An Intelligent Scheduling of Non-Critical Patients Admission for Emergency Department

EVA BRUBALLA\textsuperscript{1}, ALVARO WONG\textsuperscript{2}, DOLORES REXACHS\textsuperscript{2}, AND EMILIO LUQUE\textsuperscript{2}

\textsuperscript{1}Escoles Universitaris Gimbernat (EUG), Computer Science School, Universitat Autonoma de Barcelona, 08193 Barcelona, Spain
\textsuperscript{2}Computer Architecture and Operating Systems Department, Universitat Autonoma de Barcelona Campus Bellaterra, 08193 Barcelona, Spain

Corresponding author: Alvaro Wong (alvaro.wong@uab.es)

This work was supported in part by the Agencia Estatal de Investigacion (AEI), Spain, in part by the Fondo Europeo de desarrollo Regional (FEDER) UE under Contract TIN2017-84875-P, and in part by a research collaboration agreement with the Fundacio Escoles Universities Gimbernat (EUG).

\textbf{ABSTRACT} The combination of the progressive growth of an aging population, increased life expectancy and a greater number of chronic diseases all contribute significantly to the growing demand for emergency medical care, and thus, causing saturation in Emergency Departments (EDs). This saturation is usually due to the admission of non-urgent patients, who constitute a high percentage of patients in an ED. The Agent-based Model (ABM) is one of the most important tools that helps to study complex systems and explores the emergent behavior of this type of department. Its simulation more accurately reflects the complexity of the operation of real systems. Our proposal is the design of an ABM to schedule the access of these non-critical patients into an ED, which can be useful for the service management dealing with the actual growing demand for emergency care. We suppose that a relocation of these non-critical patients within the expected input pattern, provided initially by historical records, enables a reduction in waiting time for all patients, and therefore, it will lead to an improvement in the quality of service. It would also allow us to avoid long waiting times. This research offers the availability of relevant knowledge for Emergency Department managers in order to help them make decisions to improve the quality of the service, in anticipation of the expected growing demand of the service in the very near future.

\textbf{INDEX TERMS} Agent-based modeling and simulation (ABMS), Emergency Department (ED), response capacity, decision support systems (SDS), length of stay (LoS).

\section{I. INTRODUCTION}

Currently, the demand for emergency medical care is increasing and therefore, a hospital’s management of Emergency Departments (EDs) becomes more important. In particular, one of the most important issues to deal with is the number of patients who enter the service daily. This is becoming a major problem in EDs worldwide, since it requires the use of a large amount of resources, both human and material, which unfortunately are often very limited. In addition to the above, it should be stressed that the degree of coordination between these resources must be very high [1]. The main consequence of this situation is that the increase in the number of patients entering the service is at saturation point [2].

The result of this is that there has been an increase in the total time a patient spends in the service, from the moment the patient enters until they are discharged, which is called Length of Stay (LoS). This is the most accepted parameter in the literature, being used as an indicator of service quality. The aforementioned point of saturation can trigger a general discontentment among patients, including the feeling of being abandoned without care, limited access to emergency care and an increase in patient mortality [3].

One of the most complex areas of the hospital is the ED due to the dynamism and variability of the healthcare as well as the attention times. The operation of the service is the result of the interaction among the elements (agents) of which it is composed. The modeling and simulation of these systems, such as an ED, is one of the most powerful tools for their characterization. The simulation provides a better knowledge of their operation and it can assist decision-making to set up techniques for an optimal system operation [4], [5].

The ultimate objective of modeling and simulating a real system is to know more or find additional information.
about it. It is possible to achieve this through inference processes in the system’s variables of interest to make predictions about the behavior of these variables under different conditions, based on the information obtained from the new data generated [6], [7].

In previous studies, we developed an ED simulator based on an Agent-Based Modeling (ABM). The ED simulator has been validated within our research group in collaboration with Taulí Hospital, Sabadell. This hospital is one of the most important in Spain, attending over 160,000 patients per year. The model developed describes the behavior of ED services from the interactions between agents, as well as their interaction with the physical environment. It has been developed in NetLogo, a multi-agent programmable modeling environment for complex systems [8], [9].

Based on historical records of non-critical patients from Taulí Hospital, these types of patients do not require urgent attention and they represent a high percentage of the patients. It has been observed that saturation in the ED service is mostly due to the admission of these patients [10].

We propose scheduling policies for the entry pattern of non-critical patients consisting of relocation. From the patients expected arrival time in the input pattern, initially predicted by the historical records, this relocation will provide an improvement in the waiting time affecting all patients, and therefore, an improvement in service quality from the point of view of the patient.

The simulator is used as an evaluation tool of the result of the application of the proposed model. In fact, simulation is the only way to generate as much data as we want corresponding to different situations generated by the application of the model, regarding the scheduled input of patients into the system and the healthcare staff configuration. The analysis of this data is the way to show the improvement achieved. It would not be feasible to validate the model for the different possible situations in the real system. This is the main reason why we use simulation to obtain the necessary information for the model validation.

As a first contribution of our research, we have developed an analytical model [10], [11] to compute the throughput of the healthcare staff in an ED service, which is defined as the number of patients it can attend to per unit of time. The goal of the analytical model is to use the theoretical throughput as an index to evaluate the responsive capacity of the Staff, taking into account the configuration of the staff (admission and triage staff, number of doctors and nurses), and patient flow throughout the ED. The analytical model is based on a set of equations which takes into account not only the number of personnel, but also their level of experience and the care provided (additional tests or treatments). This model has also been validated by analysis based on records generated by simulation of the real ED.

The second contribution of this research is a proper definition of the scheduling model for non-critical patients’ admission into the service [12], as well as by their relocation with respect to the initial flow pattern supplied by the hospital’s records. The analytical model provides a way to measure the responsiveness of the healthcare staff, which will be used as a reference to carry out patient relocation. We validated the efficacy of the scheduling model based on the information from the actual records from Taulí Hospital and using the simulation to evaluate the results of its implementation.

The described research contributions offer the ED managers new knowledge about the behavior of the service, which may be relevant in decision-making, and of great interest with regards to the improvement of service quality, taking into account the expected growth in demand of the service in the very near future.

The paper is organized as follows: Section 2 presents the research objectives. In Section 3 we describe the related works. Section 4 gives a description of the main features of the ED and the simulator’s capabilities. In Section 5 we describe the analytical model for the calculation of the response capacity in the ED. Section 6 describes the scheduling model for non-critical patients’ admission. In Section 7 we present the experimental results. Finally, Section 8 closes the paper with our conclusions.

II. RESEARCH OBJECTIVES

The overall objective of the research we propose is to develop a methodology that enables us to enhance the best of attention provided in a medical ED, with a view to lessening an affected person’s length of stay inside the service, via a model for programming the admission of non-critical patients within the service.

We outline this objective in the following specific objectives:

- Set suitable indicators of system behavior which allow us to assess the impact of the proposed model.
- Typify the system in line with the demand for attention at every moment and for each viable situation. Decide on the responsiveness of the healthcare staff configuration in relation to this particular demand.
- Develop a system to permit programming the admission of non-critical patients into the service, consistent with attention capacity, and analyze the effect produced on the quality of the service based on patients’ length of stay prediction obtained by using simulation.

We will be able to analyze the effect of the model’s application, and corroborate its effectiveness, based on the values of the defined indicators or different variables of interest such as patients’ length of stay inside the service, calculated from the data generated by ED simulation.

III. RELATED WORKS

In the related literature there are studies which propose reducing the LoS, and therefore, the time that the patient is waiting to be attended (LoW). Some of the solutions that have been implemented are defined as Fast Tracks [13], [14], or other solutions known as See and Treat [15]. Rachuba et al. [16] show that diagnostic imaging services are essential to the diagnosis pathway for many patients.
Commonly, these patients need to be seen again by a doctor or emergency nurse after an X-ray has been taken to complete the diagnosis and determine the next stage in the patients’ pathway. Finally, Rachuba proposes process mapping and assessing the reduction in patients’ length of stay in ED.

Many studies try to find the elements that impact patients’ long periods of stay inside the service and its saturation [17], [18]. Others show that saturation and long waits increase the proportion of patients who leave the service without being attended by a doctor (LWBS) [19], [20]. In our research group, we looked for the best healthcare staff configuration to decrease the patients’ LoS in the service, considering constraints related to the cost of the configurations and the amount of available resources [21]. We highlight those references as well as the use of simulation to test the effectiveness of the proposed measures for improvement in the LoS of patients within the service [22]–[25].

Our proposal attempts to move a step forward by obtaining a particular method to reduce the LoS within the ED service. In this way, the improvement in the quality of the ED is achieved by changing the manner in which non-critical patients arrive at the ED and not as much by way of modifying the healthcare staff nor the physical resources. As explained in related papers, simulation gives us a way to measure quality improvement inside the ED by the application of the model.

IV. EMERGENCY DEPARTMENT MODEL

As seen in Figure 1, the ED operation consists of different stages which every patient goes through from their entry into the ED service until these patients are discharged from the service or admitted into the hospital.

For the triage stage, patients are classified according to their acuity level and they are assigned a priority. The scale of urgency implemented in the Spanish Triage service is based on the Andorran Triage model (MAT) [26] as described in Figure 2. A statistical evaluation of these records validates that the great number of patients attending the ED are non-critical patients who do not need immediate attention or they can even be outpatients.

The patients’ distribution by acuity level from historical records is described in Figure 2. A statistical evaluation of these records validates that the majority of patients attending the ED service are non-critical patients who do not need immediate attention or they can even be outpatients.

If it were possible to give information about when it was more advisable to go to the ED service, non-critical patients would probably do it when the likelihood of long waits was lower. The main beneficiaries would be these types of patients, who could take advantage of a scheduling model for admission in the ED service.

A. THE FLOW OF PATIENTS INSIDE OF THE ED SIMULATOR

We developed the ED simulator including the following agents in Netlogo: admissions and triage staff, doctors for the different hospital areas and assistant nurses for the additional tests and treatments. The agents have their own state that will be modified by the interaction with other agents, these interactions will be reflected in the global operation of the ED service.

As shown in Figure 3, when the patient arrives at the service, the simulation runs according to the flow of patients. The admissions and triage stages are the same for all patients entering the ED service. Then, patients with acuity levels 1, 2 and 3 are treated separately from the non-critical patients (levels 4 and 5), which means that they are attended by different doctors and nurses. There is also a low percentage of patients being referred to other hospital areas after the triage and others who leave the service without being seen by a doctor (due to waiting time).

In our proposal, we’re interested in analyzing the non-critical patients, patients who can be relocated in time for
their arrival at the ED service. We will take into account all patients for admissions and triage stages, but only patients with acuity levels 4 and 5 for the diagnosis and treatment stage. In this stage, the patients go through an initial medical exploration (IE) phase. A percentage of them are immediately discharged and leave the ED service after this phase, as we have illustrated by a dashed line in Figure 3. The remaining patients in the ED services go through a section of complementary examinations and/or treatment carried out by technical personnel and/or assistant nurses. After this, they return to visit the same doctor, who analyzes the additional tests or/and treatments. Finally, the patients are discharged from the ED, as shown in the same Figure 3.

Each scenario of the ED simulation is identified through a healthcare staff configuration and an input pattern of patients in the ED service (quantity and acuity level of the patients per hour). We add sensors to the ED simulator in order to obtain completely temporalized information about the output for each scenario, taking as a result information regarding the number of attended patients, Patient attention Time (PaT) and waiting time (LoW) for each patient in any stage in the ED service.

V. ANALYTICAL MODEL
The level of satisfaction with emergency care is mostly conditioned by the perception of waiting time. Furthermore, from the standpoint of service operation. The throughput of the ED is directly related to the number of patients attended and an efficient use of the ED facilities.

We are proposing a model for service characterization. This model should provide knowledge and information to make changes in the system in order to improve it. Such a model is based on the definition of a set of service quality indicators as well as a set of equations, allowing us to measure intrinsic elements of the system, such as staff activities and patient flow, as shown in Figure 3.

The equations described in this section will enable us to gain knowledge (knowledge) about the capacity of the system with regards to the service resources. The objective is to use this knowledge to design a method for the distribution of non-critical patients, changing their current entry pattern, so that their arrival at the service should be in accordance with the computed capacity of the service.

A. QUALITY SERVICE INDEXES
We start defining an index named Patient attention Time (PaT) as the total time a patient is receiving attention throughout the stages in the service (admissions, triage, doctors, additional tests or treatments) for a specific member of staff. The PaT is calculated from the sum of the times in attention in each stage, which is obtained from the ED simulator based on the statistical records provided by the hospital (Eq. 1):

\[ \text{PaT} = \sum_{i=1}^{\text{Stages}} \text{PaT}_{\text{stage}_i} \]  

\( \text{PaT}_{\text{stage}_i} \) defines the Patient Attention Time in stage i, and it is independent from the number of patients arriving at the ED service. It’s important to know that PaT is not a static value for all patients, as it depends on the path followed by each patient (not all patients need additional tests or treatment).

It’s important to note that the parameter commonly used in the literature as an indicator of service quality is the Length of Stay (LoS), which is defined as the total time a patient spends in the ED service. Unlike the previous one, the LoS not only depends on the Staff configuration but also on the number and type of patients admitted to the ED service, including the waiting time. Finally we use the Length of Wait (LoW) as an indicator. This is defined as the total waiting time of a patient throughout the service. Therefore:

\[ \text{LoW} - \text{PaT} = \text{LoW} + \text{PaT} \leq \text{LoS} \]  

\[ (2) \]

Additionally, the Equivalent Patient attention Time (EPaT) for stage i (EPaT_{\text{stage}_i}) is defined as the attention time of a patient, taking into account the possibility of working in parallel in that stage, and Equation 3 shows how it is computed. The SS_I and JS_I in Equation 3 and 5 stand for the total number of senior/junior staff in stage i respectively, and the calculation corresponds for parallelization on a pipeline model.

\[ \text{EPaT}_{\text{stage}_i} = \frac{1}{\text{SS}_i \text{PaT}_{\text{SS}_i} + \text{JS}_i \text{PaT}_{\text{JS}_i}} \]  

\[ (3) \]

The slowest stage of the staff configuration will represent the capacity at which patients can be attended in the ED service, as well the stage which can saturate the service. It is, therefore, the inverse of the equivalent attention time of the slowest stage, which will determine the number of patients that a given configuration can treat per unit of time. We define this index as Theoretical Throughput (ThP), which is the indicator for measuring patient attention capacity, which is its response capacity for specific scenarios. The ThP is represented by Equation 4.

\[ \text{ThP} = \frac{1}{\text{MaxEPaT}_i} \]  

\[ (4) \]

Indeed, the ThP for a specific stage i will be obtained by the inverse of Equation 3.

\[ \text{ThP}_{\text{stage}_i} = \frac{\text{SS}_i \text{PaT}_{\text{SS}_i} + \text{JS}_i \text{PaT}_{\text{JS}_i}}{\text{SS}_i \text{PaT}_{\text{SS}_i} + \text{JS}_i \text{PaT}_{\text{JS}_i}} \]  

\[ (5) \]

B. THEORETICAL THROUGHPUT FOR THE DIAGNOSIS AND TREATMENT STAGES
Unlike previous stages, the diagnosis and treatment stages are the most complex stages due to their non-linearity. The patients first go through an initial medical exploration (IE), which is their first contact with the doctor. There is a percentage \( p_1 \) of patients who need more tests after the IE phase and also a percentage \( p_2 \) of patients who need some treatment, which is applied and monitored by the nurses. After the test or treatment phase, the patients return to the doctor for a
final diagnosis, after completing any additional examinations (AR). The remaining patients will be discharged from the ED service directly after their first visit with the doctor. Figure 4 details the flow of patients in this phase, according to the abovementioned considerations. The number of assistant nurses, both Senior and Junior, in the staff service is presented as SN/JN respectively. The number of doctors, also Senior or Junior, is represented by SD/JD, and it is important to distinguish between:

- \( SD_{IE}/JD_{IE} \): Senior / Junior doctors in the Initial Exploration.
- \( SD_{AR}/JD_{AR} \): Senior / Junior doctors in the Analysis of Results.

In the ED services, the doctors prioritize the attention of patients after the IE, therefore, these patients will be treated when the doctor is available for the analysis of results (AR). This prevents queues on the return of patients after the additional tests or treatments.

As we described before, the ThP has been defined as an index of the response capacity for each stage. In order to calculate the ThP in the diagnosis and treatment stage, it is important to consider the mean attention time of each type of doctor. This time will depend on experience (Junior or Senior), and also on the type of healthcare they are providing, the first step (IE), or the second, consisting of the AR of a requested additional treatment. These times are obtained by the calibration of the ED simulator, and denoted by \( PaT_i^j \), which represents the mean of the Patient Attention Time for a doctor type \( i \) doing \( j \).

Then we consider:

- \( PaT_{SD}^IE \): Mean attention time of a Senior Doctor (SD) in the IE phase.
- \( PaT_{SD}^AR \): Mean attention time of a Senior Doctor in the AR phase.
- \( PaT_{JD}^IE \): Mean attention time of a Junior Doctor (JD) in the IE phase.
- \( PaT_{JD}^AR \): Mean attention time of a Junior Doctor in the AR phase.

Given these times, their inverse will provide us with the number of patients that each doctor can treat per unit time considered:

\[
\begin{align*}
P_{SD}^{IE} &= \frac{SD_{IE}}{PaT_{SD}^{IE}} = \text{Patients per minute for a SD in IE stage;} \\
P_{JD}^{IE} &= \frac{SD_{JD}}{PaT_{JD}^{IE}} = \text{Patients per minute for a JD in IE stage;} \\
P_{AR}^{SD} &= \frac{SD_{AR}}{PaT_{SD}^{AR}} = \text{Patients per minute for a SD in AR stage;} \\
P_{AR}^{JD} &= \frac{JD_{AR}}{PaT_{JD}^{AR}} = \text{Patients per minute for a JD in AR stage.}
\end{align*}
\]

Where \( SD_{IE} \), \( SD_{AR} \), \( JD_{IE} \), \( JD_{AR} \) are unknown values.

From the historical records provided by the hospital, we see that the patients can go once or more times for the tests or/and treatments, and so they visit the doctor more than once, as shown in Figure 5. The hospital statistics for non-critical patients show that the percentage of patients who need more than one test or/and treatment is very low.

There are two percentages for the complementary examinations. The first percentage \( p_1 \) of patients who, after their first visit with the doctor, require tests and the second percentage \( p_2 \) of patients that require treatment. Finally, there is a percentage \( 1 - (p_1 + p_2) \) of patients who are discharged from the ED service directly after their IE.

From the data presented in Figure 5, we can see that around 70% of non-critical patients leave the ED service directly after the IE. Therefore, 30% of patients require additional tests or treatment (\( p_1 + p_2 \)). Thus, given these percentages and the flow of patients in Figure 4, we obtain the following relations of continuity:

\[
\begin{align*}
P_{SD}^{IE} \cdot (p_1 + p_2) &= P_{SD}^{AR} \\
P_{JD}^{IE} \cdot (p_1 + p_2) &= P_{JD}^{AR} \\
SD_{IE} + SD_{AR} &= SD \\
JD_{IE} + JD_{AR} &= JD
\end{align*}
\]

The result of this linear system of equations gives us the values for \( SD_{IE} \), \( SD_{AR} \), \( JD_{IE} \), \( JD_{AR} \), and therefore, the values for \( P_{IE}^{SD} \), \( P_{IE}^{JD} \), \( P_{AR}^{SD} \), \( P_{AR}^{JD} \), for the doctor staff considered.

Once the diagnosis and treatment phases have been characterized, we can obtain the ThP for the doctors’ stage in the test and treatment phases by the sum of patients who have only been visited once by the doctor \( (P_{\text{only}IE}) \), those who have been required for complementary testing \( (P_{\text{Test}}) \), and the patients
TABLE 1. Theoretical throughput for each stage of the ED service and simulation results for the validation.

| STAFF | SIMULATION ED DATA |
|-------|---------------------|
|       | Healthcare Staff | PA (minutes) | ThP (pat/hour) | Input PA/hour | Occupancy (%) |
| ADMISSION PHASE | Junior | Senior | Junior | Senior | 20 | 21 | 22 | 23 |
|          | 3 | 0 | 8.00 | 6.00 | 22.50 | 99.72 |
| TRIAGE PHASE | 1 | 2 | 12.00 | 8.00 | 20.00 | 99.70 |
|          | 18 | 19 | 20 | 20.00 | 100.00 |

For each stage in the ED service were in accordance with the ThP obtained results are shown in Table 1. These simulation results are in accordance with the ThP obtained with the analytical model (Table 1) for all stages of the service.

Once the parameters for the Staff configuration have been configured, and according to the results in the same Table 1, a constant and homogeneous input of patients was created to ensure that the system was continuously running homogeneously to guarantee that the system has been in a steady state, after a warm-up period.

We run the simulation for three different numbers of patients entering the ED service per hour, around the theoretical ThP obtained as reference for the ThP for each stage from the system equations proposed. Afterwards, we carried out an analysis of the effect of the number of patients arriving into the service every hour using as an index the percentage of Staff time spent on attending or treating patients (which we defined as Occupancy) for each stage of the service. The obtained results are also shown in Table 1. These simulation results are in accordance with the ThP obtained with the analytical model (Table 1) for all stages of the service.

We validated [10] the analytical model to calculate the values of the ThP for admissions, triage, nursing and medical exploration stages. For the validation of the model, we used the ED simulator based on an agent-based model of the system as a sensor of the real system. The output information from ED simulation using different possible scenarios has been analyzed in order to get the information for the model validation.

The analytical model for the ThP calculation presented in this section will give us information for relocating non-critical patients, so that the theoretical ThP will be a mandatory indicator for the relocation of non-critical patients.

VI. SCHEDULING MODEL FOR NON-CRITICAL PATIENTS

The intention of the scheduling model presented in this section is to dynamically adjust the current patient pattern coming into the service to its attention capacity, in order that the flow of patients inside the service shall be in conformity with the healthcare staff available in the service at any time.

The admissions scheduling model proposed must enhance quality of care, optimize the quality perception of the attention service in relation to the population, and contribute to the sustainability of the current ED, ensuring better use of available resources. Therefore, the idea aims to improve the ED service, which is the principal entrance of patients in the healthcare services with regards to quality and user satisfaction.

The scheduling model for the entry of non-critical patients into the service is built based on the records extracted from the historical records of the hospital. On the other hand, we have defined as theoretical system throughput (ThP) the service characterization in terms of its response capacity to patients' attention. Its value is an indicator of the system capacity to absorb the demand for the service and it is a constraint in our model. It is important to keep in mind that physical space could become a bottleneck in the service, which is why we propose that, based on the throughput of the medical staff, we are able to define the capacity of the waiting rooms.

The diagram in Figure 6 gives a global view of the cycle required to dynamically acquire an appointment scheduling for the admission in the service, in keeping with the present hourly demand.

The idea is to have a recommendation plan concerning the admission of non-critical patients into the ED service. We based the model on a patient scheduling algorithm. Furthermore, we take into account the attention capacity of ED staff as a limitation, as well as the maximum delay time.
for relocation of patients as a second restriction. We based the knowledge of the system state hour-by-hour dynamically, which in turn is generated from the data obtained from the historical records and the changes made to it according to the actual demand of the ED, in relation to the access of patients and the kind of care received.

A. DISTRIBUTION OF NON-CRITICAL PATIENTS: SYSTEM STATE

We define the System State (SS) as the number of patients within the service each hour and we focus on the Patient attention Time (PaT). PaT is given as the staff time (doctor, assistant nurse or technician) spent attending the patient and the time for additional tests and/or treatments. Therefore, the SS does not take into account the Waiting Time for patients (LoW). This information provides us with a more realistic representation of what occurs hour-by-hour in the ED service.

From the hospital’s historical records, we obtain the distribution of patients arriving in the ED service, which gives us an initial approximation of the number of patients arriving each hour (historical entry patients in Figure 7). The patients that do not need additional tests or treatments are discharged by the doctor after the Initial Exploration (IE) and the patients need less than 1 hour to be attended (Direct patients), as shown in Figure 7).

Besides that, there’s a mean of 4 hours for patients who receive additional treatment, and 2 hours for those who need some tests. The SS is represented hour-by-hour, and for each hour we obtain the number of these types of patients, (test patients and treatment patients) as shown in Figure 7. The distribution probabilities of tests and treatments are provided by the hospital’s historical records.

The values of the PaT have been calculated from the calibration of the ED simulator using actual historical records from Tauli Hospital. Furthermore, it is important to note that the ED simulator contemplates a random exponential distribution to model the behavior of attention time by the staff, as well as the acuity level and the age of each patient [27]. We carried out a statistical analysis using the output of the simulation, which is given as the result of the mean PaT value for patients. This depends on whether or not they require additional treatments or tests, as we describe in Table 2.

In accordance with the values of PaT in Table 2, we have to contemplate the propagation of the patients through the following hours after their arrival in the system because they will be receiving or waiting for attention occupying an area in the ED service. Figure 7 shows test and treatment propagated patients for specific input of patients in the Historical Entry Patients row, corresponding to input pattern for a particular day from the statistical records of Tauli Hospital. For each patient we calculate the propagation time taking into consideration the mean PaT (Table 2) and for each hour i Propagated Patients $i = \text{TestPatients}_{i-1} + \text{TreatPatients}_{i-1} + \text{TreatPatients}_{i-2} + \text{TreatPatients}_{i-3}$.

Finally, the sum of Entry Patients and Propagated Patients is represented in the System State row, which is the hourly distribution of patients in the system. Therefore, the number of patients inside the system in hour $i$ is the corresponding value for $\text{SystemState}_i = \text{EntryPatients}_i + \text{PropagatedPatients}_i$. A graphical representation of the SS in Figure 7 is illustrated in Figure 8.

Here we consider that:

The bars of the graph represent the patients in the system at the arrival time as well as the hours of stay in the service. We divided the patients that arrive and are discharged after the visit with the doctor (PaT: less than one hour) from the patients who arrive and require additional treatments or tests, so we defined the latter as propagated patients according to the corresponding PaT, as shown in Table 2. As we mentioned, the entry hour for non-critical patients is illustrated in the bars, first in the bar representing their arrival and the following bars representing the hours while they are being attended by the doctor or performing the tests or treatments. We use the representation $\text{hour : patients}$ in Figure 8. It also helps to follow these types of patients in the bar chart.

The horizontal non-contiguous lines seen in the Fig.8 indicate three values that represent the attention capacity of the system (ThP), which in turn show the ideal scenario regarding patient care. If patients per hour do not surpass the ThP, then these patients should not have to wait to obtain medical attention. The value of the ThP also indicates whether it is possible or not to enhance the actual scenario by relocating patients, so that the number of patients to attend per hour becomes as homogeneous as possible and does not exceed the ThP limit value. Three possible scenarios may occur:

- If the value of the ThP is above the attention necessities, the system is oversized, then there should be no saturation and no modifications are needed.
- If the value of the ThP is below minimal service necessities, there is no alternative for patients’ relocation under ThP and the system cannot avoid saturation without modifying resource availability. However, a possible option for improving the scenario slightly is by trying to flatten
the curve, therefore reducing patients’ LoW inside the system.

- An intermediate case in which there is the option of patient relocation is when we are able to act to enhance system attention.

In the last scenario, we use the ThP value as an index for the scheduling model and it will be a constraint in the patient scheduling model, enhancing their current arrival pattern in the system, so that their arrival at the service should result in an SS in accordance with the calculated system capacity.

### B. NON-CRITICAL PATIENTS SCHEDULING MODEL

The relocation of the patients consists of the movement of the patient with respect to their arrival hour in the ED service to minimize the waiting time of the patient and consequently the LoS for all patients in the service. Only non-critical patients can be relocated and we must take into account that ThP is the maximum number of patients that the system can attend per hour without waiting times. The maximum delay time for patient relocation is 6 hours with respect to arrival hour. When the ThP has been calculated, the algorithm begins in hour \( i = 23 \) and goes backwards until the initial hour, which is the first hour in the calculated System State with a number of patients that surpasses the ThP. The algorithm performs the following steps for patient relocation:

- It determines holes (free locations for patients to move to). Move backwards hour-by-hour, until identifying the first Critical Hour (the first hour in the day with a number of patients which surpass the value of the ThP) in the calculated System State. We also identify Tentative Patients to be relocated (Figure 9).

- Depending on the relocation range (6 hours) and the restriction of ThP, the algorithm removes the tentative patients from the Critical Hour to the corresponding hour for relocation and it creates holes, removing more patients if possible. Then, the algorithm recalculates patients’ propagation for the new scenario and updates the SS, as illustrated in Figure 10.

- Once the Initial Hour is reached, the algorithm generates an update of SS in all the hours, and also of Entry Patients, which is the Schedule Entry Patients in Fig.11. In the same Figure the Schedule Patient Limits indicates the number of patients that should be admitted into the ED system per hour.

- The Appointment Scheduling Table is created (Figure 12), which describes the appointment hour for the relocated patients depending on their arrival time.

The characterization of the SS and the creation of the Appointment Scheduling Table have been developed as a
E. Bruballa et al.: Intelligent Scheduling of Non-Critical Patients Admission for ED

VOLUME 8, 2020

FIGURE 11. The System State updated (step 3).

FIGURE 12. Appointment Scheduling for non-critical patients (step 4).

FIGURE 13. System recommendation for non-critical patients into the service.

VOLUME 8, 2020

patients arriving each hour into the ED are calculated for each case study considered (called cases A, B and C), once the model for the admission has been carried out, that is to say, through the relocation of the patients, based on the System State calculated in each case. It is important to note that Area B in Tauli Hospital (the waiting room for non-urgent patients) was modeled according to the capacity of the room, in all cases defined by 52 patients (always greater than the ThP of the medical staff).

In Tables 3 to 5, the average reduction of LoW in minutes (time) is specified, as well as the corresponding reduction percentage foreseen by the simulation data, with respect to the original LoW without relocation, for all the hours that have been affected by the relocation in which a minimum reduction of 10% has been obtained, which is more than 10 minutes. LoW reduction values are shown separately for patients who are discharged after their first consultation with a doctor and do not require any test or treatment (Direct Patients), patients who need additional tests (Test Patients) and those who receive some type of treatment (Treatment Patients) respectively.

The initial relocation hour is marked in ochre in each table. The relocation effect can be extended to the previous hours, in addition to the following hours. This is because relocation can affect the waiting times of the test and treatment patients who arrived in the hours previous to the initial hour, due to its propagation. If this propagation reaches the hours in which other patients have been relocated. Some hours with atypical values in the LoW have been detected for treatment patients. Due to the outliers, the results with a reduction in the LoW have been rejected.

Finally, the expected patient input (historical input) and the input programmed by relocation (programmed input) are shown graphically for each case (Figures 14, 15 and 16). On the same graphs, the corresponding values of the average patient LoS (direct, test and treatment patients) in the service at each hour are represented, without relocation and after patient relocation, respectively.

A. CASE A

This case is characterized by a distribution of patients entering the ED corresponding to that determined by the current data from the historical records of Tauli Hospital.

Table 3 contains the details of the reduction in the LoW per hour for patients due to the relocation proposed by the model in this case. The data analyzed to obtain these values are those represented in Figure 14.

In this case, the initial hour for relocation, is hour 10 and relocation has an effect on hour 8 to hour 18. The effectiveness of relocation is noticeable during all these hours, reaching the maximum reduction value in hour 12 for all patients. The reduction in the case of direct patients only exceeds the minimum value by 10 minutes in hour 12. Even so, those hours with a reduction of 9 minutes are also shown, with these reductions, in all cases, very close to 50% compared to the original LoW value.
TABLE 3. Case A: Time (minutes) and percentage of average reduction in LoW due to patient relocation.

| Hour | Direct Minutes | Test Patients % | Treatment Patients % |
|------|---------------|-----------------|----------------------|
| 0    | -             | -               | -                    |
| 1    | -             | -               | -                    |
| 2    | -             | -               | -                    |
| 3    | -             | -               | -                    |
| 4    | -             | -               | -                    |
| 5    | -             | -               | -                    |
| 6    | -             | -               | -                    |
| 7    | -             | -               | -                    |
| 8    | -             | -               | -                    |
| 9    | -             | -               | -                    |
| 10   | 9             | 47%             | 43%                  |
| 11   | 13            | 45%             | 51%                  |
| 12   | 13            | 54%             | 49%                  |
| 13   | 9             | -               | -                    |
| 14   | 14            | -               | -                    |

For test patients, the relocation of patients has an effect in the LoW value until hour 16 and for treatment patients, the effect is extended until hour 18. In both cases, the effect of relocation on the previous hours to the initial hour is also shown. The LoW for these test or treatment patients who arrive in those previous hours could be reduced due to the fact that the subsequent hours have been modified by relocation of other patients, since they may be propagated in the system for two, three or four hours. This effect is observed in Table 3 for hours 8 and 9.

Finally, the graphs in Figure 14 show the effect of relocation on the entry of patients and their length of stay (LoS).

The relocation of patients and the consequent scheduling of their admission entails a redistribution in the arrival of these patients to the service according to the system’s attention capacity. This new situation results in an improvement in the length of stay of the patients in the service, as can be observed in the LoS lines represented on the graphs of the Figure 14.

B. CASE B

Case B is characterized by an input distribution of patients with a high concentration of patients in the central hours of the day. Table 4 contains the details of the reduction in LoW per hour for patients due to the relocation proposed by the model. The effect of relocation focuses on the 6 hours following the initial relocation hour, and it also affects the previous hours for treatment patients admitted to the system during those hours.

As already mentioned, the effect for treatment patients in these previous hours to the initial relocation hour is due to the fact that these patients remain in the service during the hours in which relocation of other patients has been carried out and that reduces their wait. An average reduction of more than 50% is achieved in all hours between hour 9 and hour 12 for the three types of patients. Once again, the positive effect of patient relocation is confirmed in this case, precisely in the central hours of the day for which the entry of patients has been considered greater.

The graphs in Figure 15 show the effect of relocation on the entry of patients and on their total time in the service (LoS).
TABLE 5. Case C: Time (minutes) and percentage of average reduction in LoW due to patient relocation.

| Hour | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|------|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Direct Patient | - | - | - | - | 30 | 46 | 55 | 66 | 40 | 53 | 58 | 54 | - | - | - | - | - | - |
| Test Patient | - | - | - | 67 | 81 | 86 | 50 | - | - | - | - | - | - | - | - | - | - | - |
| Treatment Patient | 69 | 110 | 127 | 154 | 183 | 165 | 60 | - | - | - | - | - | - | - | - | - | - | - |

FIGURE 16. Case C: Entry of patients with 17 more patients due to an accident at hour 13 without relocation (historical input) and with relocation (scheduled input). Average LoS per corresponding hour for direct, test and treatment patients without relocation and with patient relocation.

The previous case, given the greater saturation of the service in the central hours of the day without relocation. In the case of direct patients, the LoS curve flattens out almost completely. For test patients, LoS with patient relocation is maintained between 150 and 200 minutes from hour 9 (initial relocation hour), and for treatment patients, LoS with patient relocation remains below 300 minutes.

C. CASE C

In this case, we have experimented with a scenario characterized by an influx of patients with an unexpected increase in patients at a certain time of day, simulating a possible accident. The scenario contemplates an entry of 17 patients in hour 13.

The graphs in Figure 16 show the effect of relocation on the entry of patients and the total time of their stay in the service (LoS) for this case. The relocation of patients, before and after the arrival of the patients by the simulated accident, manages to improve the times of all patients in the hours affected by the accident. In return, the hours furthest from the hour of the accident, used for scheduling relocated patients to avoid saturation in critical hours, suffer an increase in waiting times. Even so, the results show a global improvement in patient waiting times.

These experimental results, obtained through the analysis of simulation data of the three case studies considered, demonstrate the positive effect of the patient relocation method and, therefore, validate the scheduling model for the admission of non-critical patients into the ED. With its application through the relocation of these patients, the LoW improves in all cases, and consequently the LoS of the patients and, therefore, the quality of care provided in the ED. Table 5 contains the details of the reduction in LoW per hour for patients due to the relocation proposed by the model for this case.

VIII. CONCLUSION

The main contribution of this research has been the definition of an analytical model for the calculation of the response capacity of a certain healthcare staff configuration in an ED. We have seen how this characterization of the system, through its ability to absorb the demand of the service, indicates the ideal situation of the state of the system in each hour and it is a reference for measuring system performance.

We proposed a scheduling model for non-critical patients which provides a method to enhance the quality of the healthcare service and it is a useful tool for the service administration in order to attend the current growing demand for emergency services.

The implementation of the model in the ED simulator and the results produced by the analysis of simulation data validate the effectiveness of the model and improve patient waiting times in the system globally and, as a result, improve the quality of healthcare. It is also important to mention that the implementation of the model inside a real ED service will be efficient to the extent that the proposed scheduler system performs on the real entry of patients, which will be dependent on the patients’ or health service users’ decisions.

REFERENCES

[1] F. Kadri, S. Chaabane, and C. Tahon, “A simulation-based decision support system to prevent and predict strain situations in emergency department systems,” *Simul. Model. Pract. Theory*, vol. 42, pp. 32–52, Mar. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1569190X13001767

[2] A. Boyle, K. Beniuk, I. Higginson, and P. Atkinson, “Emergency department crowding: Time for interventions and policy evaluations,” *Emergency Med. Int.*, vol. 2012, pp. 1–8, Feb. 2012, doi: 10.1155/2012/838610.

[3] P. C. Sprivulis, J. Da Silva, I. G. Jacobs, A. R. L. Frazer, and G. A. Jelinek, “The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments,” *Med. J. Aust.*, vol. 184, no. 12, pp. 208–212, Jun. 2006.

[4] A. M. Mancilla, “Simulation: A tool for the study of real systems,” Departamento de Ingenieria de Sistema, Universidad del Norte, Barranquilla, Colombia, Tech. Rep., 1999, vol. 6, pp. 104–112. [Online]. Available: http://rcientificas.uninorte.edu.co/index.php/ingenieria/article/view/222601443
[5] J. Pavón, M. Arroyo, S. Hassan, and C. Sansores, “Simulation of social systems with software agents,” in Proc. Actas del Campus Multidisciplinar en Percepcion e Inteligencia (CMIPI), Universidad de Castilla-La Mancha, 2006, pp. 389–400. [Online]. Available: http://samer.hassan.name/files/CMPI

[6] L. R. Izaquierdo, J. M. Galán, J. I. Santos, and R. Del Olmo, “Complex systems modelling through agent based simulation and systems dynamics,” Empiría, Revista de Metodología de Ciencias Sociales, vol. 16, pp. 85–112, 2008.

[7] M. Gul and A. F. Guneri, “A comprehensive review of emergency department simulation applications for normal and disaster conditions,” Comput. Ind. Eng., vol. 83, pp. 327–344, May 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360835215000972

[8] M. Taboada, E. Cabrera, M. L. Iglesias, F. Epelde, and E. Luque, “An agent-based decision support system for hospitals emergency departments,” Procedia Comput. Sci., vol. 4, pp. 1870–1879, May 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050911002614, doi: 10.1016/j.procs.2011.04.203.

[9] A. Arisha and G. Bobashev, Eds., A Generalized Agent-Based Model to Simulate Emergency Departments. Nice, France: IARIA, Oct. 2014. [Online]. Available: http://www.thinkmind.org/index.php/view/article&articleid=simul_2014_3_30_50128

[10] E. Bruballa, M. Taboada, A. Wong, D. Rexachs, and E. Luque, “An analytical model to evaluate the response capacity of emergency departments in extreme situations,” in Proc. 7th Int. Conf. Adv. Syst. Simulation (SIMUL), 2015, pp. 12–16.

[11] E. Bruballa, A. Wong, D. Rexachs, E. Luque, and F. Epelde, “Evaluation of response capacity to patient attention demand in an emergency department,” Int. J. Adv. Syst. Meas., vol. 10, nos. 1–2, pp. 11–22, 2017. [Online]. Available: http://www.ijarajournals.org/systems_and_measurements/sysmea_v10_n12_2017paged.pdf

[12] E. Bruballa, A. Wong, D. Rexachs, E. Luque, and F. Epelde, “Scheduling model for non-critical patients admission into a hospital emergency department,” in Proc. Winter Simulation Conf. (WSC), Dec. 2017, pp. 2917–2928.

[13] M. Sanchez, “Effects of a fast-track area on emergency department performance: Determinants of effectiveness and quality of care,” Acad. Emerg. Med., vol. 11, no. 5, p. 460, May 2004.

[14] S. W. Rodi, M. V. Grau, and C. M. Orsini, “Evaluation of a fast track unit: Alignment of resources and demand results in improved satisfaction and decreased length of stay for emergency department patients,” Qual. Manage. Health Care, vol. 15, no. 3, pp. 163–170, Jul. 2006.

[15] B. Davies, “ ‘See and treat’ or ‘see and treat’ in an emergency department,” in Proc. Winter Simulation Conf., Dec. 2007, pp. 1519–1522.

[16] S. Ruchaba, K. Knapp, L. Ashton, and M. Pitt, “Streamlining pathways for minor injuries in emergency departments through radiographer-led discharge,” Oper. Res. Health Care, vol. 19, pp. 44–56, Dec. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2211692317300619

[17] P. Yoon, I. Steiner, and G. Reinhardt, “Analysis of factors influencing length of stay in the emergency department,” Can. J. Emerg. Med., vol. 5, pp. 155–161, May 2003.

[18] N. R. Hoot and D. Aronsky, “Systematic review of emergency department crowding: Causes, effects, and solutions,” Ann. Emerg. Med., vol. 52, no. 2, pp. 126–136, Aug. 2008.

[19] L. M. Stock, G. E. Bradley, R. J. Lewis, D. W. Baker, J. Sipsey, and C. D. Stevens, “Patients who leave emergency departments without seeing a physician: Magnitude of the problem in Los Angeles County,” Ann. Emerg. Med., vol. 23, no. 2, pp. 294–298, Feb. 1994, doi: 10.1016/0196-0644(94)70043-5.

[20] T. E. Day, A. R. Al-Roubaie, and E. J. Goldlust, “Decreased length of stay after addition of healthcare provider in emergency department triage: A comparison between computer-simulated and real-world interventions,” Emerg. Med. Jr., vol. 30, no. 2, pp. 134–138, Feb. 2013. [Online]. Available: https://emj.bmj.com/content/30/2/134

[21] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, and E. Luque, “Simulation optimization for healthcare emergency departments,” Procedia Comput. Sci., vol. 9, pp. 1464–1473, Jan. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050912002827

[22] W. S. Armel, S. Samaha, and D. W. Starks, “The use of simulation to reduce the length of stay in an emergency department,” in Proc. Winter Simulation Conf., vol. 2, Dec. 2003, pp. 1907–1911.

[23] J. Wang, J. Li, K. Tussey, and K. Ross, “Reducing length of stay in emergency department: A simulation study at a community hospital,” IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 42, no. 6, pp. 1314–1322, Nov. 2012.

[24] K. W. Tan, H. Lau, and F. C. Y. Lee, “Improving patient length-of-stay in emergency department through dynamic queue management,” in Proc. Winter Simulations Conf. (WSC), Dec. 2013, pp. 2362–2373.

[25] D. J. Medeiros, E. Swenson, and C. Deflitch, “Improving patient flow in a hospital emergency department,” in Proc. Winter Simulation Conf., Dec. 2008, pp. 1526–1531.

[26] W. Soler, M. Gómez Muñoz, E. Bragulat, and A. Alvarez, “Triage: A key tool in emergency care,” Anales del Sistema Sanitario de Navarra, vol. 33, pp. 55–68, Jan. 2010.

[27] Z. Liu, D. Rexachs, F. Epelde, and E. Luque, “A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity,” Comput. Ind. Eng., vol. 103, pp. 300–309, Jan. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360835216304685

EVA BRUBALLA is currently an Associate Professor with the Computer Science School, Escoles Universitaries Gimbernat i Tomàs Cerdà, Universitat Autònoma de Barcelona (UAB), and an Associate Researcher with the Computer Architecture and Operating System Department, UAB, Spain. She has coauthored eight fully reviewed articles. Her research is on the use of simulation to improve emergency departments’ operation.

ALVARO WONG is currently an Associate Researcher with the Computer Architecture and Operating System Department, University Autonoma of Barcelona, Spain. He has worked in performance prediction of HPC applications in the ITEA 2 European Project No. 09011, research centers and industries. He has coauthored a total of 35 fully reviewed technical articles in journals and conference proceedings.

DOLORES REXACHS is currently an Associate Professor with the Computer Architecture and Operating System Department, University Autonoma of Barcelona (UAB), Spain. She has been the supervisor of 12 Ph.D. theses and has been an Invited Lecturer with Universities in Argentina, Brazil, Chile, and Paraguay. She has coauthored more than 90 fully reviewed technical articles in journals and conference proceedings. Her research interests include parallel computer architecture, parallel I/O subsystem, fault tolerance in parallel computers, tools to evaluate, predict, and improve the performance in parallel computers.

EMILIO LUQUE is currently an Emeritus Professor with the Computer Architecture and Operating System Department, University Autonoma of Barcelona, Spain. He has coauthored more than 300 technical articles in journals and conference proceedings. His major research areas are: parallel and distributed simulation, performance prediction and efficient management of multichipset-multicore systems, and fault tolerance in parallel computers.