Communications Constraint in the Air Traffic Control System

MATTHEW ROSENSHINE*
Cornell Aeronautical Laboratory Inc., Buffalo, N. Y.

An analysis of the communications constraint inherent in the present air traffic control system is accomplished by developing, validating, and exercising a Markov chain model of the communications in a terminal area control sector. The results obtained using the model indicate that the air traffic control system is already operating at, or very near, the operations rate constraint imposed by its communications capabilities. These results also indicate the increases in communications system capabilities which are required for increased instrument flight rules (IFR) operations rate capabilities if the requirements for communications remain similar to those currently in effect. It is pointed out that these requirements are quite severe, and a suggested alternative that relaxes the requirements for communications is discussed. The communications problem is particularly pressing because, although improvements and changes are occurring which may allow reduction in the required separation between aircraft (the other constraint on controlled operations rate), none appear to be emerging to relax the constraint imposed by communications limitations. If no improvements are made, it can be expected that increased traffic in the presence of the communication constraint will result in more service disruptions as long aircraft queues form and must be dissipated before normal terminal operation can be resumed.

A Markow Chain Model of Terminal Area Air Traffic Control Communications

AIR traffic control communications are quite frequent in terminal areas where traffic is relatively congested. The use of radar vectoring procedures along the relatively few approach and departure paths in use at a given time as well as the existence of holding stacks means that the message contents to successive aircraft will be largely repetitive, requiring little decision time on the part of the controllers but an appreciable amount of voice communications. The relatively small size of terminal area sectors necessitates additional voice communications to pilots to instruct them to change radio frequencies upon leaving one sector and entering another. In such a case, it is reasonable to measure the capacity of a sector manned by a single controller in terms of the controller's ability to transmit the required control messages to all the aircraft under his control, relinquish control over aircraft that have been serviced, and accept new aircraft for service as requested. In this case, the controller's capabilities can be assumed to be exceeded if the number of aircraft requiring service tends to grow beyond the airspace capacity for them. On the other hand, the controller's capabilities may be assumed equal to the task if he can reduce the number of aircraft requiring service without artificial constraints.

The pattern of arrivals requesting service is governed by some random process, as is the service time to each aircraft. The latter process is a bit more complex as it is also governed by the number of aircraft in the sector. This is true because each aircraft must be cleared in sequence to move up to a new position as the service to the "lead" aircraft is completed. (Visualize a stack of aircraft in which, when the bottom aircraft is cleared to land, all other aircraft must be notified in ascending altitude sequence that they are cleared to descend to a lower altitude and each aircraft must acknowledge beginning its descent before the next one in the stack may be notified.)

It would seem that, under the circumstances, the expected service time or time between releases of successive aircraft from the sector should be proportional to the number of aircraft in the sector. This conclusion is substantiated by data analysis for many different sector types.[1] The sector controller may then be viewed as a stochastic server and the number of aircraft in the sector as a stochastic process $i(t)$, where t represents time. Under the assumption that the arrivals of aircraft constiute a Poisson process, the observation of the stochastic process at a sequence of discrete times t_n, $n = 1, 2, \ldots$, corresponding to the release of successive aircraft from the sector forms an imbedded Markow chain.[2] This Markow chain consists entirely of transient states under quite natural conditions.

Validating the Communications Model

The process of validating the model consisted of deriving analytical results using what was felt to be an appropriate distribution of service (communications) times in conjunction with the model, and comparing these results with actual air traffic control data collected at the New York Air Route Traffic Control Center. Agreement between the theoretical and actual terminal area holding stack data is the basis on which the model is subsequently accepted as valid.

Theoretical Calculations

A wide class† of queue-size-dependent service time distributions may be defined by

$$b_{i,k}(t) = [(k\mu/i)^k/(k-1)!]t^{k-1}e^{-k\mu t/i}$$

where i = number of customers in the system, k = an integer (Erlang number when $i = 1$), and μ = service rate when there is one customer in the system. The possible variation of service time distribution implied is clear if it is noted that $b_{i,k}(t)$ may be interpreted as an Erlang-k density function with mean i/μ and the variance $i^2/k\mu^2$. The most interesting case, insofar as stacked aircraft are concerned, occurs when $k = i$. The function $b_{i,i}(t)$ is a gamma service time density that may also be interpreted as the sum of i independent

Presented as Paper 67-868 at the AIAA 4th Annual Meeting and Technical Display, Anaheim, Calif., October 23–27, 1967; submitted October 10, 1967.

* Principal Mathematician.

† The wide class is treated here because it involves no additional effort and its members can be used to answer additional questions. Only $b_{i,i}(t)$ is used in validating the model.

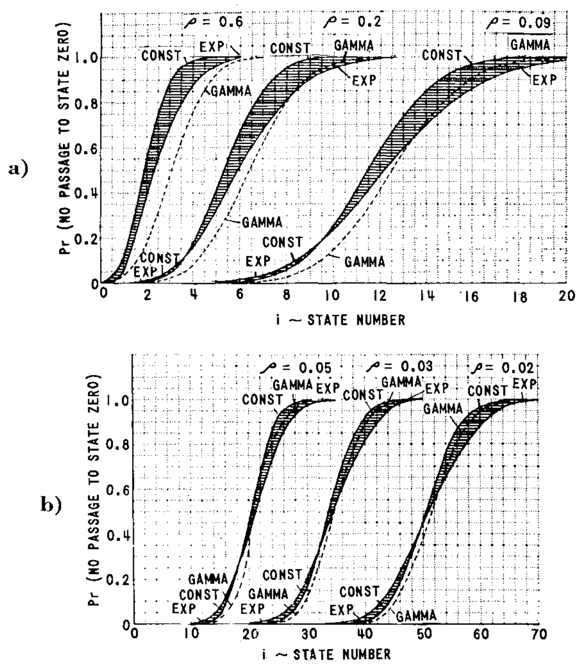

Fig. 1 Probability of no passage to state zero from state *i* (infinite queues allowed).

identically distributed exponential random variables each with mean $1/\mu$. ($1/\mu$ is the average service time to a single aircraft when there is one aircraft in the sector.)

The Markow chain transition matrix associated with the aforementioned class of service time density functions consists of elements

$$p_{ij} = \begin{cases} \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^{j-i+1}}{(j-i+1)!} b_{i,k}(t)dt & \begin{cases} i \geq 1 \\ j \geq i-1 \end{cases} \\ p_{1j} & i = 0 \\ 0 & \text{otherwise} \end{cases}$$

where λ = mean rate at which arrivals occur. These elements are interpreted as the probability of $j - i + 1$ arrivals between the completion of two successive services given that there are i customers in the system at the time of completion of the first of these two services. Evaluating the integral and replacing λ/μ by ρ yields

$$p_{ij} = \frac{(k+j-i)!}{(k-1)!(j-i+1)!} \frac{(k/i)^k \rho^{j-i+1}}{[(k/i)+\rho]^{k+j-i+1}} \quad \begin{cases} i \geq 1 \\ j \geq i-1 \end{cases}$$

For $i = 0$, $p_{0j} = p_{1j}$ because, when there are no customers in the system, it is necessary to wait for one to arrive, then begin service on it, and count the number of customers waiting in the queue when its service is complete. For $i \geq 2$ and $j < i - 1$, all p_{ij}'s are equal to zero since the largest possible reduction in the number of customers between successive service completions is one.

The queueing system model of the terminal area control sector can be used to determine the probability that the queueing system can be emptied out in the presence of continued arrivals given that there are "now" a specified number of aircraft in the queueing system. Graphs of these probabilities are shown in Figs. 1a and b for the exponential, constant, and gamma service time distributions.‡ (State j indicates a queue size of j, including the customer being served.§)

A graphic presentation of queue size limits to be observed for various probabilities of emptying the stack in the presence of continued arrivals is shown in Fig. 2 as a function of the traffic intensity parameter ρ. This graph was constructed using results for the gamma distribution shown in Figs. 1. The data points corresponding to these results are indicated on Fig. 2.

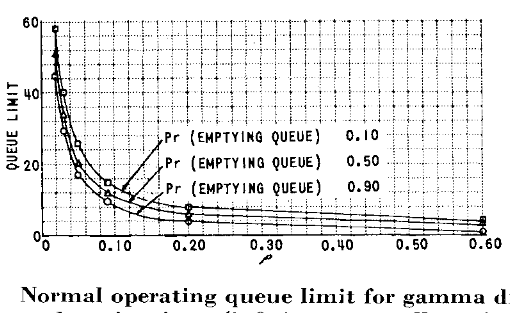

Fig. 2 Normal operating queue limit for gamma distributed service times (infinite queues allowed).

Actual Data Collection and Reduction

One of the criticisms to which collections of air traffic control communications data have been subjected is that the lengths of contacts between controllers and pilots are often a function of the time available for these contacts; when traffic is sparse, contacts tend to be long. This criticism can be circumvented by collecting data when the controller is very busy. Under such conditions, it is reasonable to assume that contact length is kept to a minimum. It was felt that some of the sectors within the jurisdiction of the New York Air Route Traffic Control Center (ARTCC) should offer the opportunity to collect communications data when the controller is busy with routine work during virtually the entire length of a relatively long time interval (perhaps 15 min). The routine nature of the work, i.e., all aircraft get similar, though not identical, instructions and requests, means that time for reflection is minimal and almost all of the time may be spent communicating. Communications data from such a sector may be used to estimate maximum channel utilization and average message lengths under the redundancies required by current air traffic control procedures with assurance that the controller and pilots are not including extraneous material in their messages as they might if time were available. (Control through the ARTCC means that all of the aircraft are operating under instrument flight rules.)

Conversations with Center personnel indicated that there were several sectors which would be satisfactory on the basis of the preceding considerations. A high-altitude (above 10,000 ft) transiton sector that holds aircraft for and feeds aircraft to Kennedy Approach Control was chosen for sampling because it represents part of the landing flow process in the terminal area. The sector is designated as sector 5B of the New York ARTCC. A map indicating the geographi-

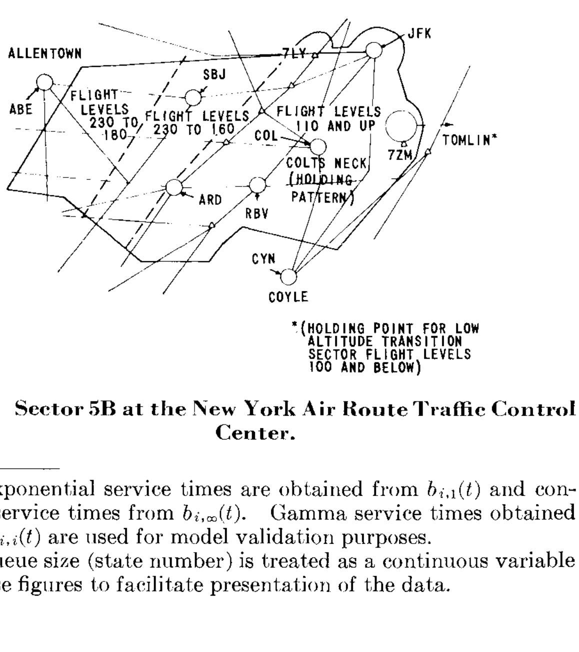

Fig. 3 Sector 5B at the New York Air Route Traffic Control Center.

‡ Exponential service times are obtained from $b_{i,1}(t)$ and constant service times from $b_{i,\infty}(t)$. Gamma service times obtained from $b_{i,i}(t)$ are used for model validation purposes.

§ Queue size (state number) is treated as a continuous variable in these figures to facilitate presentation of the data.

cal outline of the sector and the altitudes within the sector jurisdiction is given in Fig. 3. The major sector responsibility is to clear traffic from the south and west to Kennedy Approach Control. During periods when traffic is backed up from the Approach Control Sector, the sector controller controls traffic into and in the Colts Neck (at COL in Fig. 3) holding pattern.

A 1-min sample of the communications data is presented, along with 5-sec timing marks, in Table 1. The data were taken during two very busy periods of $13\frac{1}{2}$ min each so that each sample contains 162 5-sec intervals. Channel utilization was calculated based on utilization of these intervals. Most of the intervals were either entirely utilized or entirely unutilized. The remainder of the intervals were classified as fully utilized if they contained more than 2.5 sec of communications based on stop-watch timing. Intervals with less than 2.5 sec of voice content were considered entirely unutilized. The resulting channel utilizations are shown in Table 2 for the entire periods as well as for seven subperiods (six of 2 min each and one of $1\frac{1}{2}$ min) during each sample period.

Average contact¶ duration during the two sample periods was estimated by counting the number of contacts in each sample (54 in the first and 59 in the second) and using the relationship

$$\text{average contact time} = \frac{\text{channel utilization} \times \text{time length of sample}}{\text{number of contacts in sample}}$$

This resulted in average contact times of 10.8 and 9.6 sec. (A larger number of contacts recorded in the period with the lower channel utilization illustrates one of the limitations of contact duration as a measure. During one part of the second sample, the controller was getting information from several aircraft almost simultaneously, i.e., he was switching conversations between pilots very rapidly in a fashion almost analogous to a conference telephone call. However, the contact, by definition, terminates when the pilot involved in the conversation changes. Hence, a larger number of contacts are recorded than might ordinarily be employed to exchange the same information.)

The maximum rate at which aircraft can be accommodated on a single glide slope and runway under separation currently required at the glide slope is 36 aircraft per hour.[3] Assuming an arrival rate set at this value, values of ρ of 0.108

Table 1 One-minute sample of communications[a]

C	Northwest 16 radar contact. Hold southwest Colts Neck 2 33 radial. Right turns ↓ Approach time of 2 3 4 3.
P_1	Roger, hold southwest Colts Neck ↓ on 2 33 radial. Right turns. Approach time will be ↓ 2 3 4 3?
C	Northwest 16, affirmative ↓
P_2	American 48, out of 16 ↓
C	American 48, leaving 16, roger. Braniff 4 descend and maintain 1 ↓ 6000. Report leaving 17. Kennedy altimeter 2996.
P_3	2996. ↓ Braniff. We're out of 18 for 16. Will give you out of 17.
C	Roger ↓
P_4	TWA 14 out of 22.
C	TWA 14 leaving 22, ↓ roger.
P_5	TWA 172 level 14.
C	TWA 173, roger. ↓ United 8348 descend and maintain flight level 18 ↓ 0.
P_6	United 8348 descend to 180 leaving 190. ↓

[a] Legend: ↓ = 5-sec timing mark, C = controller, P_N = pilot N.

¶ A contact generally consisted of a call-up, a message transmission, and an acknowledgment. When an interruption terminated such a sequence, the completed portion was counted as a contact as was the interruption.

Table 2 Channel utilization measured during busy periods in sector 5B of the NYARTCC[a]

	Period I	Period II
Entire period	0.722	0.697
First 2 min	0.667	0.584
Second 2 min	0.667	0.833
Third 2 min	0.750	0.875
Fourth 2 min	0.708	0.791
Fifth 2 min	0.791	0.708
Sixth 2 min	0.708	0.600
Last $1\frac{1}{2}$ min	0.779	0.389

[a] Numbers represent fraction of time that channel is occupied.

and 0.096 are obtained for average contact times of 10.8 and 9.6 sec. The queue limits corresponding to these values of ρ are read from Fig. 2 as 7 and 8, if the probability of emptying the stack in the presence of continued arrivals is to remain at least 0.90.** These results agree very well with the observed maximum of seven in both cases.

The close agreement between the results of the theoretical calculations and the maxima observed in actual stack operations is the basis on which the model was accepted as valid. Having accepted the model as valid, theoretical data derived from the model may be used as a basis from which to infer some additional conclusions. This is done in the next section.

Effects of Communications Constraint on Future ATC Systems

An examination of Figs. 1 indicates that large increases in the queue limits cannot be obtained through changes in the type of distribution of message times. (Recall that interest is centered about cases in which the probability of emptying the queue in the presence of continuing arrivals is at least 0.90, i.e., probability of no passage to state zero is at most 0.10.) This means that, insofar as communications are concerned, if it is desired to increase the capacity of the system and still maintain the discipline of the current system, either fewer messages or shorter messages will be required. The effect of either of these changes is to reduce the value of ρ and, consequently, increase the queue limit (see Fig. 2).

In order to effect changes of the types suggested, the system of voice communications used in busy terminal areas will have to be altered. The precise mechanism by which such changes would be effected is by no means clear. However, it is easy to envision that high-speed automatic composition and data link transmission of messages would have the desired effect. (Note that high-speed composition would be required as well as high-speed transmission. The former requirement is often overlooked. The latter requirement, fulfilled in isolation, offers little improvement.) As devices with these capabilities will not be available in the immediate future, it is perhaps more useful to think in terms of reducing the number of messages required or reducing the redundancy required in the messages by automatically providing the controller with additional visual feedback about the behavior of the aircraft under his control. This would allow a reduction in the number of verbal contacts initiated primarily for checking purposes. It is recognized that equipment which would perform such tasks is neither easily come by nor inexpensive. They represent a new generation of equipment in the sense that they would perform some of the controller's checking activities for him as a routine matter and would bring to his attention situations in which desired objectives were not being achieved.

The potential of such devices may be evaluated by apply-

** These queue limits are for infinite capacity queues. Comparable results for finite capacity (set at $1.2/\rho$) stacks are 6 and 7

Table 3 Queue limits[a]

	System	
Sector	Current	Computer-aided
Approach control	5	12
Local control	10	17

[a] Values are approximately one less if only finite queues are allowed.

ing the model to results obtained in a Federal Aviation Agency study.[4] The basic objective of this real-time simulation study was to evaluate the general utility of digital computer techniques as an aid to terminal area air traffic control. Emphasis was placed on using digital computer calculations to enable the controller to bring an aircraft to the instrument landing system glide slope as close as possible to a previously assigned time. During this study the lengths of the individual messages on each channel were recorded. Interestingly enough, the message times recorded were not much different from those that had been recorded for the present system by Franklin Institute.[5] However, the channel utilizations recorded were considerably lower than those recorded in studies of the present system, even for arrival rates considerably in excess of those feasible under present separation requirements. It may thus be inferred that, although this computer-aided control system will not shorten message times, it will lighten the burden on the communications system by reducing the number of messages through computer transfer of routine information, reduction in the number of control commands required due to more accurate predictions of future aircraft positions, and a general reduction in required redundancy due to the capability of verifying correct reception of some messages by observing control displays.

A quantitative comparison of the effects of the computer-aided control system may be made using results presented in a recent air traffic control communications study.[6] Assuming, as before, an arrival rate of 36 aircraft per hour, the traffic intensity parameters for the approach control and local control sectors obtained for the present system are $\rho = 0.137$ and $\rho = 0.079$. For the computer-aided system the corresponding figures are $\rho = 0.070$ and $\rho = 0.049$. The queue limits obtained from Fig. 2 for these traffic intensities are tabulated in Table 3. The current approach control limit is approximately equal to the limits established in letters of agreement between some towers and the centers which they service.

There is a requirement for improving the air traffic control communications system capabilities if the traffic handling capabilities of the system in terminal areas are to be improved. This requirement follows from the controller's inability to keep up with demands for voice communications in busy terminal areas if the traffic flow is increased beyond current levels and communications requirements remain unchanged. The severity of this requirement accrues from the inability to effect such improvements by merely instituting procedural changes. The channel utilized in controlling the traffic in sector 5B of the New York Center was operated at a channel utilization of 80% over an 8-min period. Higher channel utilizations over similar time periods do not appear feasible. The conclusion, then, is that other ways must be found to transmit some of the information or information redundancy. A logical inference which can be drawn from this conclusion is that new equipment will be required to increase the communications capabilities of controllers in terminal area sectors. It is probably reasonable to introduce a design goal for such equipment and suggest that it be considered in making any decisions on the new equipment to be procured.

The limitation in the capacity of the communications system to accommodate aircraft is brought about by the necessity to impart instructions to each aircraft as if it were the only aircraft in the system, i.e., call up, transmit instructions, and receive acknowledgment for each move to a new "checkpoint." An opposite extreme, which may not be fully realizable but which presents a worthy goal, can be described as "one instruction for all aircraft," e.g., a clearance to the bottom aircraft in a stack is a signal for all other aircraft in the stack to commence descent to the next stack level. The effect of reaching such a goal would be the removal of the dependence of communications service time on the number of aircraft being controlled. This would provide an infinite queue limit, at least insofar as communications considerations are concerned. Granting that the suggested Utopian state of affairs may not be completely realizable, a design goal of one voice instruction followed by a series of data link transmissions may be a feasible substitute.

It has long been recognized that the speed of data links should be utilized in air traffic control. The repetitive and somewhat specialized nature of the communications in a terminal area may prove a very useful area in which to introduce this type of equipment. The single voice transmission would be to a "lead" aircraft but could be overheard by other aircraft in the sector. A data link acknowledgment would be followed by a data link clearance to the next aircraft in line to begin moving up to the lead position. Another data link acknowledgment would initiate a data link clearance to the next aircraft, etc. The data link could serve well in such repetitive contexts. The cooperating pilots would have to be made aware of what the positions in this arrangement are and their data link acknowledgments might include some information such as a stack level to indicate to the controller that they know where to proceed.

The need for progress in resolving the communications problem is particularly pressing because, although improvements and changes are occurring which may allow reduction in required separation between aircraft, none appear to be emerging which would relax the operations rate constraint imposed by communications limitations. If no improvements are made, it can be expected that increased traffic in the presence of the communications constraint will result in more service disruptions as long aircraft queues form and must be dissipated before normal terminal operation can be resumed.

References

[1] Smith, M. et al., "Human Factors Analysis of Voice Communications Practices in Air Traffic Control, Volume I," Rept. TM 339-84, June 1960, Convair/Pomona.

[2] Rosenshine, M., "Queues with State-Dependent Service Times," *Transportation Research*, Vol. 1, No. 2, 1967.

[3] Blumstein, A., "An Analytical Investigation of Airport Capacity," Rept. TA-1358-G-1, June 1960, Cornell Aeronautical Lab.

[4] Halverson, A. G. et al., "Investigation of Advanced Sequencing and Control Concepts in an Automated Terminal Environment, Volume I—Simulation Studies," Final Memo. Rept. 101-200 R, April 1963, Federal Aviation Agency.

[5] Felton, W. W., Fritz, E., and Grier, G. W., "Communications Measurements at the Langley Air Force Base During November and December, 1951," F-2276-1, Nov. 1952, The Franklin Institute Laboratories for Research and Development.

[6] "Future Air-Ground-Air Communication Subsystem Investigation," July 1966, Communication and Systems Inc.