A PROCEDURE FOR ALLOCATING ZONAL ATTRIBUTES TO A LINK NETWORK IN A GIS ENVIRONMENT

Thomas Jonsson, Ph.D.
Postdoctoral Fellow
Connecticut Transportation Institute
University of Connecticut
Civil & Environmental Engineering, Unit 2037
Storrs, CT 06269-2037
Voice (860) 486-2340
FAX (860) 486-2298
thoj@engr.uconn.edu

and

Lund University
Dept. of Technology & Society
Box 118, 221 00 Lund, Sweden
Voice +46 46 222 02 75
FAX +46 46 12 32 72
thomas.jonsson@tft.lth.se

Zuxuan Deng
Research Assistant
Connecticut Transportation Institute
University of Connecticut
Civil & Environmental Engineering, Unit 2037
Storrs, CT 06269-2037
Voice (860) 486-0586
FAX (860) 486-2298
zud03001@engr.uconn.edu

John N. Ivan, Ph.D., P.E.
Assoc. Professor
Connecticut Transportation Institute
University of Connecticut
Civil & Environmental Engineering, Unit 2037
Storrs, CT 06269-2037
Voice (860) 486-0352
FAX (860) 486-2298
johnivan@engr.uconn.edu

WORDS: 3646 +250*1 +250*8 = 5896

NOVEMBER 21, 2005

Revised paper for presentation at the 85th Annual Meeting of the Transportation Research Board.
ABSTRACT
This paper presents a procedure for assigning zone data to links in the context of a Geographic Information System (GIS) environment. This may be used for several applications; the one motivating this paper is the need to approximate how much secondary traffic is accessing the main street network from adjacent zones. This will then be used in models to predict the number of accidents on the main road network. The full procedure consists of three sub-procedures: 1) The splitting of links into shorter segments that either are fully located within one zone or act as a border between the same two zones for their entire length, 2) Identification of all link segments either adjacent to or interior to each zone, and 3) Allocation of the zone attributes to the links associated with each zone, according to attributes of the zones and the links, as well as other information describing the area.

INTRODUCTION
Spatial data from Geographic Information System (GIS) databases is being used more and more in transportation planning due to the convenient structure they provide for entering, viewing and manipulating spatially-oriented data. Applications in the traffic safety area have, however, been relatively infrequent to date, limited mostly to visual representation of accident locations (1,2,3). A more recent application is for use in predictive models for accidents in a road network (4, 5, 6). In some cases land use data has been incorporated into the models on a macroscopic level (7, 8).

This paper describes a procedure by which GIS land use inventories are used to identify the land development intensity associated with specific highway links to be used as input to models for predicting crashes at minor intersections on road segments. Specifically, this procedure takes land use data (population and employment by category) identified with geographic zones and allocates it to links representing the road segments passing through or abutting the corresponding zones. The resulting population and employment identified with each road segment can then be used to estimate the volume of traffic entering and exiting the main road as a measure of exposure for estimating and applying crash prediction models. Several sub-procedures have been developed: splitting of links into smaller segments when needed, identifying the links passing through and abutting each zone and calculating weights to specify how much of the zone is associated with each link. The procedure has been developed for use in predictive accident models but parts or all of the procedure can also be applied to other spatially-related transportation planning problems.

BACKGROUND
It is common practice in the realm of motor vehicle crash forecasting to estimate separate models for intersections and the segments connecting intersections. This practice has evolved due to the differences in the types of collisions that typically occur at intersections and on segments. For example, angle collisions can only occur at locations where vehicles approach one another from intersecting paths, and collisions involving turning vehicles can also only occur where vehicles turn on or off the roadway. Both of these situations are only possible at intersections or driveways. In contrast, run-off-road and head-on collisions can occur anywhere, but due to their nature are generally identified as occurring on a segment rather than at an intersection.
The attractive logic behind this approach (of estimating models separately for intersections and segments) breaks down when one attempts to define what an intersection is. The current practice typically only includes intersections between major roads, largely because these are the facilities for which traffic volumes and other road characteristics are available. The availability of traffic volume data is especially critical, since exposure has consistently been demonstrated to account for the larger part of the systematic variation in crash experience (9, 10). Not having traffic volume data available for minor intersections (those between major roads and minor roads or driveways) is problematic, because these intersections introduce the same kinds of road safety risks as do major intersections (those between major roads). Consequently, in order to accurately predict the incidence of crashes at minor intersections, information about the intensity of traffic entering and exiting the minor roads and driveways along the segment is needed.

If traffic volumes at every intersection and driveway on the road network were required, even just estimating such models would be infeasible (not to mention applying them in practice) due to the resulting enormous data collection and management load. However, it may be possible to exploit the link between land development and trip generation (11), if not to estimate these volumes, then at least to represent them as a surrogate. Thanks to the spread of electronic mapping and land use inventories organized using GIS, land development information is now available in many jurisdictions (both state and local level), not only for preparing data sets for model estimation, but also for application and prediction of crash rates.

Land use data has previously been used for predictive accident models on a macroscopic level (7, 8), but then the models have been for blocks or entire TAZs. There has also been some earlier work on approximating the effect of land use on accidents for links (12), but with only the number and types of driveways as a proxy for the land use. The procedure developed in this paper makes it possible to use the number of residents and employees by zone in the development of accident models for links.

DATA

The data used for the development and testing of this methodology was obtained from the Capitol Region Council of Governments (CRCOG), the largest of Connecticut’s regional planning agencies, serving the City of Hartford (the state capital and third largest city) and 28 towns surrounding it. CRCOG is the only planning agency in Connecticut to maintain its own travel demand forecasting model. Conveniently, CRCOG uses ArcGIS by ESRI® to manage the land use data and link network for the model, and made both the land use and road network layers available to us. Additional data about the spatial distribution of land development in the study area was obtained from the Center for Land Use Education and Research (CLEAR) at the University of Connecticut.

The data set obtained from CRCOG provides numbers of residents, retail employees and non-retail employees in each of 1122 traffic analysis zones (TAZ’s) covering the CRCOG region as well as some towns bordering the region (including some in Massachusetts, e.g., the City of Springfield). The zones have a wide range of degree of development, ranging from rural sparsely populated to very dense in the center of Hartford. FIGURE 1 depicts the area covered by the model, indicating the size and distribution of the TAZ’s. The sizes of the zones vary between 0.03 and 53.47 square miles (0.08-138 sqkm), with a mean value of 2.38 square miles (6.16 sqkm). The wide range is due to the fact that the model includes both zones located in CRCOG member towns, which are represented in detail, as well as towns outside the CRCOG region,
which are represented more aggregately. The CRCOG towns are furthermore more highly urbanized than the surrounding towns, which also contributes to their division into smaller TAZs.

The CRCOG network does include centroids connecting the TAZs into the road network, but the locations of the centroids do not necessarily correspond to the distribution of the land development in the zone. To be able to distribute the land use activity of each TAZ to the correct links, additional land cover data was obtained from CLEAR. This data is based on aerial photographs which have been digitally processed to convert them into different land cover types (13), see FIGURE 2.

FIGURE 1 Zone division for the CRCOG region and surrounding towns, generated from CRCOG data.
FIGURE 2 Land cover map including both major roads and local street network, generated from CRCOG and CLEAR data.
METHODOLOGY DEVELOPED

The proposed procedure for linking zone data to links is split into three sub-procedures:

1. Splitting of links in GIS database so that each link segment belong to the same zones for its entire length
2. Recognizing which links belong to a certain zone
3. Allocation of the zone attributes to the links associated with each zone

The development platform we are using is ArcGIS by ESRI®. ArcGIS is an integrated collection of GIS software products for building a complete GIS.

FIGURE 3 shows a generated map over a part of the CRCOG network. Even when a street acts as a boundary of a TAZ the match of the zone boundary and the link is far from perfect, since the zone boundaries and the street links were coded separately. The procedure has to take this into account and in some cases identify links as belonging to a zone when in the GIS network they are located slightly outside the zone, as in reality they coincide with the border of the zone.

FIGURE 3 Road network with TAZ boundaries, generated from CRCOG data.
Splitting of Links

In some cases a link does not border, or fall within, the same zones for its entire length. FIGURE 4 a, b and c show some examples of this. To be able to easily associate links and zones we propose a first step in which all links that are not bordering or being included in the same zone/zones for their entire length are identified and split into smaller segments that do. Each segment of the link gets the same attribute as the original link except for the link length which is recalculated.

There are two primary ways in which the situation is identified, one for the case when a link acts as a border between zones (Type 1, FIGURE 4a), and one for when the link changes from being fully within one zone to being fully within another (Type 2, FIGURE 4b). Situations can also arise when the link changes from being fully within one zone and continues on to become a border link (Type 3, FIGURE 4c), this case is handled in the same way as Type 1.

FIGURE 4 a-c Visualization of the splitting of links, Type 1: Passing several zones (a), Type 2: Intersecting zone boundary (b), Type 3: Changing from boundary link to internal link (c). The links are split at the X.

To identify links belonging to Type 1, a buffer zone is created around all corners in the polygon which denotes a zone. If a link passes through this buffer zone without having a start-end point very close to the buffer zone, it is split at the point closest to the polygon corner (FIGURE 4a). To handle links belonging to Type 2, all links intersecting a zone boundary are split at the place of the intersection (FIGURE 4b). Type 3 is handled in the same way as Type 1. Each link is only belonging to, and handled according to, one of the three types.

Connecting Zones and Links

In order to recognize which links belong to a specific zone, not only the links within the TAZ polygon need to be considered, but also those just outside it. This is due to the fact that the link network and the TAZ polygons don’t match up perfectly even when the link should be the zone border. A buffer zone of 200 feet is defined around each TAZ in which links are also identified as belonging to the TAZ. Several different buffer widths were tested to come up with an ideal buffer width; this buffer width would need to be re-estimated for different networks since the optimal buffer zone width will depend on how well the zone boundaries and links agree.

The zone buffers tested were 50, 75, 100 and 200 feet. TABLE1 shows the results of a test for 20 randomly selected zones where each buffer width has been applied and the length of links not correctly assigned to zones has been totaled (both missed links and false matches). A small number of link segments have been identified as belonging to the TAZ when they
shouldn’t; these are marked with a positive value. Conversely, the total length of links that should have been identified but were not is listed with a negative sign. An OK means all links that belong to the TAZ, and no excess links, have been identified. For no TAZ was there both missed and excess links.

**TABLE 1 Test of Buffer Widths for the Identification of Links Associated with Each TAZ**

<table>
<thead>
<tr>
<th>#</th>
<th>Random Zone Number</th>
<th>50ft</th>
<th>75ft</th>
<th>100ft</th>
<th>200ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>524</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>3</td>
<td>713</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>+400ft</td>
</tr>
<tr>
<td>4</td>
<td>274</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>5</td>
<td>354</td>
<td>-600ft</td>
<td>OK</td>
<td>OK</td>
<td>-1300ft</td>
</tr>
<tr>
<td>6</td>
<td>604</td>
<td>-2500ft</td>
<td>-1300ft</td>
<td>-1200ft</td>
<td>-500ft</td>
</tr>
<tr>
<td>7</td>
<td>895</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>8</td>
<td>281</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>+200ft</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>-3800ft</td>
<td>-3500ft</td>
<td>-2400ft</td>
<td>OK</td>
</tr>
<tr>
<td>10</td>
<td>327</td>
<td>-800ft</td>
<td>-600ft</td>
<td>-300ft</td>
<td>+400ft</td>
</tr>
<tr>
<td>11</td>
<td>934</td>
<td>-2400ft</td>
<td>-1200ft</td>
<td>-900ft</td>
<td>+200ft</td>
</tr>
<tr>
<td>12</td>
<td>82</td>
<td>-850ft</td>
<td>-700ft</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>13</td>
<td>581</td>
<td>-600ft</td>
<td>-600ft</td>
<td>-400ft</td>
<td>OK</td>
</tr>
<tr>
<td>14</td>
<td>275</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>15</td>
<td>791</td>
<td>-2600ft</td>
<td>-800ft</td>
<td>-150ft</td>
<td>OK</td>
</tr>
<tr>
<td>16</td>
<td>135</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>17</td>
<td>310</td>
<td>-3500ft</td>
<td>-3000ft</td>
<td>-2400ft</td>
<td>OK</td>
</tr>
<tr>
<td>18</td>
<td>701</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>19</td>
<td>344</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>20</td>
<td>265</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
</tbody>
</table>

The excessive links that occur for the 200 feet buffer are mostly located on the opposite side of a bordering link, joining it at a sharp angle and thus being in the buffer zone for more than just a length equal to the buffer zone (FIGURE 5).

**FIGURE 5 Example of excessive link, joining at sharp angle the bordering link at opposite side from the zone (760).**
Assigning the Land Use Data to the Links

In the third step the land use data, in the form of number of residents, retail employees and non-retail employees, is assigned to the links. The land use should ideally be distributed according to where the traffic it generates actually enters the major road network. However, the land use data is only given for an entire TAZ, and thus there is a need for a procedure where weights are assigned in more detail to the links according to the portion of the land use activity accesses the network through each link.

The first thought for assigning weights was to assign weights automatically according to the distances between the links and the centroid of the zone. This, however, did not turn out to be feasible as the centroids were discovered not to be located at the center of the development, but more or less arbitrarily.

Instead, a rather straightforward semi-automatic procedure has been used where land cover maps are checked against road maps for the different zones, and if the developed areas of a zone are homogeneously spread along the neighboring or internal links, then only the link length was used as the weight. In the case of non-homogeneously spread land use the weights are determined manually by approximating the portion of the land use area that should be associated with each of the links. The manual assignment of the land use in a zone to each link is very time consuming, and is therefore infeasible to repeat for a large number of zones. In the CRCOG network, the developed areas are fairly evenly spread in nearly all zones; of the 1122 TAZ’s only 53 warranted a manual assignment of link weights.

FIGURE 6 illustrates the process of assigning weights for non-homogeneous zones. The land cover and the minor streets are used to approximate how much of the land use exits on each link. The eastern zone boundary consists of four different links, delineated by the intersections. The southern border consists of one long link. For this zone no links have been split. If the land cover were homogeneously spread, the land use would have been assigned by weights only according to link length. In that case the southern link would have been allotted the majority of the land use. In this case the land cover has been manually examined and the weights have been set equal among all five links, as approximately the same area of developed land is connected to each link.

\[
P_{jk} = \frac{W_{ijk}}{\sum_{j' \in J_{i}} W_{j'k}}
\]

Where:
- \( J_{i} \) = Set of links adjacent or inside TAZ \( i \), \( \forall i \in I \)
- \( p_{ijk} \) = Proportion of zone \( i \) allocated to link \( j \) for land use type \( k \)
- \( w_{ijk} \) = Weight of land use \( k \) on link \( j \) for zone \( i \)

The weights are set to being equal to the link length for zones with a homogeneous land spread. For zones with a non-homogeneous land spread, the weights are assigned manually according to the portion of land cover that exits toward the link.

In FIGURE 6 it can also be seen that there is one street in the northern part of the developed area not assigned to the zone (marked white). This is a centroid connector which is
automatically excluded. The location of the centroid illustrates the previously mentioned problem with the somewhat arbitrary location of the centroids.

In the case of a link that acts as a border between two zones, the link is assigned land use activity from both zones, and the assigned land use is added together.

**FIGURE 6 TAZ with non-homogeneous land use, links in yellow are the links associated with this zone, developed areas and minor streets included in plot.**

**DISCUSSION AND CONCLUSIONS**

The procedure works well in so far as it is able to allocate the zone data to the links. It would be preferable to validate the procedure against actual data regarding how many vehicles are entering and exiting the zones through each link. This, however, requires very detailed data for which adequate resources are not available in the funded project. An estimate of the functionality of the procedure will however be available after the use of the assigned data in accident modeling in so far as how much the assigned zone data can improve the accident models.

The buffer zone used for the identification of links associated with each zone was found to perform best when set to 200 feet, this size is dependent on how well the zone boundaries and the links are in agreement and needs to be recalibrated for each dataset.

The procedure still requires some manual work for establishing the weights. It would be desirable to limit the need for time consuming manual work further. One possibility for automating the process would be to assign the land use by automatic image processing of land cover maps. The developed areas are however not always connected to the nearest major streets, thus calling for the need of a more complicated analysis where the minor street network is used to aid in the determination of which major street the land use activities generated will be exiting and entering through.
To conclude, the developed procedure provides the possibility to allocate zonal data to a link network in a semi-automated manner. Further development is still needed to enhance the procedure; for instance the calibration of the procedure against actual trip data on a detailed level. Although the CRCOG model offers a wide variety in zone and link density it is still desirable to test the procedure on other networks to test the transferability of the results.

ACKNOWLEDGMENTS
This research was funded by the New England Transportation Consortium, funded under grant number NETC 04-5, and was conducted at the Connecticut Transportation Institute. The authors gratefully acknowledge Eric Snowden at CRCOG for providing the GIS data and answering questions about it, and Dr. Lisa Aultman-Hall, Jianhe Du, and Darrell Smith at the Connecticut Transportation Institute (University of Connecticut) for assistance with analytical manipulation of the GIS data.

REFERENCES
10. Jonsson, T. Predictive models for accidents on urban links – A focus on vulnerable road users, Department of Technology and Society, Lund University, Lund, Sweden, 2005.