1. Problem Definition

Ambulance location models determine number and location of ambulances to maximize coverage of emergency calls with minimum number of ambulances. Four decades ago the first ambulance location model, location set covering model (LSCM), was presented by Toregas et al. in 1971 (1). Since then, presented models have been more sophisticated. In these forty years the models have progressed from static to dynamic, deterministic to probabilistic, and single objective to multi-objective structure. Main logic and structure of dynamic models is mainly based on the static models. There are two main differences between static and dynamic models: first, dynamic models relocate ambulances in real time. Every time a new emergency call is registered, this model is run to relocate ambulances based on the updated information such as travel time between traffic zones. Second, dynamic models try to minimize ambulance movement in cities.

Stochastic or probabilistic models consider this fact that ambulances are not always available. These changes in structure of the models show that they tried in addition to meet standard measures determined by Emergency Medical Services (EMS) Act, deal with problems occurring in real world such as limited capacity of ambulances and their unavailability when sent for a mission.

The main goal of ambulance location models is to maximize demands coverage within standard time with minimum possible number of ambulances. To the authors’ knowledge, available studies on ambulance location models have not compared logics behind previous models to understand which model is more comprehensive. In this study we recognize and introduce characteristics expected from an ideal ambulance location model. These measures, listed below, are based on EMS’ standards, operation, maintenance, and management aspects. The first six measures are based on our findings of previous models (2). We also added two more measures.

A- Coverage of high percentage of demands (95%) in the standard period
B- Considering ambulances capacity
C- Coverage of each zone by at least one ambulance
D- Considering unavailability of ambulances
E- Providing backup ambulances
F- Determination of which zone is covered by which station
G- Minimizing maintenance and operating cost
H- Dividing city to independent EMS’ regions

Measure A is based on the United States Emergency Medical Services Act of 1973 that determines 95% of requests in urban areas in 10 minutes and in rural areas in 30 minutes must be covered (2). All the models have covered this measure.

Measure B considers this fact that ambulances have a limited capacity and one ambulance cannot cover all the demands in specific time period.
Measure C makes sure that all the possible demand zones are covered not only zones with high demands.

Measure D talks about unavailability of ambulances. It helps to consider unpredictable condition in which ambulances are not available such as being out of order.

Measure E needs some more clarification. Some models (such as DSM and BACOPI) in addition to consider one ambulance for each zone have provided back up ambulances (usually with different acceptable coverage time) for zones with higher demands. But the point is if a model considers unavailability and especially capacity for ambulances, the high demanded zones will be automatically covered. So a model covering measures D and B, will also cover measure E.

Measure F is needed to check if an ambulance is overloaded. If it is not known that each ambulance is covering what demand points, it would not be clear that how many calls the ambulance is expected to cover. Also it is needed to check if emergency calls are distributed uniformly between ambulances. In addition, EMS managers need to know what ambulance have to be dispatched to cover emergency call from specific zone.

Measure G deal with maintenance and operating costs. These costs are directly related to the number of ambulances and total time or distance travelled by ambulances. Most of the previous models deal with the former, but none of them take care about the latter.

Measure H deals with operation management of ambulances. In previous models, zones covered by each ambulance either are not determined or if it is determined, there is overlap in covering specific zones specially ones with high demands. So in both cases, an EMS manager is not sure what ambulance should be dispatched.

In Table 1 previous models are adapted to the eight measures. It can be seen that in the best case five out of eight measures is covered by Araz et al.’s model. Again it is emphasized that only static models are focused in this study, because they constitute main core of logics behind dynamic models.

Table 1. The characteristics of current EMS location models

<table>
<thead>
<tr>
<th>Model</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSCM [1]</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>MCLP[3]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>TEAM[4]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>BACOPI[5]</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>BACOPII[5]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>DSM[6]</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>CMCLP[7]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Araz et al.[8]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>MEXCLP[9]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>MALT I[10]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>MALT II[10]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>REL-P[11]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>TTM[12]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>
2. Proposed Model

Regarding all measures in Table 1, the following model which is a multi-objective, probabilistic, and capacity constrained model is proposed. This model is inspired from MEXCLP and Araz et al.’s models (8 and 9).

The maximum expected covering location problem formulation (MEXCLP) proposed by Daskin (1983) is one of the first probabilistic models for ambulance location. In this model, it is assumed that each ambulance has the same probability \( q \), called the busy fraction, of being unavailable to answer a call, and all ambulances are independent. The busy fraction can be estimated by dividing the total estimated duration of calls for all demand points by the total number of available ambulances.

Thus, if zone \( i \) is covered by \( k \) ambulances, the corresponding expected covered demand is \( E_k = d_i (1-q^k) \), and the marginal contribution of the \( k^{th} \) ambulance to this expected value is \( E_k - E_{k-1} = d_i (1-q) q^{k-1} \).

Araz et al. (2005) introduced a multi-objective maximal covering location model based on Pirkul and Schilling’s CMCLP model (1997). They considered three objectives in their model: maximization of the demands covered by one vehicle, maximization of the demands covered with backup ambulances, and minimization of the total travel distance for backup ambulances. Backup ambulances provide coverage in distance greater than defined standard distance. They employed Fuzzy Goal Programming (FGP) approach to solve their multi-objective model.

Using these two models and targeting measures in Table 1, we propose the following model:

Maximize \( z_1 = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{W}_i} d_i (q_{j(k-1)} - q_{jk}) y_{ijk} \) 
Minimize \( z_2 = M \) 
Minimize \( z_3 = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{W}_i} d_i (q_{j(k-1)} - q_{jk}) t_{ij} y_{ijk} \)

Subject to
\( \sum_{j\in\mathcal{W}_i} x_{ijk} \geq 1 \) \( (i\in\mathcal{V}) \) 
\( y_{ijk} \geq y_{lk} \) \( (i\in\mathcal{V}, j\in\mathcal{W}_i, k = 2, \ldots, P_j) \) 
\( \sum_{j\in\mathcal{W}_i} y_{ijk} \leq 1 \) \( (i\in\mathcal{V}, k = 1, \ldots, P_j) \) 
\( y_{ijk} \leq x_{ijk} \) \( (i\in\mathcal{V}, j\in\mathcal{W}_i, k = 1, \ldots, P_j) \) 
\( \sum_{j\in\mathcal{W}_i} \sum_{k=1}^{P_j} x_{ijk} \leq P \) 
\( \sum_{i\in\mathcal{V}} \sum_{k=1}^{P_j} d_i (q_{j(k-1)} - q_{jk}) y_{ijk} - \sum_{k=1}^{P_j} c x_{jk} \leq M \) \( (j\in\mathcal{W}) \) 
\( y_{ijk}, x_{ijk} \in \{0,1\} \) \( (i\in\mathcal{V}, j\in\mathcal{W}, k = 1, \ldots, P_j) \)

Where: \( \mathcal{V} \) is the set of demand zones, and \( \mathcal{W} \) is the set of zones which have potential to be selected as ambulance stations. A demand zone of \( i\in\mathcal{V} \) is covered by the zone \( j\in\mathcal{W} \) if and only if \( t_{ij} \), the travel time between \( i \) and \( j \), is less than \( t \), the standard time of coverage ( \( t_{ij} \leq t \)). So \( \mathcal{W}_i = \{ j\in\mathcal{W} \mid t_{ij} \leq t \} \) shows the set of zones which can cover the demands of point \( i \) in the standard time \( t \). \( d_i \) shows demand in zone \( i \). Also, \( P \) is number
of all available ambulances, $C$ is coverage capacity of an ambulance, and $P_j$ is maximum number of ambulances that can be settled in zone $j$.

The binary variable $y_{ijk}$ is equal to 1 if and only if the area $i$ is covered by $k^{th}$ ambulance from zone $j$. The binary variable $x_{jk}$ is equal to 1 if and only if $k$ ambulances are settled in zone $j$.

$q_{jk}$, busy fraction, is inspired from MEXCLP model. $q_{jk}$ indicates unavailability of ambulances in station $j$ with $k$ ambulances. It is obvious that this coefficient depends on factors such as number of ambulances in the station and amount of demands that the station has to cover. It will be very suitable to define an appropriate relationship between these parameters, but it is to be considered that the amount of requests covered by every station is one of the output data of models and it is not definite beforehand. Maybe one of the solutions to this problem is using a recursive manner. To do so, we start with a hypothetical amount of $q_{jk}$; it is assumed that $q_{jk}$ is equal to $(\bar{q})^k$ in the first repetition, which $\bar{q}$ is average value for unavailability of ambulances which is equal for all stations. After solving the model, the amount of $q_{jk}$ can be calculated and with substitution of that in the model, it can be solved it again. This way is continued until the amount of $q_{jk}$ do not show a meaningful change.

If $E_{jk}$ be the expected amount of covered requests by $k$ ambulances in station $j$, the coefficient of $y_{ijk}$ in the goal function $z_1$ is equal to $E_{jk} - E_{j(K-1)}$ which shows increase in the amount of covered demands by adding $k^{th}$ ambulance to station $j$ (2):

$$E_{jk} - E_{j(K-1)} = d_i(1 - q_{jk}) - d_i(1 - q_{j(K-1)}) = d_i(q_{j(K-1)} - q_{jk})$$

The goal function $Z_2$ originally has the following non-linear form:

$$z_2 = \text{MinMax} \left( \sum_{i \in V} \sum_{k=1}^{P_j} d_i(q_{j(k-1)} - q_{jk}) y_{ijk} - \sum_{k=1}^{P_j} C x_{jk} \right)$$

To change $z_2$ to a linear form, we defined dummy variable $M$ accompanying with constraint 9:

$$M = \text{Max} \left( \sum_{i \in V} \sum_{k=1}^{P_j} d_i(q_{j(k-1)} - q_{jk}) y_{ijk} - \sum_{k=1}^{P_j} C x_{jk} \right)$$

In fact, $M$ indicates the maximum difference between capacity of station $j$ and the amount of request covered by that station. Therefore goal function $Z_2$ makes sure that each ambulance station is not overloaded and also total covered demands in the city are distributed fairly between all stations.

Goal function $Z_3$ shows the total time traveled to cover demands in standard time have to be minimized. This function is used to choose station $j$ from $W_i$, set of stations that can cover zone $i$ in standard time. Using $d_i$ as the weight makes the model locate station close to zones with higher demands. Also this function minimizes maintenance cost which is directly related to traveled distances.

The constraint 4 indicates that every zone like $i$ should be under coverage of at least one ambulance in a standard time period. The constraint 5 also illustrates that if zone $i$ is
under coverage of \( k \) ambulances from station \( j \), it definitely will be covered by \( (k-1) \) ambulances from the station.

The constraint 6 shows that every area like \( i \) should be covered only by one station. Therefore the model output would be independent EMS regions in which each demand zone is under coverage of only one station. This idea will be helpful for network managers to know exactly from which station ambulances must be dispatched. In constraint 7 the location of every station and the number of dedicated ambulances to it, is determined. Constraint 8 shows that number of available ambulances is limited, and constraint 10 depicts \( y_{ijk} \) and \( x_{ijk} \) as binary variables.

3. Data

To show that the proposed model is applicable in real world, the model is run for Mashhad City, Iran. Mashhad is the second populated city in Iran with around 2.5 million people. To apply proposed location model in this city, the following parameters are needed to be determined:

1. Ambulance demand in each zone
2. Travel time between zones
3. Ambulances unavailability time

To estimate parameters listed above, following data sets are employed:

1. EMS demands for one-week in the city [13].
2. Comprehensive transportation study of the city [14].

4. Method

To solve the multi-objective model, the Fuzzy Goal Programming (FGP) is employed. FGP is obtained by applying fuzzy set theory in goal programming (GP). In GP, for each objective function specific value or bond (goal) is defined and model tries to minimize sum of deviations from defined goals. GP is an extension form of linear programming (LP) and in the case that LP is infeasible, GP can provide as close as answer to the defined goals. It is very difficult especially in real world to determine the goals for objective functions. So to deal with this fuzziness and imprecision, FGP applies fuzzy set theory to GP by assigning a fuzzy membership function to each objective function. There are four popular forms for fuzzy membership function: right-sided, left-sided, triangular, and trapezoidal linear form.

First FGP method is introduced by Narasimhan (1980) and one year later Hannan (1981) made this method easier to equivalent LP. Since then many studies on FGP methods has been done and an extensive literature on this subject can be found in Chanas and Kuchta (2002). In our study, Chen and Tsai’s prioritized approach (2001) was employed to solve the proposed model.
5. Expected Results

Proposed model with models listed in Table 1 are run on Mashad City using CPLEX solver in GAMS software (19). Mashhad City has 141 traffic analysis zones (TAZ) and it was assumed that all these zones are capable to be selected as a EMS station: \( V = W = \{ 1, 2, \ldots, 141 \} \). The output of this stage is location of ambulances for each model separately. Having this outputs, calls for EMS during [7:00-15:00] time period are simulated to see which model can provide most coverage. During the simulation, it is assumed that travel time between TAZs is fixed.

After comparison the outputs of the models, sensitive analysis on proposed model’s parameters including unavailability coefficient, ambulance capacity, and travel time between TAZs will be done.

6. References

19. GAMS IDE 2.0.19.0, Module GAMS Rev 130, Lic date Jan 31, 2002, Build VIS 20.5 130