the data, the networks, and the alternatives without the intermediary steps of coding, keypunching, computer running, and printout. Instead of taking problems to the computer, perhaps planners should try bringing the computer to their problems.

REFERENCES

1. E. W. Dijkstra. A Note on Two Problems in Connection With Graphs. Numerische Mathematik, Vol. 1, 1959, pp. 269-271.
2. R. B. Dial, P. S. Loubal, E. Luther, and G. J. H. Brown. A Procedure for Long Range Transportation (Sketch) Planning. PRC Systems Sciences Co., McLean, VA, 1973.
3. P. S. Loubal and B. Mendes. STREAK Interactive Transportation Modelling Program User's Guide. De Leuw, Cather and Co., San Francisco, 1975.
4. South Bay Corridor Study—Phase II. De Leuw, Cather and Co., San Francisco, June 1977.
5. Boulder Transit Study—Analysis of Alternatives for the Boulder Valley. Denver Regional Transportation District, June 1977.
6. South Bay Corridor Study—Phase II. De Leuw, Cather and Co., San Francisco, Appendix B, Vol. 2, June 1977.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Graphics.

Creation of Urban Transportation Network Models From DIME Files

Jerome M. Lutin, Princeton University

The development of a set of computer programs that use U.S. Bureau of the Census geographic data to create urban transportation network models is reported. The process uses the Dual Independent Map Encoding (DIME) files created by the U.S. Bureau of the Census for each of the major standard metropolitan statistical areas of the United States. Manual coding of networks is reduced to a minimum, and highly detailed network models can be created. Functioning and use of the program are documented, and the way in which computer graphic displays are readily generated from DIME file data is demonstrated. Examples of computer graphic output from the program are presented.

Transportation planners have long been faced with the problem of analyzing vast amounts of data. Although the use of computers has made data handling easier and advanced mathematical techniques have provided powerful tools to test hypotheses and find underlying structures, the planner still has difficulty in using all the numerical output received through the process. As a result, the planner, working under time constraints and monetary budgets, has less opportunity to analyze alternative plans.

This paper documents the development of a methodology to create models of urban transportation networks from available U.S. census data. The methodology revolves around the use of new computer software to minimize network data collection. In addition, it provides the basis for graphic displays of transportation system planning information.

Much of the transportation planner's task involves analysis of transportation systems such as streets, bus routes, and rail lines. Dealing with these systems numerically often requires creating an abstraction of the system by using graph theory in which street or line segments are represented as links or "edges" and intersections or stations are represented by nodes. Numerical values that represent speed, capacity, distance, and other system characteristics are assigned to each link and node, and this results in a network simulation that can be analyzed by using computerized mathematical models.

MODELING TRANSPORTATION NETWORKS

Modeling a transportation network involves three stages of work. First, one must specify the network to be modeled. This involves making a number of important decisions including (a) the zonal structure or geographic subdivisions of the area being modeled, (b) the scale or level of detail of the model, and (c) the number of elements—links and nodes—to be included in the model. Second, one must prepare the data for machine processing. This includes (a) specifying link characteristics, such as length (distance), time, and capacity; (b) coding the network, which includes numbering nodes and determining nodal x-y coordinates; and (c) preparing the input data cards or records. At this point, the network model is ready for machine processing.

Machine processing of the network model includes several tasks. First, one must create a computer file, or historical record, of the network. Next, zone-to-zone routes through the system, or minimum paths, must be calculated. These files of sequential links and nodes, known as trees or vines, must be stored and verified. Often, manual searching through tables of numerical output is required to see if the computer-generated routes are reasonable and error-free. From these minimum path files, individual link travel times and costs are summed to derive zone-to-zone travel times and costs. Finally, zone-to-zone travel demand is matched to the network to determine the amount of traffic flowing over each route. Other types of network analysis are possible as well, but the procedure outlined above is the one most commonly used.

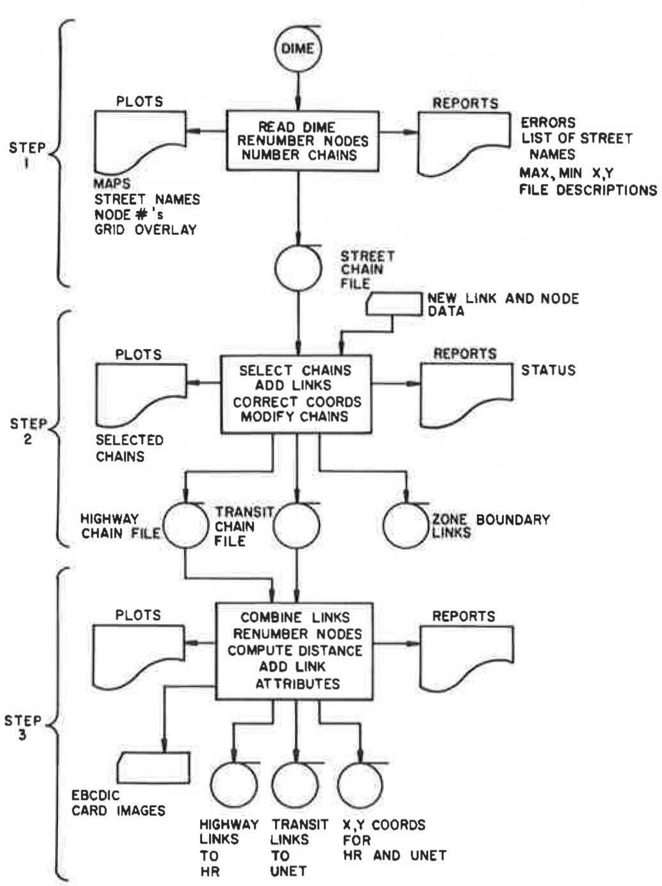
Two common difficulties have been found in creating and analyzing network models. First, a great deal of information must be collected to supply an adequate representation of a network. Urban areas typically have thousands of streets and hundreds of kilometers of bus lines. A computerized network model may have several thousand links and nodes that require thousands of data items punched on computer cards. This process, known as coding, becomes a time-consuming task that must be repeated for each network to be analyzed. Second, for each network, the analyst must examine vast amounts of computer output material that contains volumes of numbers. This complicates the task of evaluating the results of each network simulation.

Planners recognize that graphic tools, such as charts and maps, are often the best means of presenting numerical data for transit systems. For example, one may wish to use a graphic representation of flow volumes on a street network in which the width of the band indicates the number of trips that use each link of the system. This graphic presentation conveys far more information than the same data presented in tabular form.

Federally sponsored libraries of computer programs have been developed for the analysis of urban transportation systems and are currently widely used throughout the country. Some of these programs have been written or modified to produce graphic output of data. In particular, the Urban Transportation Planning System (UTPS) of the Urban Mass Transportation Administration (UMTA) and the computer programs of the Federal Highway Administration (FHWA) have been modified to produce graphic displays of networks on plotting devices attached to computers (and line printers as well).

However, the production of graphic output for transportation networks requires even more data collection. It is necessary to establish a coordinate system or grid to locate each point to be displayed. Two numerical coordinate values, one for the horizontal or x-direction and one for the vertical or y-direction, must be calculated for each point. These values must then be punched on computer cards or added to the data bank representing

Figure 1. Flow chart of network process using DIME files.



the coded links and nodes of the network. Typically, this process has been done largely by hand. The development of electronic digitizers now allows the analyst to record the network by tracing lines from a map onto a sensitized tablet that electronically records numerical x-y coordinates for each point. Even with digitizing equipment, however, the process of network creation is a laborious task.

DUAL INDEPENDENT MAP ENCODING (DIME) FILES

For the 1970 Census, the U.S. Bureau of the Census developed the Geographic Base File (GBF) and a computer program—Dual Independent Map Encoding (DIME)—that contains x-y coordinate files for most streets and many natural boundaries. The resulting areas are roughly equivalent to city blocks. DIME files are currently available for most of the more than 200 major U.S. cities that the bureau describes as standard metropolitan statistical areas (SMSAs), and the system is being expanded to include all SMSAs. If these files were properly linked to transportation planning computer programs, the result would be near elimination of the manual effort to code and describe even the most complex network. It would provide the desired graphic display information as well.

The Princeton University Transportation Program has used the DIME files to produce x-y coordinates for each bus route in Trenton, New Jersey. The files were also used in conjunction with the Bureau of the Census ADMATCH program to allocate employment records to traffic zones.

In developing the DIME files for urban areas, the Bureau of the Census adopted a format for data storage that best suited its purpose—primarily, the location and coding of street addresses for automated processing of data and census data collection by mail. Thus, the DIME file records are grouped mainly by census block and are not directly usable for describing a transportation network. Each record in the DIME file is 300 characters long and represents a block face or geographic boundary segment. Since most streets and natural boundaries are located in the DIME files, it is possible to extract required data and reconstitute a street network that is appropriate for most urban transportation planning. The main problem involves the manipulation of DIME records into "chains" of links that represent entire streets.

SOFTWARE DEVELOPMENT

This project has developed a method whereby census data can be more effectively used by planners within the framework of the Urban Transportation Planning System (UTPS). A set of three computer programs has been developed that can read the DIME files and, on user request, output a selected file of nodes and node coordinates in a variety of formats (such as tape files or punched cards). It will be possible to use these, in turn, as inputs to the UTPS network-building programs UNET and HR. The end product will be a computer package that provides the interface between census data and the UTPS package.

Through use of these programs, it is expected that the effort required to code data for many transportation planning projects can be greatly reduced. It will enable transportation planners to model and analyze more alternative transportation networks within a given time and cost budget. Basically, the provision of a software link between the DIME files and transportation network simulation models has three objectives: (a) to provide

Figure 2. Bureau of the Census block map of Trenton urbanized area.

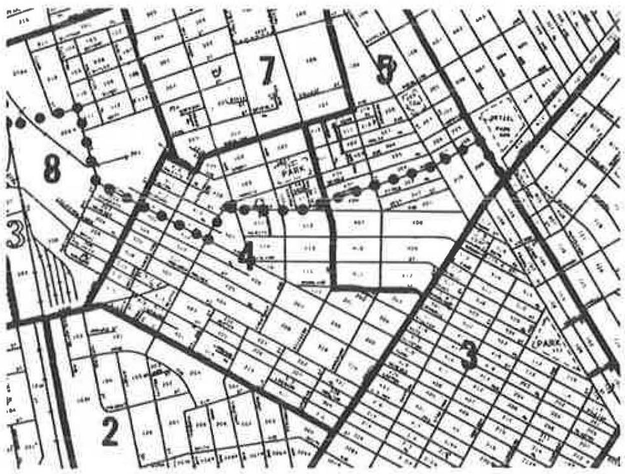


Figure 3. Computer-drawn map of Trenton DIME file.

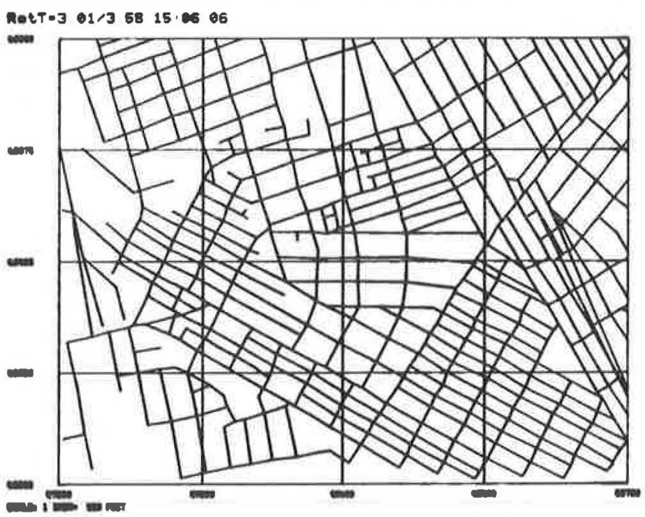
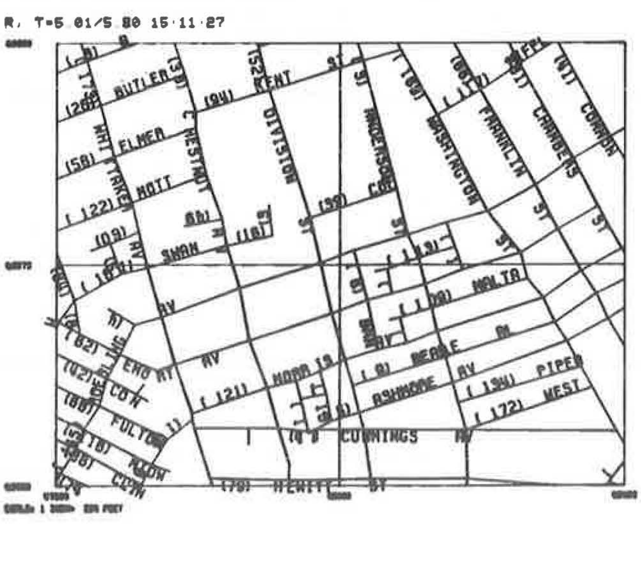


Figure 4. Enlargement of Trenton DIME file map.



a data base suitable for graphic display of analysis results, (b) to reduce data-handling requirements for transportation planning, and (c) to provide the capability for developing very detailed network models for specialized projects.

PROGRAM OPERATION

Figure 1 shows a general flow chart of the operating sequence of programs currently under development. Functioning of the programs is described below.

Step 1—DIME Editing

The first step involves reading the raw DIME files. Here, the user is provided with measures that indicate the completeness of the DIME files. Errors (tape errors, missing data, and out-of-range values) are given, counts of records are taken, street names and boundary names are listed, and maximum dimensions of the network are determined.

The program outputs printed reports plus, if desired, plots of the network, with street names, node numbers, and a planning grid overlay, for determining new coordinates. At the same time, the program creates an output file of links chained together according to street or boundary name. Each chain is assigned a number for ease of reference. All nodes in the DIME file (each record contains both a "from" node and a "to" node and four-digit node numbers by tract) are assigned unique, new node numbers. Each DIME file is grouped according to map number to correspond to the Metropolitan Mapping Series (MMS) of the Bureau of the Census. Thus, the user can select various segments of the urban area for inclusion in the network.

Step 2—Editing of Chain Files

The second step in the process involves reading the chain files produced by step 1. In this step, the user can add new links to the chain file, modify and restructure street chains, and correct node coordinates. Step 2 produces several output files of links chained together at the option of the user. For example, one can obtain a zone boundary file, a transit network file, and a highway network file. This stage also produces reports of changes and plots of the networks.

Step 3—Creation of Link Files

This stage produces output files of links in forms suitable for analysis by network-building programs. The program reads selected chain files, combines links (by deleting unused nodes), and renumbers nodes to correspond with conventions used in network-building programs. Link lengths or distances are calculated, and various attributes, such as type of facility and speed limit, can be added to each record. In many cases, default values for link characteristics are supplied to simplify network coding. Files of x-y coordinates are produced so that the output link files can be graphically displayed. The user-selected output files produced at this stage are correctly formatted for use by UMTA network processing programs HR and UNET for highway and transit analysis respectively.

PROGRAM OUTPUT

A number of output reports are produced at each stage of processing. For example, in stage 1, a comprehensive street chain directory is produced. For each chain, the report lists chain number, name, number of links, number of nodes, and the node number and x-y coordinates for each node in sequence. Other reports, including reports of errors found and a dictionary of corresponding new and old node numbers, are also produced.

However, the main value of the program lies in its ability to produce graphic output and to produce networks capable of graphic display. Figure 2 (1) shows a segment of the Bureau of the Census MMS map of a portion of Trenton, New Jersey. One can see that each street is included as well as block and tract identification, some major open areas (parks and cemeteries), and major rail lines. This map is the basis for the geocoded DIME file for the Trenton area. Figure 3 shows a computer-drawn map of the same approximate area that uses chain files extracted from the Trenton DIME file. A coordinate grid (of the same scale as that used in the DIME file) is overlaid on the map. In Figure 4, the scale has been enlarged four times, and street chain numbers and street names are both included.

These maps were drawn on the face of a Tektronix 4013 cathode ray tube (CRT) terminal by using a preliminary version of the software described above. Standard CALCOMP plotting routines were used for legends and alphanumeric characters. These figures show the great level of detail available in the DIME files.

CONCLUSIONS

Although some work remains to be done in testing and improving the programs, the work reported here has attained two major objectives:

1. The feasibility of using the DIME geocoding system as a basis for transportation network modeling has been established.
2. A methodology for processing the DIME data files has been developed.

Increasingly, transportation planners are being asked to provide more detailed and more specific analysis of transportation systems in urban areas. This requires detailed network models. Use of the software described here and the available Bureau of the Census DIME files will enable the planner to create detailed network models with a minimum of manual intervention. Although many errors and gaps have been found in DIME files, the Bureau of the Census is striving to obtain complete and correct DIME files for all SMSAs. The software described here provides for human intervention, primarily to detect and correct errors. Although the process cannot be completely automated, the amount of effort required to create a network model will diminish as the quality of DIME file data improves. In the future, it is expected that DIME-based networks, which can be readily matched to other demographic census data, will be in widespread use among transportation planning agencies.

ACKNOWLEDGMENTS

I would like to express my gratitude to Granville Paules, Linda Moore, and Robert B. Dial of the Urban Mass Transportation Administration for their support of this work. I would also like to thank Warren B. Powell, William D. Collins, and Brian P. VonHersen for their work in programming the software described above. This research was performed under a university research and training grant and a research and development grant from the Urban Mass Transportation Administration.

REFERENCES

1. Block Statistics Trenton, NJ-PA Urbanized Area. U.S. Bureau of the Census, Metropolitan Map Series, U.S. Government Printing Office, HC(3)-151, Map Sheet P-21, 1971.
2. M. A. Meyer and J. Silver. Present Status and Future Prospects of the Census Bureau's GBF/DIME CUE Program. U.S. Bureau of the Census, presented at Conference on Geographic Base File Systems—A Forward Look, Boston, April 16-17, 1974.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Graphics.

Computer Geocoding of Travel Surveys

Priscilla Sawyer Cornelio and Joseph M. Manning, Urban Transportation Systems Associates, Inc., Newton, Massachusetts

Computer geocoding of travel survey data, which involves use of a geographic base file and a series of user-oriented computer programs, is a workable and preferable alternative to manual geocoding, which is tedious and time-consuming. The basis for the computer geocoding system is the Dual Independent Map Encoding/Geographic Base File (DIME/GBF) developed by the U.S. Bureau of the Census. The DIME/GBF, which exists for all standard metropolitan statistical areas, contains detailed information on street segments, including street name, direction, range of house numbers, and census tract and block. Two programs developed by the U.S. Bureau of the Census, ZIPSTAN and UNIMATCH, are used to perform the actual geocoding. ZIPSTAN is a preprocessor program that arranges the addresses before being linked by address to the DIME/GBF by use of the UNIMATCH program. Geocoding on-board and travel surveys in the Boston area indicate that 70 to 80 percent of all addresses can be geocoded to a detailed zone level at a processing cost of $2.12/1000 addresses. Addresses not geocoded automatically are generally incomplete or contain invalid information. The basis for this system, the methods and procedures used in the Boston area, the results of the matching operation, and the costs involved in the effort are presented and discussed.

Planners and engineers frequently use travel and attitude surveys to determine existing travel patterns and latent demand for new or improved transportation facilities.