The simulation of large-scale scenarios requires high-performance simulations. MATSim, an agent-based transport simulation, is increasingly reaching limits. The traditional approach is to scale the scenarios, i.e. simulating only 10% of a population instead of 100%. This paper suggests MacroSim, a macroscopic mobility simulation module for MATSim, to overcome the current performance limits within MATSim. It uses volume-delay functions to estimate travel times for links based on experienced usages of these links. This allows to decouple agents and thus allows a parallelization of the mobility simulation within MATSim per design. A preliminary implementation of MacroSim showed promising results (7 to 50 times faster than the current mobility simulation depending on the scenario size). Given its limitations - most important no back propagation of traffic congestion - MacroSim is suggested as a complementary mobility simulation to the current implementation for cases where scenario size and simulation performance are more important than precise traffic dynamics.

1. Introduction

MATSim is an activity-based multi-agent simulation framework. It is designed for large-scale scenarios. All its components have been developed with the goal of keeping the computation burden as low as possible given certain targeted model characteristics. Thanks to this design, it has been possible to deal with scenarios of up to the order of $10^7$ agents. Yet, even if this is in itself an achievement of the project, it cannot be denied that such large-scale scenarios are associated with very long computation times. These are manageable in a MATSim development context but not necessarily compatible with practical use in transport research, unless a super-computer is available to the modeler. When dealing with very large scenarios, for example country-wide ones, it would be useful to substantially improve computation performance.

In MATSim, transport demand is modeled through a population of agents with daily activity and travel plans. The agents optimize their daily plans in a co-evolutionary process until a dynamic user equilibrium is achieved. The agents optimize the plans trough the following steps (see Fig. 1): executing the plans in a mobility simulation, scoring the...
plans on how well it performed given all other agents’ behavior and the transport environment, adapting the plans in an evolutionary way, and re-executing them. Among the MATSim main components, the transport simulation module (so called “mobsim”) requires in its current implementation\(^1\) a substantial part of the necessary computation time. It therefore appears as one of the modules where speed gains would have an impact. In its current form, the mobsim implements a queue-based model. This allows for the modeling of single vehicles - the resolution level targeted in the design of the framework. It omits, however, the representation of other details i.e. drivers behavior, crossroad dynamics, etc. - which would make the model computationally expensive.

This paper explores a speed up idea for the mobsim, introducing MacroSim. MacroSim is based on the idea of handling agents sequentially and decoupled from each other over the whole simulation. Instead of the widely used approaches (including the current implementation of the mobsim) which require the concurrent simulation of the whole model environment (system-based), in MacroSim the interactions of one agent with all the others are not explicitly modeled. They are represented at a higher abstraction level through constraints in road capacity and speed on any given road section of the network, expressed by volume-delay functions. This mechanism is not only expected to reduce the computational burden, but it can be argued that this sort of decoupling of individuals agents from the rest of the agent population and the representation of the traffic situation with volume-delay functions closely reflects the actual perception of travelers. Observations show that, when traveling, few travelers know and care about the precise state of the traffic and care even less about the other travelers except for their role as moving obstacles that one has to avoid. In other words, they are just a reason why an individual cannot advance with the speed, along the path, and with the driving style he or she would like to. Actual interaction among travelers is minimal; and if it happens, it is most often limited to informal intersection and traffic control. Direct inter-modal interaction is even rarer - being pedestrian crossings the most relevant exception.

Handling agents sequentially has the important property of allowing for massive parallelization of the transport simulation by design. Today many transport simulations face a parallelization limit. One of the main reasons is that these simulations have to concurrently simulate the full environment and the full transport system with all travelers, whether they are represented as agents or as flows, making parallelization limited by design. MacroSim is designed to overcome this limitation and aims at combining the performance advantages of volume-delay-functions, traditionally used in macroscopic transport models, like for example PTV Visum\(^2\), with the advantages of the individual agent representation of agent-based transport models, like MATSim\(^3\).

In this paper, Sec. 2 presents the idea and concept of MacroSim, Sec. 3 shows the performance gains achieved with a first simple implementation, and the sections Sec. 4 and Sec. 5 discuss and conclude the paper.

2. MacroSim

2.1. The core idea: an individual-based mobsim

MacroSim suggests a fundamental change in the simulation approach. In this paper, simulations which are considering the whole agent population concurrently are called system-based simulations, whereas the innovative approach proposed for MacroSim is defined as individual-based simulation (see Sec. 1). Individual-based in the sense that each agent and the environment in which it acts (which in principle includes all other agents) are “decoupled” and the agents are handled individually (Fig. 2). The agents influence each other only indirectly via their usage of resources (e.g. link capacities and facility capacities). When an agent’s plan is executed, the usage of any resource is recorded. After all agents are fully handled, the full record of these usages are transformed into average usage times (e.g. travel
Counts: Representing network and resource usage

Update: Resource 1: update travel time / usage time for each time bin with volume-delay function
Resource 2: ...
Resource 3: ...

Counts: Representing network and resource usage

Average travel and usage times

Simulation:
Agent 1: execution, scoring and replanning with average travel times and average resource usage times
Agent 2: ...
Agent 3: ...

Simulation:
Agent 1: execution, scoring and replanning with average travel times and average resource usage times
Agent 2: ...
Agent 3: ...

Average travel and usage times

Fig. 2. Overview of the MacroSim concept.

times for links) through volume-delay functions (Appendix A). These usage times are then used in the next simulation iteration - MacroSim, as part of MATSim (see Sec. 2.2) still relies on the iterative, co-evolutionary approach of MATSim. Through this mechanism, the duration of the activities or trips is determined for each agent, resulting in new usage patterns. If an agent can be handled separately from all other agents, the scoring and replanning (see Fig. 1) become also independent of the other agents and can be done immediately after the simulation of an individual agent (more details in Sec. 2.2). Furthermore, because the usage of resources can be calculated for each resource (e.g. network link or facility) separately, this step can be executed fully parallel too. In summary, this results in a fully parallel transport simulation per design.

2.2. Integration in MATSim

In MATSim, the mobsim executes plans and does the actual transport simulation (Fig. 1). Today, the default mobsim uses a queue model: Links are represented as queues and vehicles that enter a link line up at the end of the queue. Once a vehicle reaches the front of the queue - which is immediately if there is no queue - and has already spent on the link the equivalent of the free-flow travel time for that link, it is forwarded to the next link if the flow capacity of the present link allows for it and if the next link has enough free storage capacity. The flow capacity defines for a link how many vehicles per time interval are allowed to pass, the storage capacity the maximum number of vehicles which the link can accommodate concurrently. Intersection dynamics are not considered and transport controls like traffic lights and inter-modal interactions like pedestrian crossings are optional and turned off by default.

This queue model represents a very good compromise between simulation performance and level of detail. Additionally, it also allows for back-propagation of traffic congestion: If a link is fully congested, no vehicle is allowed to enter the link, which propagates the congestion to the upward links. Like many other transport simulations however, the queue model also requires a concurrent simulation of the full agent population and of the full simulated environment. This becomes an issue if large populations, e.g. entire regions and countries with millions of agents, should be simulated. The current practice is to scale down the model. For example, instead of modeling 100% of a country’s population, only 10% is modeled. The idea is that comparable system dynamics will arise if the transport system’s capacities are scaled down proportionally. While experience shows that this approach results in valid results for car traffic, it becomes difficult if public transport is explicitly included in the simulation (scaling of public transport vehicle traffic dilutes the public transport schedule) and problematic if also shared modes (scaling the fleet reduces the number of opportunities per area) and slow modes (scaling the number of traffic participants reduces the number of interactions) are explicitly simulated. As such modes become more and more a topic of transport research, the simulation of non-scaled scenarios becomes more and more a priority. The MATSim community has recognized the requirement for further performance improvement of MATSim, but as different past efforts showed, the possibilities within the current architecture seem limited (for a recent review see Nagel). It is in this context, in which MacroSim with its parallelization by design is suggested as a possible solution and as an extension for the MATSim framework.
MacroSim is planned to be integrated into MATSim primarily as a mobsim (Fig. 3). As the agent is handled separately from all other agents, however, the plan can be scored during the execution and, because the score is thus immediately available, immediately replanned (Fig. 2). Therefore, one agent has to be handled only once per iteration and a transition between different modules (simulation, scoring and replanning) is not required anymore. In consequence, events, which are used to communicate between different MATSim modules, are also not required anymore. This is important as moving events between modules is suspected to be part of the reason, if not the main reason, why a further performance improvement of MATSim within the current architecture is difficult. On the other hand, events are also the output of any MATSim simulation plus enable the flexible integration of extensions into MATSim. To keep these advantages, an approach inspired by another special implementation of MATSim (P-Sim) is planned for MacroSim: MacroSim is integrated into MATSim as a mobsim. Each mobility simulation with MacroSim consists of several sub-iterations without events. Only the last sub-iteration creates events, which are then returned to the larger MATSim framework, and the next full MATSim iteration starts.

2.3. Agent simulation in MacroSim

In MATSim, a plan consists of alternating activities and legs. Legs define travel between the location of two activities. In MacroSim, any execution of an activity or leg determines the start time of the following activity/leg, which determines when a resource is used and ergo - with volume-delay functions (see Appendix A) - what the usage time of this resource is (see Sec. 2.1), which in turn determines the duration of the activity/leg. Scoring is done with the execution of the activities and legs. After the full plan is executed, replanning can immediately be done for this agent.

2.3.1. Leg handling

If the leg is traveled by a transportation mean which is not explicitly handled by the traffic simulation (in the MATSim jargon a teleported mode), the arrival time and the score can immediately be calculated and returned. If the mode is actually simulated, the simulation determines for each step of the leg the flow-dependent travel time at the start time of the step, which determines the start time of the next step and so on. The usage-dependent travel times are calculated with volume-delay functions based on the usage of the resource (e.g. link) at that time in the previous simulation iteration. The full iteration through all steps of a leg results in the full duration of the leg and therefore the arrival time.

For the mode car one resource usage (step) is the passing of one link. Volume-delay functions for different link types (including e.g. pedestrian crossings) and intersection types (including e.g. traffic signals) exist (see Friedrich for a recent introduction into the calculation of link capacities for different link types). This allows to include complex network characteristics with minimal additional calculation time. If larger areas, for example entire countries, are simulated, volume-delay relationships for street traffic can potentially be represented with macroscopic volume-delay
functions for entire zones based on macroscopic fundamental diagrams (MFD)\textsuperscript{11,12,13}.

Using volume-delay functions provides a coarser representation of traffic as compared to the current mobsim. In agent-based transport models like MATSim, however, where one is interested in the modeling of individual transport behavior, this might not necessarily be a disadvantage: In surveys, people estimate their travel times with an offset of several minutes\textsuperscript{14}, round times, and the estimate is influenced by their overall perception of the trip. It can be assumed that the people do also not use more precise estimations when planning their everyday life.

For public transport, one resource usage (step) represents one ride with a public transport vehicle. The duration can be calculated based on an updated schedule. The updated schedule is the original schedule updated with delays based on a full MacroSim simulation of all public transport vehicles (as for the car mode) immediately before the simulation of the agents. This also adds the traffic caused by the public transport vehicles to the usage of the network links. Capacity restraints in public transport vehicles can be considered in the scoring step through vehicle-passenger-load-dependent generalized costs which are calculated with similar “volume-delay functions”.

2.3.2. Activity handling

Given a start time, simulation parameters, agent preferences, facility usage, and opening-times of the activity facility, the duration and therefore also the end time of the activity are calculated.

To incorporate the capacity restraints of activity facilities, adjusted volume-delay functions are suggested (see Appendix A). They allow to calculate the additional activity time and generalized costs of using these facilities and resources under capacity restraints. For example in shops or for leisure activities, other customers represent a reason why the experience might be better (e.g. bars) or worse (e.g. shops) and take longer than without other customers. These aspects can well be represented with positive or negative volume-delay functions which adjust activity times and scores based on the usage of this resource in the previous simulation iteration.

2.4. Shared systems

In shared transport systems, the availability of the service depends on the activities of other users. If every agent is simulated separately, this implicit coordination between the agents is not possible anymore.

Two solutions are suggested here. The first is to use functions for zones (e.g. 100m times 100m, municipalities, or car-sharing stations), that return the waiting time to access the system / being served by the system depending on the number of agents that used the system and the number of empty vehicles in this zone at this time in the previous iteration. Once the waiting time is over, the agents can be moved through the system as with mode car, just that the traffic they cause is not recorded as link usage. This traffic plus any overhead traffic that the shared system might cause are added to the traffic counts in a post-processing step where the requests are assigned to vehicles and the vehicles are routed. This solution implicitly deals with some of the conceptual problems that arise with the simulation of shared services in MATSim\textsuperscript{15}, for example what the agent should do if he is not served within a reasonable time. By working with expected waiting times, the solution again also reflects that people have only a coarse feeling about the state of the transport system.

The second proposed solution can be applied when small user groups share a vehicle. A possible scenario is a household sharing a vehicle. In such cases, it is suggested to identify these small user groups in a pre-processing step and then simulate these agents together.

3. Performance Experiments

A preliminary, simple implementation of MacroSim was tested against the standard mobsim of MATSim. The tested scenarios were a minimal standard MATSim test scenario called ”PT-Tutorial” (sec. 3.1) and a small size full scenario for the area of Sioux-Falls, US-SD (sec. 3.2). The test simulations were run without replanning to isolate the run time effects of the mobsims. MacroSim was run with 10 sub-iterations for each full MATSim iteration. Both simulations were run on three threads on the same platform (a business laptop) with a comparable background workload. The performance results presented in Table 1 are averages over 7 iterations of a single run.
3.1. **PT-Tutorial**

The "PT-Tutorial"\(^1\) consists of a street network of 16 nodes and 48 links and a public transport network of 4 stops and 2 lines. The scenario has a population of 900 agents with 1090 public transport and 1138 car trips. The traffic occurs with distinct morning and afternoon peaks. The agents conduct 1800 home activities, 900 work activities and 428 shopping activities.

One full mobsim simulation with the traditional mobsim, the q-sim, took 0.911 seconds. A full mobsim simulation with the new mobsim took 0.207 seconds. Of these 0.207 seconds 14% (0.029 s) were used to process the events and return them to MATSim. One sub-iteration with MacroSim took 0.018 seconds, which is 2% of the duration of the original simulation. So without event processing, MacroSim was about 50 times faster than the current standard mobsim.

3.2. **Sioux-Falls**

The Sioux-Falls scenario\(^2\) consists of a street network of 1’810 nodes and 3’359 links and a public transport network of 150 stops and 5 lines. Slow modes were teleported. 84’110 agents with 47’699 public transport, 109’076 car, and 11’445 slow mode trips build the population. The traffic occurs with distinct morning and afternoon peaks. The agents conduct 168’220 home activities, 56’904 work activities and 27’179 shopping activities.

One full mobsim simulation with the traditional mobsim took 8.870 seconds. A full mobsim simulation with the new mobsim took 13.503 seconds. Of these 13.503 seconds 5% (0.659 s) were used to process the events and return them to MATSim. One sub-iteration with MacroSim took 1.284 seconds, which is 14.5% of the duration of the original simulation. So for this full scenario MacroSim (without event processing) was about 7 times faster than the current standard mobsim.

<table>
<thead>
<tr>
<th>Table 1. Performance Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT-Tutorial</td>
</tr>
<tr>
<td>Street network - Nodes</td>
</tr>
<tr>
<td>Street network - Links</td>
</tr>
<tr>
<td>PT network - Stops</td>
</tr>
<tr>
<td>PT network - Lines</td>
</tr>
<tr>
<td>Agents</td>
</tr>
<tr>
<td>Car trips</td>
</tr>
<tr>
<td>PT trips</td>
</tr>
<tr>
<td>Slow mode trips</td>
</tr>
<tr>
<td>Run time full mobsim - q-sim</td>
</tr>
<tr>
<td>Run time full mobsim - macrosim</td>
</tr>
<tr>
<td>Run time event processing - macrosim</td>
</tr>
<tr>
<td>Run time sub-iteration - macrosim</td>
</tr>
<tr>
<td>Run time reduction mobility simulation</td>
</tr>
</tbody>
</table>

4. **Discussion**

MacroSim is proposed as an idea how to overcome the current performance limitations in MATSim. It deals with each agent separately by making use of the observation that interaction between traffic participants is usually reduced to perceiving each other as obstacles that prevent free-flow movement. Volume-delay functions are used to represent this relationship and return expected travel times.

A first and simple implementation of MacroSim was able to achieve substantial performance gains - 7 to 50 times faster - without yet making explicit use of the parallelization potential. Even though the performance gains were
smaller for a larger scenario, these tests still give indications of what could be possible with a full and proper implementation of MacroSim. One should keep in mind too that for now the effort put on the performance of MacroSim has been extremely limited compared to that put on the current standard mobsim. This means that, in a way, the speed up observed is purely due to the idea behind MacroSim, but it is expected that there is still a lot of room for improvement through better implementation.

The performance gains come at a cost however. Giving up the detailed transport simulation of the current MATSim mobsim means that traffic dynamics are not represented with the same level of realism. This concerns first of all back-propagation. In the current proposal of MacroSim, back-propagation of congestion across links is not possible. The consequences should be limited, however, as a recent MATSim study of the Zurich area showed that in the dynamic user equilibrium most often no to very few congestion occurs. Nevertheless, potential users of MacroSim should be aware of this limitation and what this means for their studies.

A second limitation of MacroSim is the reduced compatibility to any existing MATSim modules and extensions that are based on the exchange of events. MacroSim creates events and returns them to the MATSim framework only every $n$-th iteration. This limitation is acceptable, however, when a short computation time has high priority, as the approach allows to overcome the performance ceiling imposed by the generation of the events.

Aware of these limitations, MacroSim is not proposed as a full replacement of the current queue-based traffic simulation in MATSim. It is proposed as a complement, which can be used for applications where traffic dynamic representations comparable to macroscopic transport models like PTV Visum are sufficient and performance is of more importance. This applies to most application-oriented (as opposed to extending-oriented) applications of MATSim. One important example for such applications is the creation and calibration of base scenarios.

The introduction of a macroscopic transport model into an agent-based transport simulation is a very timely topic because it captures two main tendencies in the field:

On the one hand, macroscopic transport simulations, like for example PTV Visum, face more and more challenges how new shared transport modes, which tailor their service to individual users, like for example car-sharing or ride-sharing, can be integrated into a model that only simulates average flows, not individual agents.

On the other hand, agent-based transport models like for example MATSim, could handle those new modes which require a representation of individuals. But the larger the scenarios become, the more the parallelization and computational performance limitations become an issue.

MacroSim, as proposed in this paper, is an idea how to overcome these limitations for certain use cases. A full implementation of MacroSim will show if the expected performance gains are realizable and if all planned features are possible.

5. Conclusion

The current architecture of MATSim has reached its limits in computational performance and parallelization. New, also radically new solutions are required and encouraged. In this context, MacroSim, an agent-based macroscopic mobility simulation for MATSim, is a concept to make MATSim massively parallel by design and to give up events, which are suspected to be one of the major bottlenecks, at least for the most part of the simulation. The performance gains come with new limitations in the detail of the transport simulation and the possible applications of MacroSim. Instead of detailed transport system dynamics, average expected travel times are used as in macroscopic transport models. Based on observations on the perception of traffic by the people however, it is argued that these changes should not necessarily make the demand modeling less realistic.

Acknowledgements

The authors would like to thank Daniel Scherer for the collaboration in developing the initial ideas for MacroSim and the anonymous reviewers for their valuable feedback.
Appendix A. Volume-delay functions

A central concept for MacroSim are volume-delay functions. In transport, volume-delay functions return for a link with a given capacity the expected travel time dependent on the current link load. The two most prominent examples are the functions proposed by the Bureau of Public Roads\(^\text{19}\) and by Davidson\(^\text{20}\). For example the function by the Bureau of Public Roads\(^\text{19}\) has the following form:

\[
t_{\text{cur}} = t_0 \cdot \left(1 + \alpha \cdot \left(\frac{q}{q_{\text{max}}}\right)^\beta\right)
\]

with \(t_{\text{cur}}\) the current travel time, \(t_0\) the free-flow travel time, \(q\) and \(q_{\text{max}}\) the current and the maximum load (capacity), and \(\alpha\) and \(\beta\) shape parameters.

Other functions have been proposed that also take into account for example different intersection designs or pedestrian crossings. For a recent introduction into the calculation of link capacities for different link types see Friedrich\(^\text{10}\).

In MacroSim, volume-delay functions are also suggested for other non-link capacity-load-performance relations. For example the generalized cost of using a public transport vehicle depending on the number of people in versus the capacity of that vehicle, or for example the generalized cost of shopping or leisure activities depending on the number of people using that shop or leisure facility versus the facility’s capacity.

For situations, where the shape and/or the parameter values for volume-delay relationships are not available, they can be derived from the according situation in the current implementation of MATSim. This assures that at least the macroscopic behavior of the current MATSim implementation is reproduced.

References

13. Knoop, V. Large scale urban traffic operations: theory, empirics and control. Presentation; IVT Seminar; ETH Zurich, Zurich; 2016.
15. Bösch, P.M. Required autonomous vehicle fleet sizes to serve different levels of demand. Presentation; VSP-Seminar; TU Berlin, Berlin; 2016.