Building a Realistic Data Environment for Multiagent Mobility Simulation

Feirouz Ksontini[1,2], Mahdi Zargayouna[1], Gérard Scemama[1], Bertrand Leroy[2]

[1] Université Paris-Est, IFSTTAR, GRETTIA
Boulevard Newton, Champs sur Marne
F-77447 Marne la Vallée Cedex 2, France
[2] Institut VeDeCom,
Versailles, France
{hamza-mahdi.zargayouna,gerard.scemama}@ifsttar.fr
{feirouz.ksontini,bertrand.leroy}@vedecom.fr

Abstract. Transport systems are increasingly complex and are made of more and more connected entities. It becomes critical to develop micro-simulation tools to understand the new transport systems dynamics. However, the data for building mobility simulation are quite hard to get, and simulations on new areas are not easy to set up. In this paper, we propose methods for building a realistic data environment for multi-modal mobility simulators. We also propose a method to integrate travel patterns (patterns of travelers' origins and destinations). The methods presented in this paper can be used when dealing with new areas for which we have few and incomplete data.

1 Introduction

The development of new information and communications technology contributes to the emergence of a new generation of multimodal real-time services that assist travelers throughout their trip (mobile devices, localized vehicles, trackable goods, etc.). These information flows increasingly impact travelers behaviors and the traffic generated by their local decision making. In this context, dynamic simulation of travelers mobility is a crucial step for understanding, analyzing and predicting this evolving dynamics of transport networks.

The multiagent paradigm is relevant for the simulation of urban transport systems [1]. It indeed facilitates an approach by analogy in the transport domain which one of the objectives is the coordination of distributed entities [2]. This is why the multiagent approach is often chosen to model, solve and simulate transport problems [3]. The authors in [4] list several reasons for the privileged use of multiagent systems in these applications, such as the natural and intuitive problem solving, the ability of autonomous agents for the modeling of heterogeneous systems, the ability to capture complex constraints connecting all problem-solving phases, etc. Indeed, the concept of an agent is well suited for the representation of travelers in transit or road traffic scenarios [5]. They are autonomous entities which are situated in an environment, adapt their behaviors to the dynamics they perceive and interact with others agents in order to achieve specific goals. For Parunak [6], "Agent-based modelling is most appropriate for domains characterized by a high degree of localization and distribution", which is the case for complex and dynamic transport applications.

We have designed and implemented the multimodal travel simulator SM4T (Simulator for Multiagent MultiModal Mobility of Travelers) in the context of the EC-funded project Instant Mobility [7]. The simulator allows for the understanding and the prediction of future status of the networks and it can be also used for testing new applications that track individual travelers. For instance, it has been used to evaluate the impact of individualized real-time data on the behaviors of traveler agents [8]. SM4T is a fully agent-based tool for multimodal travelers mobility. It enables for the rapid prototyping and execution of simulations for several kinds of online applications. The application simulates the movements of travelers on the different transport modes and networks while taking into account the changes in travel times and the status of the networks. Since it assumes the continuous localization of travelers, SM4T can notably simulate and evaluate the impact of a wide range of community transportation apps, such as user-submitted travel times and route details, community-based driver assistance, community parking, etc.

However, multimodal mobility simulators such as SM4T need a lot of data about the considered geographic region and the different existing transport modes. For instance, SM4T has been deployed on the city of Toulouse (France) with the geographic data of the city, the description of the road network, the description of all the available public transport network, the timetables of the vehicles, etc. These data were made available to us from the operating support system of public transport operators and from road transport support systems of road transport operators. This great amount of data of different nature limits the applicability of multimodal mobility simulators to new areas. To this end, this paper focuses on building a realistic data environment for multimodal mobility simulators in the absence of access to these proprietary data. We also propose a method to integrate travel patterns (patterns of travelers' origins and destinations) in these simulators.

The remainder of this paper is structured as follows. In section 2, we discuss the choice of the simulation platform and previous proposals for travelers mobility simulation. In section 3, we briefly present SM4T. Section 4 presents the methods to use when dealing with a new area to simulate. In section 5, we describe some experiments before to conclude and describe some further work we are conducting.

2 Related Work

There exists several multiagent simulators for travelers mobility. For instance, MATSim [9] is a widely known platform for mobility micro-simulation. However, the mobile entities in MATSim are passive and their state is modified by central modules, which limits its flexibility and its ability to integrate new types of (proactive) agents. Transims [10] simulates multimodal movements and evaluates impacts of policy changes in traffic or demographic characteristics. AgentPolis [11] is also a multiagent platform for multimodal transportation. The proposals of this paper might profit to these platforms since they also require the same kind of information than SM4T. SM4T has been developed on top of Repast Simphony [12]. In the context of transportation applications, one main choice criterion for the simulation platform is its ability to create geospatial agent-based models, i.e. its ability to integrate and process geographic data. Among all the available simulation platforms that would fit with our requirements (e.g. Gama [13]), Repast is the most mature one.

3 The SM4T Simulator

The purpose of SM4T is to represent travelers (drivers and passengers) and transport means (public transport vehicles and private cars) in a micro-level and to simulate their dynamic movements and their interaction (tracking, planning requests, plans update, etc.). In the following, we briefly describe the simulator, a more detailed description can be found in our previous paper [7].

The multiagent system is made of planner agents, car agents, public transport vehicle agents and traveler agents. The planner agents compute the best road itinerary for the car agents and the best multimodal itinerary for the traveler agents based on the latest status of the networks. The calculation method has been detailed in [14]. A planner agent is created when an agent request is submitted to the system, and leaves the system right after.

Each car agent has an origin and a destination when created, which are chosen randomly. The agent asks the planner for the best itinerary between his origin and his destination. At each simulation tick, the car agent checks if he has reached his destination. If so, he leaves the simulation.

The origins and the destinations of the public transport vehicle agents are provided by the predefined timetables. When created, each vehicle agent infers his itinerary from his timetable. If there are passengers onboard, they are moved to the same coordinates at the same time by the vehicle agent. That means that, when they are onboard a vehicle, traveler agents delegate the control of their movements to the vehicle agent. While the vehicle agent has not reached his destination, he travels at each tick the allowed distance, following his current speed. When the vehicle reaches a stop, he searches among his onboard travelers who has to leave at this stop. Then he searches among the waiting travelers at the stop who has to take him.

As for car agents, the origin and destination of the traveler agent are chosen randomly. When they are not walking, traveler agents do not travel on their own, but share rides with others, which are responsible of their movements. The traveler agent alternates between walking and waiting for a vehicle.

4 Network Data

As for any multimodal mobility simulator, SM4T needs a minimal set of data to function properly. The minimum input data (xml files) of the simulator are:

- the road network,
- the public transport network,
- the transfer mapping,
- the timetables of the public transport vehicles,

Optionally, the simulator might use travelers profiles and use them to infer certain agents characteristics (pedestrian speeds for instance). In the first version of SM4T, all these data were made available to us from the transport operators of the considered area (Toulouse, France). When considering new areas, these data have to be approximated otherwise.

4.1 Road Network

The road network is a description of the roads, crossroads and driving directions. Apart from the geographic description of the roads, mobility simulators need to have a mapping between traffic flows (vehicles/hour), the traffic density and the speeds. Indeed, to make vehicles move in a realistic way, they should not always move in free-flow speed, but should slow down when traffic becomes dense.

When dealing with a new territory, the geographic description of the network can be found in the form of free editable maps (such as OSM). However, the mapping between the number of vehicles and the speeds is generally missing in these maps. To make vehicles move in a realistic way, we approximate this mapping by analogy from data that we have about other areas. The objective is to have a realistic triangular fundamental diagram of traffic flow that gives a relation between the flow q (vehicles/hour) and the density k (vehicles/km). The fundamental diagram suggests that if we exceed a critical density of vehicles k_c, the more vehicles there are on a road, the slower they will be. Here is the equation we use to model this phenomenon:

This equation is parametrized with α the free flow speed on this road, β the congestion wave speed and k_c the critical density. As $v = \frac{q}{k}$:

$$v = \begin{cases} \alpha & \text{if } k \leq k_c \\ \frac{-\beta(k-k_c)+\alpha k_c}{k} & \text{if } k > k_c \end{cases} \tag{1}$$

Thus we can define a cost function that returns a travel time per distance units ($1/v$) in function of the number of agents $|A_e|$ on this edge:

$$cost(|A_e|) = \begin{cases} \frac{1}{\alpha} & \text{if } |A_e| \leq k_c \\ \frac{|A_e|}{-\beta(|A_e|-k_c)+\alpha k_c} & \text{if } |A_e| > k_c \end{cases} \tag{2}$$

When a car agent joins an edge of the network, his speed is calculated following the equation above. Another method for defining a realistic behavior of

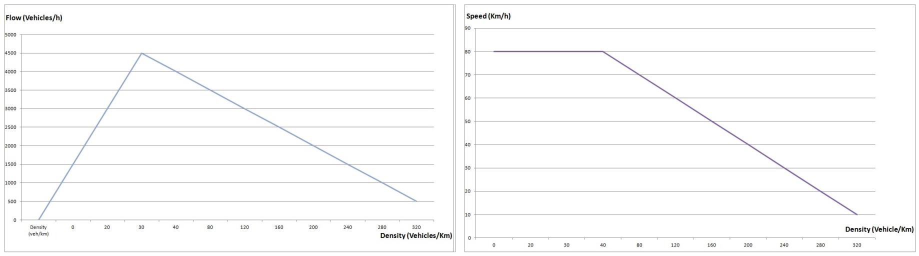

Fig. 1. Fundamental diagram with $\alpha = 112{,}5$, $\beta = 12{,}5$ and $k_c = 40$. (left). Speed in function of density (right).

cars flow is to specify drivers behaviors that depend of the surrounding cars behaviors, generally using a car-following model. Such model would need however a great number of parameters. Including car-following behaviors in SM4T is one of our ongoing works.

4.2 Public Transport Network

A public transport network is composed of two elements. The first element is the network, which is described by a set of transport lines, each of them composed of a set of itineraries, each itinerary is composed of a sequence of edges, each edge has a tracing in the form of a sequence of pairs $\langle longitude, latitude \rangle$, and is composed of an origin node and a destination node. Finally, every node is defined by its name and coordinates. The second element is the timetables of the vehicles, which describe the paths of the vehicles and the corresponding visit times. Each timetable is then a sequence of pairs $\langle stop, time \rangle$.

When not available from the public transport operator, these data are quite hard to recreate. However, increasingly, we can find some open data describing the network in terms of lines, but without the itinerary details, and without the geographic tracing between the stops. To approximate the geographic tracing between stops, we define the following procedure, using the road network defined in the previous subsection:

1. find the road transport edge to which the origin and destination stop belong. If a stop doesn't belong to any road, find the closest road to it,
2. compute a road shortest path between the two roads,
3. get the geographic tracing of the shortest path and add it as a tracing of the edge.

After several tests, we have verified that this method gives a quite accurate description of the public transport network.

To recreate the timetables, the frequencies of the lines have to be found. The general public web portals of the operators can be used. By submitting transport requests at several times of the day (peak and off-peak), frequencies can be approximated. We create vehicles departure times following the description of the lines accordingly, and with visit times at the stops that are coherent with the geographic tracing that we have defined in the previous step.

4.3 The transfer mapping

The transfer mapping is a table informing about the stops of the network for which a transfer by foot is possible and the road transport nodes that are reachable from the stops. This mapping is very important because passengers start and end their trip on the road network (while the main part of their trip is on the public transport network). They thus have to pass from one network to another. To recreate this file, we start at each stop of the public transport network and look for all the reachable stops and nodes (crossroads) that are at most 500 meters away.

4.4 Travel Patterns

A travel pattern clusters the considered geographic region in zones and describes the number of persons asking to leave or to join each region. Travel patterns are very important because they allow to have a simulation that mimics more realistically the mobility behaviors of cars and passengers. In order for the simulator to integrate travel patterns, we propose the following procedure.

	Z_1	Z_2	Z_3
Z_1	0	8	4
Z_2	4	0	2
Z_3	2	0	0

Table 1. Example of D matrix ($N = 3$)

	Z_1	Z_2	Z_3
Z_1	0	4	2
Z_2	2	0	1
Z_3	1	0	0

Table 2. Example of S matrix with $|A| = 10$ and $|M| = 20$

Let the travel pattern for the considered period of time in the form of a matrix $D = \{(d_{ij})\}$ of dimension $N \times N$ with N the number of regions; d_{ij} is the number of persons traveling from zone i to zone j. For instance, in table 1, the number of travelers in the travel pattern is $M = 20$, 8 are going from zone 1 to zone 2, 4 are going from zone 2 to zone 1, 2 ar going from zone 3 to zone 1, etc. However, the number of actually simulated agents A is not necessarily equal to the number of persons M in the travel patterns. Our objective is to generate non-deterministic simulated origins and destinations that are proportional to the travel pattern. To this end, we first create a matrix $S = \{(s_{ij})\}$ of dimension $N \times N$; s_{ij} is the number of simulated agents that will be traveling from zone i to zone j, $s_{ij} = d_{ij} \times \frac{A}{M}$, where A is the number of simulated agents and M the number of actual travelers in the pattern, i.e. $M = \sum_{i=0}^{N} \sum_{j=0}^{N} d_{ij}$. Based on S, a dynamic mapping table P is created (cf. Table 3). The P table maps intervals with $(zone_{origin}, zone_{destination})$ pairs. The length of the interval is proportional to the current relative weight of the zones pairs in S. When an agent is generated with an origin belonging to Z_1, and a destination belonging to Z_2, the (Z_1, Z_2) cell in S is decremented and P is updated accordingly (cf. Table 3).

Interval	Origin-Destination	Interval length calculation
$[0\%, 40.00\%[$	$Z_1 - Z_2$	$\frac{4}{10}$
$[40.00\%, 60.00\%[$	$Z_1 - Z_3$	$\frac{2}{10}$
$[60.00\%, 80.00\%[$	$Z_2 - Z_1$	$\frac{2}{10}$
$[80.00, 90.00\%[$	$Z_2 - Z_3$	$\frac{1}{10}$
$[90.00\%, 100\%[$	$Z_3 - Z_1$	$\frac{1}{10}$

Table 3. Example of P table

Interval	Origin-Destination	Interval length calculation
$[0\%, 33.33\%[$	$Z_1 - Z_2$	$\frac{3}{9}$
$[33.33\%, 55.55\%[$	$Z_1 - Z_3$	$\frac{2}{9}$
$[55.55\%, 77.77\%[$	$Z_2 - Z_1$	$\frac{2}{9}$
$[77.77, 88.88\%[$	$Z_2 - Z_3$	$\frac{1}{9}$
$[88.88\%, 100\%[$	$Z_3 - Z_1$	$\frac{1}{9}$

Table 4. P after the choice of (Z_1, Z_2)

At each tick of time, the simulator generates a number of new traveler agents for which we have to define an origin and a destination. For each new traveler agent, a random number $\rho \in [0, \ldots, 1]$ is chosen and P is used to choose the origin and destination zones. Let's say that $\rho = 0.15$, then following the P table in Table 3, the $Z_1 - Z_2$ pair is chosen for the agent. The origin is then chosen randomly in Z_1 while the destination is chosen randomly in Z_2. The new P table is given in Table 4 and the probability to chose that pair of zones again becomes lower ($\frac{3}{9}$). This way, even if the origin and destination are chosen nondeterministically, the chosen origin and destination zones remain proportional to the travel pattern all along the simulation.

However, travel patterns are usually defined after long surveys on big geographic regions, while the simulations often concern smaller areas. That means that we will have a huge D matrix with zones pairs that mostly do not concern the considered area. We could simply consider the submatrix $D' \subset D$ with zones pairs in the considered area, but in this case, we would have underestimated volumes in D'. Indeed, four cases for the values $d_{ij} \in D$ are possible:

1. both i and j are in the considered area
2. i is the considered area but not j
3. j is the considered area but not i
4. neither i nor j are in the considered area

For the first case, d_{ij} are simply copied in d'_{ij}, since these volumes completely concern the considered area. For the second case, we should add d_{ij} to a certain cell d'_{ik} where k is in the considered area. To do so, we execute a shortest path from the centroid of zone i to the centroid of zone j and report the sequence of zones $i, \ldots, k, \ldots, j$ that a traveler going from i to j would visit: k is the last zone in the considered area. The volumes in d_{ij} are then added to d'_{ik}. A similar procedure is followed for the third case, where we add d_{ij} to the cell d'_{kj} where k is the first zone in the zones shortest path sequence $i, \ldots, k, \ldots, j$. For the last case, we could think of ignoring them, since neither the origin nor the destination zone are in the considered area. However, some travelers, even if they are not departing from nor arriving in the considered area, could pass by the considered area, and should therefore be considered in the simulation because they impact traffic. Again, we execute a shortest path and report the sequence of zones $i, \ldots, k, \ldots, l, \ldots j$ that a traveler going from i to j would visit: k is the first zone in the sequence that is in the considered area, while l is the last zone in the sequence that is in the considered area; d_{kl} is added to d'_{kl}.

Figure 2 summarizes the transformation process of the travel pattern. On the left, we have an example of the three last cases of travelers flows: an incoming flow to the considered area, an outgoing flow and a traversing flow. On the right, we have the result of our procedure with a restriction of the travel pattern to origin and destination zones of the considered area.

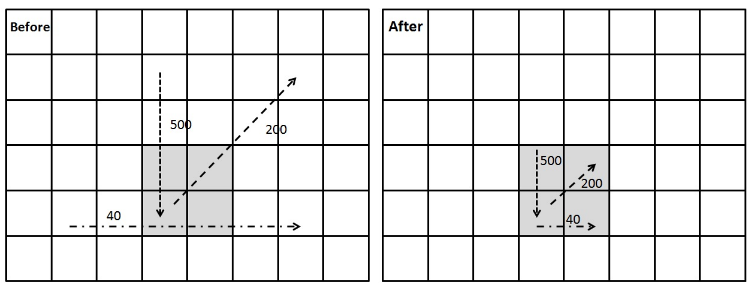

Fig. 2. Restriction of the travel pattern to the considered area (grey zone)

5 Experiments and results

We demonstrate the use of the SM4T simulator on a new area, for the Paris-Saclay region, France, for which we have built a realistic data environment. For public transport network, we have partially relied on open data[1] with GTFS[2] format. GTFS defines a common format for public transportation schedules and associated geographic information. These data represent all the Ile-de-France, we have extracted those that match with the simulated region. We have also travel patterns for all the Ile-de-France[3]. We have used all the procedures described in this paper to build the data environment of the simulator. The result is a road network with 23,594 roads, a public network with 7,431 stops and 7,431

[1] provided by the STIF : the organizing Authority of sustainable mobility in the Ile-de-France

[2] General Transit Feed Specification

[3] Provided by LVMT Lab from MODUS model which was developed in collaboration with DRIEA-IF

edges. We have simulated 18,178 buses, 30,000 cars and 30,000 passengers. A screenshot of our current simulation is in Fig. 3 (blue squares are buses, pink circles are cars and black crosses are pedestrians).

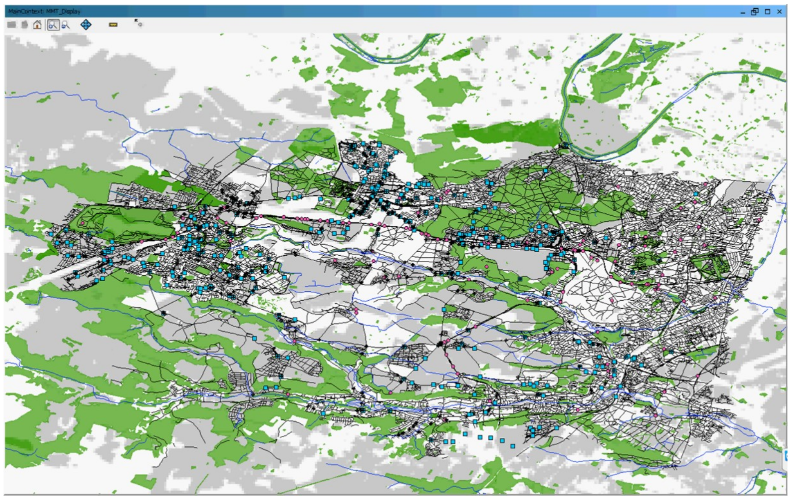

Fig. 3. Simulation Execution (cars: pink circles, travelers: black crosses, buses: blue squares)

The validation of the simulated dynamic status of the network is very important. For a multimodal traffic simulator, the simulated traffic status have to be confronted with the real traffic. This is however a very difficult process, because the data are difficult to get from the transport operators, and general public real-time information only concern main streets and highways and generally no information about public transport vehicles positions and occupancy is provided. We are currently investigating methods to validate the simulation results.

6 Conclusion and Perspectives

In this paper, we have focused on the reconstruction of input data needed for a multimodal mobility simulator and built a realistic data environment for it. This realistic data environment could be used by any other mobility simulator. We have applied these methods on the SM4T simulator with a new territory to simulate: the Paris-Saclay region. Now that we have a running simulation on this territory, the next step that we want to explore consists on testing new services on it. For instance, we are investigating the provision of an autonomous vehicles service and evaluate its impact on the surrounding traffic.

References

1. Badeig, F., Balbo, F., Scemama, G., Zargayouna, M.: Agent-based coordination model for designing transportation applications. In: Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on, IEEE (2008) 402–407
2. Zargayouna, M., Balbo, F., Scémama, G.: A multi-agent approach for the dynamic vrptw. In: Proceedings of the International Workshop on Engineering Societies in the Agents World (ESAW 2008),, Springer-Verlag (2008)
3. Davidsson, P., Henesey, L., Ramstedt, L., Tornquist, J., Wernstedt, F.: An analysis of agent-based approaches to transport logistics. Transportation Research Part C - Emerging Technologies **13**(4) (2005) 255–271
4. Bazzan, A.L., Klügl, F.: A review on agent-based technology for traffic and transportation. The Knowledge Engineering Review **29**(03) (2014) 375–403
5. Bessghaier, N., Zargayouna, M., Balbo, F.: An agent-based community to manage urban parking. Advances on Practical Applications of Agents and Multi-Agent Systems (2012) 17–22
6. Parunak, H.V.D., Savit, R., Riolo, R.L.: Agent-Based Modelling vs. Equation-Based Modelling: A Case Study and Users' Guide. In: Proceedings of Workshop on Modelling Agent Based Systems (MABS98), Paris (1998)
7. Zargayouna, M., Zeddini, B., Scemama, G., Othman, A.: Agent-based simulator for travellers multimodal mobility. Frontiers of Artificial Intelligence and Applications **252** (2013) 81–90
8. Zargayouna, M., Othman, A., Scemama, G., Zeddini, B.: Impact of travelers information level on disturbed transit networks: a multiagent simulation. In: Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, IEEE (2015) 2889–2894
9. Maciejewski, M., Nagel, K.: Towards multi-agent simulation of the dynamic vehicle routing problem in matsim. In: Proceedings of the 9th International Conference on Parallel Processing and Applied Mathematics - Volume Part II. PPAM'11, Berlin, Heidelberg, Springer-Verlag (2012) 551–560
10. Nagel, K., Rickert, M.: Parallel implementation of the transims micro-simulation. Parallel Computing **27**(12) (2001) 1611–1639
11. Jakob, M., Moler, Z., Komenda, A., Yin, Z., Jiang, A.X., Johnson, M.P., Pechoucek, M., Tambe, M.: Agentpolis: towards a platform for fully agent-based modeling of multi-modal transportation (demonstration). In: International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4-8, 2012 (3 Volumes). (2012) 1501–1502
12. Tatara, E., Ozik, J.: How to build an agent-based model iii – repast simphony. In: Applied Agent-based Modeling in Management Research, Academy of Management Annual Meeting, Chicago. (2009)
13. Taillandier, P., Vo, D.A., Amouroux, E., Drogoul, A.: Gama: A simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In: PRIMA. Volume 7057 of Lecture Notes in Computer Science., Springer (2012) 242–258
14. Zargayouna, M., Zeddini, B., Scemama, G., Othman, A.: Simulating the impact of future internet on multimodal mobility. In: 11th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2014, Doha, Qatar, November 10-13, 2014, IEEE Computer Society (2014) 230–237