THE SIMULATION OF PASSENGER MOVEMENTS THROUGH A TRANSIT STATION

Peter A. Fausch,[1] David Dillard,[1]

and John F. Hoffmeister.[2]

[1] Barton-Aschman Associates, Inc.

[2] Peat, Marwick, Mitchell & Co.

ABSTRACT

The simulation model being developed is a discrete-system, event-oriented representation of the movement of individual passengers through a transit station. The model is stochastic in nature with the entities of the system being the individual passengers whose movement through various activities in the station give rise to the events which drive the simulation.

Station activity subsystems, such as the ticketing areas, passenger movement areas and platform, are represented by links, nodes, and areas.

The outputs from the model are Time Impedance Measures (e.g., Walking Times, Time spent in queue, and Total In-Station Time) and Occupancy or Density Levels (sq. ft. per person) in the movement and queueing areas of the station.

The model being developed is part of the New Systems Requirements Analysis Program of the Urban Mass Transportation Administration.

ACKNOWLEDGEMENTS

The material for this paper was developed from the following documents prepared for the U.S. Department of Transportation, Urban Mass Transportation Administration, Office of Research, Development and Demonstration.

1. General Functional Specifications for a Transit Station Simulation Model, prepared by Barton-Aschman Associates, Inc. and Peat, Marwick, Mitchell & Co., November 3, 1972.

2. Detailed Technical Specifications for a Transit Station Simulation Model, prepared by Barton-Aschman Associates, Inc. and Peat, Marwick, Mitchell & Co., July 1973.

The work was prepared under the direction of Dr. Robert B. Dial, Chief of the New Systems Requirements Analysis Program.

I. OBJECTIVES OF THE STATION SIMULATION MODEL

There is no universally accepted list of station design objectives. Walking time, waiting time, total in-station times, and space standards per person are among some of the more important variables used in connection with determining design objectives.

The principal value of the simulation model in the design process will be to evaluate specific station layouts by generating variables of the type described above. More specifically, the simulation model will determine four basic types of design data for a given layout:

1. The walking times, time in queue, and total in-station times for individuals, and individual paths for specific movement areas or for the entire station.

2. The pedestrian occupancy (sq. ft. provided per person) in specified areas of the station.

3. The utilization, defined as the fraction (or percentage) of time some entity (e.g., a fare collection gate) is engaged.

4. The distribution of these variables for comparison against design standards or levels of service standards.

The simulation model can also aid in laying out and visualizing space relationships in a transit station. In this sense, it can serve as a design tool, as well as an evaluation tool.

II. STRUCTURE OF THE SIMULATION MODEL

The first step in developing the simulation model was to identify the station system. Figure 1 illustrates the pedestrian and vehicle flow patterns in a typical station system including the station system boundary with its environment (i.e., areas external to the simulation), general activity areas, and specific functional areas (i.e., free area and paid areas).

The station system, illustrated in Figure 1, suggested that the first step in modeling the station system was to subdivide the system into a series of sub-systems where each sub-system represents a basic type of functional activity. The blocks in the flow diagram actually identify these functional sub-areas (e.g., fare collection areas).

The second step is to identify the system entities, attributes, activities, and events that occur in each of these sub-systems. Table 1 describes station system and subsystem terminology.

Within the station sub-systems there are two basic entities -- persons and vehicles. The attributes of each of these entities are of two types -- exogenous and endogenous. The exogenous attributes are properties or parameters which the person or vehicle carries with him into the station system which are not subject to change (e.g., desired walk speed, access mode, egress mode, handicapped status, etc.). Endogenous variables are subject to change within the system (e.g., waiting time, time in queue, in-system time, etc.).

The dynamic feature of the station system model is the activities that occur within each functional area of the station -- that is, the activity within the area performed by persons and vehicles. The genesis of an activity is an "event" which may start or stop an activity. An event triggers the changes in endogenous variables which describe the status of the system. The generation of events is the basic orientation of the model. Figure 2 describes the activity in a typical functional area of a station system in terms of entities, activities, and events.

The station activity sub-systems are represented in terms of links, nodes and areas. Figure 3 illustrates the node, link, and area convention. In this convention, the pedestrian flow pattern and associated areas in any functional area of a station system are represented by nodes which are dimensionless and represent queue devices, decision points and points where arrivals or departures are created or destroyed. A link or a node can be one-way or two-way depending on the area and pedestrian flows to be modeled.

The node and link convention illustrated in the example in Figure 3 and described further in Table 2 provides the framework for describing all of the important activities, events, and interactions within functional activity areas of any station system. The basic node and link concept provides the user the flexibility to add or combine links and thereby control the level of detail used to describe a station system. It also provides the framework to develop an efficient data processing technique utilizing proven techniques developed in other transportation models. Finally, by layint out the station in terms of functional areas, links, nodes, and associated movement areas, the user must "think through" the operation of the station. This is an effective and rigorous design and evaluation tool all by itself.

The node and link convention provides the physical description of the station system from which the system image is created. The system image is the set of numbers which describe the state of any system at any instant in time. Associated with the system image are the process routines and mathematical operations which change the system image.

There is a system iamge associated with each link in a station system. The system image for a link includes the following information.

1. The total number of persons in the area associated with the link.
2. The number of persons that are actually in queue at the "downstream" node of the link.
3. The number of persons "in movement" on the link.
4. The density of people (i.e., area per person) in the area associated with the movement area of the link.
5. The number of people in the queueing area of the "downstream" node.

The routines which create the system image produce a set of time statistics. This information includes:

1. The total time spent in the area represented by the length of the link (referred to as in-system time where in this case the system in one link).
2. The time spent in queue at the downstream node.
3. The walk time (other than in queue).

Path Choice Model

The procedure chosen to model individual passenger path choice through a station can be though of as a modified, continuous parameter, dynamic Markov chain where the transition probabilities (i.e., probabilities of selecting alternate paths at a decision point) from node to node within the station are dynamically updated as a function of congestion effects within the station. Since a given passenger must reach a specific destination node, corresponding to a specified line or mode destination within the station, it is a dynamic Markov chain with absorbing states, where these latter states correspond to destination nodes.

The path choice model involves two basic types of computations which can be broken into (1) computations before the simulation begins, and (2) computations during the simulation. The computations before the simulation establish preliminary values for network characteristics and they reflect the behavioral assumption that the passenger in the station has a basic familiarity with the station layout although he can't see the congestion more than one link ahead. The computations performed during the simulation reflect the congestion effects one link ahead of the passenger.

Thus, the path choice algorithm models the non-optimal behavior of passengers within the station (i.e., not all passengers choose the shortest path from their origin to their destination within the station). It also models the probability of selecting alternate paths at a decision point (i.e., transition probabilities from node to node), within the station as characteristic of a homogeneous dynamic Markov chain, while minimizing the amount of user-specified input needed, but allowing the user the flexibility to specify input

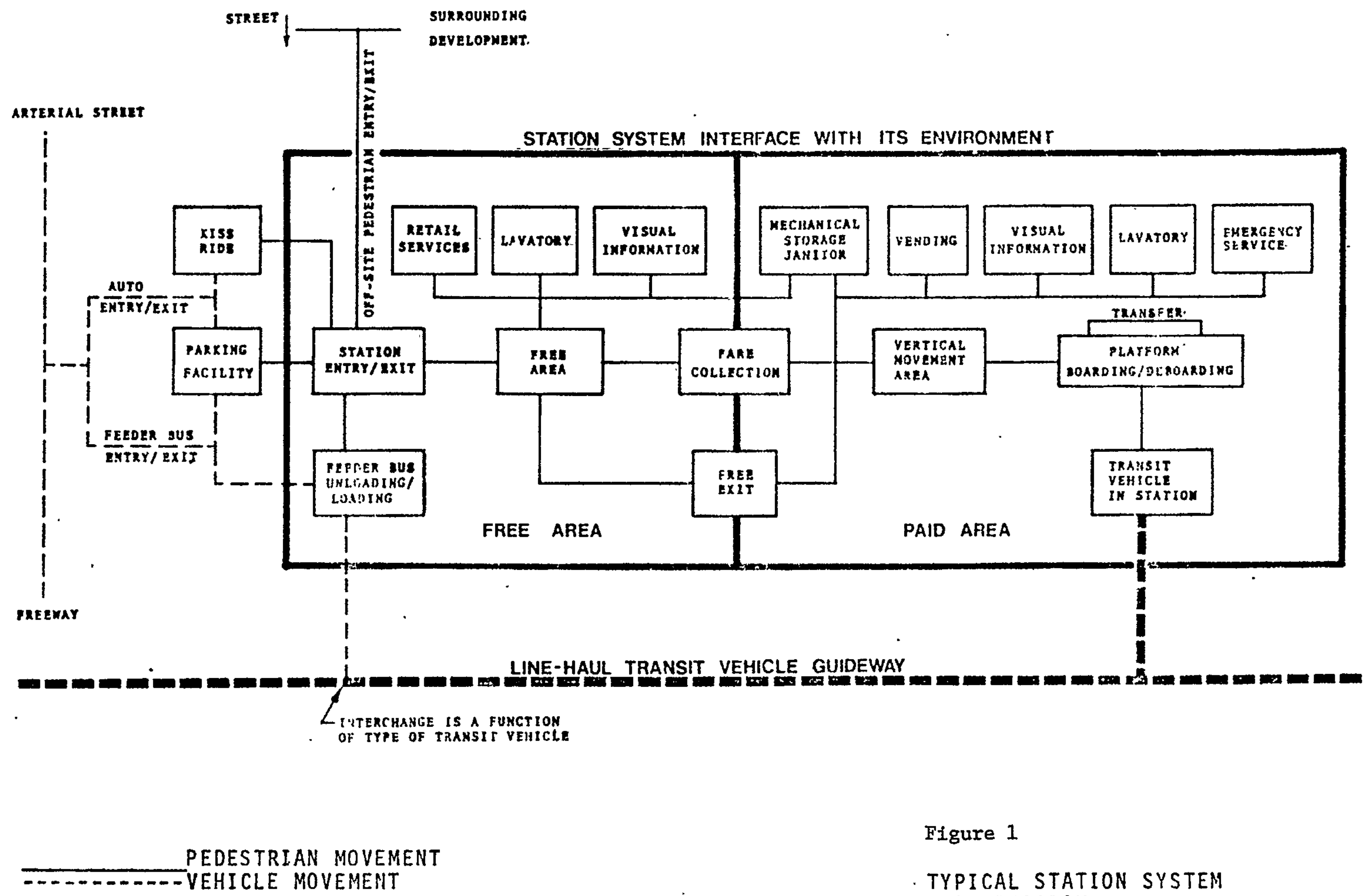

Figure 1

TYPICAL STATION SYSTEM FLOW DIAGRAM

TABLE 1

SYSTEM CONCEPT DESCRIPTION

System or Sub-system	Entities	Examples	Attributes of Entities: Exogenous	Attributes of Entities: Endogenous	Types of Activities and Events Performed by Entities
Transit Station	Activity Sub-systems	entrance area, fare collection area, stairs/escalator area platform area ancillary area	type linkage function configuration boundaries length, queue shape influence, queue discipline	number of people in module, number of people in queue, space provided per person	define specific functional areas within station system
Activity Sub-systems	People	old people young people handicapped people people with slow walk speeds, people with fast walk speeds	desired walk speed, line haul route, destination, handicapped status	path choice, walk time, total in-system time, time in queue	arrival in activity area movement in activity area waiting in queue area "serviced" by queue device departure from queue device slow down other people
	Vehicles	line haul transit vehicle, feeder bus vehicle, kiss and ride vehicle, park and ride vehicle, elevator	mode, line, number of doors, number of persons deboarding	door open time, boarding time required, deboarding time required, departure time, total, in-system time	vehicle arrives, vehicle stops, doors open; people get out, people get in, doors close, vehicle leaves

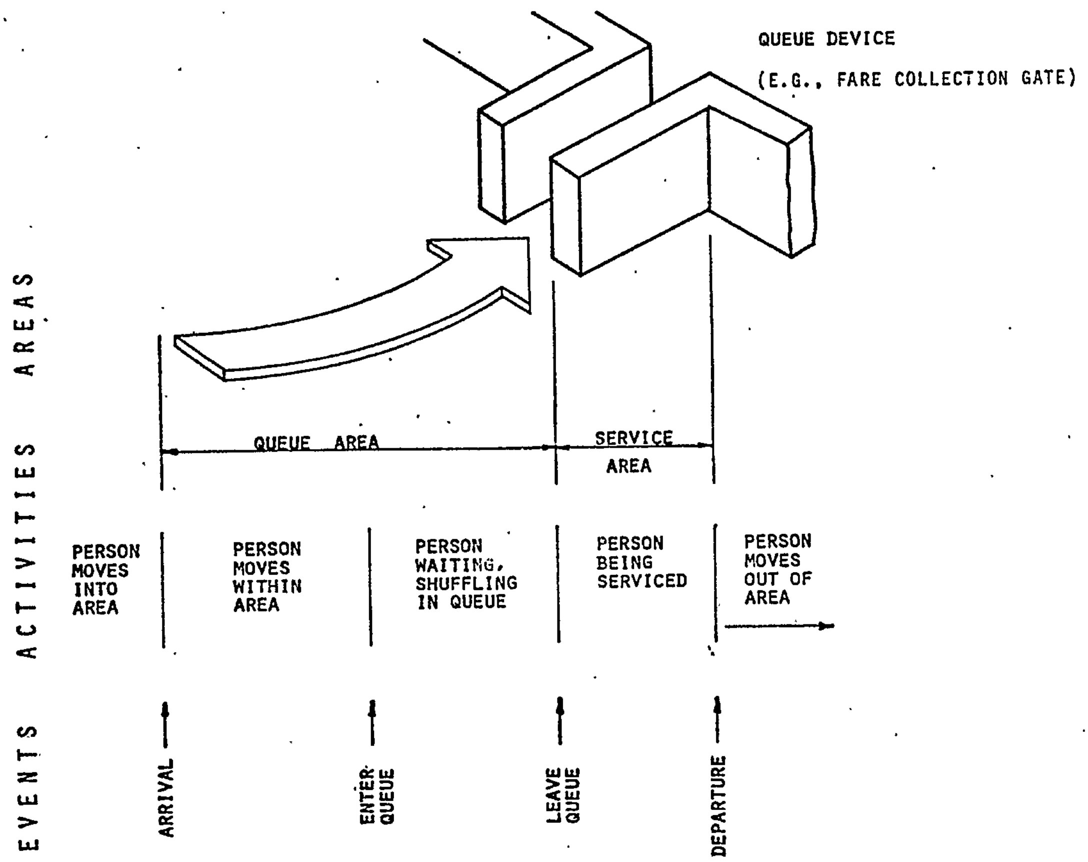

FIGURE 2

RELATIONSHIP OF ENTITIES, ACTIVITIES, AND EVENTS

TABLE 2

NODE AND LINK ATTRIBUTES

	Exogenous	Endogenous
Node	-description (e.g., fare collection gate) -number identification -designated queue area -queue device description -point where persons are generated or "destroyed" -type of service time distribution -mean, variance of service time distribution -type of arrival time distribution -mean, variance of arrival distribution	-number in queue area -occupancy of queue area
Link	-nodes -description -length -one-way or two-way -movement area -accepts handicapped? -inbound or outbound -speed of operation (if escalator or moving walk) -does it compete for passengers with other links?	-number of persons on link -area per person in area associated with link

percentages at nodes in the station network in order to divert passengers on either "efficient" or "inefficient" paths to ancillary facilities, such as phone booths, concessions, rest rooms, restaurants, newsstands, etc.

III. COMPUTER PROGRAM ORGANIZATION

A. List Structures

In the computer program developed to implement the station simulation, an individual person is represented by a series of attributes which are represented in the core of the computer as one record consisting of a number of fields which actually contain the individual attributes. The attributes include exogenous attributes of the individual (e.g., destination, handicap status, desired walk speed and red flag attribute) plus endogenous attributes (e.g., those characteristics which change as the passenger moves through the station).

As the individual proceeds through the station system; enters links, enters queues, leaves queues, and is processed by a queue device, his movement is reflected by moving the pointers rather than physically moving the entry around the core. The records, with pointers which relate to them, create a linked list. There are two types of lists--a linked events list and a queued events list.

In the computer program, progress of an individual through a station system is marked by a series of events. These events correspond to the transition points between activities, and they actually cause the change in state which modifies the system image. The principal events of significance are the arrival of an individual at a node (i.e., an individual entering a queue), an individual leaving a queue, the processing of an individual, and an individual departing from a node. The simulation algorithms are basically built around determining the time required for a particular event and identifying the next event to occur for the particular individual being processed. The actual simulation is accomplished by "walking down" the linked events list and processing those in the list. Whenever the event time attribute of an individual is changed, his new "position" in the list is determined by the value of the event time followed by a reordering of the list in ascending values of the event times.

Since each person on the event list has an assigned "next event" time, there are no conditions blocking the execution of these events and thus, these events are also called "unblocked" events.

Whenever an individual cannot be serviced by a queue device, the "next event" time cannot be predicted. Thus, no record of this individual can appear in the events list. The individuals for which no future event times exist are assigned to queued events lists. Thus, every person in the station will be represented by the appearance of a pointer, to either the events list or queued events list.

In summary, there are two lists to be developed in the simulation--an events list, and a queued events list. Each individual in the station will be represented by the appearance of his attribute table pointer in either the events list or the queued events list, but not in both.

B. Overall Program Organization

The overall program is built around a number of processing steps among which there are four main sequential steps which describe the overall structure of the computer program. Each of the four processing steps involves a series of subprograms which perform the specific functions necessary to accomplish the main processing steps. The main processing steps are presented in Figure 4 and include initialization, input, simulation, and output reports. The main program logic controls all the sequences of operation to be performed including calling of subprograms (except where program control must be delegated to the subprogram level), and termination of a processing run.

IV. APPROACH TO MODEL VERIFICATIONS

The approach to be used in the internal and external validation of the station simulation model has been patterned after the framework established by Emshoff and Sisson from procedures suggested by Herman (4). The suggested approach involves the following sequence.

1. Internal Validation. Tests should be performed to determine that the internal operations of the model perform exactly as intended.

2. Sub-system Testing. Tests should be performed on key sub-system elements (e.g., links and nodes representing particular activity sub-systems) to determine that the model provides a reasonable prediction of the values of model variables when compared with historical data.

3. Planning Community Review. Results of the system testing should be reviewed by members of the professional planning community, knowledgeable about station operation, to obtain agreement as to the reasonableness of the structure, variables, and accuracy of the model.

4. Operational Demonstration. A potential user should have the opportunity to explore the use of the model to become familiar with its abilities and to examine its impacts on the design decisions.

5. Use as Design Tool. The model should actually be used to aid design decisions and careful records should be kept of its predicitons and the accuracy of actual results.

V. OUTPUT REPORTS

The output reports to be developed by the simulation model can be grouped into two general groups--stationwide statistics and link and node statistics.

A. Stationwide Statistics

There are four types of output reports which present overall station statistics for various types of information.

1. The first type of output oriented toward overall station statistics is a presentation of basic system operating characteristics in numerical order. For a link, the basic output is maximum number of persons that were in the area associated with the specific link at any one instant during the simulation period; the maximum pedestrian density that occurred in the movement area (sq. ft. per person) at any instant during the simulation period; and, the total number of persons that were assigned to the link during the simulation period (e.g., the hourly volume on the link). For a node, the basic output is the maximum number of persons in queue at the node at any instant during the simulation period; the usage (i.e., the maximum number of people in the queue area at any one instant during the simulation period expressed as a percent of the capacity of the queue area); and, the total volume through the node.

2. In order to allow the user t6 easily identify the most critical areas in terms of system usage, the output reports in the previous section can be reformatted in a second output series where they would be printed out in descending order of people density for both links and nodes on which he saves statistics during the simulation period in order to minimize core storage and computer running time, this output report will allow the user to identify those links and nodes in which saving detailed statistics will have some value.

3. To evaluate the overall station operation, the user may request summaries of the overall station walk time, time in queue, and total time in the station system.

B. Link and Node Statistics

For selected links and nodes in the system, the user will want specific density and impedance time characteristics. Based on a preliminary evaluation of critical station areas, or his experience on previous runs as to critical areas of the station, the user of the model will select specific link and node output reports for these purposes.

C. Additional Options and Capabilities

1. "Checkpointing"

   Termination of the simulation occurs when one of the following conditions is met.

   a. End of simulation period.

   b. Number of persons outside queue area exceeds user-specified limit (in percent).

   c. Occupancy in any movement area less than user-specified limit.

In all cases, termination would be considered a "checkpoint" and the user would receive the output statistics plus the checkpoint file necessary for preloading the network on a future run. At restart after the checkpoint, the user would have the option of adding, deleting, or changing data input values used in the previous run, plus the option to modify station loadings for the next run.

In the case where a processing run began with a checkpoint file created on a previous run, the statistics associated with those persons used to "fill" the network would not be included in summary statistics. This would assure that output summary statistics would not be transported from one run to another but would be created independently for each run.

Checkpoint termination triggered by situations where simulation output values exceeded some specified limit would essentially reflect an "out of control" situation rather than some undesirable level of operation which should be allowed to occur to experience the full range of values which occur during the simulation period.

D. Application of Statistical Analysis

1. Mean, Variance, and Confidence Intervals

Most of the output reports summarize output values in terms of a mean, variance, and confidence interval. The values which are used to calculate these statistics are collected over a period of time as a series of values $X_j(t)$, where j indexes the output series at various pointe of interest in the station. Since the values used to calculate the output statistics are generated by a stochastic time-dependent process, the values in the time series will be correlated with each other. In this case, it is necessary to use more sophisticated statistical methods to estimate the variance of the output statistics and the confidence interval around the mean value, which accounts for the correlation. A finite auto-regressive technique is used to represent the auto-correlated behavior in the time series. The station simulation model uses an auto-regressive statistical routine to generate the following statistics for any series of user-specified output values.

a. The sample mean;

b. The sample population; variance

c. The sample size used to calculate a and b;

d. The lower confidence point of the confidence interval for the mean; and

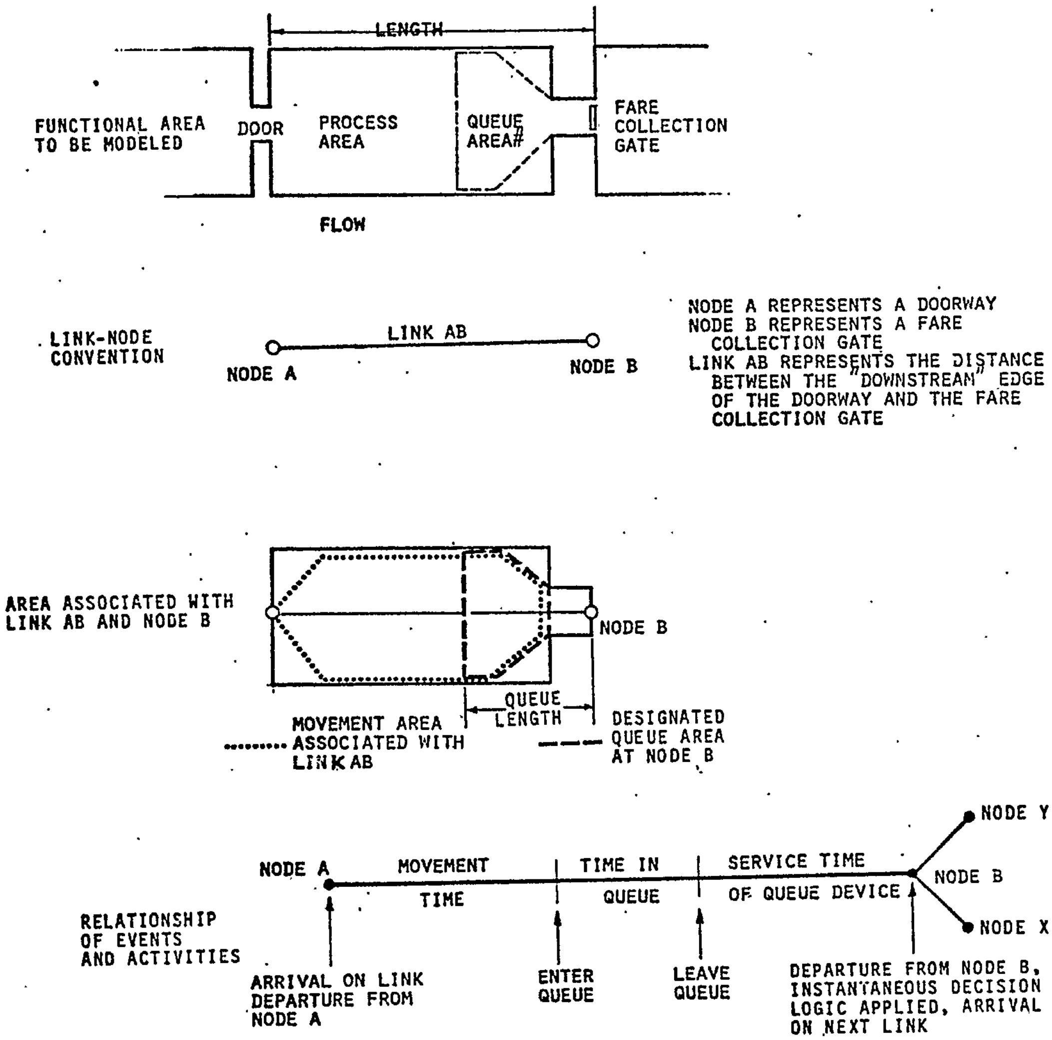

FIGURE 3

DEFINITION OF NODES, LINKS, AND AREAS FOR SIMPLE ONE-WAY FLOW ACTIVITY AREA

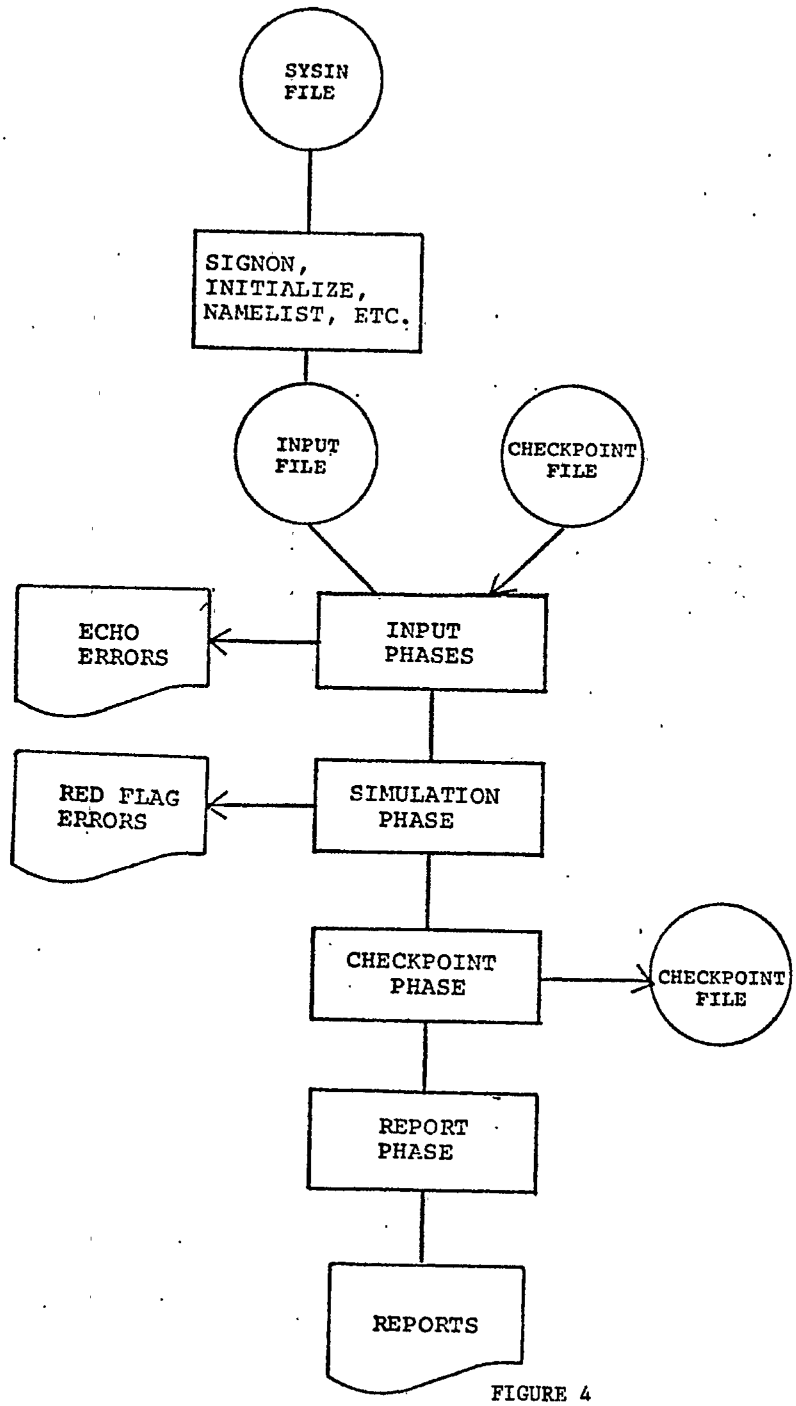

FIGURE 4

OVERALL SYSTEM FLOW CHART

e. The upper confidence point of the confidence interval for the mean.

2. Initial Bias

It is intuitively obvious that the output statistics generated by the model when the station is first filling with passengers do not represent statistical stability in terms of operating characteristics. Because of their dependence on the initial conditions; observations near the beginning of the simulation period are not representative of the process of interset and their inclusion in the calculation of the mean makes this quantity a biased estimator of the true mean value. However, as the number of observations used to calculate the mean becomes very large, the bias goes to zero since the early observations become less influential on the average. Thus, it is desirable to identify the number of observations, X*, within the total number of observations generated by the model, which should be discarded such that the values remaining represent observations less dependent on initial conditions and thus, essentially independent. The statistical routine used in the model identifies the number of observations, X*, which must be discarded from the total number of observations in order to assure that the output statistics are not biased by the initial conditions.

3. Sample Size

The analysis of output statistics using the statistical methods outlined above can provide valuable input into the proper total running time that should be used to generate statistically reliable outputs. Specifically, the user may find that shorter running times than expected can be used while preserving a specified level of statistical accuracy (e.g., measured by the size of the confidence interval). Obviously, this can reduce the cost of running the model.

4. Correlation Analyses

The output from two or more runs can be utilized as input to a user-supplied subroutine and the statistical analysis routine to measure the correlation between two time series at different points in the station (e.g., queues at two different points in the station). If the two time series are highly inter-correlated, then changing parameters (e.g., service times) at one point in the station will affect what happens at another point in the station. This capability to measure the correlation, and thus, stationwide correlations, should aid treatly in assessing the value of passenger flow metering concepts which may be proposed.

VI. INPUT

The basic inputs required to simulate a typical transit station include input to describe overall station operation, station arrival data, station structure, and output options.

A. Overall Station Operation

In order to describe the overall station operation, the user must supply values for the following basic station parameters: number of arrival/departure nodes; number of ordinary nodes, number of links, starting time of the simulation, and length of the simulation period. Also required to describe overall station operation is the distribution of total person arrivals to the station. The user must supply a distribution whose independent value is time and dependent value is cumulative number of persons that have arrived by all nodes at all arrival/departure nodes. It is from this distribution that the number of passengers that arrive on a specific vehicle at a specific time during the simulation is determined.

B. Station Arrival Data

For each arrival/departure node in the station, the user must specify data which describes the node; and, the volumes of persons which use this node to reach other arrival/departure nodes in the station.

C. Station Structure Data

The most difficult task for the user will be to translate a given station layout into a format that can be used to prepare input. This task will require that the user first designate links, nodes, and areas to represent the facilities and movement areas of the station. Here the decision will be made as to what level of detail is desired.

D. Output

Here the user must specify the number of the link or node on which additional statistics are required and the type of output report desired.

BIBLIOGRAPHY

1. American National Standard, Specifications for Making Buildings and Facilities Accessible to, and Usable by, the Physically Handicapped, 1971

2. Conway, R. W, Johnson, B. M. & Maxwell, W. L., Some Problems of Digital Systems Simulation.

3. Dial, R. B., "A Multipath Traffic Assignment Model", Highway Research Record #369, Washington, 1971.

4. Emshoff, J.R. & Sisson, R.L., Design and Use of Computer Simulation Models, MacMillan Co., 1970.

5. Fishman, G.S., Concepts and Methods in Discrete Event Digital Computer Simulation, Wiley & Sons, June 1973, Sec. 9.4.

6. Fishman, G.S., "Estimating Sample Size in Computing Simulation Experiments", Management Science, Vol. 18, No. 1, September 1971.

7. Functional Specifications for a Station Simulation Model, Dept. of Transp. Urban Mass Transp. Admin., New Systems Requirements Analysis Program, Contract No. DOT-UT-20020, Barton-Aschman Associates, Inc. & Peat, Marwick, Mitchell & Co.,Nov 3,72.

8. Fruin, J.J., Pedestrian Planning and Design, Metropolitan Association of Urban Designers and Environmental Planners, 1971.

9. Gafarian, A.V., Walsh, J.E., Statistical Approach for Validating Simulation Models by Comparison with Operational Systems - Illustrated for Traffic Flow, Int'l Fed. of Operational Research Societies, Boston, 1966.

10. Gerlough, D.L., Huber, M.J., An Introduction to Traffic Flow Theory, ASCE draft report.

11. Gordon, Geoffrey, System Simulation, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

12. Hankin, B.D., Wright, R.A., "Passenger Flow in Subway", Operational Research Quarterly, Vol. 9, No. 2, June 1958, pp. 81-88.

13. Hannan, E.J., Multiple Time Series, John Wiley & Sons, New York, 1970.

14. Naylor, T.A., Finger, J.M., "Verification of Computer Simulation Models", Management Science, Vol. 14, No. 2, October 1967.

15. Navin, F.P.D., Wheeler, R.J., "Pedestrian Flow Characteristics", Traffic Engineering, Vol. 19, No. 9, June 1969, pp. 30-33, 36.

16. Pritsker, A., Kiviat, Philip J., Simulation with GASP II, Prentice-Hall, 1969.

17. Regional Plan Association, draft of unpublished report.

18. Station Simulation Symposium, Dept. of Transp. UMTA, New Systems Requirements Analysis Program, Contract No. DOT-UT-20020, Peat, Marwick, Mitchell & Co., December 1972.

19. Walsh, John E., Handbook of Nonparametric Statistics II., Van Nostrand Co., Inc., 1965.

20. Wohl, Martin, Martin, B.V., Traffic System Analysis For Engineers and Planners, McGraw-Hill, 1967.