Proceedings of the 2011 Winter Simulation Conference
S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds.

DESIGN OF CENTRALIZED AMBULANCE DIVERSION POLICIES USING SIMULATION-OPTIMIZATION

Adrian Ramirez-Nafarrate
John W. Fowler
Teresa Wu

Arizona State University
699 S. Mill Avenue
Tempe, AZ 85281, USA

ABSTRACT

Ambulance Diversion (AD) has been an issue of concern for the medical community because of the potential harmful effects of long transportation; however, AD can be used to reduce the waiting time in Emergency Departments (EDs) by redirecting patients to less crowded facilities. This paper proposes a Simulation-Optimization approach to find the appropriate parameters of diversion policies for all the facilities in a geographical area to minimize the expected time that patients spend in non-value added activities, such as transporting, waiting and boarding. In addition, two destination policies are tested in combination with the AD policies. The use of diversion and destination policies can be seen as ambulance flow control within an emergency care system. The results of this research show significant improvement in the flow of emergency patients in the system as a result of the optimization of destination-diversion policies compared to not using AD at all.

1 INTRODUCTION

Ambulance Diversion (AD) are the periods when overcrowded Emergency Departments (EDs) request ambulance services to bypass their facilities (GAO 2003). The Center for Disease Control and Prevention reported that the mean annual hours on AD in metropolitan areas was 404 hours from 2003 to 2004 (CDC 2006).

The medical community strongly recommends to avoid or minimize the periods on AD because of the potential harmful effect of longer transportation on the health status of the patient (ACEP 1999a; ACEP 2008). However, there is evidence suggesting that not diverting ambulances may increase the waiting time and the number of patients boarding in EDs (Massachusetts Nurse Newsletter 2009).

On the other hand, the effect of ambulance patients on the operations of EDs determines the effectiveness of the system in many performance measures. Thus, ambulance patients tend to receive higher priority to start treatment than walk-ins because of their severity conditions. In addition, ambulance patients also have longer treatment times and larger admission probability than walk-ins. Therefore, the allocation of the patient affects the flow of other patients.

This paper proposes the centralized design of AD policies combined with effective destination policies as a mechanism of ambulance flow control. The destination policy determines the hospital destination of a patient when there are more than one open hospital in the region. The objective is to minimize the time that patients who require emergency assistance spend in suboptimal treatment (a.k.a non-value added time). The results show the potential of designing effective destination and diversion policies to smooth the flow through different stages of care.

2 LITERATURE REVIEW

Empirical studies regarding ambulance diversion exists in the literature. These type of papers identify the main causes to divert ambulances and show efforts to design system-wide policies that enable the reduction of the diversion episodes in a region.

A survey conducted by the CDC (2006) revealed that the main causes for initiating AD episodes are the lack of beds in inpatient units, the high number of patients waiting in the ED and the complexity of ED cases. These factors are also identified in other publications. Furthermore, an additional cause that is consistently named in other publications is the high number of patients boarding, which are the patients in the ED waiting for an open bed in an inpatient unit (ACEP 2008; Pham et al. 2006).

Empirical studies to design AD guidelines are available in the literature. The objective of these guidelines is to minimize the amount of time spent on AD in a region. For example, Vilke et al. (2004a) designed a plan to observe the effect of AD in two hospitals. The authors found a reciprocating effect. Thus, when one hospital goes on diversion, it is very likely that the neighboring hospital starts diverting ambulances within a short period. An enhanced project that comprised multiple hospitals had the objective of redesigning the guidelines to start AD. These new guidelines are more restrictive and the results show a significant reduction of AD hours in the region (Vilke et al. 2004b).

Two similar studies were conducted in other regions. The first of them introduces a new AD protocol for a county with 600,000 inhabitants and 10 hospitals (Asamoah et al. 2008). The new protocol constrained the time on AD to only one hour out of every eight. The mean number of hours on diversion in the system was reduced by about 82%. The second study is a project carried out during three years involving 17 hospitals (Patel at al. 2006). Similarly, this study is based on the redesign of AD guidelines that restrict the causes to initiate AD and limits the duration of the episodes. The results show a reduction of the hours spent on AD in the system by about 75%.

Even though these results show a significant reduction of AD in a system, which increases the accessibility to emergency care, these publications do not provide information about the impact of reducing AD in other measures, such as waiting time. It is known that AD can relieve congestion from an ED and reduce the average waiting time within a facility (Ramirez, Fowler, and Wu 2010). Therefore, restricting the use of AD might have undesirable effect if appropriate actions are not implemented to reduce congestion. Furthermore, “No AD” laws approved in some regions might cause a rise in the waiting time and patients boarding (Massachusetts Nurse Newsletter 2009).

On the other hand, few analytical studies on the use of AD in multiple hospitals are available in the literature. For example, Hagtvedt et al. (2009) models the AD decision of two hospitals using an analogy of the Prisoner’s Dilemma and introduces a payoff function that includes a penalty for diverting ambulances. The authors found that a centralized system to control AD episodes is needed given that voluntary cooperation might not be a robust approach.

Deo and Gurvich (2011) analyze the effect of AD in two hospitals using a model based on queuing networks. The authors use the average waiting time in each ED as the performance measure of their interest. They observed that a centralized AD can be Pareto improving compared with not allowing AD in the system. In addition, they introduce the rule that initiates AD when all the beds in the ED are occupied. These results suggest that AD can bring benefits to the system if the guidelines are properly designed based on a centralized system. However, their models do not include important aspects observed in emergency systems, such as non-stationary arrivals, transference from EDs to inpatient units and distance to other hospitals.

This paper proposes a model based on simulation to analyze the effect of AD and destination policies in the flow of patients in an emergency care system. Furthermore, a simulation-optimization approach using genetic algorithms is introduced as a method to design effective AD policies from a centralized perspective. The objective is to find Pareto improving policies that reduce the average-patient non-value added time in each hospital of a system.

3 EMERGENCY CARE DELIVERY SYSTEM MODEL

This research is based on a discrete-event simulation model of an emergency care delivery system (ECDS) that includes multiple hospitals and ambulances delivering patients to their EDs. There are three main modules that allows the execution of the simulation and the evaluation of the decision policies: emergency patient generation, ambulance destination decision and hospital simulation.

The emergency patient generation module creates patients that require transportation to an ED. This module assigns a random location that represents the departure point from the ambulance

Then, the ambulance destination decision evaluates the options related to the open hospitals that can receive the new patient. The decision of where to take the new patient is based on a destination policy. After determining the destination hospital, the module schedules the arrival of the patient taking into account distance and a random variable associated with velocity.

The hospital simulation module executes the events related to the operations of each hospital, including ambulance and walk-in arrivals, end of treatments, direct admissions, etc. In addition, it keeps track of the status of the hospital to initiate AD if the conditions of the policy under evaluation are satisfied. Figure 1 presents an overview of the model.

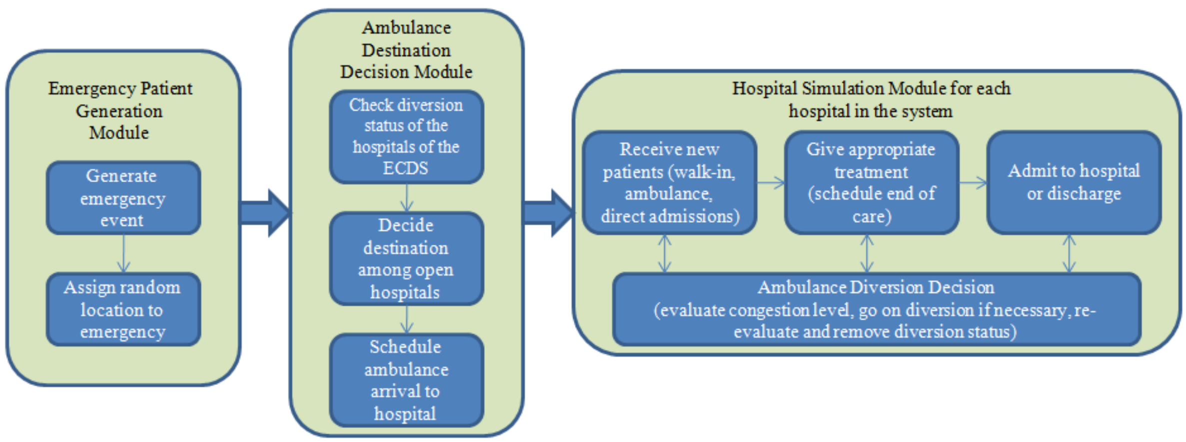

Figure 1: Overview of the simulation model of an ECDS

Each hospital in the model has a similar structure. They include an ED and one inpatient unit (IP). Upon arrival to the ED, a severity level is assigned to the patient. This level determines the priority for being placed in a bed. The probability of receiving a specific severity level depends on the arrival mode. Patients arriving by ambulance are more likely to have higher severity level than walk-ins. The duration of the treatment in the ED depends also on the severity level. Higher severity implies longer treatment. After ending treatment in the ED, the patient can be discharged or be admitted to the IP unit. The admission probability is also correlated with the severity level. The IP unit receives direct admissions as well. If a patient in the ED requires admission, but the IP does not have available beds, then the patient waits in the bed of the ED (boarding) until an IP bed opens up. In addition, patients in the ED can leave without treatment (LWOT) if they have waited for a long period. Figure 2 depicts the patient flow in each hospital included in the model. Appendix A describes the input data for the generic hospital used in the model.

Two assumptions are made regarding the accessibility to emergency care in the system. First, if the arrival of an ambulance with a patient is scheduled to the destination hospital and that hospital goes on diversion before the arrival event takes place, then the patient is still accepted in the hospital. This avoids redirecting patients while they are on the road. The second assumption avoids having all the hospitals in the system on diversion at the same time. Thus, if the last open hospital reaches the condition to go on diversion, then all the hospitals go off diversion. This type of practice is commonly used in real settings (AEMS 2000).

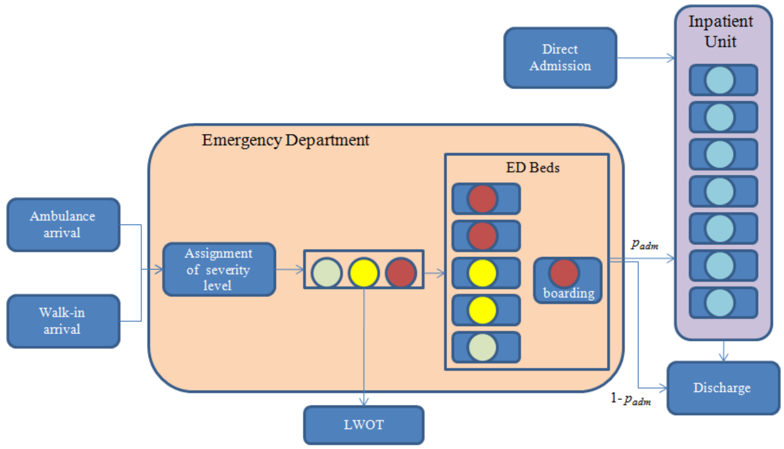

Figure 2: Patient flow within the hospitals included in the model

4 CENTRALIZED DESIGN OF AD POLICIES

This paper introduces a simulation-optimization approach to allow a centralized design of AD policies that enables the improvement of the hospitals' performance in the system. Particularly, this research is based on the use of genetic algorithms (GA) to define the parameters of the AD policies for each hospital in a system and it uses the simulation model described in Section 3 to evaluate the performance of the set of policies.

Thus, the GA chromosome represents the union of AD policies for all the hospitals in the system. Therefore, the recombination and mutation of chromosomes enables the exploration of sets of policies that can improve the performance in the ECDS. Then, simulation evaluates the fitness of each chromosome by obtaining the average-patient non-value added time per hospital. Finally, the evolution of the GA allows finding the set of policies that have the best performance after a finite number of generations. Figure 3 depicts the process of the centralized design of AD policies proposed.

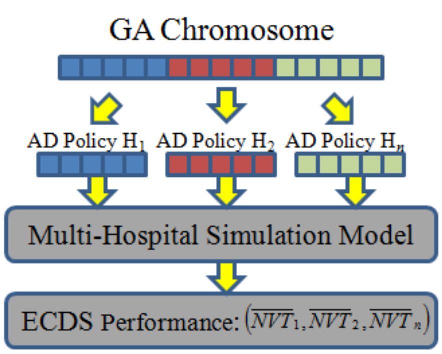

Figure 3: Centralized design of AD policies using simulation-optimization

4.1 Performance Evaluation

The hypothesis of this research is based on the assumption that a smart allocation of ambulance patients, through ambulance diversion and destination policies, can reduce the time that patients spend in suboptimal treatment at different stages of care. Therefore, the performance of the ECDS is evaluated by the vector that contains the average-patient non-value added time for each hospital in the system: $(\overline{NVT}_1, \overline{NVT}_2, \dots, \overline{NVT}_n)$, where n is the number of hospitals in the system.

The average-patient non-value added time of a hospital is a measure that includes transportation, waiting in the ED and boarding; and it is computed in the following form:

$$\overline{NVT}_i = p_i^a \bar{T}_i + \sum_{k=1}^{5} w_{k,i} \bar{W}_{k,i} + p_i^{adm} \bar{B}_i \tag{1}$$

where,
$\overline{NVT}_i$= Average patient non-value added time in hospital H_i
p_i^a= Fraction of ambulance arrivals to hospital H_i
$\bar{T}_i$ = Average transportation time of ambulance patients received at hospital H_i. This includes patients whose final destination is H_i, and patients diverted from H_j to H_i, for all $i \neq j$
$w_{k,i}$ = Weight given to the average waiting time of patients with severity level k in hospital H_i
$\bar{W}_{k,i}$ = Average waiting time of patients with severity level k in hospital H_i
p_i^{adm} = Fraction of ED patients admitted to hospital H_i
$\bar{B}_i$ = Average boarding time in hospital H_i

The first term on the right hand side of the equation takes into account the average transportation time of ambulance patients only. The second term is a weighted average, based the on severity level, of all the patients in the ED, except the patients that left without treatment (LWOT). The third term includes all the patients that were admitted from the ED. Note that the components related to waiting and boarding considers walk-in patients.

Although the average proportion of ambulance arrivals to EDs is 15% (CDC 2010), the characteristics of their patients can cause important disruptions to the flow of patients in a hospital. Therefore, an effective design and combination of AD and destination policies could smooth the patient flow, which will be observed as an improvement on the performance vector.

4.2 Ambulance Diversion and Destination Policies

This paper compares three types of AD policies, each of them combined with two types of destination policies.

Ambulance Diversion Policies:

1. No AD. Ambulance diversion is not implemented if this type of policy is used.
2. Simple AD. The simple AD policy initiates a period of AD when all the beds in the ED are occupied.
3. Optimized Single-Factor AD Policy (SF AD). This type of policy observes if a particular state variable reaches a threshold in order to initiate an AD episode. The optimal thresholds are obtained via GA.

Ambulance Destination Policies:

1. Nearest Hospital (NH). With this policy, the patient is transported to the nearest hospital from the emergency location.
2. Least Crowded Hospital (LCH). The patient is transported to the hospital with the fewest number of patients waiting in the ED.

The first type of AD policy reflects the situation in some regions and the recommendations of the medical community. The simple AD policy is suggested by Deo and Gurvich (2011) as Pareto improving policies in their queuing analysis. The optimized single-factor AD policy is the core of this research. It is based on a proposed structure for the policy that takes into account one of the main crowding variables and it includes parameters to reevaluate and remove the diversion status. The next section provides a deeper explanation of this type of policy.

The guidelines of EMS suggest to take a patient to the nearest appropriate hospital (ACEP 1999b). Therefore, this research evaluates two types of destination policies related to this recommendation. The nearest hospital policy could improve the first term of Equation 1, while the second policy could have an effect of the second and third term.

4.3 Encoding of Single-Factor AD Policies

The AD policies proposed in this paper are based on the observation of one of the main variables that triggers the diversion status in practice. These state variables are:

NQ_i: Number of patients waiting in the ED of hospital H_i.
NB_i: Number of patients boarding in the ED of hospital H_i.
$NIPB_i$: Number of beds available in the IP unit of hospital H_i.

Ramirez, Fowler, and Wu (2010) presented a bi-criteria analysis of single-factor policies based on these variables for a single hospital. Those policies comprises two parameters to set or remove/reevaluate the diversion status. Hence, the reevaluation could be continuous or at discrete points.

This paper improves the definition of an AD policy by considering three parameters: (Don, $Doff$, Δt). The Don parameter represents a threshold on a state variable to set the diversion status on. The $Doff$ parameter is another threshold on the same state variable to remove the diversion status. Δt is the reevaluation frequency after going on diversion. Diversion status can be removed only at a reevaluation point.

Since three state variables are considered in this research, then there are three types of single-factor AD policies, one for each state variable. Therefore, the length of the AD chromosome that contains the AD policies of all the hospitals in the ECDS is $10n$, where n is the number of hospitals. The subchromosome that represents the AD policy of a particular hospital is depicted in Figure 4.

Gene	1	2	3	4	5	6	7	8	9	10
Variable	P_i	U_{NQi}	L_{NQi}	Δt_{NQi}	U_{NBi}	L_{NBi}	Δt_{NBi}	L_{NIPBi}	U_{NIPBi}	Δt_{NIPBi}

Figure 4: Chromosome partition that represents an SF AD policy in one hospital

The first gene describes the type of factor to consider in the policy of hospital H_i. Thus, $P_i = 1$ implies that AD policy of hospital H_i is based only on number of patients waiting in the ED (NQ_i); $P_i = 2$ indicates that AD is based on the number of patients boarding (NB_i); and $P_i = 3$ means that AD is based on the number of beds available in the IP unit ($NIPB_i$). Therefore, the execution of an SF AD policy requires values for three parameters. If the policy is type 1, then the parameters are in the genes 2, 3 and 4. If it is type 2, then the genes of interest are 5, 6 and 7. If the type is 3, then the related genes are 8, 9 and 10.

The first of the three parameters that define an SF AD policy is a threshold that triggers the diversion status (Don parameter in genes 2, 5 or 8). Thus, if policy is type 1, then the hospital H_i sets the diversion status when $NQ_i \geq U_{NQi}$. If it is type 2, then H_i goes on diversion when $NB_i \geq U_{NBi}$. If it is type 3, then diversion is set when $NIPB_i \leq L_{NIPBi}$. After going on diversion, the state of the system is reviewed every Δt time units, represented by genes 4, 7 and 10 for policies type 1, 2 and 3, respectively.

The $Doff$ parameter (represented in genes 3, 6 and 9) is a threshold used to decide the removal of the diversion status at a review point. If policy is 1, then the diversion status is removed if $NQ_i \leq L_{NQi}$. If policy is type 2, then diversion is removed if $NB_i \leq L_{NBi}$. If the policy is type 3, then diversion is removed if $NIPB_i \geq U_{NIPBi}$. Note that for all the policies the threshold U is greater than the threshold L. Policy type 3 has the U and L interchanged because of the meaning of the state variable (number of available beds in the IP).

An example of an SF AD policy is: "set the diversion status if there are at least 15 patients waiting in the ED, reevaluate every hour after going on diversion and remove the diversion status if there are 5 or less patients waiting". This policy is encoded as:

Gene	1	2	3	4	5	6	7	8	9	10
Variable	1	15	5	60	Null	Null	Null	Null	Null	Null

In this example, genes 5 to 10 can take any value and the simulation code does not take them into account because the first gene specifies the type of policy.

4.4 Multi-Objective Genetic Algorithm

This research implements the NSGA-II algorithm (Deb et al. 2002). This algorithm selects the survivor chromosomes based on a front number and a crowding distance. The front number of a specific policy *P* is related to the number of policies which dominate *P* (domination count). The nondominated policies of a set of policies have front number equal to one. Then, policies in front one are removed from the total set and the process repeats. The new set of nondominated policies is assigned to front two. The process repeats until a front number is assigned to all the policies. The crowding distance is related to the diversity of the policies. The crowding distance of a specific policy *P* is an estimation of the perimeter of the cuboid formed by the nearest policy neighbors of *P*. The policies with larger crowding distance are more likely to be included in the parent selection since diversity encourages exploring areas with low density of policies.

5 CASE STUDY: ECDS WITH THREE HOSPITALS

The AD and destination policies presented in Section 4.2 are used in a case study that comprises three hospitals. Two configurations of random locations are presented; one of them assumes that the ECDS is in a 10x10 squared-miles area (Random 1), while another assumes that the area is 20x20 squared-miles (Random 2).

The hospital built for this research is a generic hospital that incorporates data from published papers and other sources. Based on the generic hospital, two configurations of relative size are used in the experimentation: one assumes the same relative size (1:1:1) and another assumes different sizes for all the hospitals, one of them has 10% more arrivals than the generic and another 20% more arrivals than the generic (1:1.1:1.2). The combinations of scenarios and strategies are summarized in Table 1.

Table 1: Scenarios and strategies used in the experimentation process.

Scenarios		Strategies	
Location (H_1, H_2, H_3)	Relative Size (H_1: H_2: H_3)	Diversion Policies	Destination Policies
Random 1: (1.7, 9.2), (4.8, 3.8) & (8.5, 7.3)	1:1:1	No AD	NH
Random 2: (19.2, 6.4), (6, 10.5) & (12.3, 18.9)	1:1.1:1.2	Simple AD	LCH
		Optimized SF-AD	

The results for the case study are shown in Table 2. It includes the average-patient non-value added time per hospital for each strategy, the sum of the non-value added time in the system, the standard deviation and the percentage of time spent on diversion in each hospital. More than one solution can be seen for the SF AD strategies because the multi-objective GA can produce multiple Pareto solutions.

For each scenario, the strategies that allow AD outperform the No AD strategy. Thus, AD can reduce the total average-patient non-value added time. Furthermore, the SF AD policy proposed in this paper produce Pareto improving solutions in most of the scenarios. The Simple AD policy have better performance than No AD, but the SF AD is better than the Simple policy.

Table 2: Results from the experimentation process

Scenario	Strategy	Average-Patient NVT (mins)			Total NVT	Std. Dev NVT	Percentage of time on diversion		
		H1	H2	H3			H1	H2	H3
Random 1 1:1:1	No AD - NH	21.28	155.63	45.29	222.21	71.65			
	Simple AD - NH	25.55	86.00	44.03	155.59	30.98	6.33	22.43	11.36
	SF AD - NH	67.26	52.54	27.58	147.38	20.06	8.77	33.38	21.55
	SF AD - NH*	20.47	101.09	36.96	158.52	42.59	10.45	23.43	16.64
	No AD - LCH	54.86	41.47	36.74	133.08	9.40			
	Simple AD - LCH	37.12	43.32	40.73	121.17	3.12	6.24	10.01	7.67
	SF AD - LCH	36.48	42.78	37.67	116.93	3.35	7.57	0.82	5.41
	SF AD - LCH*	53.56	36.56	29.76	119.88	12.26	9.50	32.10	26.37
Random 1 1:1.1:1.2	No AD - NH	25.24	273.94	160.93	460.11	124.52			
	Simple AD - NH	46.87	197.28	144.58	388.73	76.32	2.24	22.63	19.05
	SF AD - NH	71.34	155.17	148.48	374.99	46.59	12.92	31.63	25.43
	No AD - LCH	132.93	123.55	119.53	376.01	6.88			
	Simple AD - LCH	125.77	126.18	123.65	375.61	1.36	3.35	5.61	5.95
	SF AD - LCH***	117.39	106.46	111.13	334.98	5.48	2.89	18.71	13.25
Random 2 1:1:1	No AD - NH	34.12	144.28	26.26	204.66	65.99			
	Simple AD - NH	38.97	81.91	32.92	153.79	26.71	10.45	22.31	8.24
	SF AD - NH	43.66	75.64	33.92	153.21	21.82	10.55	24.57	8.90
	SF AD - NH*	26.37	104.73	25.56	156.66	45.48	13.89	22.68	11.21
	No AD - LCH	58.40	43.88	37.62	139.90	10.66			
	Simple AD - LCH	42.06	46.15	41.27	129.48	2.62	7.56	10.42	6.84
	SF AD - LCH**	34.34	36.89	40.07	111.30	2.87	9.71	15.58	3.40
Random 2 1:1.1:1.2	No AD - NH	50.57	317.08	131.81	499.47	136.60			
	Simple AD - NH	83.50	238.65	137.97	460.12	78.71	3.46	18.11	12.57
	SF AD - NH**	83.18	226.75	123.35	433.28	74.07	3.92	18.14	12.31
	SF AD - NH	155.39	189.56	103.21	448.17	43.49	5.59	23.71	17.65
	No AD - LCH	168.27	156.33	150.16	474.76	9.21			
	Simple AD-LCH*	159.98	153.01	149.22	462.20	5.46	3.41	4.52	4.65
	SF AD - LCH	76.71	61.11	290.24	428.05	128.02	44.70	55.61	3.30
	SF AD - LCH*	151.06	131.86	149.27	432.19	10.61	10.67	20.80	15.48
	SF AD - LCH***	153.12	144.20	141.33	438.66	6.15	3.60	9.91	5.70

*Dominates No AD strategy; **Dominates Simple AD strategy; ***Dominates No AD and Simple AD strategies

Regarding the destination policies, LCH hospitals outperform NH. In addition, LCH can balance the performance across hospitals, reducing the standard deviation of the average patient non-value added time.

The reduction in the average-patient non-value added time using AD policies depends on the destination policy implemented. Thus, if NH is used, then the Simple AD policy reduces the total average-patient non value-added time by 19.56% and the SF AD reduces it by 22.64%. If LCH is used, then the reduction produced by Simple AD and SF AD are 4.79% and 13.33% respectively. In any case, the gaining of using intelligent diversion is significant compared to No AD.

An important aspect to highlight related to the destination policy is that the case study assumes an urban area for the hospitals. If the analysis is conducted in a rural area, it is very likely that contradictory results would be observed. However, AD is not recommended in rural areas because the significant increase in transportation would cancel the potential benefits of AD and it would jeopardize the health condition of the patients.

6 CONCLUSIONS

This paper presented a centralized design of AD policies using GA and simulation to evaluate the performance. The AD policies are combined with destination policies in an ambulance flow control framework that allows the allocation of ambulance patient in an ECDS. The findings suggest that the centralized design of diversion policies and effective destination rules can reduce the time that patients spend in suboptimal care, including the walk-in patients.

The results of the experimentation show that the proposed SF AD policies outperform not diverting at all and they produce better results than a simple AD policy. These observations imply that an intelligent strategy for implementing AD in a region can smooth the patient flow in the entire ECDS. The improvement depends on the destination policy that accompanies the AD policy. The LCH destination policy outperforms NH. However, this might hold only in urban settings, like the one assumed in this research.

The centralized design of AD policies assumes a high level of collaboration among hospitals to improve the patient safety. This might be an issue in the real setting, but the efforts shown in the literature suggest that healthcare organizations are willing to work together to bring benefits to their systems.

Even though the model used for this research considers a generic hospital built from data obtained in the literature, the effectiveness of the proposed methodology does not depend on the model but in the encoding of the AD policy and the evaluation of the system.

Future extensions of this research include the optimization of multiple-factor AD policies, the evaluation of a destination policy that includes transportation and crowding factors simultaneously and the optimization of the destination policy using simulation-optimization.

A INPUT DATA FOR GENERIC HOSPITAL

The generic hospital described in Section 3 was built using C++ with information published in different sources. The main sources of are: Cochran and Bharti (2006), Cochran and Roche (2009) and CDC (2010). Figure 5 depicts the pattern of the arrivals considered for the model. The pattern observed in the arrival rate is also identified in other papers and official reports across the United States (Burt, McCaig, and Valverde 2006; CDC 2008). Note that ambulance arrivals comprise 15% of all the arrivals to an ED according to Figure 7. This percentage is consistent with the averages published by the CDC (2010). Green (2006) proposes a set of arguments to assume Poisson process for the arrivals to healthcare systems. Hence, this paper assumes Poisson process for all its arrivals. In order to schedule ambulance arrivals to an appropriate hospital, the transportation time is estimated by $\tau \times M(l, H_i)$, where $M(l, H_i)$ is the Manhattan distance between the emergency location and the selected hospital, and τ is the transportation time per mile. This paper assumes a distribution for τ, such that $\tau \sim$ Normal(1.25, 0.5). This implies that the average transportation time is 1.25 minutes per mile, which is similar to the data presented by Google Maps as transportation time per mile in Maricopa County, AZ. The severity level assigned to each patient depends on the arrival mode. Table 3 presents the percentages of each severity level.

The mean treatment time per severity level is shown in Table 4. This paper assumes that the treatment time follows an Erlang distribution with shape parameter of 3.

After ending treatment in the ED, the patients can be admitted to the IP unit with a probability that depends on the severity level. These probabilities are presented in Table 5. The overall admission percentage is 15% which is in the range the average seen in metropolitan areas in the United States (CDC 2010).

Direct admissions to IP occur according to a Poisson process with a mean of one admission per hour, which is similar to the total external arrival rates of the hospital analyzed in Cochran and Bharti (2006). The treatment time in the IP is also assumed to be an Erlang distribution with shape parameter equal to 3 and a mean of four days, which is similar to the data found by Cochran and Bharti (2006) and close to mean length of stay in IP units according to the CDC (2010).

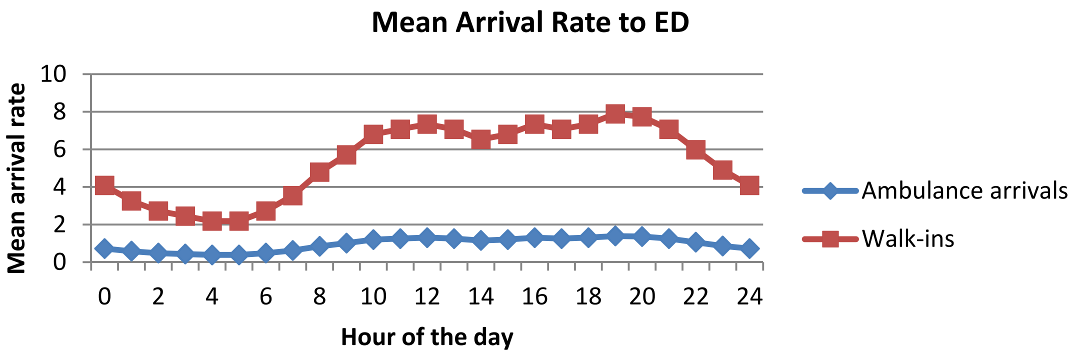

Figure 5: Arrival rate to a single ED

Table 3: Percentages for severity mix

	Arrival Mode		
Severity Level	Ambulance	Walk-Ins	Overall
1	15	2	3.95
2	42	16	19.90
3	30	40	38.50
4	10	30	27.00
5	3	12	10.65
Overall	15	85	

Table 4: Mean Treatment Times in the ED

Severity Level	Mean Treatment Time (min)
1	273
2	273
3	140
4	106
5	30

Table 5. Admission probabilities to IP

Severity Level	Admission Percentage
1	70
2	34
3	10
4	5
5	3
Overall	15

In order to model the LWOT patients, this paper incorporates an approach presented by Miller, Ferrin, and Shahi (2009). The LWOT routine consists of removing patients from the queue if they have not been placed in a bed within 24 hours. This paper assumes that LWOT patients go home or visit a non-emergency physician; therefore, they are not scheduled to arrive to another hospital in the model.

The hospitals in the model have 20 beds in the ED and 200 IP beds. The number of beds considered for the ED is similar to the median in the United States (CDC 2006) and the size of the IP unit is suitable for a medium-size hospital.

The simulation length for the research is fixed to six months after a warm-up period of one month and ten replications per strategy are considered. These parameters were defined after a set of pilot runs to obtain precise estimation of the performance measure of interest. In addition, Common Random Numbers (Banks et al. 2010) are used to expose the different strategies to similar conditions and reduce the noise among them.

REFERENCES

ACEP (American College of Emergency Physicians). 1999a. "Guidelines for Ambulance Diversion." Accessed May 7, 2011. http://www.acep.org/content.aspx?id=30038.

ACEP (American College of Emergency Physicians). 1999b. "Emergency Ambulance Destination." Accessed May 7, 2011. http://www.acep.org/content.aspx?id=29196.

ACEP (American College of Emergency Physicians). 2008. Emergency Department Crowding: High-Impact Solutions. Accessed May 7, 2011. www.acep.org/WorkArea/DownloadAsset.aspx?id=50026.

AEMS (Arizona Emergency Medical Systems, Inc.). 2000. "Transfer of Care. Guideline Statements on Prehospital Diversion." Chapter 8 in *Red Book*. Accessed May 7, 2011. http://www.aems.org/aems/redbook/chap8.pdf.

Asamoah, O. K., S. J. Weiss, A. A. Ernst, M. Richards, and D. P. Sklar. 2008. "A Novel Diversion Protocol Dramatically Reduces Diversion Hours." *The American Journal of Emergency Medicine* 26:670-675.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2010. *Discrete-Event System Simulation*. 5th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Burt, C. W., L. F. McCaig, and R. H. Valverde. 2006. "Analysis of Ambulance Transports And Diversions Among US Emergency Departments." *Annals of Emergency Medicine* 47(4):317-326.

CDC (Centers for Disease Control and Prevention). 2006. "Staffing, Capacity, and Ambulance Diversion in Emergency Departments: United States, 2003-04." In *Advance Data from Vital and Health Statistics* 376.

CDC (Centers for Disease Control and Prevention). 2008. "National Hospital Ambulatory Medical Care Survey: 2006 Emergency Departments Summary." *National Health Statistics Reports* 7.

CDC (Centers for Disease Control and Prevention). 2010. "National Hospital Ambulatory Medical Care Survey: 2007 Emergency Department Summary." In *National Health Statistics Report* 26.

Cochran, J. K., and A. Bharti. 2006. "A Multi-Stage Stochastic Methodology for Whole Hospital Bed Planning under Peak Loading." *International Journal of Industrial and Systems Engineering* 1:8-36.

Cochran, J. K., and K. T. Roche. 2009. "A Multi-Class Queuing Network Analysis Methodology for Improving Hospital Emergency Department Performance." *Computers & Operations Research* 36:1497-1512.

Deb, K., A. Pratap, S. Agrawal, and T. Meyarivan. 2002. "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II." *IEEE Transactions on Evolutionary Computation* 6(2):182-197.

Deo, S., and I. Gurvich. 2011. "Centralized vs. Decentralized Ambulance Diversion: A Network Perspective." *Management Science* 57(7):1300-1319.

GAO. 2003. *Hospital Emergency Departments: Crowded Conditions Vary among Hospitals and Communities.* United States General Accounting Office. Report to the Ranking Minority, Committee on Finance, U.S. Senate.

Green, L. V. 2006. "Queuing Analysis in Healthcare." In *Patient Flow: Reducing Delay In Healthcare Delivery*, edited by R.W. Hall, 281-307. Springer's International Series.

Hagtvedt, R., P. Griffin, P. Keskinocak, M. Ferguson, and F. T. Jones. 2009. "Cooperative Strategies to Reduce Ambulance Diversion." In *Proceedings of the 2009 Winter Simulation Conference*, edited by

M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, 1861-1874. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Massachusetts Nurse Newsletter. 2009. "State's 'No Diversion Policy' Is Putting Strain on Massachusetts Hospitals." April 2009:8-9.

Miller, M., D. Ferrin, and N. Shahi. 2009. "Estimating Patient Surge Impact in Several Regional Emergency Departments." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 1906-1915. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Patel, P. B., R. W. Derlet, D. R. Vinson, M. Williams, and J. Wills. 2006. "Ambulance Diversion Reduction: The Sacramento Solution." *American Journal of Emergency Medicine* 24:206-213.

Pham, J. C., R. Patel, M. G. Millin, T. D. Kirsch, and A. Chanmugam. 2006. "The Effects of Ambulance Diversion: A Comprehensive Review." *Journal of the Academic Emergency Medicine* 13(11):1220-1227.

Ramirez, A., J. W. Fowler, and T. Wu. 2010. "Bi-Criteria Analysis of Ambulance Diversion Policies." In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan, 2315-2326. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Vilke, G. M., L. Brown, P. Skogland, C. Simmons, and D. A. Guss. 2004a. "Approach to Decreasing Emergency Department Ambulance Diversion Hours." *Journal of Emergency Medicine* 26(2):189-192.

Vilke, G. M., E. M. Castillo, M. A. Metz, L. U. Ray, P. A. Murrin, R. Lev, and T. C. Chan. 2004b. "Community Trial to Decrease Ambulance Diversion Hours: The San Diego County Patient Destination Trial." *Annals of Emergency Medicine* 44(4):295-303.

AUTHOR BIOGRAPHIES

ADRIAN RAMIREZ-NAFARRATE is a PhD candidate in Industrial Engineering in the School of Computing, Informatics and Decision Systems Engineering at Arizona State University. His research interests include modeling, simulation and analysis of healthcare delivery systems. He received a MS in Manufacturing Systems at ITESM and a BS in Industrial Engineering at Universidad de Sonora, both in Mexico. His email address is adrian.ramirez@asu.edu.

JOHN W. FOWLER is a Professor in the Operations Research and Production Systems group of the School of Computing, Informatics and Decision Systems Engineering at Arizona State University. His research interests include modeling, analysis, and control of manufacturing and service systems. He is a Fellow of the Institute of Industrial Engineers and is the SCS representative on the Board of Directors of the Winter Simulation Conference. He is an Area Editor of the T*ransactions of the Society for Computer Simulation International*, an Associate Editor of *IEEE Transactions on Semiconductor Manufacturing*, and Editor of *IIE Transactions on Healthcare Systems Engineering*. His email address is john.fowler@asu.edu.

TERESA WU is an Associate Professor of the School of Computing, Informatics and Decision Systems Engineering at Arizona State University. She received her Ph.D. in Industrial Engineering from the University of Iowa in 2001. Her current research interests include: distributed decision support, distributed information system, healthcare informatics and data fusion.. Professor Wu has over 40 articles published (or accepted) in such journals as *International Journal of Production Research, Information Science, IEEE Transactions on Engineering Management, ASME: Journal of Computing and Information Science in Engineering*. She serves on the Editorial Review Board for *International Journal of Production Research, IEEE Transactions on Engineering Management, Computer and Standard Interface, International Journal of Electronic Business Management*. Her email address is teresa.wu@asu.edu.