DIGITAL COMPUTER SIMULATION STUDIES

STUDIES OF INFORMATION NETWORKS

Kevin D. Reilly
Information Scientist
Institute of Library Research
University of California
Los Angeles, California

Digital computer simulation models for analysis of information networks are being developed. Concepts from a variety of sources (2,3,6,8) provide the basis for formulation of the models. Several of the major aspects that are incorporated include: user behavior; scheduling within the information processing center; (machine-readable) file organization and maintenance; and distribution of conventional library resources. A single comprehensive model for the entire network is also being developed to allow for detailed study of special portions of the system, followed by a study of the effects of changes in the sub-system upon the entire system. Using a higher level programming language (GPS/360) was deemed to be essential because of the necessity of model change and adaptation imposed upon us in this area. Modelling must allow for many different detailed systems designs (corresponding to the variety of plans) and for changes in any single plan many of which cannot be foreseen.

Central to the user behavior model are the development of information needs in time, the decisions the user makes in determining whether the information network can meet a particular need, and the reaction of the user to system response. Time development of the need is assumed to be an independent variable. Associated with each need is the time within which the need must be met (NT). Whether the network is deemed to be the best alternative for meeting the need depends on: (1) competition from other solutions and matters of convenience (2) timing factors. The probability of consideration to use the library (CON PROB) and the estimation of service (EST) change depending on timing factors (e.g., comparison of NT with actual service time, AST) and convenience in service.

An early version of these concepts was discussed previously. (3) Figures 1-3 provide an illustration of a portion of the results when these concepts were implemented. The need time (NT), expected service time (EST), and actual service time (AST) are all distributed quantities; NT and AST remain fixed in time whereas EST changes. Figure 1 illustrates how, with proper choice of weighting factors in determining each new value of EST (in terms of previous EST and current AST values), the user-averaged EST distribution is seen to become very similar to the AST distribution. Meanwhile, in the same exercise of the model CON PROB is seen to level off at roughly the same value (approximately, .700) for several different starting values (see Figure 2). Figure 3 illustrates the numbers of: needs developed by the users; considerations to use the library network; actual requests; system responses meeting users' needs. The number of requests presented to the system is a small fraction of the number of needs that are developed. The number of successful requests, however, is a very high proportion of the number of requests made. Various possibilities for the rules used in changing CON PROB and EST, correlation between tardiness and inconvenience in service etc. provide sufficient complexity to warrant detailed study of this portion of the model.

The model for scheduling the information center retrieval activities is built upon a previous effort. (1) Additions to this model are made to handle input into the material transportation system of the library network. Distributions of total processing time (the time from request to delivery of materials) are determined.

Of particular concern in the file organization and maintenance subsystem model are data organizations for a system in which the request and update requirements include: massive batch processing on a daily, weekly, etc. basis; on-line demand processing during each day for requests with highest priority; and small-batch processing in which moderate to high priority requests are handled at intervals throughout the day. Activity organization of the file is essential. Factors discussed are those relating to costs, turn-around time for requests, update frequency, effect of user behavior (an adaptive element), and matters relating to emergency requests. To the extent that this model overlaps some of the features of the scheduling model a comparison can be made between the automatic scheduling system of this model (in which requests, though in different queues, are theoretically in contention for immediate service) and the rigid, externally imposed, scheduling system of the other model.

The material distribution model introduces classification schemes for the types of users, materials, and requests. Users are classified according to their status, their point of entry into the system, and their roles. Materials are grouped according to subject area of knowledge. (For the University of California library network, the group of about 200 areas of study in which formal degree programs are recognized form the basis of classification whereas for a science-technology-industry network, a convenient grouping might be the fields recognized by the National Research Council of the National Academy of Sciences.) Requests are typed as one of the following: factual inquiries; requests for a single material item; requests for a survey of materials on a subject; an exhaustive survey of materials in a subject area. (4) The model implies a need to specify the strength of libraries in terms of a probability measure for finding materials of a certain class. It leads also to the need for assessment of interdisciplinary activity. (An experimental project for obtaining such data, utilizing information available in three separate (machine-readable) files is a part of the experimental phase of the model.) An outline for an application of some of these concepts is presented elsewhere. (7) Other model uses include determining the effects on system behavior of the distribution of materials and of the communications hook-ups.

REFERENCES

1. Blunt, C., et al. A General Model for Simulating Information Storage and Retrieval Systems. Report No. AD 636 435, Defense Documentation Center, Washington, D. C., 1966.

2. Brown, G., et al. EDUNET: Report of the Summer Study on Information Networks. New York, J. Wiley, 1967.

3. Carter, L., et al. National Document-Handling Systems for Science and Technology. New York, J. Wiley, 1967.

4. Goodman, A., et al. Final Report: DOD User-Needs Study, Phase II--Flow of Scientific and Technical Information Within the Defense Industry. Anaheim, California: North American Aviation - Autonetics Division, 1966.

5. Hayes, R. M., and Reilly, K.D., The Effect of Response Time Upon Utilization of an Information Retrieval System--A Simulation. a paper presented at the Operations Research Society of America Annual Meeting, May 31-June 2, 1967, available from the Institute of Library Research, UCLA.

6. Kent, A. (ed.). Library Planning for Automation. Washington, D. C., Spartan Books, 1965.

7. Reilly, K. D., "Outline for a Simulation Study of the California State Library Network". Part V of reference 8.

8. Specification of a Mechanized Center for Information Services for a Public Library Reference Center. Los Angeles, California, University of California, Institute of Library Research, 1968.

ACKNOWLEDGEMENT

The work described in this paper was supported in part by NSF grant GN-422.

Computing time was provided by the UCLA Campus Computing Network.

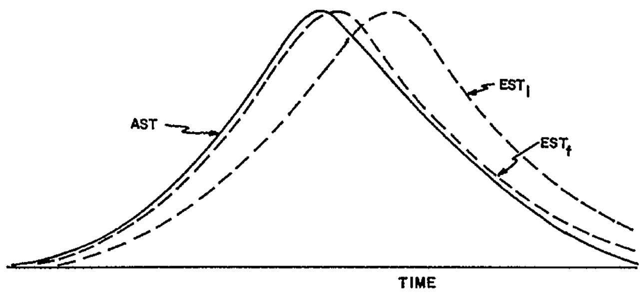

FIGURE 1

DISTRIBUTION OF ACTUAL SERVICE TIME (AST) (AVERAGED OVER VARIOUS NEED TYPES), EXPECTED SERVICE TIME AT THE BEGINNING (EST_i) AND AT THE END OF SIMULATED INTERVAL (EST_f).

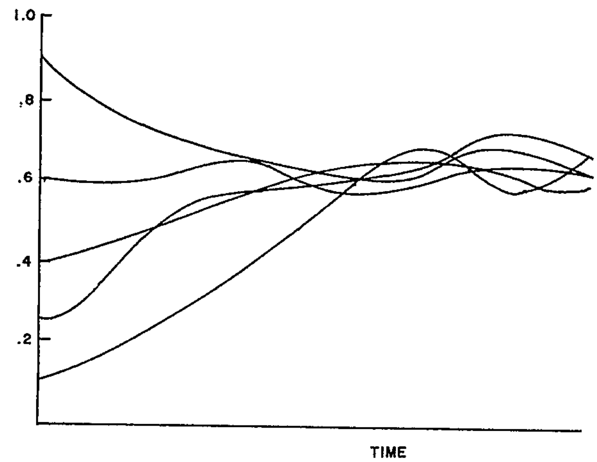

FIGURE 2

PROBABILITY OF CONSIDERATION TO USE THE LIBRARY NETWORK (AVERAGED OVER TYPES OF NEEDS AS WELL AS OVER THE USERS) AS A FUNCTION OF TIME, WITH DIFFERENT INITIAL VALUES.

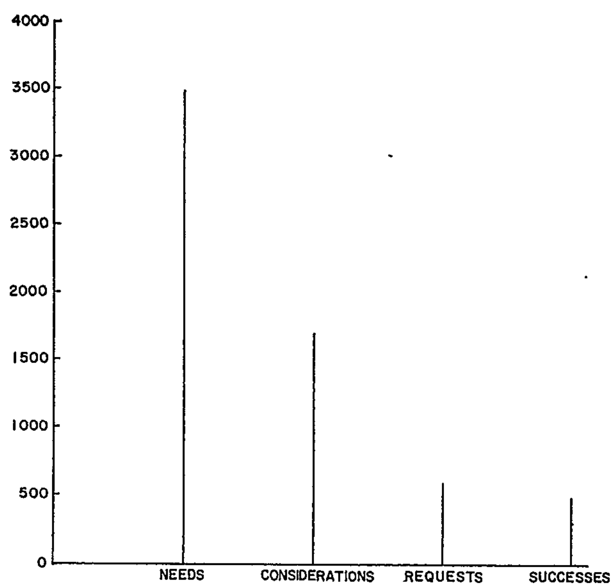

FIGURE 3

NUMBER OF NEEDS, CONSIDERATIONS TO USE THE LIBRARY, REQUESTS AND RESPONSES MEETING THE NEEDS.