Integrating GIS and CAD for Transportation Data Base Development

Jin-Yuan Wang and Jeff R. Wright

An accurate and reliable transportation network data base is a prerequisite of a successful service route design system. The focus of this research is the design and development of an interactive procedure for use in the reclassification and verification of the U.S. Geological Survey's 1:100,000 digital line graph data for use in a system that automatically designs rural snow and ice control routes for the Indiana Department of Transportation. Details of the digital line graph data format and conversion routines are presented.

An important though often neglected factor in the development of strategies for delivery of public services is the design of service routes. The importance of effective route design in the public sector relates not only to the cost of service provided, but also to a variety of intangible benefits such as equity and quality of service. The quality of service provided by public institutions is often difficult to quantify (*1*). In contrast to the private sector, where engineering management objectives are usually specified in terms of economic efficiency, government agencies strive to provide the best level of service possible as measured by public welfare and safety. These performance criteria are generally difficult to quantify for most public service activities. An important factor, however, in the quality of services provided is the planning and management of those services. For an institution such as the Indiana Department of Transportation (InDOT), a key factor in effective planning and management is the ability to design efficient route configurations for the delivery of services. The removal of snow and ice from the intrastate highway system is a good case in point.

In developing a strategy for winter season snow and ice removal, the goal of InDOT is to provide efficient service within the constraints on available resources (plowing and abrasive spreading equipment, sand and salt supplies, and manpower). Although holding down overall cost is a primary consideration, the safety of the public is the major objective. Public safety in this context has two distinct but related components: (a) the condition of the road surface, and (b) the performance of the snow removal fleet during the operation. An effective snow removal operation is one that provides rapid and orderly snow removal and abrasive application without excessive interference with public transportation activity (*2*).

Many public sector engineering management problems may be formulated and solved using network-based models (*3*). Among the most complex problems are those that involve the routing of service vehicles for such items as trash collection, police patrol, road painting and cleaning, roadside weed control, and snow and ice control. Although a rich set of analytical methodologies has evolved to solve these problems, the data required for their applications are often enormous. For the design of snow and ice control routes on rural highways, for example, complete network distance and adjacency data are needed, including precise configuration information for all major intersections. The effort required to collect, verify, and digitize these data on an ad hoc basic generally precludes their use in model development and problem solving.

Although ad hoc efforts to develop data bases to support these models are not cost-effective, many general data collection efforts are being conducted that contain these data and in digital form. For example, the U.S. Geological Survey (USGS) presently provides complete digital line graph (DLG) data at a scale of 1:100,000 for the entire United States. Data layers at this scale exist for such man-made features as hydrography, roads, railroads, and transmission systems. Magnetic tapes containing these data may be purchased for any region for interest.

The focus of this research is the design and implementation of an intelligent data filter for use in conditioning these data in a cost-effective manner so that the resulting spatial data base is suitable for use in the automated design of service routes.

DLG DATA CONSIDERATION

For the application of vehicle routing over the state highway network, a complete description of the network is required. This description must include all major intersections in each district, adjacency information that indicates which intersections are joined directly by roads, and the lengths of each of these roads. Planning for snow and ice control requires additional information, such as the width in lanes of every roadway, and more complete descriptions of intersections, including features such as turn lanes or traffic control devices. Currently, the most effective means of assembling such data has been to produce a network by reading the information from a highway map, and then to add additional arcs to the graph to account for multiple lanes. On a statewide basis, this process could prove prohibitive, and the accuracy of the resulting maps could be less than desired.

Digital Map Data Sources

A major effort of this research has been to explore available digital map data sources, and to find the best methods to

School of Civil Engineering, Purdue University, West Lafayette, Ind. 47907.

convert these data to a usable network representation. One class of geographic data considered was a raster, or grid, format. Data gathered by aerial or satellite photography is in this form. A raster format is impractical for a network application because of the large amount of processing required to extract the roadway information from the data grid. Also, to obtain accurate distances along roads, the sampling size of the raster format must be fairly small, or the errors in calculation would be multiplied many times over.

The most promising format for this application is a vector format in which roadways are represented as polylines, that is, a set of short segments connecting two intersections. Many basic functions are simplified, especially those of finding adjacent intersections and road lengths. The source for vector data thus far has been the USGS's DLG format. Roadway information is available from this source for the state of Indiana at 1:100,000 scale. As part of the National Mapping Program, the USGS is creating this data base by digitization of existing 1:24,000 scale archive maps. Current estimates from the agency indicate that the coverage of the United States in 1:24,000 scale digital maps is only 2 percent complete for hydrography and transportation information. However, to complete a data base for the 1990 decennial census, the USGS has produced a set of digital 1:100,000-scale maps for transportation and hydrographic information. Because this research project was primarily concerned with representation of the highway network, the available 1:100,000-scale DLG format data is adequate. The critical information for use of these data in a geographic information system is the availability and expense of the distributed data, their logical format, and their accuracy. These issues will be discussed in the following sections.

The DLG data are a vector representation of the environment. Map information is distributed by the National Cartographic Information Center of the USGS in two formats: standard and optional. The primary difference in the formats is the internal measurement units used. In the standard format, all coordinates are given in pure integer values, which can be mathematically transformed into geographic coordinates. The USGS recommends use of the optional format distribution of the data, in which all coordinates are recorded in Universal Transverse Mercator (UTM). The optional format also includes a more complete set of linkages between node, line, and area information.

The structure of the distributed DLG file can be described as a heading section identifying the geographic location of the cell, followed by one or more categories of data (see Figure 1). The map heading includes the name, scale, creation and modification dates, and corner registration points of the map. This heading also includes edge matching information, indicating the extent to which adjacent maps are compatible. Each category within the DLG file begins with a record stating the category name, and the number of node, line, and area records comprising that category. Although the capability exists to represent 15 categories within map files, currently distributed maps contain one category per file. As a result, the files are much more easily managed.

The node, line, and area records are designed to provide the maximum amount of usable network information in as compact a form as possible. This information includes an internal identifier, an integer code unique to other nodes within the map in as compact a form as possible. Node records currently contain the following information:

```
USGS-NMD  DLG DATA - CHARACTER FORMAT - 06-10-86 VERSION
WATSEKA, IL IN                15 1984,      100000  S15
 .000000000000000D+00   .000000000000000D+00
 .10000000000D+01  .00000000000D+00  .00000000000D+00  .00000000000D+00
SW     40.750000  -87.250000      478895.54  4510824.90
NW     40.875000  -87.250000      478935.34  4524698.35
NE     40.875000  -87.125000      489466.18  4524675.85
SE     40.750000  -87.125000      489446.70  4510802.36
ROADS AND TRAILS     0   253   253 010   152   152 010   403   403   1
N   1  478895.54  4510824.90      2       0    0
 -385   386
N   2  478935.34  4524698.35      2       0    0
   -5    6
N   3  489466.18  4524675.85      2       0    0
   -9   10
N   4  489446.70  4510802.36      2       0    0
A   1  484185.94  4517750.36     38    0    1    0    0
 -386 -385  383  381  379  378  375  374  371  372  356  201
  179  156  124  104   72   54   39    9   10   -1    2   -3
   -4   -6   -5  -22  -47  -50   52  -71  -91 -111 -135 -166
 -229 -366
    0    0
A   2  484185.94  4517750.36      4    0    0    0    0
  386 -369 -384  385
A   3  484185.94  4517750.36      4    0    0    0    0
    8    5    6   -7
A   4  484185.94  4517750.36      4    0    0    0    0
   13  -10   -9  -12
L   1    5    6    1    6           2    0    0
  486938.89  4524686.33  488335.89  4524683.35
L   2    5    7    7    1           2    0    0
  486938.89  4524686.33  486161.65  4524687.99
L   3    8    7    1    8           2    0    0
  483621.64  4524690.88  486161.65  4524687.99
L   4    9    8    1    9           2    0    0
  480512.69  4524700.06  483621.64  4524690.88
```

FIGURE 1 Sample DLG data file structure.

- Coordinate pair: a pair of real numbers that represent the UTM coordinate that locates the node.
- Length of line segment list: an integer value representing the number of map lines incident to the node.
- Length of attribute list: an integer value stating the number of attribute codes assigned to the node.
- Line segment list: a list of integers that represent the internal identifiers of each line incident to the node.
- Attribute list: a possibly empty list of pairs of integers that describe the purpose of the node.

For an application of a network analysis algorithm, a node record must contain only the incidence list. However, to implement a geographic data display of these data, all coordinate information must also be maintained. Line information is stored in a format similar to that of the node. Each line record contains the following data:

- Internal identifier.
- Starting node: an integer representing the internal identifier of the starting node.
- Ending node: an integer representing the internal identifier of the ending node.
- Left area: an integer representing the internal identifier of the area to the left of the line, when traversed from start to end node.
- Right area: an integer representing the internal identifier of the area to the right of the line, when traversed from start to end node.
- Length of coordinate list: an integer value stating the number of points which constitute the line between the start and end nodes.
- Length of attribute list.

- Coordinate list: a list of pairs of real numbers that represent coordinates through which the line passes.
- Attribute list: a possibly empty list of pairs of integers that describe the purpose of the list.

The bounding area codes can be used to identify whether the line includes part of the neatline. A neatline represents the boundary of a map, and serves to clip all lines so that the map represents a regularly shaped area. In order to implement a network analysis algorithm, the distance between adjacent nodes must be known. Because the coordinates of data points in the optimal distribution format DLG is the UTM system, distance between adjacent nodes can be calculated as the total length of all segments making up a line. The attribute information should also be included in a problem data base, as the algorithm must be able to consider what feature of a map the line represents. Once again, to represent the map information in graphic format, the coordinate information must be maintained.

Area records are stored in the same structure as node records. The primary differences between the meanings of the various fields lie in the meanings of the line segment list. For an area record, the number of line segments making up the boundaries of the area is stored, rather than the length of the incidence list of a node. Also, the number of islands within the area is stored in an area record. This procedure allows the area to contain holes that could not otherwise be consistently represented in the DLG format. Network analysis does not require any of the area information from the map. For other tasks, however, this information could be managed and manipulated exactly as that of the node and line data.

Accuracy of Distributed Data

The USGS does not quantify the accuracy of the DLG maps that it distributes. Because the maps are produced from original archived maps, few allowances can be made for changes to the map area since the last revision of the source data. Future goals of the National Mapping Project are to standardize all published maps, and to implement a 5- to 10-year cyclic inspection process. Most necessary changes to the map data base would be made after the entire national data base is digitized. Critical changes to the map data base may be made by the Survey on special request. Alternatively, the data user could attempt to modify his copies of the map data. The USGS does, however, make an effort to ensure that the digitized map correctly represents the source map.

The DLG data users guide (*4*) states that manually digitized source maps have a resolution of 0.001 in., with an accuracy of no less than 0.005 in. If the source map was digitized automatically, the resolution is 0.0013 in. The guide further states that the digitized data are plotted and compared to the source material. The plots are checked for positional accuracy and completeness of the graph. Attribute information for the map is entered manually using the original source map as a reference, and then checked by software to guarantee that only valid codes were entered. This process does not, and could not easily check the entered attributes to verify that they provide a necessary and sufficient description of the line. Topological fidelity or planarity of the data is verified by software. This procedure guarantees that no lines cross except at nodes, and that any line entering a node ends at that node. Edge matching of positions and attributes is said to be performed whenever the adjacent cells are available in the data base. Experimental experience with these data will indicate that this edge-matching procedure may not be sufficient.

DLG DATA AND TRANSPORTATION NETWORK FOR VEHICLE ROUTING

The use of USGS DLG data for route design and analysis is hindered by two distinct factors. First, the DLG data set includes considerably more information than needed (about county roads and walking trails, etc.). Next, the information provided by DLG data may not be adequate for routing purpose. These concerns are discussed in detail in the following subsections.

Important Characteristics of a Transportation Data Base

In developing a transportation network data base, a number of important characteristics must be considered independent of the positional accuracy and reliability of the data. These data base elements include the following concepts.

Geometric Relationships

An accurate coordinate system is the fundamental element necessary to represent the topologic relationships among roads and intersections.

Highway Classifications

In most of the cases, highways are categorized in terms of their importance. Higher-class highways are supposed to receive higher quality of services. For example, InDOT uses average daily traffic (ADT) as the basis for highway classifications (*3*). There are three different classes used by InDOT, plus a special class called "no service" (to specify road segments that can be traveled but that receive no service).

Number of Road Lanes

One of the most important considerations during the route design procedures is the number of lanes of a road, because each lane should receive a minimum level of service, especially for street sweeping and snow and ice control.

Directions

The direction of traffic flow is, of course, an important consideration in snow and ice control on the intrastate highway system. Maintenance equipment must generally follow existing traffic patterns. For example, in Figure 2 suppose that the path indicated by road segments linking (sequentially) Node 7 to Node 6 to Node 5 to Node 0 to Node 8 to Node 9 to Node 10, and path Node 4 to Node 3 to Node 0 to Node 1 to Node 2 represent two Interstate segments with an overpass crossing at Node 0. In this configuration, path Node 2 to Node 6, path Node 5 to Node 4, path Node 4 to Node 8, and path Node 9 to Node 2 are entrance and exit ramps. Although original Survey data make no direction distinctions, it is clear that travel along a path Node 2 to Node 6 to Node 5 would not be possible, and that a vehicle traversing the path Node 2 to Node 6 must necessarily visit Node 7 next.

The over- or underpassing intersection is another concern. For example, Node 0 in Figure 2 represents an over- or underpassing intersection. Therefore, the path Node 5 to Node 0 to Node 3 is not possible. One way to handle this situation is to delete this over- or underpassing node. (However, this solution usually implies loss of the planarity and the geometric relationship of the graph.) The restricted direction method mentioned earlier can also address this situation. For the ramp area shown in Figure 2, the following statements specify the over- or underpassing situation of Node 0.

- Node 5 is the only feasible destination if travel is from Node 8 to Node 0.
- Node 3 is the only feasible destination if travel is from Node 1 to Node 0.
- Node 8 is the only feasible destination if travel if from Node 5 to Node 0.
- Node 1 is the only feasible destination if travel is from Node 3 to Node 0.

Turn-Around Point

The possibilities for making turns at a particular node (location) is important from the standpoint of a route designer. Any intersection of two roads is considered a candidate for making a turn. Also, any median crossovers existing on highways may be other possible candidates. Information regarding the locations and maneuverability at such locations are important for any routing procedure.

Data Compactness

Another serious concern of applying network-flow-based models to the real-world routing problems is the computational intractability caused by the excessive data size. It is desirable to represent the service area by using a minimum amount of data to reduce computational complexity.

Limitations of Survey 1:100,000 DLG Data

Although Survey DLG data provide excellent coordinate systems to represent geometric relationships among roads and nodes (intersections), the USGS DLG data have some shortcomings. The USGS DLG data set includes considerably more data than needed for route design. A portion of the transportation network as represented by USGS DLG data for an area of Indiana is shown in Figure 3. The left panel displays a complete Survey 1:100,000 data set for a representative area in Indiana including all road map neat lines and boundaries. Each road is made up of sets of vectors, with nodes positioned at each vector intersection. Each vector and node is assigned a specified classification according to Survey conventions. The right panel displays only those roads and intersections of interest in the design of snow and ice control routes for the same area. Most of the vector and node information has been removed. In general, approximately 94 percent of the original Survey data are unnecessary for this application.

In addition, USGS DLG data cannot address most of the characteristics listed earlier. For example, the number of lanes, crossover positions, and restricted directions are lacking in the existing USGS data or are not be represented appropriately. In order to overcome shortcomings, an interactive filter program has been developed that is introduced in the next section.

AN INTERACTIVE FILTER FOR DLG DATA

In order to overcome the limitations of the USGS 1:100,000 DLG data set in support of automated route design, an in-

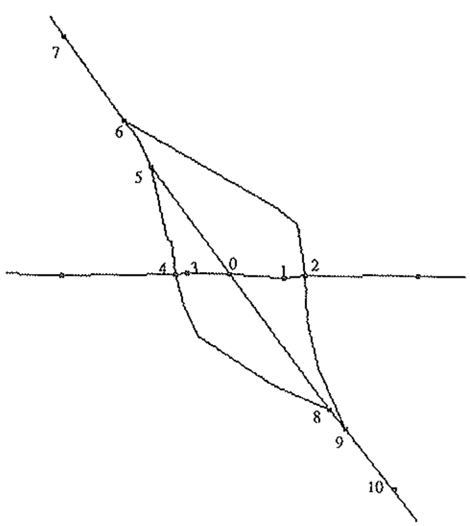

FIGURE 2 The DLG presentation of a ramp area.

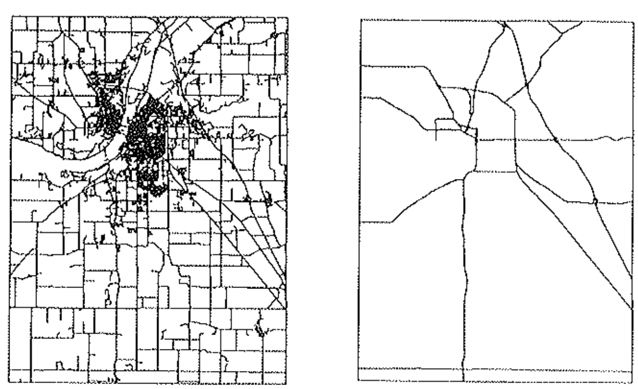

FIGURE 3 DLG data reduction for route design GIS.

teractive data filter has been designed and implemented. This system is designed for use by transportation engineers within the maintenance division of InDOT with the goal of developing a comprehensive state transportation data base that can be accessed by computerized route design algorithms. The structure of this interactive data filter is shown schematically in Figure 4. It consists of three separate functions: (a) input of the original DLG data files (b) interactive reclassification of road segments and node (intersection) objects, and (c) output of reclassified data files in DLG format. This filter uses USGS DLG format as its input and output format. That is, the output data of this filter program can be read by any software that can accommodate DLG data format.

Data Input

The original USGS DLG files are used as the input of the data filter program. There are two steps that must be considered during this stage.

Maps Selection

USGS divides each 30-minute area into 4 or 16 separated files, and refers to each file by a name. By pointing a mouse cursor at any map panels, the user selects the appropriate files for filtering, and those files will subsequently be displayed as for interactive manipulation. A representation of such a mechanism is shown in Figure 5.

Neat Line and Maps Merge

Boundary lines (called neat lines) are used by USGS to locate edge nodes and to maintain the topologic continuity among DLG files. Because it is necessary to merge the selected files into a single map, some of the neat lines are redundant and need to be removed. The concept of the neat lines removal and map merge is shown in Figure 6. Notice that some of the neat lines still need to be kept for the future possible merge.

Network Object Classification

The ultimate goal of this data filter program is to produce a suitable data set that can represent the network properly. Road and node specification play two major roles in this data filter program. An interactive method is presented for the purpose of road and node reclassification.

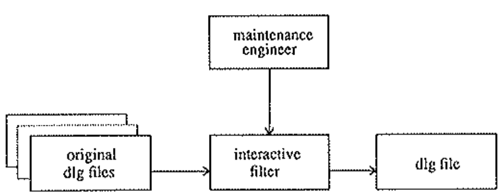

FIGURE 4 Schematic of the data filtering process.

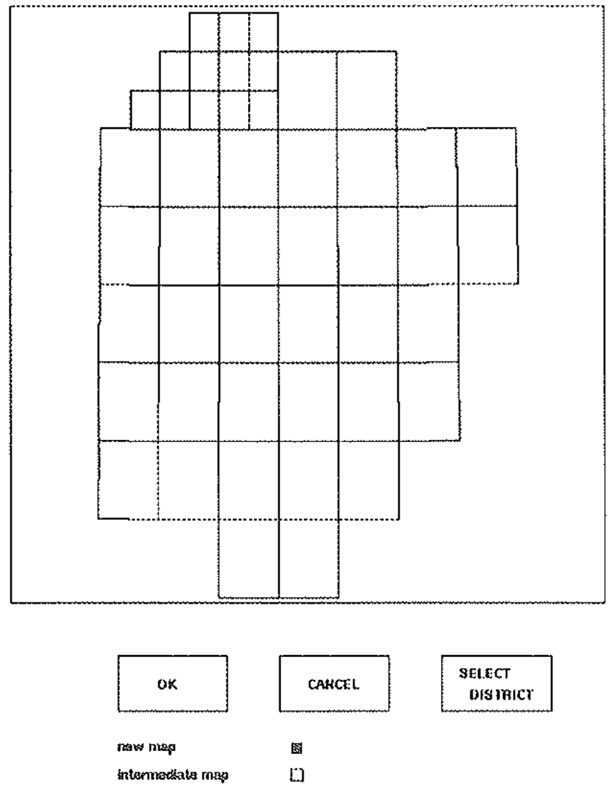

FIGURE 5 Map selection stage.

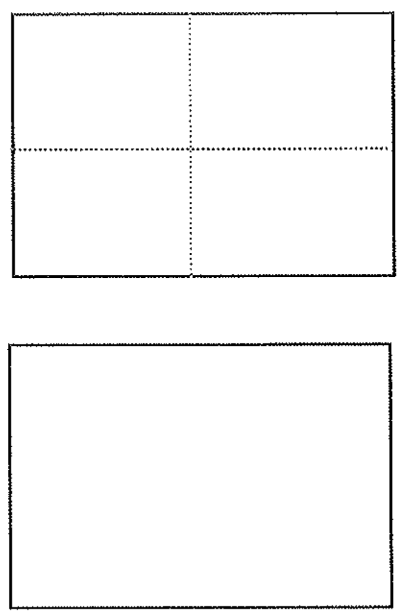

FIGURE 6 *Top*, before neat line removal, *bottom*, after removal.

Road Classifications

The sequence of activities conducted during road reclassification is shown in Figure 7. In this application, five major road classifications are required: Classes 1, 2, and 3, based on ADT data; no-service roads, which do not require service

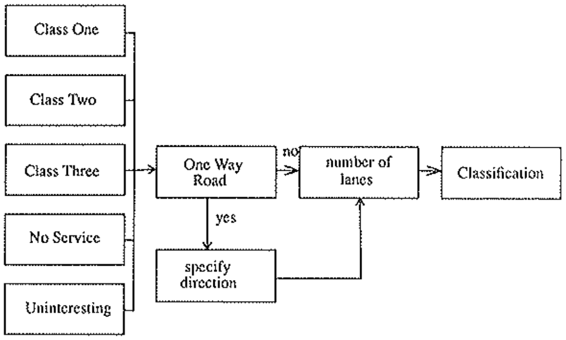

FIGURE 7 Road classification.

but can be used to connect two highways; and all other (uninteresting) roads. Separate functions are provided for each group of classifications.

Restricted Directions

The restricted direction is specified by selecting three nodes in a row. For example, a restricted direction is declared if we select three nodes, Node 1, Node 2, and Node 3, as a sequence. This means that if travel is initiated from Node 1 to Node 2, continued travel to Node 3 is feasible.

Node Classifications

The sequence of activities conducted during node reclassification is shown in Figure 8. Node classification is used to label nodes and to indicate maneuver restrictions at individual nodes such as turn-around restrictions. In addition, it may be necessary to add nodes to the network representation such as to add locations not contained in the USGS data files. Node classification may be important for vehicle maneuverability in snow and ice control operations.

There are five different functions provided for node specification: unit location, no turn around, add new node, label node, and uninteresting. The unit location function is used to designate the location of a storage or maintenance facility. The no-turn node is used to specify locations where a U-turn

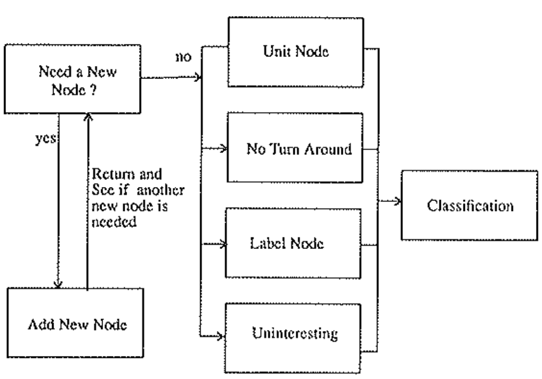

FIGURE 8 Node classification.

is not allowed. Label node is used by users to mark any node for future reference.

Utilities

In addition to the basic road and node classifications, this data filter also provides several convenient functions to help user conduct the filtering task. These capabilities include:

- Zoom: zoom in or out of a certain area.
- Highlight: highlight two certain road simultaneously in order to help users identify certain roads.
- Information: show information of a certain road or a particular node.
- Showme: show the current status of classification.
- Check direction: verify restricted directions.
- Save: saves current configuration every 20 min.

Output

The output stage of the interactive data filter consists of two components: (a) a debugging function, and (b) a routine that prepares the completed data set for subsequent use. The debugging function is designed to verify each road and node segment and to report inconsistencies to the user. The data preparation routine writes the completed map to a standard USGS DLG-3 format. The only difference between the final data representation and the original is that the major and minor attribute codes have been changed, and that the area information is discarded because it is not needed to accomplish route design functions. The files may now be read and manipulated by any program environment that can accommodate this format.

CONCLUSION

The routing of vehicles for snow and ice control is perhaps the most difficult task of public sector routing problems, yet the data required for solving this problem are not significantly different than those required for other service planning and management problems. Consequently, accurate and complete representation of the physical network suitable for snow control vehicle routing will be of use to maintenance engineers for a wide range of applications such as scheduling and routing of mowing, painting, weed control, facilities and equipment servicing, inspection, and possibly some pavement maintenance activities.

Experience has demonstrated that the USGS 1:100,000 data represent an excellent and cost-effective source of transportation network data for use in applications such as route design and evaluation. The necessary filtering and reclassification of those data can be achieved with interactive computer support routines used by maintenance personnel knowledgeable about the snow and ice control operation and the physical features of the network. In Indiana, field engineers at InDOT district offices are completing this activity, and the quality of the resulting network representation is excellent. In addition to data conditioning, inaccuracies and deficiencies in the original USGS data are easily identified and repaired.

REFERENCES

1. D. H. Mark and R. Stricker. *Routing For Public Service Vehicles.* ASCE, Urban Planning and Development Division, Dec. 1971, pp. 165–178.
2. *The Indiana Department of Highways 1985–86 Snow Packet.* Public Service Office, Indiana Department of Transportation, Indianapolis, 1985.
3. B. L. Golden. *Introduction to Recent Advances in Vehicle Routing Methods.* In *Transportation Planning Models*, M. Florian, ed. Elsevier Science Publishing Co., Inc., N.Y., 1984, pp. 383–418.
4. *Digital Line Graphs from 1:100,000 Scale Maps, Data Users Guide 2*, U.S. Geological Survey, Reston, Va., 1985.

Publication of this paper sponsored by Committee on Winter Maintenance.