Optimizing emergency medical assistance coordination in after-hours urgent surgery patients

Marin Lujak, Holger Billhardt, and Sascha Ossowski

University Rey Juan Carlos, Madrid, Spain

Abstract. This paper treats the coordination of Emergency Medical Assistance (EMA) and hospitals for after-hours surgeries of urgent patients arriving by ambulance. A standard hospital approach during night-shifts is to have standby surgery teams come to hospital after alert to cover urgent cases that cannot be covered by the in-house surgery teams. This approach results in a considerable decrease in staffing costs in respect to having sufficient permanent in-house staff. Therefore, coordinating EMA and the hospitals in a region with their outhouse staff with the objective to have as fast urgent surgery treatments as possible with minimized cost is a crucial parameter of the medical system efficiency and as such deserves a thorough investigation. In practice, the process is manual and the process management is case-specific, with great load on human phone communication. In this paper, we propose a decision support system for the automated coordination of hospitals, surgery teams on standby from home, and ambulances to decrease the time to surgery of urgent patients. The efficiency of the proposed model is proven over simulation experiments.

1 Introduction

Most hospitals that perform emergency surgery service provide also after-hours surgery for urgent patients whose conditions are not critical but might result in increased probability of morbidity or mortality. Out-of-hours is a period which is generally defined to be between 6 PM to 8 AM weeknights and the whole weekend, even though the definition might vary from one hospital to another. The growing demand of simultaneous multiple patients for emergency medical assistance (EMA) and urgent surgery treatment provided by hospitals puts a strong focus on the combined EMA and hospital surgery treatment coordination effectiveness and efficiency. The management of the hospital network and the emergency medical assistance in each region, city or town is challenged to deal with the seemingly conflicting objectives of fast, efficient and effective urgent patient response minimizing total system cost and maximizing the quality of care.

In this paper, we develop a decision-support system for the coordination of EMA and hospitals for after-hours urgent surgery patients. We assume that there are multiple hospitals available for urgent cases surgery treatment and for each hospital there is a sufficient number of in-house surgery teams needed to care for in-house and emergency patients safely. A surgery team consists of the individuals needed to adequately staff one operating room (OR) (e.g., a surgeon, an anesthetist, two nurses and a nurse anesthetist). Furthermore, we assume that there is a number of surgery teams on standby from home, coming to hospital after alert. The savings in staff expenditure between having sufficient staff in-house for urgent cases in respect to having them taking call from home might be considerable [5]. Hence, a good balance between the efficiency and the flexibility in hospital and EMA network management is a prerequisite for providing optimal care to patients.

The decision-support system proposed in this paper is based on the coordination of the assignment of idle ambulances to pending patients, and a simultaneous assignment of ambulances assisting patients in-situ to adequate hospitals together with the assignment of standby out-of-hospital surgery teams to the same. The multi-objective optimization of arrival times of multiple actors is solved for the minimization of patients' surgery waiting times. Responding to a possibility of occurrence of multiple simultaneous patients and based on the relative positions of the patients, surgery teams, and available hospitals, our approach is based on a system's view, not concentrating only on minimizing single patient delay, but concentrating on the system best solution in respect to the (temporal and spatial) multitude of patients. Simulated emergency scenarios demonstrate the efficiency of the coordination procedure and significant decrease in the urgent patients waiting time to surgery treatment.

This paper is organized as follows. In Section 2, we describe the the State-of-the-Art practice in the EMA coordination for urgent surgery patients. In Section 3 we formulate the EMA coordination problem for urgent surgery patients arriving by ambulance. Section 4 describes briefly the proposed multi-agent architecture with the modified auction algorithm for EMA urgent surgery coordination. Section 5 contains simulation results comparing the proposed coordination approach and the benchmark urgent surgery coordination procedure first-come-first-serve. We draw conclusions in Section 6.

2 State-of-the-Art practice and related work

The emergency medical system for the assistance of urgent surgery patients is made of the following participants: out-of-hospital patients, hospitals with after-hours urgent and emergency surgery option, Medical Emergency Coordination Center (ECC), ambulances staff, and standby out-of-hospital surgery teams. Usually, each hospital has assigned to it one or more out-of-hospital standby surgery teams positioned at alert outside hospital and obliged to come to the hospital in the case of emergency. The reason for their outside hospital position are staffing costs which make a large portion of costs in surgical care services [6]. Significant cost savings can be achieved by increasing staffing flexibility [3] and assignability to multiple hospitals.

The standard approach used in most of out-of-hospital after-hours urgent surgeries is the following. Patients are diagnosed in the place of emergency: at their momentary out-of-hospital location or at a health center without after-hours urgent surgery option. In both cases the ECC applies First-Come-First-Served (FCFS) strategy and locates the nearest available (idle) ambulance with ALS and dispatches it to pick up the patient. The use of ambulance for urgent patients is proven to increase patient chances in respect to the use of private transportation. The concrete example is infarct treatment [18] where ambulance should be considered a place for initial diagnosis, triage and emergency treatment since pre-hospital triage in the ambulance reduces infarct size and improves clinical outcome [16]. After the ambulance arrives to the scene and diagnoses the urgency at patient's momentary out-of-hospital location, ambulance confirms the diagnosis to the ECC which has a real time information of the states of ambulances. ECC sequently applies FCFS strategy for hospital assignment by locating the nearest available hospital with operating room working after-hours. The hospital then alerts the closest surgery team of the urgent surgery case.

The process for urgent surgery treatment coordination usually used in the ECCs is manual and the management is based on case by case principles with high human load necessary for telephonical arrangements to find a solution. This can significantly worsen the total delay time for patients awaiting surgery. In the case of a simultaneous presence of multiple urgent patients, hospitals and surgery teams located in multiple sites, support for optimized EMS coordination based on information updated in real time is necessary for efficient surgery planning and scheduling.

There is a vast Operations Research and Multi-Agent Systems literature in medical emergency assistance coordination. There exist different ambulance deployment, relocation and dispatch models, e.g., [9], operating room planning and scheduling, e.g., [3], and patient scheduling solutions, e.g., [15]. The proposed methods are mostly based on queuing theory, simulations and mathematical programming, e.g., [8, 13, 14, 17].

Henderson in [8] outlines some of the key challenges EMS providers face, such as traffic congestion, increasing call volumes, hospital diversion, and increasing transfer times at emergency departments. Ingolfsson in [10] surveys research on planning and management for emergency medical services. In [1], Bandara et al. study optimal dispatch of paramedic units to emergency calls to maximise patients' survivability. Their computational results show that dispatching the closest vehicle is not always optimal and that dispatching vehicles considering the priority of the call leads to an increase in the average survival probability of patients.

Emergency medical assistance literature is abundant also in the multi-agent system community, e.g., [4, 11]. Domnori et al. in [4] discuss the fitness of agent-based applications to managing healthcare emergences and large scale disasters and their application to problems where the main challenge is coordination and collaboration between components. López et al. in [11] propose a multiagent system using an auction mechanism based on trust to coordinate ambulances for emergency medical services. The auction mechanism here is based on three individual patient priority cases, where the winning ambulance is the one with the best estimated arrival time and a good trust degree. Lujak and Billhardt in [12] proposed an organization-based multi-agent application for emergency medical assistance (EMA) based on a distributed relaxation method for the assignment problem called auction algorithm [2] and the mechanism based on trust. The experiments results confirm the reduction of the average response times of EMA services.

Considering out-of-hours emergency surgery, in [19] the balance between hospital costs and patient safety was examined to determine the optimal size of emergency surgery teams that are on-call after-hours, including medical and nursing staff. The study found that the use of defined procedure-based safety intervals to plan on-call rosters can reduce the number of staff rostered on-call without jeopardising patient safety.

The key premise of this argument is that fewer nighttime staff will be sufficient if patients wait a little longer for surgery, but not so long as to exceed safety intervals.

For the ambulance assignment problem, not infrequently applied dispatching method is first-come-first-served (FCFS) policy which is the method temporally discriminating patients and not considering the availability of hospitals or hospital staff. However, different centralized and distributed Operations Research optimization methods can be applied for the multi-agent task allocation and coordination problem encountered in this context. Since in this scenario, scalability, robustness and flexibility are of outermost importance, distributed methods, such as auction algorithm [2] are of preference. To the best of our knowledge, the literature on integrated mutual coordination of EMA, multiple hospitals operating rooms and out-of-hospital surgery teams is lacking which is the reason why in this paper we propose an integrated solution model for this problem.

3 Problem formulation

In this paper, we treat the problem of after-hours out-of-hospital urgent surgery patient assignment to ambulance assistance, and consecutive patient transfer to adequate hospital with minimal waiting time for surgery. We assume that after transferring a patient to hospital, ambulance is redirected to the base station where it waits for the next emergency patient call.

In Figure 1, we present patient delay time components:

- Call response and resource assignment time spent by the ECC (the time of analyzing the problem and giving it the highest priority category deciding on the ambulance and the hospital assignment);
- Mobilization of the ambulance and transportation time of the ambulance from its momentary position to the patient;
- Time of patient assistance in situ by ambulance staff;
- Transportation time of the ambulance with the patient to the hospital;
- Transportation time of the surgery team members from their momentary positions to the hospital;
- Expected waiting time due to the operating room occupancy of other prior pending patient(s) in the hospital (if any).

Hesitation of patients to search for medical help together with the delays which are the result of the manual centralized coordination of multiple actors in EMA sometimes might average several hours and thus can prevent the early application of life-saving procedures and contribute substantially to a diminished effectiveness of surgery treatment. In the case of multiple simultaneous pending patients, the right combination and the individual choice of the ambulances to be assigned can significantly improve overall patients' chances and reduce the resulting morbidity and mortality. After a patient gets assisted in situ by ambulance staff, individually minimal expected time to surgery is the time resulting from the following three parameters, Figure 1:

- transportation time of the ambulance with the patient from the initial patient location to assigned hospital,

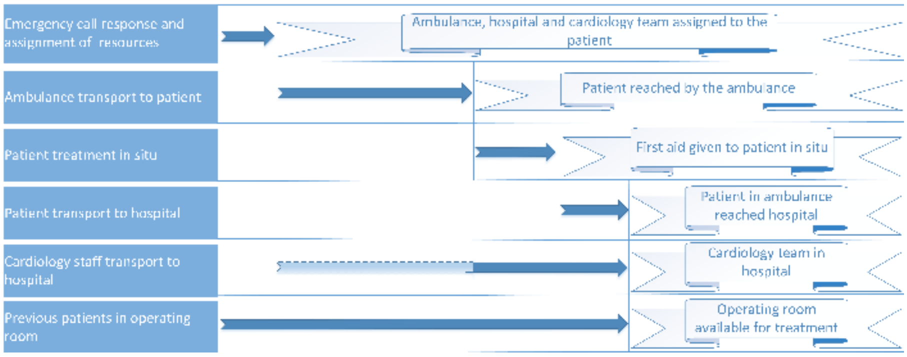

Fig. 1: Temporal sequence of six medical emergency events necessary for PCI treatment

- transportation time of the surgery team to the same,
- expected waiting time until the operating hall gets available.

The patient's and surgery team's arrival time to the hospital depends on their distance from the hospital and the driving conditions on the road. The availability of the operating hall depends on the previous patients (if any) booked for the operation hall with higher or equal urgency level to the patient in question.

In the case of multiple simultaneously appearing urgent patients, the objective is to find the minimum of the sum of all the patient delay times such that the system results in as high utilitarian value as possible. The objective is, therefore, twofold:

- to assign ambulances to simultaneous pending patients such that the assignment results in the minimum average time of transport of ambulances to simultaneous patients momentary locations considering their individual maximum allowed waiting times,
- to assign ambulances with patients to hospitals minimizing the combined times of patients transport to hospitals, and arrival times of surgery teams positioned outside hospitals, such that the difference between the expected arrival times of patients and surgery teams to hospitals is minimum.

In the following, we give the multi-agent system model and the mathematical programming definition of the problem inspired by [7].

Multi-Agent System representing EMS Considering a time horizon made of T time periods, given are four distinct agent sets. Let $\Xi = \{\xi_1, \dots, \xi_{N_p}\}$ be a pending patient set. Let $\Psi = \{\psi_1, \dots, \psi_{N_c}\}$ be a set of surgery teams, each one made of at least one surgeon, one anaesthetist, two nurses and an anaesthetist assistant. Let $A = \{a_1, \dots, a_{N_a}\}$ be the set of identical, capacitated ALS ambulance vehicles to be routed and scheduled to assist patients based on one-to-one assignment and let $H = \{h_1, \dots, h_{N_h}\}$ be after-hours urgent surgery-capable hospitals. Furthermore, all agent sets are represented by points in the plane. N_p, N_c, N_a, N_h and N_b represent (not necessarily equal) cardinality of each set respectively. Agents initial coordinates are positioned, w.l.o.g., in a square environment $E = [0, l]^2 \subset \mathbb{R}^2$ of side length $l > 0$. The abbreviation $p(t)$ is used for the position of any kind of agent at time $t = 1, \ldots, T$; e.g., $p_a(t) \in E$ being the position of agent $a \in A$ at the beginning of each time period $t = 1, \ldots, T$, where T is the last period of the planning time horizon we are interested in.

Mathematical Formulation We concentrate on the problem of the minimization of the average total delay time of urgent patients to get surgery treatment. No patient should be discriminated positively or negatively for his/her location. In the case that there is only one pending patient in the system, then the best ambulance is the ambulance which will arrive in the shortest time possible and the problem is to find ambulance $a \in A$, surgery team $\psi \in \Psi$ and hospital $h \in H$ that in combination minimize patient $\xi_k \in \Xi$ time to hospital:

$$\min_{a \in A} t(a, \xi_k) + \min_{h \in H_{av}} \left(\max_{h \in H_{av}} \Big(t(\xi_k, h), \min_{\psi \in \Psi_{av}} t(\psi, h) \Big), \min \rho_{h,\xi_k} \right), \qquad (1)$$

where hospital h_{ξ_k} chosen for patient ξ_k, $k = 1, \ldots, N_p$ is

$$h_{\xi_k} = \arg\min \left(\max_{h \in H_{av}} \Big(t(\xi_k, h), \min_{\psi \in \Psi_{av}} t(\psi, h) \Big), \min \rho_{h,\xi_k} \right), \qquad (2)$$

and Ψ_{av} is a set of available surgery teams and H_{av} set of available hospitals with necessary equipment. Furthermore, $t(x, y)$ is travel time from position x to position y and ρ_{h,ξ_k} available time periods of hospital h for patient ξ_k and $\min \rho_{h,\xi_k}$ is a first time period hospital h will be free for patient ξ_k. The objective for each patient $\xi_k \in \Xi$, thus, is to choose a triple $\langle a, h, \psi \rangle$ minimizing Equation 1.

From the global point of view, multiple-patient problem is to assign patients in order to optimize the global waiting time for the treatment for all patients, i.e., find assignments of $a \in A$ and $h \in H$ such that:

$$\min_{a \in A} \sum_{k=1}^{N_p} t(a, \xi_k) + \min_{h \in H} \sum_{k=1}^{N_p} \left(\max_{h \in H_{av}} \Big(t(\xi_k, h), \min_{\psi \in \Psi_{av}} t(\psi, h) \Big), \min \rho_{h,\xi_k} \right). \qquad (3)$$

Waiting time or patient delay is the sum of the time needed for the arrival of the ambulance to the patient, and the minimum value between the maximum of the arrival time of the patient to hospital and the arrival time of surgery team to the same (if not in-situ), and the minimum waiting time due to the pending patients booked for the operation room before patient ξ_k.

4 Solution approach

To improve the response times of the emergency management system towards urgent surgery patients, we present the dynamic resource assignment model for ambulances, surgery teams and hospitals assignment to patients performed over iterative combinatorial auctions, e.g. [2, 12]. The proposed solution is founded on the collaborative multi-agent system (MAS) organizational structure and MAS coordination model with four classes of agents seen as autonomous and independent decision makers. There exists a determined sequence of steps and message exchanges which is performed in order to resolve each urgent surgery case. The agents are described based on their characteristics and states as follows:

Patient: Each patient agent $\xi \in \varXi$ represents a real pending urgent surgery patient in the medical emergency assistance. When calling ECC, from his/her initial location, he/she gets assisted in-situ by ambulance crew, and gets transferred to hospital where he/she receives the urgent surgery treatment. Each patient is described over a tuple $\xi = \{p_\xi(t),\ \varDelta_\xi, t_\xi^{in}\}$, where $\varDelta_\xi$ is patient $\xi \in \varXi$ status which can be: pending patient waiting ambulance ξ_{wa}, being assisted in-situ ξ_{ais}, moving in ambulance to hospital ξ_{ath}, in hospital ξ_{inh}, and t_ξ^{in} is patient ξ detection time. The latter is defined as the time when the ECC is informed about the incident. New patient requests continuously unfold over time and must be assigned in real time to ambulances.

After-hour urgent surgery capable hospital: Hospital agents $h \in H$ collaborate with ambulances and emergency coordinator to receive patients for treatment. Furthermore, they are responsible of managing and coordinating their operation room(s) together with the assignable surgery team(s). Hospitals can be described over a touple $h = \{p_h,\ \rho_{h,\xi}\}$, where p_h is the position and $\rho_{h,\xi}$ is the temporal availability of hospital $h \in H$ for patient $\xi \in \varXi$. It is assumed that each hospital has a booking list for urgent and emergency surgery, i.e., information of the availability of the operation room within some future time. Hospitals have at the disposal the updated assignability of surgery teams $\rho_{\psi,h}(t)$ at every time period $t \in T$.

Ambulance: Ambulance agents $a \in A$ represent ALS ambulance vehicles (ambulances with advanced life support) together with their relative ambulance human crews. Ambulances communicate to ECC agent for patient and base station assignment and to hospitals for patient transfer. Furthermore, each ambulance is described over the touple $a = \{p_a(t),\ \ v_{avg}^{[a]}, s_a(t), b_a(t)\}$, where $p_a(t)$ is the current position at time period $t \in (1, \ldots, T)$ and $v_{avg}^{[a]}$ is the average velocity of ambulance a. $s_a(t)$ is its estimated end-of-service time with the current patient, if any. The dummy value -1 is used when the vehicle is free. $b_a(t)$ indicates the destination, i.e., the next station at which the ambulance vehicle will stop. Ambulances statuses can be: idle ambulance a_i, moving to incident position, a_{ip}, and ambulance moving to a hospital, a_h. At every time period t, idle ambulances a_i are considered for commitment to pending patients ξ_{wa}, and in case no patient assignment is made, they remain at their last assigned position.

In our model we assume that after arriving at patient location, the vehicle cannot be redirected elsewhere until transferring the patient to the hospital. However, at any time before getting to the patient location, the vehicle can be dispatched elsewhere.

Medical Emergency-Coordination Center: ECC receives emergency calls from patients and assigns the ambulance and hospital for each case, thus performing the high-level management of the urgent surgery logistic procedure.

Surgery team: $\psi \in \varPsi$ is responsible of the urgent surgery treatment. The team's members are positioned outside of hospital, generally at different locations, and move towards assigned hospital when needed. The combined arrival time to the hospital is the highest value of the members' arrival times. The touple which describes each surgery team is $\psi = \{p_\psi(t),\ \rho_{\psi,h}(t),\ b_\psi(t)\}$ where $\rho_{\psi,h}(t)$ is the temporal availability of surgery teams $\psi \in \Psi$ in hospital $h \in H$. It is assumed that each surgery team has its expected time of arrival to the hospital based on their momentary position $p_\psi(t)$ and the position of a hospital). The status of a surgery team can be: idle ψ_i, moving to an assigned hospital ψ_{mh}, in the assigned hospital ψ_{ih}. In general, the team can be assignable to different number of hospitals. Therefore, binary vector $\rho_{\psi,h}(t)$ expresses the assignability of the team for each time period $t \in T$ and for each hospital $h \in H$. Traditionally each surgery team is assigned to one hospital only. However, staff utilization and patient assistance can be significantly improved if all the regional surgery teams are at the disposal of all the region's hospitals.

4.1 Auction algorithm

The relaxation method for the assignment problem called auction algorithm [2] is used to resolve the problem of the assignment of ambulances, hospitals and surgery teams to urgent surgery patients. Auction algorithm is a coordination mechanism guaranteed to find the best assignment solution for the system; furthermore, it is an effective method for solving the classical assignment problem. It admits an intuitive economic interpretation and is well suited for implementation in distributed and decentralized computing systems as is the one in emergency medical assistance. Moreover, it is an iterative procedure related to a sales auction where multiple bids are iteratively compared to determine the best offer for the system, with the final sales going to the highest bidders. The original form of the auction algorithm is an iterative method to find the optimal prices and an assignment that maximizes the net benefit in a bipartite graph, the maximum weight matching problem (MWM). This algorithm was first proposed by Dimitri Bertsekas [2].

In auctions, it is important that the number of bidders is equal or higher than the number of bided objects. This is why, if the number of patients is lower than the number of ambulances, and the number of hospitals, then patients bid for ambulances and hospitals in the iterative auction algorithm based on the starting available patient assistance time described above. Otherwise, i.e., if the number of ambulances is lower than the number of patients, then ambulances bid for patients, and similarly higher number of hospitals bid for lower number of patients.

In the patient hospital assignment, we consider all pending patients who called the ambulance and are waiting for ambulance or are in the process of ambulance assistance and/or arriving to hospital but still haven't reached the same. In the following, we present the algorithm steps for hospital patient assignment. The assignment of ambulances to patients is performed in a similar way.

In each iteration

- Each hospital receives updated pending patients virtual prices (those are dual variables of the primal problem).
- Each hospital gives a bid based on the virtual prices of the patients.
- The hospital with the highest bid wins in the momentary iteration.
- If at the end of the bidding, all the patients received at least one bid and the bidding hospitals don't bid for the same patients, then there are no more unassigned patients.

- The algorithm updates the patients' prices and continues in iterations until all the patients are assigned and all the conflicts are resolved.

If each surgery team is assignable to more hospitals, preferably all, then the combinatorial result of multiple assignments gives a globally optimal solution while if each team is assigned only to one hospital, this can limit significantly the arrival time of the team to the hospital and therefore, in the case of unexpected prolonged arrival times, jeopardize the urgent surgery success.

The additional parameters of the simulation algorithm are N_{ξ}^{sim} being the total number of patients in the simulation and $N_{\xi}(t)$ representing the number of patients assisted in hospital until time period $t \in [1, \ldots, T]$. The complete simulation algorithm for emergency medical assistance of angioplasty patients follows the proceeding steps.

At each time $t \in [1, \ldots, T]$
While $N_{\xi}(t) \leq N_{\xi}^{sim}$

- assign all pending patients ξ_{wa} to idle ambulances a_i using auction algorithm;
- assign patients moving in ambulance to hospital ξ_{ath} to hospitals and standby surgery teams considering the arrival times of the teams;
- move ambulances a_{ip} to unassisted patients one step*;
- move ambulances with assisted patients to hospitals one step*;
- move surgery teams outside hospitals to assigned hospitals one step*;
- when a patient gets assisted in hospital, inform ECC of the availability of ambulance;
- introduce new patients based on the frequency of patient appearance.

* the step is calculated based on the average ambulance velocity and the duration of a time period.

5 Simulation experiments

In this Section, we describe the simulation setting, experiments, and results. The average patient waiting times resulting from the proposed optimized reassignment model are compared with the same based on the First-Come-First-Served principle used actually in many medical emergency coordination centers (e.g., SUMMA 112 in Madrid, Spain) and described previously.

In the simulation model, we follow a mesoscopic view of the emergency medical system and without loss of generality, ambulance velocities are set to an average system value. Together with the simplification of substituting the function of road travel time $t(\mathbf{x}, \mathbf{y})$ between positions $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ in Euclidean 2-space with Euclidean distance $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{2}(x_i - y_i)^2}$, we convert the road time minimization problem to Euclidean distance minimization problem which is independent of a road network structure different for each city and region.

Simulation setting We test the proposed strategy of optimized reassignment of ambulances and hospitals to patients looking at the average patient waiting times in the case of multiple pending patients and compare it with the benchmark First-Come-First-Served (FCFS) strategy of patient assignment. To demonstrate the scalability of our solution and possible application to small, medium and large cities and regions, in all of the experiments we vary the number of ambulances with ALS from 5 to 100 with increment 5 and the number of hospitals from 2 to 50 with increment 2. For simplicity and without the loss of generality, the number of surgery teams in each experiment equals the number of hospitals. The medical emergency system together with patients is positioned in the environment which dimensions are $[0, 50]^2 \subset \mathbb{R}$.

Each simulation is run over 300 patients. The number of experimented setup configurations combining different numbers of ambulances and hospitals with surgery teams sums up to 500. For each configuration, we simulate 5 instances of different random positions of ambulances, hospitals, surgery teams, and patients. Patients' positions are modeled based on the uniform distribution while patients' appearance frequency varies from low (1 new patient every 10 time periods) over medium (1 new patient every 2 time periods) to saturated one (1 new patient appearing in every time period). Time period can be considered here as a minimum time interval in which the assignment decisions are made; usually it is from 1 to 15 minutes.

In the proposed optimized reassignment model, surgery teams can be dynamically (re-)assigned to any hospital in every time period depending on the actual patient demand. Furthermore, we assume that the hospitals have at the disposal sufficient number of operating rooms so that the only optimization factor from the hospital point of view is the number of available surgery teams. If there are more patients with the same urgency already assigned waiting for treatment in the same hospital, they are put in a queue.

In the proposed model, surgery teams re-assignment to hospitals is performed as soon as an idle ambulance arrives to a pending patient. The former is made having in consideration all idle surgery teams, available hospitals, and new patients assisted by ambulances but still out of hospital.

For the surgery team arrival times to hospitals, we tested two assignment strategies: the first one minimizes the sum of the differences between the patients and the surgery teams arrival times to hospitals at the global level, while the second one concentrates only on the minimization of the arrival times of surgery teams to assigned hospitals independently of the arrival times of the assigned patients to the same.

We present the results of the latter since it gives significantly lower patient waiting times in all of the performed experiments. Even though the former considers a time window between surgery teams and patients arrival times, thus increasing the available time for surgery teams to arrive to the hospital when the patient has still not arrived, this strategy showed inferior to the minimization of arrival times of surgery teams without the reference to the assigned patients times. The reason for this is that without forecasting capabilities of new patients, the system is myopic towards new patients frequency appearance and positions and on the long run, the system suffers significant delays.

In the following, we present the results of the simulation tests.

Simulation results In the experiments, we test the performance of the proposed optimized reassignment strategy in respect to the FCFS benchmark model. For each out of 500 configurations, we use 5 instances of different patient, surgery team, hospital and initial ambulance coordinates. We compare the average patient waiting time of the proposed optimized reassignment method $\bar{t}_{OR}$ with the same of the benchmark FCFS model $\bar{t}_{FCFS}$. Relative performance function P of the proposed in respect to the benchmark model is calculated as:

$$P = \frac{\bar{t}_{FCFS} - \bar{t}_{OR}}{\bar{t}_{OR}} \cdot 100 \ , \ [\%]. \tag{4}$$

The simulation results of the performance function P for the three simulated cases of frequency of patient appearance of 1, 5, and 10 patients over 10 time periods are presented in Figures 2, 3, 4, and Table 1. The Figures show the increase of performance in average as the number of hospitals increases from slightly negative values up to more than 1000 $\%$ as seen in Table 1.

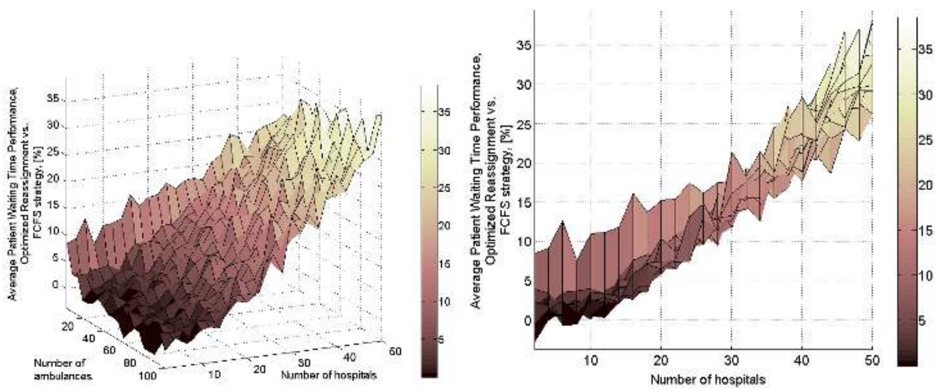

Fig. 2: Average patient waiting time performance of optimized reassignment strategy in respect to the FCFS strategy $[\%]$ for the frequency of appearance of 1 new patient every 10 time periods.

Observing the performance dynamics in respect to the varying number of hospitals, it is evident from Figures 2, 3, and 4 that with a relatively low number of after-hour urgent-surgery available hospitals, optimized reassignment gives similar results to the FCFS method. As the number of the hospitals increases, the performance increases in average up to the maximum of 38,52$\%$ for the frequency of patient appearance of 1 new patient every 10 time periods, Figure 2, and up to more than 1000 $\%$ in the cases with higher frequency of patient appearance, Figures 3 and 4.

Looking at the optimized reassignment performance dynamics in respect to the varying number of ambulances, in Figures 2, 3, and 4, two regions are evident: the first

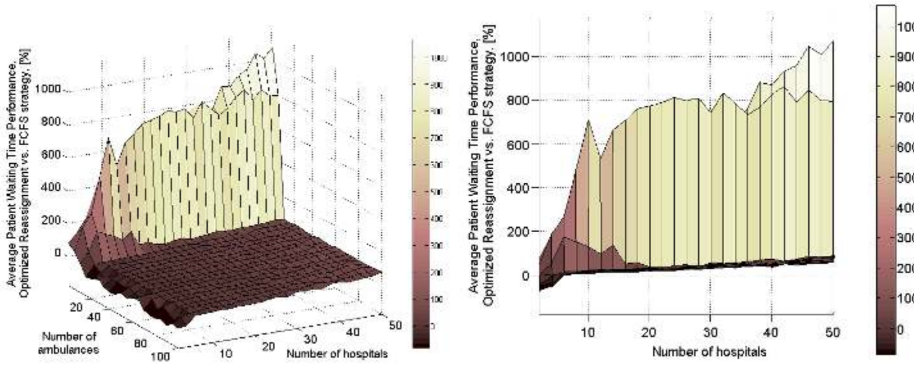

Fig. 3: Average patient waiting time performance of optimized reassignment strategy in respect to the FCFS strategy [%] for the frequency of appearance of 1 new patient every 2 time periods.

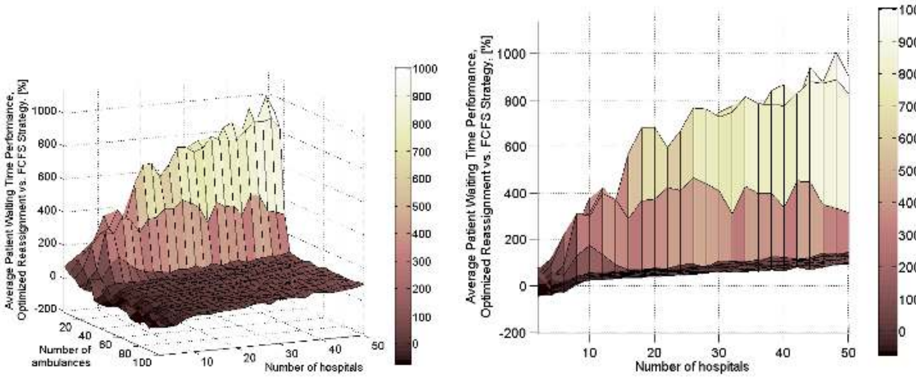

Fig. 4: Average patient waiting time performance of optimized reassignment strategy in respect to the FCFS strategy [%] for the frequency of appearance of 1 new patient every time period.

one with very low number of ambulances where the performance of the optimized reassignment is significantly better than the FCFS method, and the other region where the values do not change significantly in respect to the change of the number of ambulances. The performance values of the first region steeply decrease to the steady values of the valey region. It can be seen that as the frequency of patient appearance increases, thus the size of the region of significantly higher performance when the number of ambulances is low, increases starting at frequency 1/10 with 5 ambulances, in 5/10 frequency going up to 10 ambulances, and in frequency 1/1 arriving to 20 ambulances, Figures 2, 3, and 4. This implies that the optimized reassignment performance in respect to the FCFS method when the number of ambulances is low improves as the frequency of patient appearance increases.

Table 1: Experiments minimum and maximum values of performance function P

Frequency of patient appearance	1/10	5/10	10/10
P min.value, [%]	-2,92	-80,47	-74,62
P max. value, [%]	38,52	1067,6	1004,1

From Figures 2, 3, and 4, it is also visible that when the number of hospitals is low, minimum values of the optimized reassignment method performance increase as the patient appearance frequency increases. The number of hospitals for which the first two cases show strictly positive performance is 8, while for the case 3, it is 10. Proportionally to the increase of the number of hospitals, there is a constant increase of performance up to the maximum values as seen in Table 1.

Furthermore, as can be seen from Figure 2, when the frequency of new patient appearance is relatively low, 1 over every 10 consecutive time periods, the performance of the proposed optimized reassignment method increases in average proportionally to the increase of the number of hospitals. However, when the number of hospitals is relatively low, i.e., lower than 8, the optimized reassignment approach does not necessarily give a better patient waiting time solution. The reason is that by reassignment of surgery teams, they move from one hospital to the other, and are in the time of travel unavailable for patient assistance which worsens the patient waiting time. However, when the number of ambulances is relatively low, (lower than 20), the reassignment approach gives better results since with high number of ambulances, their geographical distribution compensates for the availability of surgery teams at all hospitals at all times and no additional combinatorial technique is necessary to improve the assignment performance. With lower number of ambulances and hospitals, since ambulances are not equally distributed in the area, the reassignment method compensates for their unequal distribution thus giving better results. This tendency is even more emphasized in the cases of higher new patient appearance frequency as seen in Figures 3 and 4 reaching up to more than 1000 % of improvement, Table 1.

6 Conclusions

In this paper, we proposed a heterogeneous multiagent system coordination model that facilitates a seamless coordination among the participants in the emergency medical assistance for the minimization of delay times of after-hours urgent surgery patients. The proposed model implies the change of the current functioning based on a manual coordination through communications via phone calls, towards an automated coordination process where the basic decisions are taken (or proposed) by software agents. The proposed multi-agent system model enables a better control of the availability of stand-by surgery teams and gives a decision making tool for ambulance and hospital assignment.

In order to reduce the transfer and waiting times of after-hours urgent surgery patients, we integrated in the multiagent model a multi-objective optimisation tool based on iterative auctions for the minimization of ambulances and surgery teams arrival times. The proposed solution results in the provably increased flexibility and responsiveness of the emergency system.

Simulation results prove the efficiency of the proposed solution resulting in significantly lower urgent surgery waiting times. The proposed auction mechanism enables spatially and temporally optimized resource assignment in the cases with multiple patients.

In real life, assumption on ambulances' equal velocities cannot be made so the estimation of arrival times should be made based on the more sophisticated methods and tools as, e.g., Google maps. For the usage of our technology, ambulances should have a GPS and a navigator for localizing the patient and navigating the way to him/her. ECC should have a digitalized map with localized ambulances, patients and hospitals and hospitals should have a receptionist service or personnel for admittance of patients.

As a future work, we plan to develop heterogeneous MAS coordination model for participants in Emergency Medical Assistance with integrated future patients forecast over a receding horizon.

Acknowledgements This work was supported in part by the Spanish Ministry of Science and Innovation through the projects "AT" (Grant CONSOLIDER CSD2007-0022, INGENIO 2010) and "OVAMAH" (Grant TIN2009-13839-C03-02) co-funded by Plan E, and by the Spanish Ministry of Economy and Competitiveness through the project iHAS (grant TIN2012-36586-C03-02).

Bibliography

[1] D. Bandara, M. E. Mayorga, and L. A. McLay. Optimal dispatching strategies for emergency vehicles to increase patient survivability. *International Journal of Operational Research*, 15(2):195–214, 2012.

[2] D.P. Bertsekas. The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of Operations Research*, 14(1):105–123, 1988.

[3] Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932, 2010.

[4] Elton Domnori, Giacomo Cabri, and Letizia Leonardi. Coordination issues in an agent-based approach for territorial emergence management. In *Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on*, pages 184–189. IEEE, 2011.

[5] MD Franklin Dexter. Influence of staffing and scheduling on operating room productivity. *Operating Room Leadership and Management*, pages 46–66, 2012.

[6] Francesca Guerriero and Rosita Guido. Operational research in the management of the operating theatre: a survey. *Health care management science*, 14(1):89–114, 2011.

[7] Ali Haghani, Huijun Hu, and Qiang Tian. An optimization model for real-time emergency vehicle dispatching and routing. In *82nd annual meeting of the Transportation Research Board, Washington, DC*, 2003.

[8] Shane G Henderson. Operations research tools for addressing current challenges in emergency medical services. *Wiley Encyclopedia of Operations Research and Management Science*, 2011.

[9] Shane G Henderson and Andrew J Mason. Ambulance service planning: simulation and data visualisation. *Operations research and health care: a handbook of methods and applications*, 70:77–102, 2004.

[10] Armann Ingolfsson. Ems planning and management. In *Operations Research and Health Care Policy*, pages 105–128. Springer, 2013.

[11] Beatriz López, Bianca Innocenti, and Dídac Busquets. A multiagent system for coordinating ambulances for emergency medical services. *Intelligent Systems, IEEE*, 23(5):50–57, 2008.

[12] Marin Lujak and Holger Billhardt. Coordinating emergency medical assistance. In *Agreement Technologies*, pages 597–609. Springer, 2013.

[13] Matthew S Maxwell, Mateo Restrepo, Shane G Henderson, and Huseyin Topaloglu. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22(2):266–281, 2010.

[14] Bożena Mielczarek and Justyna Uziałko-Mydlikowska. Application of computer simulation modeling in the health care sector: a survey. *Simulation*, 88(2):197–216, 2012.

[15] Jonathan Patrick, Martin L Puterman, and Maurice Queyranne. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research*, 56(6):1507–1525, 2008.

[16] Sonja Postma, Jan-Henk E Dambrink, Menko-Jan de Boer, AT Gosselink, Gerrit J Eggink, Henri van de Wetering, Frans Hollak, Jan Paul Ottervanger, Jan CA Hoorntje, Evelien Kolkman, et al. Prehospital triage in the ambulance reduces infarct size and improves clinical outcome. *American heart journal*, 161(2):276–282, 2011.

[17] Abdur Rais and Ana Viana. Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18(1):1–31, 2011.

[18] L Song, H Yan, and D Hu. Patients with acute myocardial infarction using ambulance or private transport to reach definitive care: which mode is quicker? *Internal medicine journal*, 40(2):112–116, 2010.

[19] Jeroen M van Oostrum, Mark Van Houdenhoven, Manon MJ Vrielink, Jan Klein, Erwin W Hans, Markus Klimek, Gerhard Wullink, Ewout W Steyerberg, and Geert Kazemier. A simulation model for determining the optimal size of emergency teams on call in the operating room at night. *Anesthesia & Analgesia*, 107(5):1655–1662, 2008.