

# TRANSIMS Training Course at TRACC

Transportation Research and Analysis Computing Center

## Part 4

### Generating Activities for Members of TRANSIMS Synthetic Households

**Dr.-Ing. Hubert Ley**

Transportation Research and Analysis Computing Center

Last Updated: December 9, 2009

Unit **4**



TRACC - TRANSIMS Training Course

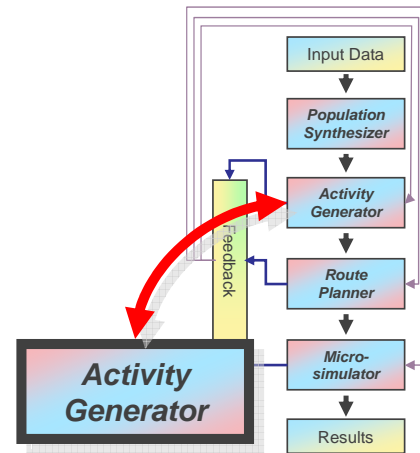
## Contents

- Introduction
- Data Sources and Framework
- Basic Methodology
- General Description of the Activity Generator
- Survey Data Preparation
- The Survey Household File
- The Survey Population File
- The Survey Activity File
- The Survey Weights File
- Household Matching
- Household Type Scripts
- Household and Person Matching
- Location Choice – Anchor Locations & Primary Activities
- Location Choice – Location Choice for Activities
- Estimating Travel Conditions



## The TRANSIMS Activity Generator

- The activity generator is used to determine the “demand” for trips that will be loaded onto the road and transit network
- The activity generator uses the synthetic population generated in the previous step to assign transportation actions to all individual people in the given households
- Additional data is needed to assign individual activities to all simulated individuals
  - The main input data is a detailed activity survey that is representative for the simulated area
  - The data consists typically of a few thousand records derived from telephone surveys

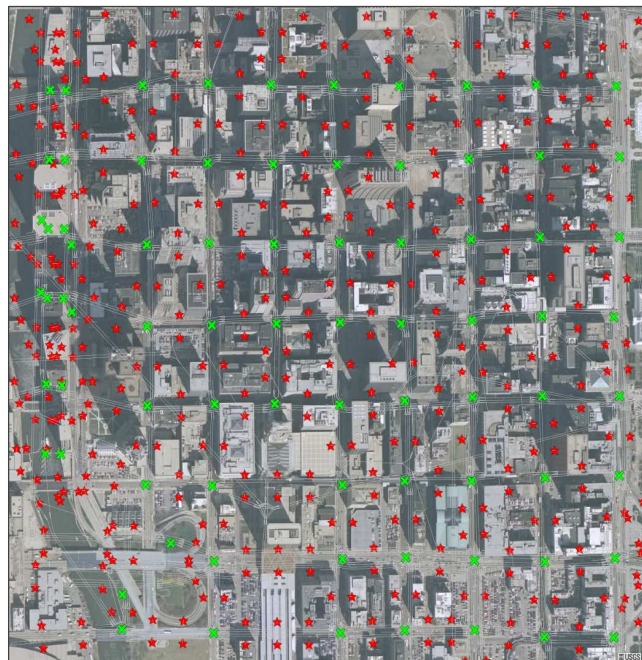


- The records capture all activities of interviewed individuals throughout the day, e.g.
  - 7am to 8am breakfast
  - 8am to 9am drive to work
  - 9am to 1pm work
  - ...

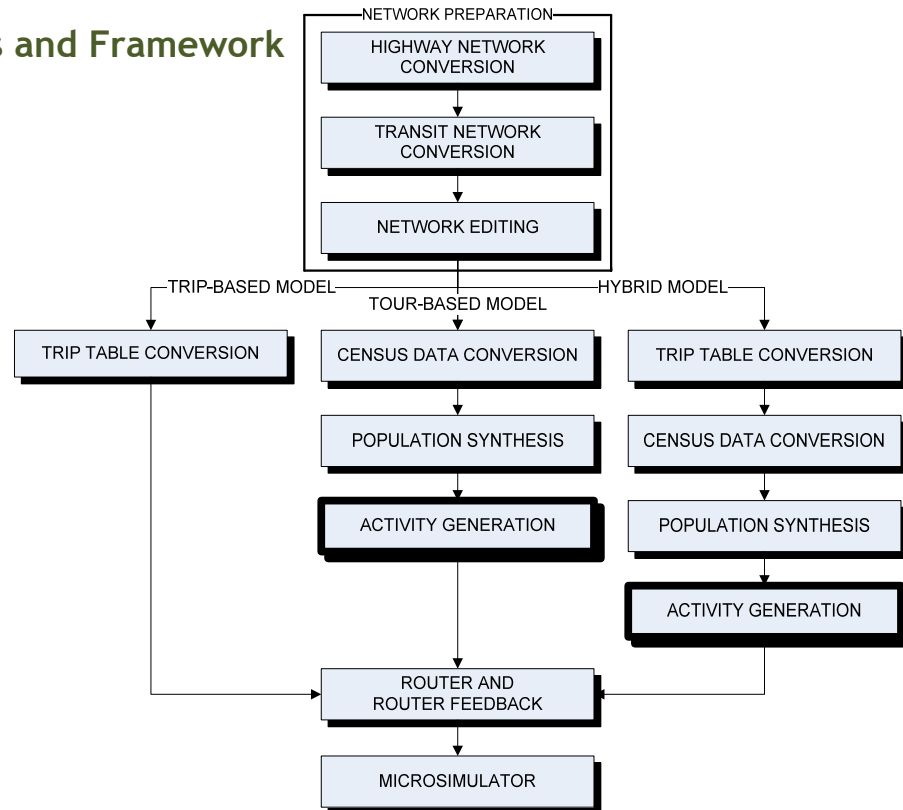


## Introduction

- The activity generator develops a list of activities for each member of a synthetic household over a 24-hour period.
- The **ActGen** program performs the following tasks:
  - Assigns activities to each household member,
  - Determines start and end-time of activities over the course of a day,
  - Selects a location for each activity, and
  - Defines the travel mode used to travel to that location.



## Data Sources and Framework



5

## General Description of the Activity Generator

- From the **TRANSIMS** perspective, travel is a means of accomplishing activities at different locations.
- “Activities” in this context consist of the things people do over the course of a typical day, such as work, shop, attend school, eat, and sleep.
- The activities can be performed at home or at other locations.
  - Activities performed at home are labeled as “at home” activities, while activities performed outside the home are collectively grouped as “tours”.
  - A tour comprises all the activities that take place from the time a person leaves home until the time that person returns home.
  - A person can conduct several tours throughout the day alone or with other household members.
  - A person may also perform one or more sub-tours during a tour. A sub-tour is a series of activities that start and end at a given location. The work location is often the anchor point for sub-tours.

6

## General Description of the Activity Generator

- Some activities are considered flexible or discretionary, while others are fixed or mandatory.
  - Work and school, for example, are likely to be mandatory activities that have fixed start and end times.
  - Shopping activities, on the other hand, are typically flexible, with more variability in location and duration.
- **TRANSIMS** enables the user to define the time and flexibility constraints for each activity type or purpose.
  - The program considers this information together with the travel time between activity locations when constructing a realistic activity schedule for each household member.
  - If several household members need to coordinate their activities or their travel (for example, to drive a child to school), those constraints are considered



## General Description of the Activity Generator

- The **ActGen** program uses a household activity survey to define the activity patterns, activity schedule, and travel modes for each synthetic household member generated by the **PopSyn**.
- A household type model allows the synthetic household to be matched with a survey household based on household demographic attributes.
- The activities for each person in the survey household are then assigned to an appropriate person in the synthetic household.
- The household activities are allocated to activity locations based on the home location and location choice models.
  - Location choice models are typically defined by activity purpose (work, shopping, etc...) and travel mode (walk, drive, etc...).
  - The default model uses the straight-line distance between potential activity locations and known locations, along with location attraction factors, to select one of the potential locations for the activity.
- Location choice scripts and zone to zone skims can be used for more sophisticated models.



## Survey Data Preparation

- The survey data are provided in four files:
  - A survey household file
    - The household file defines the number of persons and vehicles in the household, as well as any additional household attributes needed to define household type and location choice (such as number of workers, income, and area type)
  - A survey population file
    - The population file includes a data record for each person in the household; these records typically identify the person's age, gender, work status, and other attributes
  - A survey activity file
    - The activity file includes the sequence of activities performed by each household member over the course of a day. The purpose, start time, end time, travel mode, vehicle number, number of passengers, and location is provided for each activity
  - A survey weights file



9

## The Survey Household File

- For example, household 200010 has 2 persons, one of which is a worker. The household owns 2 vehicles and belongs to income category 11 (of 14 categories in total).
- The location field has no impact on the activity generation process since synthetic households have already been assigned to an activity location in the **PopSyn** process.

HHOLD	PERSONS	WORKERS	VEHICLES	INCOME	LOCATION
200007	1	1	1	14	-1
200009	1	0	0	1	-1
200010	2	1	2	11	-1
200015	1	0	0	14	-1
200020	1	0	0	2	-1



10

## The Survey Population File

- For example, household 200010 has 2 records, one for each person
- Person 1 is a 25 years old (AGE = 25) male (GENDER = 1) who works (WORK = 1)
- Person 2 is a 28 years old (AGE = 28) female (GENDER = 2) who does not work (WORK = 2)
- RELATE is a 1990 US Census acronym that indicates the relationship of a household member to the householder. It has no impact the matching process within **ActGen**.

HHOLD	PERSON	AGE	GENDER	WORK	RELATE
200007	1	29	2	1	1
200009	1	82	2	2	1
200010	1	25	1	1	1
200010	2	28	2	2	1
200015	1	82	1	2	1
200020	1	77	1	2	1



11

## The Survey Activity File

- Introduction of several lookup tables used in the survey activity file
- Purpose Codes**
  - Purpose 0 indicates that the person is at home (required)
  - Other codes can be assigned more flexibly, e.g.

- The **Vehicle Codes** are defined as follows:

**0 No auto**  
**1 Auto, driver**  
**2 Auto, passenger**

**0 Home**  
**1 Work**  
**2 Shop**  
**3 Visit**  
**4 Social/Recreation**  
**5 Other**  
**6 Serve Passenger**  
**7 School**  
**8 College**



12

## The Survey Activity File

- The **Mode Codes** are defined as shown in the list on the right:
- The **Passenger Code** represents the number of passengers besides the driver
- **Location Codes** have meaning only for a given household. They simply tell the program to change location or to return to a previously defined location.
  - For example, the home location is represented by location code 1. Each time the program finds a “1” in the location field, it returns to the activity location where the household is located.

1	Walk
2	Drive
3	Bus
4	Rail
5	Park-and-Ride Outbound
6	Park-and-Ride Inbound
7	Bicycle
8	Magic Move*
9	School Bus
10	Two-Person Carpool
11	Three-Person Carpool
12	Four-Person Carpool

\* Magic move: travel with non-household members



## The Survey Activity File

HHOLD	PERSON	ACTIVITY	PURPOSE	START	END	DURATION	MODE	VEHICLE	LOCATION	PASSENGER
200010	1	2	3	11:10	11:25	0:15	2	2	2	1
200010	1	3	2	11:50	13:20	1:30	2	2	3	1
200010	1	4	5	13:40	14:30	0:50	2	2	4	1
200010	1	5	2	14:45	15:00	0:15	1	0	5	0
200010	1	6	0	15:21	16:01	0:40	2	2	1	1
200010	1	7	1	16:05	22:30	6:25	2	2	6	1
200010	1	8	0	22:45	24:00:00	1:15	8	2	1	1
200010	1	9	0	24:00:00	27:00:00	3:00	1	0	1	0
200010	2	1	0	0:00	11:10	11:10	1	0	1	0
200010	2	2	3	11:10	11:25	0:15	2	1	2	1
200010	2	3	2	11:50	13:20	1:30	2	1	7	1
200010	2	4	5	13:40	14:30	0:50	2	1	4	1
200010	2	5	2	14:45	15:00	0:15	2	1	5	1
200010	2	6	0	15:21	15:51	0:30	2	1	1	1
200010	2	7	6	16:05	16:06	0:01	2	1	6	1
200010	2	8	0	16:20	19:00	2:40	2	1	1	1
200010	2	9	0	19:00	21:00	2:00	1	0	1	0
200010	2	10	0	21:00	24:00:00	3:00	1	0	1	0
200010	2	11	0	24:00:00	27:00:00	3:00	1	0	1	0



## The Survey Weights File

- These weights are used in the matching process to select a survey household for each synthetic household. Therefore among survey households with the same type, the probability of selecting a survey household is proportional to its survey weight.

200007	0.008636767
200009	0.021074192
200010	0.011871295
200015	0.007345929
200020	0.008132641
200029	0.145824963



## Household Matching

- The activity assignment process starts by matching each synthetic household to a survey household.
  - Synthetic households have been created and placed by **PopSyn**
- The characteristics of the household are used to classify the household type. The program then
  - identifies all of the survey households with the same household type
  - randomly selects one of those households, based on survey weights
  - assigns its activities to the synthetic household
- Household type codes can be hard coded into the household data files or defined by a household type script.
- Household type scripts use the same variables defined for the household file such as household size, number of vehicles owned, and number of workers to assign a specific household type code to the household.





## Household Type Scripts

- Some of the TRANSIMS tools use a special scripting language to modify tabular input data on the fly based on user-specified algorithms.
- If the survey household file and the synthetic household file use different variable names, then a separate household type script must be used for each.
- The following example shows a simple household type script that uses the household size and number of workers to define six household types.

```

IF (Household.PERSONS > 2) THEN
  IF (Household.PERSONS > 3) THEN
    IF (Household.WORKERS == 2) THEN
      RETURN (1)
    ELSE
      RETURN (2)
    ENDIF
  ELSE
    IF (Household.WORKERS == 1) THEN
      RETURN (3)
    ELSE
      RETURN (4)
    ENDIF
  ENDIF
ELSE
  IF (Household.WORKERS == 1) THEN
    RETURN (5)
  ELSE
    RETURN (6)
  ENDIF
ENDIF

```



17

## Household and Person Matching

- Given the household type, the **ActGen** tool randomly selects a household from the survey dataset with the same type.
- It then gathers the members of the survey household for the person-matching process.
  - The program matches each person in the synthetic household to the most appropriate person in the survey household.
  - This is done using the age, work status, and gender fields in the population files.
  - Each person is first assigned to an age-group, as shown on the right.
- If the survey household does not have a person in the same age-group and with the same worker status and gender as a person in the synthetic household, **ActGen** will attempt to identify the best match by age-group and then work status.
- The program then retrieves the out-of-home activities from the survey activity file and assigns the activity sequence, activity schedule, and travel mode to the synthetic household member.

<i>Age Group</i>	<i>Age Range</i>
<b>1</b>	<b>0–4</b>
<b>2</b>	<b>5–11</b>
<b>3</b>	<b>12–15</b>
<b>4</b>	<b>16–20</b>
<b>5</b>	<b>21–64</b>
<b>6</b>	<b>65+</b>



18

## Location Choice - Anchor Locations & Primary Activities

- The location choice model uses “anchor” activities to select the location for each activity in a tour
  - Basically, two previously defined locations, a “previous” and a “subsequent” anchor location, are needed to select the location for the next activity in the tour.
  - Initially the home location is the only anchor location.
  - The program then scans all the activities included in a tour that start and end at home to identify the primary activity or reason for the tour.
  - If an activity type is identified as an anchor activity in the **ActGen** control file, it will be considered first.
  - If the tour has multiple anchor activities, the program will select the anchor activity with the longest duration as the primary activity.
  - If the tour does not include any anchor activities, the program will consider the activity with the longest duration to be the primary activity for the tour.



19

## Location Choice - Location Choice for Activities

- Once the primary activity is identified, the program uses the home location as both the “previous” and “subsequent” anchor location for locating the primary activity.
  - The time at which the trip leaves home is used to determine the impedance for the outbound trip
  - and the time at which the trip leaves the primary activity is used to determine the impedance for the return trip
- Once the primary activity is located, the program will locate the activities that take place between home and the primary activity based on the distance to home and the distance to the primary activity location.
- After the first intermediate stop is located, it becomes the “previous” anchor location for locating the next stop on the tour.
- The process continues until all of the activities are located.



20

## Estimating Travel Conditions for the Location Choice Model

- The **ActGen** program provides two ways for estimating the travel conditions for the location choice model.
  - The first is the default method, which uses simple distance-based calculations as a proxy for travel time.
    - The distance-based model is at least 10 times faster than a skim-based script model. It requires far less data, and the data takes less time to prepare. On the other hand, it does not consider time of day, mode, or path — considerations that will be very important in any detailed study in a complex region
  - The second is a more advanced approach, which uses zone-to-zone skim files and a modeling script to calculate the probability of selecting a given location.
    - Because skim data can be defined by mode and time of day and then adjusted by household attributes (such as income) or by activity purpose, the skim files can be used to implement very sophisticated location choice models.



21

## The Default Location Choice Model (Simple Model)

- “Location\_Weight” is an optional field in the activity location table representing the relative attractiveness of each individual location (e.g. representing employment) (default value is 1.0)
- Different fields can be defined in the activity location table for different activity types
- “Location\_Weight\_Factor” is a key that can be specified to scale the field values by a fixed factor (default value is 0.0)
- “Distance” is the computed distance between previous and subsequent anchor locations and a given location “n”
- A “Mode\_Distance\_Factor” can be used as a scalar parameter to adjust the distance based on the operation characteristics of each mode.

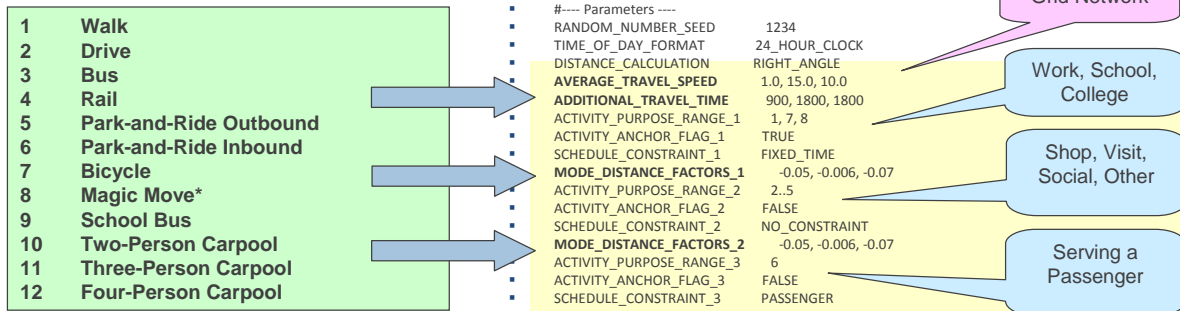
$$Utility_n = Location\_Weight_n \times Location\_Weight\_Factor_n \times \exp [Mode\_Distance\_Factor_{mode} \times (Distance_{previous \rightarrow n} + Distance_{n \rightarrow subsequent})]$$



22

## A Sample Control File

- Distance-based model is too simple, but allows for debugging of survey data and requires significantly less preparation than the more extensive skim-based methods



23

## Credits and Acknowledgements

- GIS visualization materials were mostly developed at Argonne based on the TRANSIMS tools developed by AECOM for USDOT
- Chicago road and transit network data used in some of the examples was provided by the Chicago Metropolitan Agency for Planning
- USDOT provided the funding for the development of these training materials
- USDOT provided the funding for the TRACC computing center and the resources necessary to perform these training session
- Many of the viewgraphs are directly derived from AECOM work sponsored by USDOT FHWA

24