seeing is augmented by the sensing of the signals. An interesting property of the signals is that instead of defining the space to be allocated based on the location of the train (which is usually uncertain), the resource is instead defined by the location of the signals. In other respects, however, the allocation mechanism remained unchanged: trains used

the blocks as they came to them, waiting if the signals indicated that they were already in use.

While block working makes it less likely to run into a broken-down train or one approaching from the opposite direction, it requires communication between signalmen. The signalman at the entrance to a block needs to know when the previous train leaves at the other end, as well as when a train enters from the opposite direction. Block working was originally made possible by the use of the telegraph between stations. Later innovations reduced the need for direct communication. For example, automatic signalling indicates that a train is in a particular block by directly sensing the presence of the train on the tracks. These sensors are now tied directly to the signals, providing an engineer with immediate information about traffic ahead. Many railroads have undergone an interesting transition driven by the technology: track was originally single, then doubled where needed to handle additional traffic and later, returned to single track as the increased coordination capability made coordination of the limited resource feasible. Summary

In summary, train traffic relies on communication between controllers. Engineers cannot see far enough ahead to make see-and-avoid feasible, and therefore must rely instead on signals indicating the absence of conflicts. For the purposes of allocation, the track is divided into fixed blocks. To avoid disasters, the communications between the stations has become increasingly positive, meaning that a message must be sent to say the track is clear rather than that it is occupied.
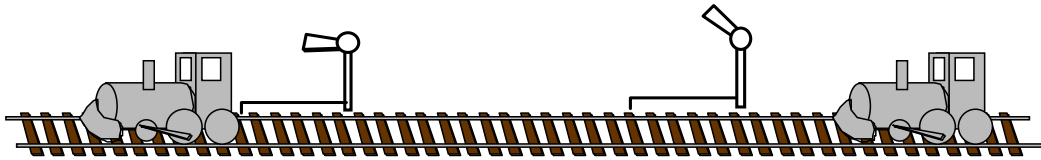
Figure 8. Signals indicate the status of the track ahead in block working.

Ships

The third system we will discuss is marine traffic. Avoidance of collisions in ships is interestingly similar to that for automobiles, since it is also relies primarily on see-and-avoid. Ships do not move very quickly and visibility is usually unrestricted, so there is usually time to think about what is happening and plan a course of action to avoid collision. However, ships cannot stop or change courses quickly (or perhaps not at all), so action has to be taken well in advance. Because of the many forces acting on a ship, adherence to precise paths is not routinely possible (NRC, 1994, p. 197), especially in confined and shallow waters such as in harbours, making navigation difficult and increasing separation requirements.

See-and-avoid

As with automobiles, the primary method of collision avoidance is see-and-avoid, as shown in Figure 9, and summarized in Table 4. Because the seas are open to all, coordination mechanisms are a result of international agreements rather than corporate directive. Collision avoidance is governed by an international agreement called the *International Regulations for Preventing Collisions at Sea*, often referred to as the *Rules of the Road* (Tate 1976).

Table 4. Summary of collision avoidance mechanisms for marine traffic.

| Step	Marine traffic	Marine traffic with radar and radio
Step 1—Resource needed	Seaway	Seaway
Step 2—Resource available	Seaway—ships are difficult to stop and maneuver precisely	Seaway—ships are difficult to stop and maneuver precisely
Step 3—Possible conflicts	Ship on collision course, detected by sight or sound	Ship on collision course, seen on radar
Step 3a—Priority	Determined by rules of the road	Determined by discussion between masters
Step 3b—Alternative resources	Bear right, slow down or stop	Bear right, slow down or stop
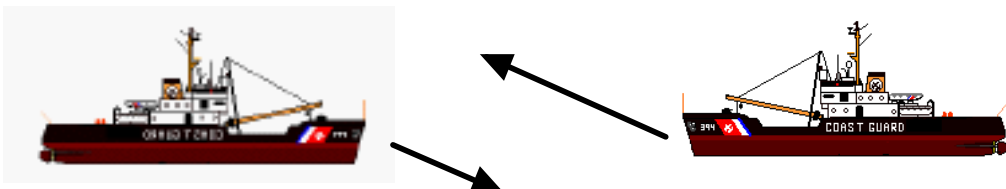
Figure 9. See-and-avoid for collision avoidance in marine traffic.

Seeing. The rules of the road impose a legal requirement for ships to keep a proper visual lookout (rule 5) and specify lights to be carried at night to ensure that others can determine a ship's relative headings, size, etc. (Part C of the rules). When visibility is impaired, for example, in a fog, ships are required to sound fog signals to warn other ships of their presence (rule 35). Following these rules is intended to ensure that step 3, looking for conflicts, is always feasible. Nevertheless, the majority of collisions between ships have been attributed to failures to keep a good lookout (Transportation Safety Board of Canada 1991).

Avoiding. To avoid, the rules require that ships maintain a "safe speed" (rule 6), that is, a speed that allows for maneuvering or stopping in time to avoid a collision. Because fog makes seeing harder, a ship's safe speed is reduced in a fog, especially when a fog signal is heard (rule 19). If a ship's master determines that a collision seems possible, the rules of the road (in Section II of Part B) determine which ship must give way (called the "burdened" ship) and which has priority (called the "privileged" ship). In general, the overtaken ship or the ship crossing from the right has priority, as shown in Figure 10, although different rules apply for sailing vessels and for numerous categories of ships restricted in their ability to maneuver (such as fishing vessels, mine sweepers, etc.). Applications of these rules allows two masters to determine who should act first to avoid a potential collisions without their having to communicate (in other words, masters should be able to independently perform steps 3a and b and come up with the same result).

The rules are written in terms of two ships meeting. A National Research Council report notes that "Interactions involving more than two vessels… are more complicated", so "[i]n such cases, considered a special circumstance by the [rules], the precise rules give way to prudent seamanship and are followed only as is practical and prudent" (1994, p. 53). Such situations are rare at sea, but common in a crowded harbour situation, as discussed below.
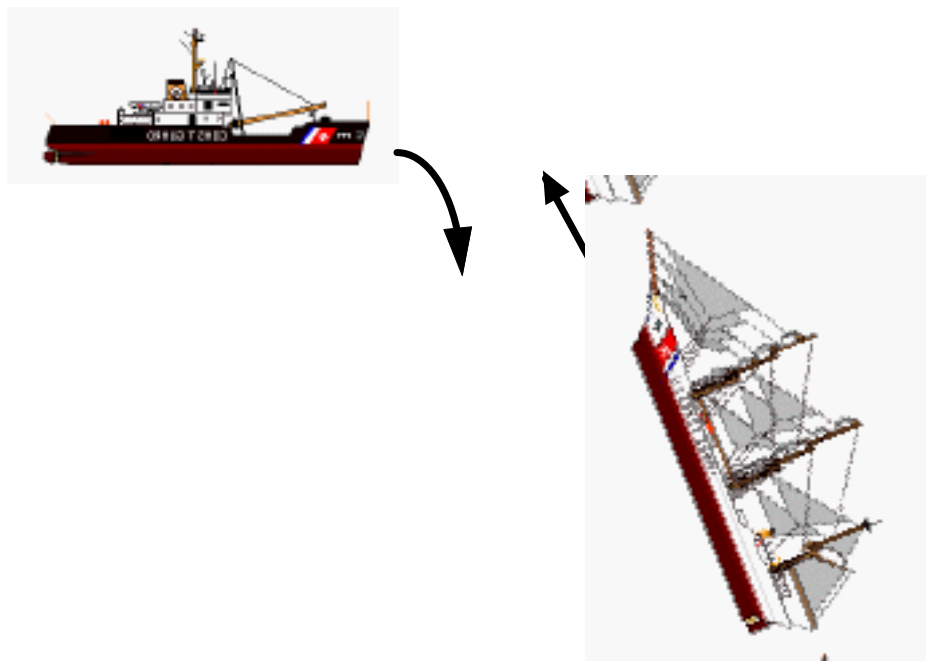

Figure 10. Example of priority rules: the sailing ship on the right is stand-on and maintains course, while the steamship on the left is give-way and changes course to the right to pass behind.

Dedicated resources for high density traffic

To facilitate traffic, some resources are dedicated to particular uses, although to a much lesser extent than with automobile traffic. First, ship traffic keeps to the right in narrow channels (rule 9). In high-density traffic areas, a traffic separation scheme may be used (rule 10), which involves lanes allocated for traffic going in each direction, as shown in Figure 11. These lanes have the same benefit of reducing the number of possible conflicts. The rules of the road require that these be crossed at right angles so the intent of a vessel cannot be mistaken. However, because of constraints of space, it is not always possible to have separate lanes in all high-traffic areas.

Allocation by managers

While there are no cases of strict control of marine traffic as in air traffic control (NRC 1994, p. 186), there are a few examples of stretches of water that permit only one-way traffic and for which queues are managed. One obvious example are locks, which operate in one direction at a time; traffic waiting for the lock queues until the lock is available, as shown in Figure 12. During heavy traffic, ships might have to wait several cycles for a chance to enter, or, as with the Suez and Panama Canals, locks might be dedicated to one direction during certain times and the opposite at other times. Another example is the Mississippi River at Algiers Point in New Orleans, which is too narrow for two ships to pass safely during high water when the current is strongest (NRC 1994, p. 171). During these times, the equivalent of stoplights is used. A controller in a position to see both sides of the strait controls a set of traffic lights, allowing ships into the channel from each direction alternately. Both of these are cases where a manager allocates the resource.
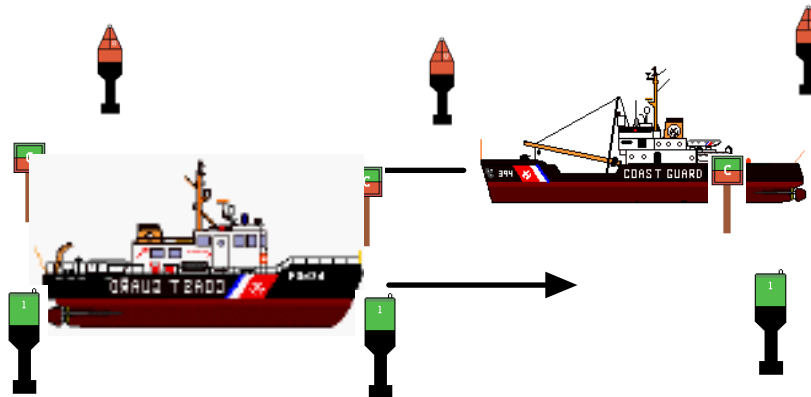

Figure 11. Dedicated resources (one-way traffic lanes) reduce conflicts in marine traffic.

Figure 12. Explicit allocation of a narrow channel. Ships wanting to transit the channel wait until it is allocated to them, in this case as signalled by the traffic lights.

Kevin Crowston

As well, there are a number of Vessel Traffic Services (VTS), shore-based systems that provide traffic and navigation information to vessels in their area of responsibility. Current VTS direct maneuvers only in emergency conditions (NRC 1994, p. 202). VTS are widely used in Europe, and in some cases, coordinate movements in a restricted area, "taking into account waterway physical limitations, berth availability, priority of movement, potential congestion points, and other factors" (NRC 1994, p. 204). In other words, such a VTS essentially pre-allocates the congested space. Even with these systems, though, controllers do not attempt to tell the ships how to navigate but rather provide the information needed for the ship's master to decide.

Changes enabled by new technology

The rules of the road codify centuries of marine tradition. However, their use has been somewhat affected by two more recent inventions, radio and radar. A key point of the rules is to ensure that ships behave predictably, thus allowing others to determine how they can avoid a collision without having to communicate. For example, while the burdened ship is required to change course to avoid a collision, the privileged ship is equally obliged to maintain its course so the burdened ship can determine a safe course (rule 17)[3]. To warn other ships of maneuvers, the rules include explicit instruction for the signals to give in various circumstances to communicate intentions (rule 34). For example, one whistle indicates a course change to the right, and two whistles, to the left. A ship can also signal that it is unable to maneuver to avoid a collision, which places the onus to avoid the collision on the master of the other ship.

Modern communications technology makes it possible to communicate in more detail than with whistles. Ships over a certain size are now required to monitor a particular radio frequency, so two masters can converse

directly to clarify what each will do. As the NRC report notes, "necessary arrangements for safe interactions normally are coordinated by radio" (1994, p. 53). In other words, all of step 3—detecting and resolving a conflict—might be done by direct communication, with the rules as a fall back. Even with the radios, however, there no guarantee that the relevant ships will communicate. For example, ships sometimes broadcast their intentions "in the blind", i.e., without knowing if others will hear them (1994, p. 54). Different ships may use different frequencies. As well, the report also notes that, "difficulties can arise when it becomes necessary to communicate in greater detail than can be accommodated through basic conning commands" and "no common language has been adopted" (1994, p. 48).

Second, ships can use radar to "see" some other ships and plot their courses, especially in fog. However, radar has limitations. It will not pick up all vessels (or icebergs), nor work through heavy precipitation. In particular, returns from low-lying vessels can be blocked by waves. Furthermore, a radar display can be misinterpreted, leading to what are called "radar-induced collisions" (Phillips-Birt 1971, p. 302). For example, with radar it is more difficult to determine the heading of a ship dead ahead. On the radar scope, such a ship will appear to be closing from ahead, but the display can be interpreted in two ways: as a ship moving in the opposite direction, closing head-on or as a slower-moving ship on the same heading being overtaken. Visually, these two situations can be distinguished by observing the position of the ship's sidelights. Such a misinterpretation was responsible for the collision in 1956 of the *Andrea Doria* and the *Stockholm*. The two were approaching nearly head-on when the master of the *Andrea Doria* misinterpreted the radar as indicating a passing situation and in trying to widen the separation turned into the path of the *Stockholm* (Phillips-Birt 1971, p. 302–3).

Summary

To summarize, collision avoidance depends on see-and-avoid, with a set of rules to give priority to one ship or another in various situations based on relative positions. As well, the rules ensure that ships behave

[3] If the burdened ship does not act in time, the privileged ship must still take whatever action is necessary to avoid the collision.

predictably in situations of potential conflict. There is a minor use of dedicated resources for some traffic, similar to lanes in automobile traffic, and of explicit allocation of shared resources by a manager.

Aircraft

In this section, we will discuss air traffic and resource allocation mechanisms associated with air traffic control (ATC). Aircraft differ from the vehicles in the previous sections because, first, they can maneuver in three directions, but second, they cannot remain aloft indefinitely or easily slow down or stop, unlike all of the other vehicles mentioned. ATC is also interesting because it has in some ways the most highly developed set of coordination mechanisms. In low-traffic situations, the main principle for separation is "see-and-avoid". In high-traffic situations, see-and-avoid is augmented or even replaced with direction from a central controller working as part of the air traffic control (ATC) system. In part, space is still used first-come-first-served, although particular spaces are reserved for particular types of operations. However, the detection and resolutions of conflicts is done by a central manager rather than in a distributed fashion.

See-and-avoid—Visual flight rules

When aviation started in the early 1920's, there was no air traffic system to speak of (Nolan 1990, pp. 2–4) and collision avoidance relied on see-and-avoid, as shown in Figure 13. As the Transportation Safety Board of Canada puts it: "For the see-and-avoid principle to be effective, it is necessary that a pilot be able to detect aircraft by visual means, recognize collision geometry based on visual cues, and react correctly, and in sufficient time, to avoid a mid-air collision" (1995). Since airplanes in this early era did not fly at night or in weather where visual cues were not available, "see-and-avoid" techniques were always feasible.

See-and-avoid for aircraft is basically the same as see-and-avoid for automobiles and ships. Airspace is allocated first-come-first-serve with conflicts resolved using simple heuristics. Aircraft converging head-on each divert to their right to avoid conflict. To reduce conflicts, opposite direction traffic flies at different altitudes—odd multiples of 1000' eastbound and even multiples westbound—a dedicated resource assignment similar to traffic lanes, as shown in Figure 14.
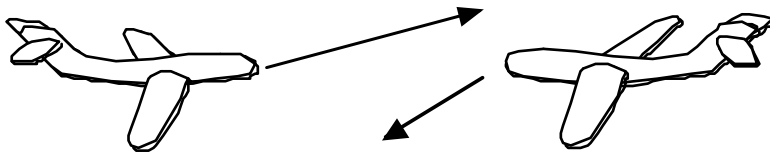

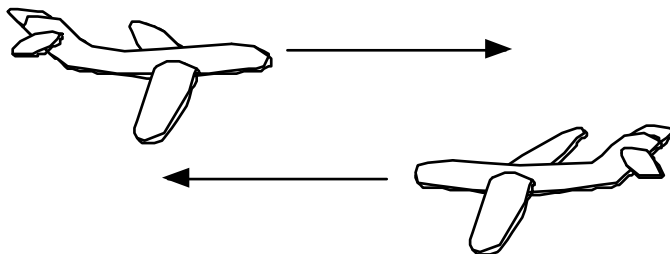**Figure 13. See-and-avoid in aviation traffic.**

Figure 14. Dedicated resources (opposite direction traffic at different altitudes) reduce conflicts in aviation traffic.

See-and-avoid techniques form the basis of the visual flight rules (VFR). Even today, a large percentage of flights are uncontrolled, meaning that pilots are expected to follow these rules and take responsibility for their own navigation and separation. VFR has the advantage of low cost since no additional coordinators are needed, as for automobile traffic. In most cases, it is not even necessary for the pilots to be in radio contact, although radio self-announce procedures have been developed to improve coordination by augmenting "see" with "hear". However, as traffic increases both seeing and avoiding become more difficult and different mechanisms are needed.

Allocation of runways

The first resources to be explicitly managed were runways. For safety reasons, only one aircraft may use a runway at a given time, causing potential resource conflicts when many aircraft want to land or takeoff. While en route, aircraft have a vast amount of available airspace to work with, but planes converge at airports and on the few available runways. Even today, runways are the bottleneck in the air traffic system.

At uncontrolled airports (the vast majority), aircraft wanting to land descend to a standard altitude, enter at a standard point into the "traffic pattern" (a rectangular circuit around the airport, leading to the runway) and thus join a queue for the runway. In other words, the runway is allocated first-come-first-served for landings. The use of the pattern funnels possibly conflicting traffic into a predictable path, as with lanes in other transportation systems, thus making it easier to see potentially conflicting traffic. Departing aircraft simply wait for an open spot in the pattern before taxiing on to the runway and taking off. This mechanism is summarized in Table 5. This system can break down if it is unclear in which direction (or on which runway, if there is more than one) an aircraft should land or takeoff. In other words, the distributed mechanism requires that all pilots agree in step 2 on which resource they are to use. Therefore, airports employ visual indicators of the preferred runway and direction or provide suggestions by radio. Nevertheless, disasters can happen if these indications are ignored. For example, a commuter airliner and a King Air aircraft collided on November 19, 1996 in Quincy, Illinois when the King Air started to take off on one runway while the commuter was landing on an intersecting runway. The system is also stressed if planes have very different flying speeds and overtake each other in the queue.

To manage the demand for runway time, the first air traffic controllers started work. A controller, standing near the arrival end of the runway, determined which aircraft would be allowed to land next or opened space in the pattern for aircraft to depart and signaled these decisions by waving flags. In other words, the distributed first-come-first-serve runway allocation mechanism was replaced by managerial decision making centralized in the controller.

Table 5. Summary of collision avoidance mechanisms for aviation traffic on runways.

Step	Aviation traffic on runways	Aviation traffic with controller
Step 1—Resource needed	Runway	Runway
Step 2—Resource available	Depends on wind; indicated by display on the ground	Depends on wind; communicated by controller
Step 3—Possible conflicts	Other aircraft in pattern to land or taking off, seen or heard on the radio	Other aircraft in pattern to land or taking off
Step 3a—Priority	Aircraft ahead or lower has priority	Determined by controller; communicated by flags, light gun or radio
Step 3b—Alternative resources	Continue on downwind leg or circle; wait, if on the ground	Continue on downwind leg or circle; wait, if on the ground

As well, the controller identified the direction for landing or takeoff, crucial since airports at that time were literally landing fields that could be used in any direction.

Currently, most aircraft are equipped with two-way radios, which are required for flights in busy areas. Each controller communicates on a particular frequency; pilots tune their radios as necessary to communicate with the appropriate controller. The controller issues instructions on this common channel to individual aircraft, telling pilots to turn to particular headings, climb or descend to particular altitudes, or placing restrictions on their operations (e.g., a minimum altitude or maximum speed) to prevent conflicts or keep the aircraft inside the controller's area of responsibility. To ensure mutual comprehension, the international language of air traffic control is English, and all pilots and controllers must be able to speak and understand a basic standardized vocabulary (unlike the situation in marine traffic).

When arriving aircraft are within approximately 5 miles of the runway, they are cleared for landing by the tower controller, who ensures that only one airplane will be on the runway at a time, as shown in Figure 15. Departing aircraft are similarly cleared on to the runway to take off. If two aircraft are in jeopardy of occupying the same runway, the local controller remedies the situation by deviating one or both of the aircraft. For example, if when an aircraft is landing the previous aircraft is still on the runway, the approaching aircraft will be told to "go around" rather than land (of course, the pilot should notice such a conflict and reach the same conclusion independently). This mechanism is summarized in Table 5. Special rules have been developed to ensure separation of aircraft using intersecting or parallel runways, where nearly simultaneous operations are possible. At particularly busy airports, landing and takeoff slots must be reserved in advance and in a few places can even be bought or sold.

ATC also controls ground movement at most large and medium-sized airports. The ground controller's primary responsibility is to assure that no two aircraft attempt to occupy a runway or portion of taxiway at the same time. When an airplane or other vehicles must cross an active runway (i.e., a runway used for takeoffs and landings) the ground controller must receive approval from the local controller. In some towers, this approval is indicated by a physical token for the runway; the tower controller hands the token to the ground controller, indicating that the ground controller has control of the runway and the tower controller can not issue take-off or landing clearances.
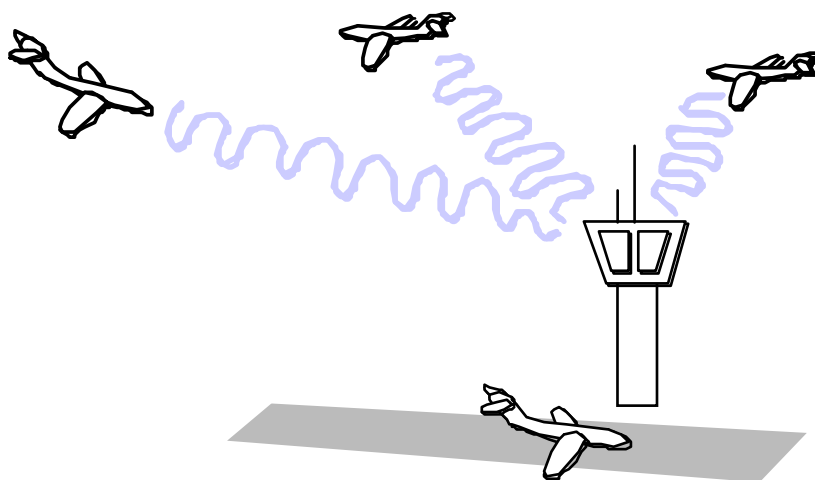

Figure 15. Explicit allocation of a scarce resource by a controller. Controllers track the position of aircraft from position reports and allocate airspace or runways to particular aircraft.

Allocation of en route airspace

Once an aircraft is airborne, it flies to its destination. The only means of navigation in the early days of aviation were visual references, which uncontrolled flights still use. The pilots of these aircraft are responsible for their own separation using see-and-avoid, as summarized in Table 6. However, pilots using instruments can fly in conditions where visual references are not available and where they cannot see other traffic. Even in good conditions, at high speeds it is impossible to see traffic far enough in advance to be able to be able to avoid a collision. Because pilots in these conditions cannot see to use see-and-avoid, a controller instead gather information and issues instructions to each pilot, in what is called controlled flight.

All controlled flights file a flight plan indicating the route and altitude to be flown, thus informing the controllers of their intentions. As the flight continues, its progress is tracked against the plan. The original air traffic system tracked flights manually. As the flight progressed, controllers tracked the position of each flight on a map of the area. Estimated positions (calculated from the flight plan) were updated with position reports radioed in from each aircraft. If a potential conflict was detected, the controller would attempt to contact the pilot by telephoning a radio station near the estimated position. During good weather, the controller would only advise the pilot of a possible conflict; in bad weather, the controllers would issue instructions to ensure the separation. As aircraft became faster, it was impossible to see oncoming traffic in time, making controlled flight necessary even in good weather.

Because of the inaccuracies in position reports, controlled traffic was separated by at least 10 minutes or about 50 or more miles. Since the position of the aircraft was uncertain, a large amount of space was allocated to ensure that there would be no overlap. As traffic grew, the system reached capacity, causing delays. Some pilots would choose to fly in uncontrolled areas, avoiding the delays, but taking responsibility for their own separation. As well, uncontrolled non-commercial traffic would frequently operate in the same airspace. Both of these problems led to mid-air collisions, leading eventually to the decision that all flights above a certain altitude and in busy airspace should be controlled.

To solve the capacity problems, controllers started to use radar, first installed in 1956. Radar allowed controllers to directly observe the position of aircraft under their control, as shown in Figure 16, thus permitting closer spacing. As well, remote communications outlets were created, allowing controllers to communicate directly even with distant aircraft, thus providing quicker response to instructions for evasive actions.

Table 6. Summary of collision avoidance mechanisms for aviation traffic in en route airspace.

| Step	Aviation traffic	Aviation traffic with controller
Step 1—Resource needed	Airspace	Airspace
Step 2—Resource available	Airspace ahead at appropriate altitude, within navigational capacity of aircraft	Airspace ahead at appropriate altitude, within navigational capacity of aircraft and under controller's authority
Step 3—Possible conflicts	Other aircraft on collision course; seen by pilot	Other aircraft on collision course; determined by controller from radar or radio position reports
Step 3a—Priority	Aircraft to right has priority; both deviate if head on	Determined by controller and communicated by radio
Step 3b—Alternative resources	Deviate to right	Deviate to right, climb or descend
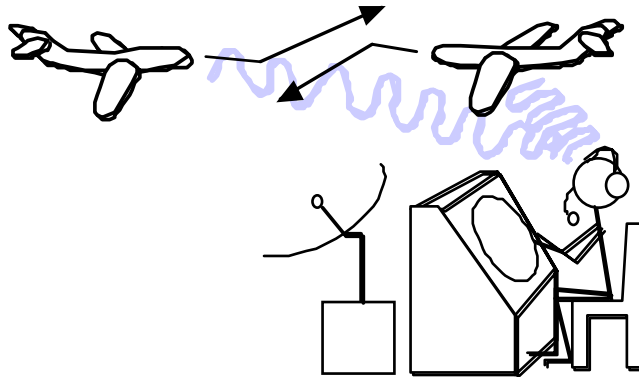
Figure 16. Explicit allocation of resources by a controller. Using radar, a controller can directly determine the position of aircraft and issue directions to avoid conflicts.

At first, these radar systems only covered high altitudes, but they were eventually extended to cover the areas around most busy airports, allowing most flights to be monitored from take-off to landing.

Controllers can also point out possibly conflicting traffic to uncontrolled VFR pilots, allowing the pilots to decide how to avoid. As the Transportation Safety Board of Canada notes, "a pilot who had been alerted to the presence of another aircraft was eight times more likely to see the aircraft than was a pilot who had not been alerted" (1995). Traffic on a collision course is particularly hard to see: because its relative position is not changing it will appear to be stationary against the sky. Interestingly, radar makes the controllers more important, because it increases the information asymmetry—the controller can see the location of all aircraft, but the information cannot be quickly conveyed to individual pilots.

Changes enabled by new technologies

Aircraft traffic continues to increase but there are also rapid changes in the technology available for managing the load. As technology makes new information exchange and provision possible, we expect corresponding changes in the coordination mechanisms. Changes in traffic control are closely related to changes in navigation technology. Currently, aircraft usually follow airways, which concentrate traffic at certain hot spots. Aircraft using newer navigation systems, such as gyroscopic positioning or GPS, can identify their position without the need for ground-based beacons. These aircraft are therefore not restricted to airways but can fly directly to their destination. Direct routes can be good for fuel efficiency and flying times or for avoiding bad weather. However, direct routing makes collision avoidance more complex for controllers, since aircraft can enter and leave an en-route sector at any point and might conflict anywhere, rather than at a few known hot spots. As well, since GPS is potentially more accurate than ground-based radio navigation, its use could have the effect of concentrating traffic flying the same route into a smaller area, increasing the chance of conflicts (Transportation Safety Board of Canada 1995)

A more recent development is Traffic Alert and Collision Avoidance System (TCAS), an on-board computer system used in large aircraft to monitor nearby aircraft and issues an alert and possible avoidance instructions if any come too close. Figure 18 shows an example display, indicating the relative positions of nearby aircraft. Possible conflicts are detected by TCAS as well as or instead of by a controller, as shown in Figure 17, and summarized in Table 7[4]. In the most

---

4 If the TCAS and the controller both notice a conflict, the pilot is supposed to follow the instructions of the TCAS system. A collision over Switzerland occurred in 2002 when one pilot obeyed the TCAS but the other obeyed the controller.

recent version, TCAS II, computers on the two conflicting aircraft can communicate to ensure that each selects complementary avoidance strategies (e.g., one climbs while the other descends).

Table 7. Summary of collision avoidance mechanisms for aviation traffic with TCAS.

| Step	Aviation traffic with TCAS
Step 1—Resource needed	Airspace
Step 2—Resource available	Airspace ahead, within navigational capacity of aircraft
Step 3—Possible conflicts	Other aircraft on collision course; determined by TCAS from transponder returns
Step 3a—Priority	Determined by TCAS and communicated to pilot on display
Step 3b—Alternative resources	Deviate to right, climb or descend
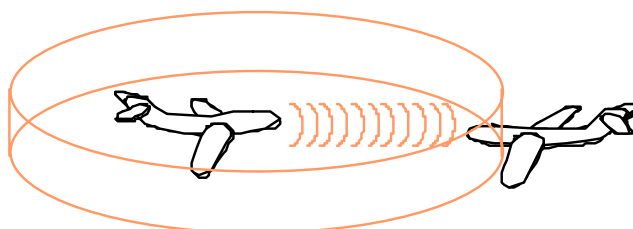
Figure 17. See-and-avoid with a collision avoidance system. Pilots can determine the position of conflicting traffic using equipment on the aircraft.
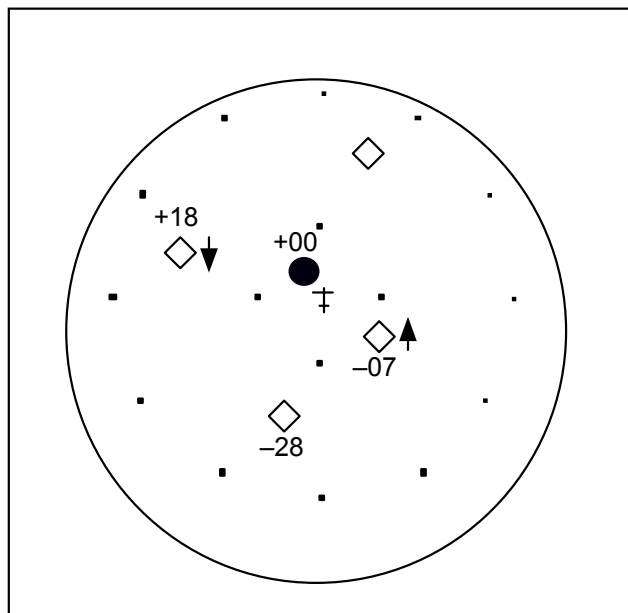

Figure 18. Example of a collision avoidance display, showing possibly conflicting traffic. The cross in the centre represents this aircraft; open diamonds are other aircraft with relative altitude and vertical direction; the closed circle just above and to the left of the cross indicates an aircraft within 30 seconds of collision.

The combination of direct navigation and traffic detection forms the basis of an initiative called "Free Flight", which would allow aircraft to use the equivalent of see-and-avoid even in instrument conditions, by using the technology to extend the range of "see" and to suggest appropriate actions to "avoid". Eventually, radar might be replaced or augmented by continuous broadcast of aircraft positions, as determined by on-board navigation equipment; such a system would allow all pilots to observe the position of all aircraft, information now available only to the controller.

Summary

Air traffic control displays a broad range of collision avoidance mechanisms. VFR flights use see-and-avoid, a distributed conflict-detection based resource allocation mechanism. Faster flights and flights in bad weather cannot see, so use space allocated by a controller. However, technological developments suggest the possibility of improving the ability of a pilot to "see" conflicting traffic, enabling a return to a distributed mechanism.

DISCUSSION

In this section, we will briefly summarize the mechanisms discussed above then discuss similarities and differences among the various systems.

Summary of mechanisms

Looking across the various tables, we can see similar problems arising in all four systems and a small number of mechanisms used across systems to address them. At some level, all four systems dedicate resources to reduce conflicts. For example, confining automobile traffic to lanes means that drivers need to worry about crossing traffic only at intersections, not at every point. Some systems pre-allocate shared resource in advance. For example, train traffic is carefully scheduled on to tracks to avoid other trains. Finally, all use some variant of see-and-avoid.

Similarities

The various control systems discussed above effectively manage the allocation of a scarce resource, namely space, to possibly conflicting uses, namely vehicles. In so doing, these mechanisms address several common issues.

Distributed yet reliable

First, transportation systems are particularly interesting because they work in a distributed fashion for the most part yet are highly reliable.

Managing infinitely divisible resources

Second, a key part of the system is the various strategies for managing an infinitely divisible resource like space. In some parts of the system, the space is divided into blocks used by one vehicle at a time (e.g., runways or stretches of track), but for the most part, space is managed by providing separation rather than ownership of a block (e.g., en route control or automobile traffic).

While few organizations manage space in quite this way, common infinitely divisible resources include time, either of people or equipment (or for that matter, of use of a chunk of space, as defined above) and money. In most cases, time is managed by breaking it into number of standard sized units and allocating those, e.g., scheduling a conference room or piece of equipment by the hour. Collision avoidance procedures suggest that the possibility of instead providing separation around the actual use of the resource. For example, tables in a restaurant are often managed by allocating them to patrons as they arrive. Such an approach might result in greater efficiency, but has the cost of communicating and planning with arbitrary times and possible fragmentation of the available time blocks.

Meta-control

Third, many transportation systems use some kind of meta-control to ensure that the amount of conflicting traffic is low enough that simple distributed collision avoidance schemes are practical. Each coordinator is viewed as a resource itself, and the system managed to ensure that none of the controllers is overloaded. For example, all systems pre-allocate some space to reduce the number of possible conflicts sufficiently for see-and-avoid to be practical. In some cases, capacity

is explicitly managed by preventing vehicles from entering the system if that would cause congestion. In an organizational context, the total load on a resource might be kept intentionally at some fraction of its theoretical capacity. The cost of this reduced output would be offset by the savings from easier scheduling of the use of the machine.

Common training and certification

Fourth, in many cases conflicts are resolved in a distributed fashion, as each operator takes a pre-agreed action. Note that these agreements work because all operators are similarly trained to act predictably. This view is consistent with Pfeffer's (1978) description of socialization as a coordination mechanism. In most of the transportation systems described, operators must be certified: a driver's license is required to drive a car, a pilot's license to fly, etc. A significant part of the training for these licenses is learning the conventions and expected behaviours (e.g., stopping at stop signs, flying a traffic pattern, etc.) as well as the common language (e.g., "see-and-avoid" or "master" of a ship). In other words, the choice of coordination mechanisms is due in part to social norms, reinforced by training and because they work. On the other hand, lack of standardized training is considered a serious obstacle to the introduction of more formal traffic control in marine traffic (NRC 1994, p. 200). Most organizations similarly require some kind of training in how resources are to be allocated, although it is usually not provided as formally.

Role of technology in the evolution of the systems

Fifth, in all systems the evolution of collision avoidance schemes provides an interesting example of the way technology can change the tradeoffs between coordination mechanisms. For example, in air traffic control, no communications was possible initially, so see-and-avoid was the only feasible allocation mechanism. The use of radar and radio made it possible for the controller to observe positions and tell pilots what to do, resulting in a centralization of control. Most recently, increased computer power and communications makes decentralization again possible, with

technologically enhanced see-and-avoid. Many railroads have been able to eliminate double tracks by better coordination of a single track.

More generally, technology has been used to augment both seeing and avoiding. For example, radar and railroad signals both make it possible for an operator to detect the presence of traffic that's beyond the range of human vision. Advanced cruise control and TCAS make it possible to react to such information more quickly and thus avoid a collision. As well, better communications technology allows information to be shared more easily. For example, a controller can communicate traffic information or instructions to a pilot; ships' masters can talk directly to determine intentions and work out maneuvers.

Common frame of reference

Finally, actors can determine what others will do in part because they share a common frame of reference that determines priority and actions. In other words, because gravity provides a common direction, everyone agrees on right and left and therefore who has priority and which way to turn.

Note that using gravity as a reference means up and down are the same for everyone, which limits avoidance actions to turns. However, in close quarters situation for aircraft, turning actually increases the chance of a collision, because turning requires raising a wing, which increases the cross-section of the target. The TSBC notes, that, "once the aircraft are inside the range of approximately 10 seconds to impact, the pilot should employ an altitude change only" (Transportation Safety Board of Canada 1995). Unfortunately, there is no way for pilots to decide who should climb, dive or maintain altitude without communications, which is only possible in such a short time with automated systems such as TCAS. Given the need for a common frame of reference, it is interesting to consider how collision avoidance might be done in deep space, where there is no gravity. Such a situation is shown in Figure 19; these two ships, approaching head-on but upside-down with respect to each other, will collide if they both turn to their right. Unfortunately, with no

common frame of reference, there is no way to turn that will guarantee avoiding the collision.

In an organizational context, these procedures suggest the need for ways to create a common frame of reference in order to establish priority and make distributed resource allocation effective (Crowston and Kammerer 1998). For example, if all employees understand how their activities add value for customers (and which customers and which needs are most important), they might be able to such evaluations independently.

Differences

Despite the similarities discussed above, there are many differences between the various systems.

Language

First, each system has unique language for describing resources and methods. For example, see-and-avoid is an aviation term; car drivers do not use the term, nor do ship masters. One advantage of our analysis is that the framework helps clarify which points are common and which are unique.

Locus of responsibility

Second, the systems assign responsibility for separation differently. Air traffic control and trains rely on controllers for separation (although the pilot is still ultimately responsible), while for ships and automobiles, the operators are directly responsible. Both trains and aircraft have centralized collision avoidance systems because of problems being able to see far enough ahead in order to avoid collisions. In the case of trains, the problem is the long distances needed to stop and the inability to maneuver. In the case of planes, the problem is the need to operate in clouds as well as at high speeds. There are some Vessel Traffic Services in marine operations, which attempt to provide additional information, but these generally do not attempt to replace the master's control.

Formality of mechanisms

A related difference is in the formality of the coordinating mechanisms. In automobile traffic, the mechanism may be as simple as a casual glance to see that the lane is clear, while aviation traffic requires requesting and receiving permission from a controller. However, even within a particular system, different mechanisms may be appropriate at different times. For example, when traffic is low, pilots may request a block clearance, which allows them to move freely within a range of altitudes, e.g., to avoid turbulence.

Cost

A final difference is that the mechanisms exhibit a wide range of cost and performance. Various kinds of cost can be distinguished: the cost of establishing and running the system, and the cost incurred by each vehicle as it progresses. On the one extreme, the distributed system used for automobiles has few central controllers and generally low overhead. On the other, the centralized system for controlling aircraft has a very high cost.
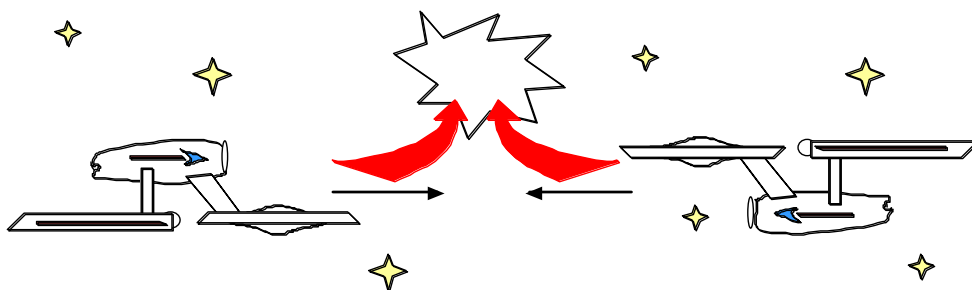

Figure 19. Collision avoidance without a shared frame of reference. Because one is upside-down with respect to the other, these two starships approaching head-on will not avoid collision if both turn to their right.

Performance

Conversely, the systems provide different levels of performance. Performance can be considered at different levels. From the point of view of an individual vehicle operator, the goal of a collision avoidance system is obviously to avoid collisions and the relevant performance measure is how many collisions occur[5]. At this level it is difficult to assess performance of collision avoidance schemes because actual failures of the systems (i.e., collisions) are relatively rare and there is little data about near misses or even non-fatal accidents for modes of transportation other than aviation. As well, differences in actual accident rates likely depend on a multitude of factors such as traffic density, further complicating the comparison. However, having said all that, automobile traffic collision avoidance is probably least reliable and aviation, most.

Performance can alternately be viewed from the point of view of a system designer, concerned with maximizing the capacity of the system given an acceptable probability of collision. More sophisticated systems allow a greater number of vehicles to operate without increasing the chances of a collision. For example, as the technology on ships improves, ships can avoid collisions more easily, so the capacity of a harbour or waterway increases. Similarly, proposals to increase the capacity of highways have included enhanced cruise control and collision avoidance systems for cars.

Discussion of similarities and differences

The comparison summarized above raises the question of why these differences? For example, why are aircraft rigidly controlled while automobile traffic is not, even though the underlying technology is more flexible, the available airspace around an airport much bigger and the number of vehicles small compared to a stretch of highway?

Some reasons may be technical. An obvious difference is how flexible and controllable the vehicles are. Trains and automobiles are both rigid because they can only go where the road goes, thus creating contention for the limited amount of roadway, while aircraft and ships are not similarly limited except in tight quarters, and therefore can resolve conflicts more flexibly. On the other hand, ships cannot regularly adhere to a precise path through the water (NRC 1994, p. 187) or stop under all circumstances (NRC 1994, p. 194) and aircraft cannot stop at all. These differences can be seen in steps 1 and 2 of our framework, where different resources are identified as needed and possible.

There are also differences in the level of performance demanded. For example, aviation is held to a much higher standard of reliability than automobile traffic. This difference might be because of the large number of people in the air and on the ground who would be affected by the collision of two jumbo jets. It may also be because few people fly themselves, and so demand controls on pilots they do not accept themselves as drivers. Finally, as the NRC notes, some difference may be due to history. Aviation developed relatively recently and makes use of newer technologies, while "marine operations are steeped in tradition and are highly fragmented from a systems perspective, affecting acceptance of technological change" (NRC 1994, p. 187). Coordination theory does not explain these differences, but it does provide a framework in which to discuss them.

CONCLUSIONS

To conclude, we will analyze a different kind of organizations to suggest organizational resource allocation methods that might be useful in transportation and to explore the implications of transportation systems for other kinds of organizations.

Alternative approaches to coordinating resource allocation

The key claim of coordination theory is that having identified a dependency and associated coordination mechanism, new processes can be generated by considering alternative coordination mechanisms. For

---

[5] Of course, collisions might also be considered as another cost of the system.

example, there is interest in the marine community in the possibility of a traffic control system patterned after air traffic control (NRC 1994). In this section, we will consider how ideas from transportation systems could be used in another kind of resource allocation.

To parallel the analysis developed above, we will consider another NSNC resource, specifically how computer-support personnel might be managed. Such personnel are in high demand in many organizations. Furthermore, as organizations have implemented personal computer and client/server systems, they have distributed the need for such expertise. It is therefore interesting to consider how their time might be reliably managed in a distributed fashion. This analysis can be extended to the allocation of the time of other professionals or more generally the time of any NSNC resource, such as tools or equipment.

Dedicated resources. Following the analysis above, the first possibility is that we might allocate particular resources to different uses, thus reducing or eliminating the chance of conflict between those uses.

For computer support personnel, this approach would give each unit that requires computer support its own support person, the well-studied question of centralization or distribution of resources. Of course, there are a number of trade-offs to consider in deciding between distributed and centralized support, of which reducing contention is but one. For example, a centralized group might provide better monitoring of and career paths for technical personnel, factors that are irrelevant to road space. However, shared resources also allow for load balancing between units if the level of help needed fluctuates. Moving back to the transport domain, the analogy would be to lanes that reverse directions depending on the volume of traffic in different directions.

Pre-allocation. The second approach discussed above for allocating resources is to schedule them. This approach is obviously applicable for computer support personnel as well. For example, routine maintenance, installations, etc. would likely be handled by making an appointment.

See-and-avoid. The final approach is some kind of dynamic allocation. Following the algorithm above, computer users first decide that they need some kind of technical assistance. They then determine who might help them, that is, what kinds of resources are available. Finally, they check if that person is free and ask them to come help if so. If not, they either wait, find someone else to ask or, if their problem takes precedence, interrupt the current task.

Because computer problem solving is more time consuming than driving through an intersection, few of the mechanisms discussed above seem directly applicable. For example, stop signs serve to allocate an intersection, but they would be cumbersome if the average car took 15 minutes to pass. One exception might be the use of controllers to explicitly allocate resources. For example, many organizations have a help desk, which acts as an initial screen for problem calls. If the problem required personal attention, the help desk personnel might direct the call to a local computer-support person if available or to a backup in a nearby division otherwise.

Technological support. As organizations implement communications technologies with higher capacities, the desirable coordination mechanisms will likely also change. For example, in many settings, the best way to tell if an expert is available is to call or visit in person. Active badge systems might allow individuals in need to help to quickly determine the status and whereabouts of different possible helpers, thus speeding up the search. If support-personnel could consult with users without having to be physically present (e.g., by using a screen sharing program), the time separation between jobs could be reduced, thus increasing productivity.

Summary

In this paper we have considered how space, a non-consumable non-shareable (NCNS) resource is allocated in transportation systems. Three approaches are used: dedicated resources for particular uses to reduce contention, pre-allocation via a schedule, or dynamic allocation, either with see-and-avoid or by a central controller. Interestingly, new technologies have allowed a shift from one

mode to another by improving an operator's ability to see and avoid conflicting traffic. These approaches were illustrated in case studies of car, train, ship and aircraft collision avoidance and seem to be applicable to other similar resources, such as the time of human experts.

ACKNOWLEDGEMENT

REFERENCES

Bedell, P.A., "BFGoodrich's Skywatch," in: *AOPA Pilot*, 1997, pp. 127–128.

Blythe, R., *Danger Ahead: The Dramatic Story of Railway Signalling,* Newman Neame, London, 1951, 132 pp.

Crowston, K., "A taxonomy of organizational dependencies and coordination mechanisms" in Malone, T.W., K. Crowston and G. Herman (eds.), *The Process Handbook*, MIT Press, Cambridge, MA, 2003, pp. 85–108.

Crowston, K., and E. Kammerer, "Coordination and collective mind in software requirements development," *IBM Systems Journal*, 1998, 37:2, pp. 227–245.

Malone, T.W., and K. Crowston, "The interdisciplinary study of coordination," *Computing Surveys*, 1994, 26:1, pp. 87–119.

National Research Council (U.S.). Committee on Advances in Navigation and Piloting, *Minding the Helm: Marine Navigation and Piloting,* National Academy, Washington, DC, 1994, 500 pp.

Nolan, M.S., *Fundamentals of Air Traffic Control,* Wadsworth, Belmont, CA, 1990, 560 pp.

Owen, W., E. Bowen, and the Editors of Life, *Wheels,* Time, New York, 1967, 200 pp.

Pfeffer, J., and G.R. Salancik, *The External Control of Organizations: A Resource Dependency Perspective,* Harper & Row, New York, 1978, 300 pp.

Phillips-Birt, D., *A History of Seamanship,* Doubleday, Garden City, NY, 1971, 319 pp.

Tate, W.H., *A Mariner's Guide to the Rules of the Road,* Naval Institute, Annapolis, MD, 1976, 135 pp.

Transportation Safety Board of Canada, "[No title]," M91C2004 (Available at http://www.tsb.gc.ca/en/reports/marine/earlier/m91c2004/m91c2004.asp).

Transportation Safety Board of Canada, "[No title]," A95H0008 (Available at http://www.tsb.gc.ca/en/reports/air/1995/a95h0008/a95h0008.asp).

Treat, J.R., N.S. Tumbas, S.T. McDonald, D. Shinar, R.D. Hume, R.E. Mayer, R.L. Stansifer, and N.J. Castellan, *Tri-level Study of the Causes of Traffic Accidents: Final Report, Executive Summary*, DOT/HS 805 099, Bloomington, IN: Institute for Research in Public Safety, 1979, 75 pp.

Tufte, E.R., *The Visual Display of Quantitative Information,* Graphics Press, Cheshire, CT, 1983, 197 pp.

Zwaneveld, P.J., *Railway Planning: Routing of Trains and Allocation of Passenger Lines*, PhD Thesis, Erasmus University Rotterdam, 1997, 194 pp.

AUTHOR

Kevin Crowston joined the School of Information Studies at Syracuse University in 1996. He received his A.B. (1984) in Applied Mathematics (Computer Science) from Harvard University and a Ph.D. (1991) in Information Technologies from the Sloan School of Management, Massachusetts Institute of Technology (MIT). Before moving to Syracuse he was a founding member of the Collaboratory for Research on Electronic Work at the University of Michigan and of the Centre for Coordination Science at MIT.

Professor Crowston has published articles and book chapters in the area of information systems and new organizational forms. His Ph.D. dissertation, Towards a Coordination Cookbook: Recipes for Multi-

agent Action, won the International Centre for Information Technology (ICIT) Thesis Prize for best dissertation in Information Systems in 1991 and was a runner-up for the International Conference on Information Systems thesis prize. He is a co-principal investigator on an NSF grant for "Tools for Inventing Organizations: Toward a Handbook of Organizational Processes".

His current research interests focus on new ways of organizing made possible by the use of information and communications technology. He approaches this issue in several ways: empirical studies of coordination-intensive processes in human organizations; theoretical characterizations of coordination problems and alternative methods for managing them; and design and empirical evaluation of new kinds of computer systems to support people working together. A specific example of the final interest is the application of document genre to the World-Wide Web.