Discrete Event Simulation of the Transport Logistics of the Austrian Red Cross

O. Koch, othmar@fsmat.at
Vienna University of Technology
Wiedner Hauptstrasse 8-10, A - 1040 Vienna, Austria

…. discusses an ARENA/SIMAN discrete simulation model for the analysis of transport logistics

…. gives scheduling heuristics and discusses different parameter choices used in the scheduling process

…. compares different scenarios corresponding to different possible organizational structures of the rescue organization

Abstract

We discuss a simulation model used in the analysis of the transport logistics of the Austrian Red Cross rescue organization. The emphasis is on the details of modelling the scheduling of ambulance service. Also, the parameters associated with different criteria for the performance of the system are discussed. Implications of the simulation results on decision making and structuring of the organization show the validity of the model and also give hints on possible improvements of the coordination of transports. The main aim is the discussion of the efficiency of central coordination as compared to decentralized planning.

1. Introduction

The aim of the analysis of transport logistics described in this paper is to examine possibilities to improve the efficiency of ambulance service for the Austrian Red Cross. To conduct the study we chose to utilize a discrete simulation model. Classical approaches for the optimization of the transport of goods seemed inappropriate for our purpose. The mathematical tools for the analysis of transport problems discussed for example in Domschke (1989) or Jándy (1967) cannot capture the dynamic situation at full but rely on average (or possibly stochastic) demands and supplies. In our opinion the detailed analysis of time critical aspects of the scheduling of emergency and other transports, restricted availability of transporters (ambulances) and limited personnel resources can best be modelled by simulation. Our simulation model was implemented in ARENA / SIMAN.

The SIMAN simulation engine turned out to be the appropriate tool for our purpose, and the ARENA system provided a comfortable developing environment.

However, we refrained from using any of the ARENA modules, but restricted ourselves to the elements of the SIMAN simulation language. The technical reference we relied on is Pegden et al. (1995).

A detailed description of our model and some hints at the implementation are given in section 2. We concentrate on the model of the traffic network, transportation system and the implementation of heuristics to ensure (near-) optimal efficiency of the coordination of patient transports. In section 3 we give the results of our simulation. Particular emphasis is on the choice of model parameters, which influence important quantities like the average waiting time of patients or total mileage required to carry through the transports. Important conclusions for decision-making and the implications of a comparison of central coordination with decentralized organization are discussed.

2. The simulation model

The mathematical model of the underlying traffic network, quite naturally, is an undirected, weighted graph, where the (positive integer) weight assigned to each edge represents the length in kilometres of the road connecting the places, or hospitals, represented by the vertices of the graph. The data for the construction of the network for the area under consideration was retrieved from electronic route planning services giving the road lengths of connections. Thus, in most cases we did not have to rely on estimates based on the Euclidean distances between locations discussed for example in Probol (1979).

Note that altogether the network consists of about 300 nodes and 1400 links connecting them. A graphical representation of the graph is given in Figure 1. The two nodes denoted by **H** represent cities with a number of hospitals. Both places are not part of the area we are discussing. The remaining graph is divided into the three subareas **Area 1**, **Area 2**, and **Area 3**, which will be discussed later. Large dots represent places with Red Cross stations.

The structure of a traffic network can be implemented quite easily in SIMAN as a *network* consisting of *intersections* (and associated *stations*, or submodels), which also enables the use of *guided transporters* navigating on the graph.

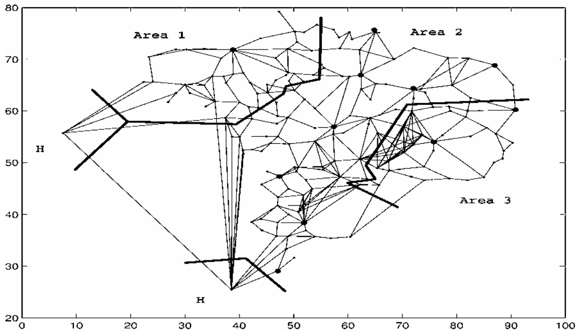

Figure 1: The traffic network

The use of *guided transporters* in a *network* implies the possibility to use the graph algorithms integrated in SIMAN to solve shortest path problems, see Pegden et al. (1995). For the ambulance service we use three different kinds of transporters with different transport capacities and demands on operating personnel:

1. *Krankentransportwagen* (KTW) can carry up to three patients, one on a stretcher and two on sedan chairs. For the different types of patients, see below. A KTW requires two operators.
2. *Behelfskrankentransportwagen* (BKTW) takes up to four patients, which have to be able to walk of their own accord, and only require one driver.
3. *Notarztwagen* (NAW) only carries one patient, require three persons to operate and are used for emergencies only.

The ambulances are stationed in special parking positions representing the locations of the Red Cross stations in the area, cf. Figure 1. Note that the average speed of transporters is assumed to be 60 km/h in general, while for emergency transports 90 km/h is permitted and in cities or villages an average of only 30 km/h is prescribed. The transporters are routed using special *driver entities* which are responsible for the acquisition of operating staff (see below), control of the free capacity, and updating and executing the planned route. To avoid deadlocks or „traffic jams“, a *relinquish block* is used upon every activation of a transporter to enable the guided transporters to pass each other by uninhibited.

The staff required to operate an ambulance is a *resource* with a capacity (number of ambulance staff on duty) that is governed by a schedule (depending on time of day and day of week). There is a different pool of rescue staff with an individual schedule for every Red Cross station.

Assignment of personnel to a transport is *first come first serve*. The mechanism applied when a change in capacity cannot be effected immediately will be discussed later.

The patient data used to drive the simulation is read in from a file containing the entry time of the patients into the system, that is, the time of the first request for ambulance service, the place (vertex of the network) of entry and the destination of the requested transport. Additionally, the patient type is read in from the file, with four different categories for this attribute:

- Emergency patients, who require preferential treatment, see below.
- Regular patients who are still able to walk of their own accord.
- Regular patients who have to be carried (and transported) on a stretcher.
- Regular patients who have to use a sedan chair.

The data was collected for a three month period (January to March 2001) and slightly adjusted to avoid exceptions due to holidays disrupting the duty roster of ambulance personnel. Altogether, a data set of 14174 patients was used to drive the simulation.

When a patient enters the system, the patient type is determined first. For an emergency patient, an admissible ambulance is assigned to carry through the transport as fast as possible. An ambulance is considered admissible if it has free capacity to transport a patient on a stretcher, is not assigned to another emergency transport and personnel resources to operate the ambulance are available. Also, we have to ensure that the closest available transporter is indeed close enough to be efficient, so we require that the approach will take no longer than the current waiting time of the patient, unless the distance to be covered is shorter than 5 kilometres.

For the NAW, we allow an approach that is twice as long because it can provide more appropriate help in case of an emergency. Thus, the longer a patient is already waiting for ambulance service, the more ambulances are admissible for the transport. If the patient entering the system is not an emergency, he/she is assigned to a waiting queue until a suitable ambulance is determined for the transport. The heuristic for a routing strategy that provides efficient coordination of tours while not creating unacceptable conditions for waiting patients is described below.

A transporter waiting at its parking position checks for an emergency transport every minute. If no such transport is requested, the queue of patients waiting for transportation is searched for a suitable task every T=15 minutes. The parameter T is critical for the performance of the system (explained in section 3).

A transport is assigned if the approach to the closest patient's entry station is shorter than the maximum of u_1=12 minutes and r=0.75 times the current waiting time of the patient. Thus, an ambulance is assigned if the tour implies only a short approach from the parking position or if the patient has been waiting for an intolerably long period of time. Note that the longer a patient has been waiting, the more likely an ambulance is assigned even if this means an inefficient detour for the Red Cross ambulance service. We will discuss the effects of this feature in sec. 3, where we will pay special attention to the choice of the parameter r which controls the ratio of tolerable waiting time for patients and admissible detour for ambulances.

When an ambulance reaches a node along the network, any pickup and drop-off actions appointed for the respective station are performed. To model this process, the transporter is delayed to allow for loading time. The duration of this delay varies stochastically according to a triangular distribution with minimum 3, mode 5 and maximum 7 minutes, cf. Kelton et al. (1998). After loading and/or unloading patients, the planned route is updated according to following rules:

- If an emergency transport is being carried through or was recently assigned, the transporter moves to the next station of its route directly on the shortest path through the network. Thus, an emergency transport is inserted at the first position into the planned route and undertaken immediately.
- If the schedule of ambulance personnel has changed and the number of operators available according to the schedule is smaller than the number actually used, no new patients are assigned to the tour, the tour is completed and the transporter moves back to its parking position and releases the operating staff.
- Otherwise, the waiting queue of patients not yet assigned a transporter is searched for a possibility to coordinate any of the requested patient routes with the planned route of the transporter such that no intolerable detour results. To this end, for every patient in the queue the data of entry station and destination are inserted into the transporter's planned tour at every possible combination of positions until an admissible tour is found.

In this context, criteria for an admissible tour:

- The transporter's free capacity is sufficient to carry through the transport from entry station to destination, even if additional patient pickups and drop-offs are scheduled during the tour.
- The detour for the transporter in kilometres is shorter than the maximum of u_1=12 minutes and r=0.75 times the current waiting time of the patient in minutes, but in any case less than 20 km.
- The detour for every patient assigned to the same tour of the transporter as compared with a direct transport from his/her entry position to the destination is smaller than u_2=10 kilometres.
- The currently planned tour contains no more than nine patient pickups and drop-offs.
- If no admissible route is found for a patient, the procedure is repeated for the remaining patients in the waiting queue.

If no drop-off or pickup is currently scheduled, but the capacity of ambulance personnel is sufficient to carry through further transports, the ambulance returns to its parking position, taking the shortest path but pausing at every node along the way to check for new tasks.

To illustrate our simulation model, we display a screenshot from a small demo version of our program in Figure 2. This model only contains 10 nodes from the actual network and a reduced number of ambulances and personnel for easier graphical representation. The 10 nodes of the network comprise 5 Red Cross stations and 2 hospitals. Currently, 1 out of 3 KTW is operating, while the 2 BKTW and 1 NAW are waiting for assignments. Consequently, 2 out of 6 currently available ambulance personnel are busy. The queue of patients not assigned an ambulance contains 2 patients, while 1 patient is waiting for pickup.

Figure 2: Screenshot from a demo version

3. Simulation results

In this section we discuss the results of our simulation study. First, we analyze the effect that the choice of the system parameter r, which controls the ratio of the tolerable waiting time for patients and the admissible detour for ambulances, has on the time the patients spend in the system, from the first request for a transport until drop-off at the destination. Moreover, it is shown that the choice of the parameter T has interesting implications for the actual planning process.

The role of the parameters u_1 and u_2 is far from trivial. The choice of the values u_1=12 and u_2=10 is based on extensive testing to find the configuration providing optimal exploitation of synergies. The resulting parameters were found to imply the least total number of kilometres that the ambulances have to cover (and thus, minimal transportation costs). The tests leading to this conclusion cannot be discussed here in detail.

The main goal of our analysis is the discussion of the possible improvements if the status quo of decentralized scheduling is replaced by central coordination. To this end, we compare the situation where, for the three subareas shown in Figure 1, the routes in every area are coordinated using the respective resources only, with centralized planning in the whole area. Finally, we show that our generic rules for tour planning lead to a behaviour which very well corresponds with the actual layout of the traffic network, demonstrating both the validity of our model and the practicability of the location of the Red Cross stations. We conclude with a few remarks on the validation of our model.

Short waiting time vs. low transportation costs

Here, we discuss two scenarios demonstrating the ambivalent role that the optimization of transport schedules has for the patients waiting for a transport. In **Scenario 1**, we choose the parameter r=0.75, while in **Scenario 2**, we define r=0.5. The intuitive meaning of this choice of parameters is the following: In **Scenario 1**, a detour for an ambulance is more readily accepted in order to reduce the waiting time of the patients, while in **Scenario 2** optimal transport planning in terms of minimizing transportation costs for the rescue organization is emphasized. Indeed, the results given in Table 1 show that the choice of the parameter r indeed influences the simulation results in the expected way. The characteristic values of the system's behaviour which (given in Table 1) are the following:

- *transfer*: Transfer time for each patient.
- *load*: Time spent for pickup and drop-off.
  This includes the loading times for other patients during the transport.
- *wait*: Waiting time.
- *TIS*: Total time spent in the system.

We distinguish between the values for emergencies */em* and other patients */pat*. In the simulation run, a total of 942 emergency and 13,232 other transports were carried through, with some variability in the measured characteristic values. Thus, Table 1 gives the quantities' mean values and the 95% confidence intervals. The values for **Scenario 3** also given in Table 1 are discussed later. The last row contains the total distance required by all transporters to carry through the requested tasks.

Obviously, the choice of the parameter r does not influence the emergency transports in a statistically significant way. This is indeed the behavior we expect. It is interesting to note that the load time for emergency patients amounts to approximately 10 minutes, the value we should expect if the transport is carried through directly without loading or unloading another patient in the meantime, thus verifying the simulation model's correct behavior. Also quite naturally, the transfer and load times are not affected by the parameter r. The wait time, however, increases when we choose the parameter such as to reduce detours for the rescue organization. Indeed, the paired t-test with confidence level 95% (see Law and Kelton (1991)) gives a difference of 1.64 ± 0.85 min. The increase in the waiting time also causes a difference in the total time in the system of 1.78 ± 1.05.

	Scenario 1	Scenario 2	Scenario 3
transfer/em	16.68±0.76	16.66±0.75	16.66±0.75
transfer/pat	38.37±0.59	38.42±0.56	37.80±0.57
load/em	9.96±0.07	9.98±0.06	10.04±0.06
load/pat	13.41±0.16	13.50±0.13	12.52±0.13
wait/em	8.32±0.69	8.44±0.81	8.48±0.64
wait/pat	35.75±0.76	37.38±0.92	28.87±0.81
TIS/em	39.34±1.27	39.47±1.31	39.62±0.6
TIS/pat	87.53±0.87	89.31±1.05	79.18±0.95
Tot. mileage	694,166	685,750	739,907

Table 1: Characteristic values of patient transport

Finally, we try to prove that the difference in the waiting time can indeed be attributed to the choice of the parameter r. We distinguish two different factors which add up to the total waiting time of a patient: The time from the request for a transport until the assignment of an ambulance, and the time until the assigned ambulance arrives for picking up the patient. These values were determined independently (batched together for emergency and other patients, however). The difference in the latter quantity is not significant at confidence level 95% and amounts to 0.55 ± 0.72. The time until a suitable transporter is assigned, however, differs by 2.24 ± 0.29. Thus, the parameter choice induces a statistically significant difference in the total time in the system, which is to be attributed to the difference in the waiting time until a transporter is chosen for the tour. However, the advantages for patients if we choose r=0.75 are compensated to some extent by the reduction in the total travelling distance for the rescue organization if we set r=0.5. The latter choice of this parameter enables more synergies to be made use of in the coordination of tours and thus the total mileage amounts to 8,416 kilometres less in **Scenario 2** than in **Scenario 1**.

Improving the coordination of transports

In this section we discuss the choice of the parameter *T* and demonstrate how to improve the coordination of tours. This parameter controls the frequency at which a new possible tour is scheduled for an ambulance that is idle at the rescue organization's quarters. Naturally, this parameter cannot be chosen too large, because this would mean a long waiting time for patients until there is even an attempt at assigning a suitable transporter. We found that the choice *T*=15 minutes was optimal. In fact, if the parameter is chosen as too small, an undesirable effect is observed. An ambulance is assigned to the first requested, admissible transport. If *T* is larger, there are more possibilities to choose from in order to determine an optimal transport. To illustrate this interesting point, consider **Scenario 3**. In this case, we choose *T*=3. Table 1 gives the results which show the effects this choice has on the patients. Apparently, there is some advantage for the patients in **Scenario 3** as compared to **Scenario 1**. The total time in the system differs by 8.35 ± 1.00. This effect is again attributable to the waiting time. In this case, however, we found both the time until the assignment of a transporter and until the pickup of the patient to show a statistically significant advantage for **Scenario 3**. Still, the model with *T*=3 is unusable for our purpose, because it makes efficient coordination of tours impossible. Indeed, the travelled distance for the rescue organization increases unacceptably by 45,741 km. Also, the behaviour becomes more random and erratic, so we cannot draw clear conclusions from the results for this scenario.

Central coordination vs. decentralized planning

Here, we consider two different ways of coordinating ambulance service. We restrict ourselves to the parameter choice from **Scenario 1**, for **Scenario 2** the results are similar. So far in this paper we always assumed tour planning to take place on a global scale, where all the tasks and available resources in the whole area given in Figure 1 are considered simultaneously for the scheduling process. Now, we want to compare the situation with the results for decentralized coordination. In this second scenario, the resources associated with the three different subregions **Area 1**, **Area 2** and **Area 3** from Figure 1 are used exclusively to carry through transports associated with the respective area. The required data was available separately for each subarea; in fact it was accumulated especially for the purpose of simulating central coordination. The statistically significant differences between the two scenarios were determined by lumping the data from the subareas together into one data set and comparing this data set with the results for the whole area using the paired *t*-test at confidence level 95% (results for the three subareas in Table 2).

	Area 1	Area 2	Area 3
transfer/em	22.10±2.13	15.64±0.87	19.50±2.64
transfer/pat	46.05±0.72	34.74±0.67	39.57±1.04
load/em	9.81±0.21	10.02±0.08	9.98±0.29
load/pat	13.68±0.34	13.02±0.21	13.46±0.23
wait/em	10.28±1.42	8.97±0.92	11.83±3.91
wait/pat	55.08±2.52	33.73±1.45	48.09±1.60
TIS/em	49.46±3.56	39.46±1.66	48.18±5.98
TIS/pat	114.8±2.65	81.49±1.87	101.12±2.1
Tot. mileage	167,581	373,314	161,617

Table 2: Characteristic values for decentralized planning

For both the transfer and load times of emergency as well as of other patients, no statistically significant overall difference between central and decentralized coordination can be observed. However, there is apparently some slight disadvantage if **Area 1** or **Area 3** is considered separately. This is compensated for by the favourable results for **Area 2**, however.

The waiting time for patients, on the other hand, is improved by 5.29 ± 0.94 min if we consider central coordination for the whole area. This small difference can entirely be attributed to the waiting time until an ambulance is assigned. Indeed, this time factor differs by 5.09 ± 0.40 between the two scenarios, whereas the waiting time from the time a tour is assigned until the patient is picked up shows no significant difference. Note that the waiting time for decentralized planning is longest in the smaller regions **Area 1** and **Area 3**. This is obviously due to the restricted flexibility in planning if coordination is reduced to a smaller scale. There is no significant difference in the waiting time for emergencies for both scenarios.

Finally, we observe that the difference in the waiting time also has an influence on the total time a patient takes from the first request for an ambulance until drop-off at his/her destination. There is an advantage of 4.75 ± 1.19 min if central coordination is considered. Curiously, there is a statistically significant difference for emergency patients as well. The advantage of 1.88 ± 1.83 min can be neglected, however, and is a rather random effect, obviously. We conclude that there is a slight advantage for patients if we use central coordination instead of decentralized planning, but we are more interested in the gain in efficiency this implies for the Red Cross ambulance service. We find that central coordination reduces the total mileage by 8,346 km. This reduction by 1.19% seems rather insignificant, however.

Analysis of the tours

In this section, we discuss the interrelations between the location of the Red Cross stations and organization units and the tours resulting from our scheduling heuristics.

The region we consider, cf. the graph of the traffic network given in Figure 1, is subdivided into three subareas associated with a number of Red Cross stations. These stations are depicted as the 11 larger dots in Figure 1. We are interested in the question whether the cooperation between these three organizational units is very strong for an optimal coordination of transports.

Let us discuss the results given in Table 3. For the three subareas, we give the number of starting points and destinations of individual patients which were transported by an ambulance from the respective area. We exclude the NAW from this discussion because it is associated with the whole region from Figure 1. „Pickup A1" - „Pickup A3" and „Drop-off A1" - „Drop-off A3" denote the total numbers of pickups and drop-offs, respectively, that were carried through in **Area 1**, **Area 2**, and **Area 3** („A1", „A2", and „A3") by ambulances associated with the respective subareas („T1" … transporters from **Area 1**, etc.). „Pickup H" and „Drop-off H" denote the same quantities for transports to and from hospitals, and „Pickup O" and „Drop-off O" refer to transports leaving the area under consideration.

	T1	T2	T3
Pickup A1	1.350	283	9
Pickup A2	225	4.331	399
Pickup A3	5	414	1.229
Pickup H	323	4.770	532
Pickup O	3	130	13
Total Pickup	1.906	9.928	2.182
Drop-off A1	329	837	42
Drop-off A2	230	3.251	333
Drop-off A3	1	924	288
Drop-off H	1.299	4.534	1.482
Drop-off O	47	382	37
Total Drop-off	1.906	9.928	2.182

Table 2: Analysis of the tours

Obviously, the vast majority of ambulances operate in the area they are associated with. As concerns pickup, this trend is quite distinct. There is some amount of interchange between **Area 1** and **Area 2**, and between **Area 2** and **Area 3**, but not between **Area 1** and **Area 3**. This is no surprise, as **Area 2** separates the other subareas. Moreover, **Area 1** and **Area 3** are most easily accessible via a freeway passing through **Area 2**. So especially for transports to and from hospital (the majority of hospitals is situated at the bottom corner node denoted by **H** of the graph given in Figure 1), service of **Area 2** by ambulances from **Area 1**, and more noticeably, from **Area 3** is quite natural.

In general it could be observed that places served regularly by ambulances from a different area are found in special topographical situations, favouring access from an adjacent area. Not surprisingly, pickup from hospital is an important factor as well. This effect is much stronger even for drop-off of patients. Still, the results show that for the remaining drop-offs, a tendency to stay in the same area can be observed. For the apparent synergies when entering a different subarea, the same factors seem to be important as in the case of pickup. It is not possible to conduct a more detailed analysis of the tours here. However, when considering the precise location of the nodes in the traffic network, it can be inferred that the tours in our simulation reflect the topographical situation very well and that ambulances keep in the area they are associated with unless a special topographical situation suggests to serve an adjacent subarea.

3.5 Validation of the model

Validation of the model proved quite difficult for lack of complete data, so we touch the issue only briefly. There are no records on transfer times, waiting times etc., so we can only estimate from experience that the simulation results indeed give the right order of magnitude. The mileage is subject to bookkeeping, however. Unfortunately, we only have reliable data for the total distances covered by all the patients, without taking into account the possibility to transport more than one patient at a time. From partial data available for some subareas, however, we reckon that the true value is overestimated by about 40-50% by the value on record. This value for the area under consideration gives a total of 891,139 km. Thus, the value 694,166 km from our simulation reflects the correct order of magnitude and we accept the model to work dependably.

References / Acknowledgement

[1] *Domschke, W.,* Logistik: Transport, Volume 1, Oldenbourg, München, 3rd edition, 1989.

[2] *Jándy, G.,* Optimale Transport- und Verkehrsplanung, Physica-Verlag, Würzburg, 1967.

[3] *Law, A.M. and Kelton, W.*, Simulation Modeling and Analysis, McGraw-Hill, New York, 2nd edition, 1991.

[4] *Pegden, C., Shannon, R., and Sadowski, R.,* Introduction to Simulation Using SIMAN, McGraw-Hill, 2nd edition, 1995.

[5] *Probol, M.G.,* Tourenplanung mit EDV unter besonderer Berücksichtigung der Erfordernisse eines Unternehmens des Eisen- und Baustoffgroßhandels, Ph.D. Thesis, Braunschweig, 1979.

Acknowledgement. I would like to thank Dr. Heinz Weigl from „ESLA Solutions" for providing the ARENA software package and for valuable hints on technical and modelling aspects