Model for Optimum Deployment of Emergency Repair Trucks: Application in Electric Utility Industry

K. G. Zografos, C. Douligeris, and L. Chaoxi

The uninterrupted supply of electricity is an important criterion for measuring the quality of service in the electric utility industry. An effective method for reducing service unavailability is to reduce the response time of emergency repair trucks (ERTs). An integrated methodological framework for the optimum deployment of ERTs is presented. The proposed methodology is based on an iterative procedure that involves the partitioning of a given geographical area into service territories and the subsequent simulation of the emergency repair operations within each service territory. A real-world problem is used to demonstrate the applicability of the proposed methodology. The results of the case study suggest that increasing the number of available ERTs from one to two decreases the response time by 64 percent, whereas an increase from two to three can provide only an additional 8 percent reduction. The reduction is mainly attributed to reduced dispatch time. Another significant observation is the fact that during heavy load, the nearest-neighbor dispatching policy provides better performance than the first-come, first-served policy. The proposed model is used by an electric utility company as a tool for determining the required number of ERTs and their service territories in such a way as to achieve predetermined service unavailability objectives.

Managers of emergency repair operations in electric utility companies frequently face decisions related to the optimum deployment of their emergency repair fleets. The major objective of emergency repair operations is that of minimizing service unavailability (I,2). Service unavailability can be decreased by reducing (a) the frequency of power interruptions, (b) the number of customers affected per interruption, and (c) the duration of the interruption. The first two approaches are related to the design properties and maintenance policies of the power distribution network. The third approach relates to the deployment of emergency repair trucks (ERTs).

This paper is motivated by the problem of deploying the ERTs of a large utility company. The main focus is the reduction of the duration of electric power interruptions through the optimum deployment of ERTs. In particular, this paper will address the issue of determining the optimum number of ERTs and their service territories in such a way as to achieve a predetermined threshold value of power restoration time, balance the workload of ERTs, and provide uniform level of service to customers.

K. G. Zografos, Department of Civil and Architectural Engineering, University of Miami, Coral Gables, Fla. 33124. C. Douligeris, Department of Electrical and Computer Engineering, University of Miami, Coral Gables, Fla. 33124. L. Chaoxi, Transportation Laboratory, University of Miami, Coral Gables, Fla. 33124.

Properties and Characteristics of ERT Deployment Problem

The properties and characteristics of the ERT deployment problem reflect the operations of a large electric utility company. This utility provided to us data covering its ERT deployment operations from July 1, 1988, to July 1, 1989. Although the particular values of the data represent the operating characteristics of this company, the general emerging patterns are typical of the emergency repair operations of any large utility company in the United States (3).

The demand for emergency repair services is created from incidents causing power interruptions that occur randomly in time and space. When a power interruption occurs, one or more customers call a service center to report service unavailability. On the basis of the customers' calls, a work order (ticket) is created by a computerized system, in this application called Trouble Call Management System (TCMS). This ticket is assigned to an emergency repair truck for further investigation and repair. The ERTs are mobile servers and at the time of dispatch can be located anywhere in a designated repair district.

The time elapsed between the arrival of a service call and the power restoration is called service restoration time (SRT). For analysis purposes the SRT is divided into the following components: ticket creation time ($T_0$), dispatch ($T_1$), travel time ($T_2$), and repair time ($T_3$).

$T_0$ is equal to the time interval between the placement of the call and the generation of the ticket. $T_0$ is controlled by the TCMS and is almost constant with an average value of about 10 min. $T_0$ does not depend on the spatial and temporal distribution of calls (3,4).

$T_1$ is equal to the elapsed time between the creation of a ticket and the assignment of the ticket to the first available ERT. $T_1$ depends on the workload assigned to each ERT.

$T_2$ is defined as the elapsed time between the ticket assignment and the arrival of the ERT at the scene of the incident. $T_2$ is a function of the shape and the size of the service area and the travel speed of the ERT (3,5).

$T_3$ is defined as the time interval between an ERT's arrival at the scene of the incident and the power restoration. The duration of the repair time depends on the type of incident, that is, severity of the problem, type of failing equipment, day versus night repair, adverse versus favorable weather conditions, and the training and expertise of the repair personnel (3).
The deployment of ERTs falls into a general category of dynamic vehicle routing problems that possess the following characteristics:

- Probabilistic demand over time and space,
- Probabilistic distribution of service times,
- Mobile servers, and
- Minimization of total system time (waiting, travel, and service time).

The problem of the optimum deployment of emergency response units has been extensively covered in the literature. Particular emphasis has been paid to the allocation and dispatching of police patrol units, fire engines, and ambulances in an urban environment (5,6).

Common to all these problems is the fact that demands vary stochastically, both temporally and spatially. Several methods have been used for solving this problem. We can categorize the models as queueing, travel time, geometric, and simulation models (7).

Queueing models have been developed to determine the number of units required to be on duty in order to achieve a threshold value of dispatch delay. In queueing models, calls for an emergency unit arriving at a service center are placed in queue and served on a first-come, first-served basis, or according to a priority system, or in more complex models according to the dispatching policy at hand—an example of which is a model's considering a minimum and a maximum number of units required to serve a specific call (7,9). Although these models provide a clear insight in the queueing phenomena of spatially distributed servers, they cannot be applied to solve the previously defined ERT deployment problem, because they assume predetermined districts and they require multiple-server dispatching and preventive patrolling.

Travel time models are more appropriate when the degree of congestion in the system is low and the travel time dominates the system response time (7). In the development of travel time models, the fact that units travel from a depot or from a place in the district to the place where they are needed is taken into account. In particular, the geography of the area, the travel speeds in the appropriate directions, and the location of the units are considered (7,11). Although these formulations are closely related to the ERT deployment problem, they cannot be applied directly for its solution, because they are limited by their first-come, first-served dispatching policy and the requirement that servers return to their depots after servicing the incident.

In geometric models the spatial distribution of the calls combined with the location of the units and the corresponding traveling speeds is taken into account to validate the empirical data (12,13). These models are not dynamic and stochastic in their formulation.

Simulation models have been used to study the effect of proposed administrative, organizational, and technological changes on the performance of emergency response systems. In simulation models a more realistic picture of the events that take place can be achieved since many factors can be taken into account and their effect on the performance of the system can be studied (6). Existing simulation models are problem-specific, and none of them has been developed in the context of emergency services in the electric utility industry.

A variety of objectives has been proposed for the optimum deployment of emergency service restoration units. Minimization of response time, equal workload for all units, uniform performance in all service territories, and minimum average delay have been used to evaluate the performance of various emergency response systems (7).

Essential to the optimal allocation of the emergency units is a districting method that considers the distribution of the calls for service in time and space (15).

The location of mobile servers within a service area is also a problem related to the design of emergency response systems. The location of the server affects its response time, which is the dominant component of the total system time under light workload conditions (3,14). Location models for the deployment of emergency vehicles have been proposed by Charnes and Storbeck (16), Daskin and Stern (17), Tozgaras et al. (18), and Brandeau et al. (19). These models deal only with the travel time aspects of the emergency response system, assuming that a service unit is always available to respond to a call in its area of responsibility. Location models considering the queueing aspects of emergency response systems have also been proposed in the literature (20,21).

Most of the existing models have been motivated by applications that have substantial organizational, operational, and technical differences from the emergency repair operation of electric utility companies. For instance, in the electric utility industry environment there is no preventive patrolling as in police operations (8,9), there is no multiple-vehicle dispatching as in fire and police operations (5), and the service unit does not necessarily return to its home base after an incident as in medical emergency systems and fire protection operations (12).

These differences preclude the wholesale application of existing emergency response models to the electric utility service restoration operations (22). Therefore, a model that reflects the service restoration operations of electric utility companies should be developed.

PROPOSED METHODOLOGICAL FRAMEWORK

The problem of the optimum deployment of ERTs is a complex resource allocation problem that presents serious ana-
lytical difficulties for its solution. For methodological reasons the problem can be decomposed into two interrelated nested subproblems: designating response areas (districting) and determining the number of ERTs required to be on duty in each district per shift.

Given the complexity and magnitude of real-world problems—large and dispersed geographical areas and the temporal, spatial, and priority distribution of calls—it is not possible to develop an exact optimization algorithm for the service restoration problem. Therefore, an iterative procedure combining optimization and computer simulation is proposed as a solution. The basic modules and the logic of the proposed method are shown in Figure 1.

The inputs of the proposed procedure are (a) the arrival rate of repair calls, (b) the travel time between geographic entities (atoms) of the study area, and (c) the number of ERTs that should be available per shift (N).

The initial module of the proposed methodology involves the solution of the districting problem. The objective of the districting problem is to partition the area under study into areas of primary responsibility, that is, service areas, so as to achieve some level or combination of levels of service. Analysis of the existing operations (22) led to the conclusion that the entire area should be partitioned into homogeneous service areas according to work load of the ERTs and area covered by each unit. The basic assumption behind the consideration of such criteria is that service areas with “similar” characteristics should have similar performance.

After the solution of the districting problem, the method proceeds by simulating the service restoration operations in each of the generated service areas. At this stage, a comparison of the performance of the service restoration operations is performed by comparing the simulated values with preestablished target values. If the performance of the system is not satisfactory, the number of ERTs (N) is increased by one and the entire process is repeated. The process stops when the preestablished threshold values are achieved.

**Model for Designing Emergency Repair Districts**

In its general form, the districting problem can be expressed as follows:

Given an area consisting of a number of elementary spatial units (atoms) with a given level of activity, determine nonoverlapping contiguous clusters of atoms (districts) that “optimize” a set of objectives.

The districting problem has been used extensively in the literature in order to design political districts (15,23,24), school districts (25–27), residential refuse collection districts (28), sales territories (29, p.469; 30,31), and inspection and repair territories (32,33).

Two basic approaches have been used for the solution of the districting problem: implicit enumeration and heuristic algorithms. The first approach formulates the districting problem as a 0-1 optimization problem and uses an implicit enumeration technique for its solution (15). This method involves two stages: the generation of feasible districts and the selection of the optimum districting pattern. Although this method is mathematically rigorous, its computational burden is very high. The second approach has been suggested in the literature for the solution of large-scale districting problems (22,32,34). This approach is based on a generalized formulation of the transportation problem.

A modification of the latter approach is used for the mathematical formulation of the emergency repair districting problem. The following notation should be introduced before the mathematical presentation of the model:

\[ N = \text{number of centers or service territories}; \]
\[ M = \text{number of atoms}; \]
\[ I = \text{set of service territories } I = \{1,2,...,N\}; \]
\[ J = \text{set of all atoms } J = \{1,2,...,M\}; \]
\[ X_{ji} = \text{workload of the } j\text{th atom assigned to the } i\text{th center}; \]
\[ t_{ij} = \text{separation (travel time) of service territory center} \]
\[ A_{i} = \text{area of atom } j; \]
\[ AT_{i} = \text{area of service territory } i; \]
\[ A = \text{average area of a service territory}; \]
\[ P_{i} = \text{workload of atom } j; \]
\[ \begin{align*}
PT_i &= \text{workload of service territory } i; \\
\bar{P} &= \text{average workload of a service territory}; \\
\lambda_j &= \text{number of calls originating at atom } j; \\
\bar{S}_j &= \text{average repair time of calls originating at atom } j; \\
\alpha_1 &= \text{maximum allowable percentage deviation of workload of a service territory from the average service territory workload } \bar{P}, \quad 0 \leq \alpha_1 \leq 1; \\
\alpha_2 &= \text{maximum allowable percentage deviation of the area of service territory from the average area of a service territory } \bar{A}, \quad 0 \leq \alpha_2 \leq 1; \\
m_{ij} &= M_0 \text{ (very large number)} \text{ if atom } j \text{ belongs to an enclave and service territory } i \text{ is not the neighboring area of the enclave}, \text{ and } 1 \text{ otherwise.}
\end{align*} \]

Since the following relations hold,
\[ \begin{align*}
PT_i &= \lambda_j \bar{S}_j \\
\bar{P} &= \sum_{j \in J} X_{ij} \\
\alpha_1 &= \frac{\sum_{i \in I} PT_i}{N} \\
\alpha_2 &= \frac{\sum_{i \in I} X_{ij}}{N} \\
\bar{A} &= \frac{\sum_{i \in I} \bar{A}}{N} \\
\end{align*} \]

The mathematical expression of the model can be written as follows:

Minimize
\[ F = \sum_{i \in I} \sum_{j \in J} X_{ij} \]

Such that
\[ \sum_{i \in I} X_{ij} = P_i \quad j \in J \quad (2) \]
\[ (1 - \alpha_1) \bar{P} \leq \sum_{j \in J} X_{ij} \leq (1 + \alpha_1) \bar{P} \quad i \in I \quad (3) \]
\[ (1 - \alpha_2) \bar{A} \leq \sum_{j \in J} X_{ij} \leq (1 + \alpha_2) \bar{A} \quad i \in I \quad (4) \]
\[ X_{ij} \geq 0 \quad (5) \]

Step 1

Determine the number of required service territories (\( N \)). If this is the initial iteration of the algorithm, then \( N \) is determined through the minimum number of required service territories \( N_{\min} \) corresponding to service day \( d \) and shift \( s \). Otherwise, the number \( N \) used in the previous iteration of the algorithm is increased by one (i.e., \( N = N + 1 \)).

Step 2

Select an initial set of \( N \) atoms to be the centers of the \( N \) service areas.

Step 3

Use the formulation of the linear program 1–5 to find assignments of atoms to service territory centers. An atom may be split into more than one part, and each part may be assigned to a different service territory.

Step 4

Within each service territory that results from Step 3, find the center that minimizes the weighted travel cost. The following formulas can be used to determine a possible initial set of coordinates of the adjusted centers \( (\bar{x}_i, \bar{y}_i) \) (note that \( (x_{i\text{ref}}, y_{i\text{ref}}) \) is the reference point for calculations of distances in service territory \( i \)):

\[ \bar{x}_i = \frac{\sum_{j \in J} (x_{i\text{ref}} - x_j) X_{ij}}{\sum_{j \in J} X_{ij}} \quad \forall i \in I \quad (6) \]
\[ \bar{y}_i = \frac{\sum_{j \in J} (y_{i\text{ref}} - y_j) X_{ij}}{\sum_{j \in J} X_{ij}} \quad \forall i \in I \quad (7) \]

Given the fact that the \((\bar{x}_i, \bar{y}_i)\) coordinates calculated by Equations 6 and 7 are optimum for the Euclidean distance metric and not for the Manhattan metric (movement is only allowed...
on the x- and y-axes) used here for the calculation of the travel matrix, a search procedure is employed to determine the optimum location of the centers for the next iteration of the algorithm (22). The search procedure starts from the best among the previous center and \((\bar{x}, \bar{y})\) and stops when no better center can be found (22).

**Step 5**

Check if there is an enclave. If an enclave exists solve Problem 2–5 with the following objective function:

Minimize

\[
F = \sum_{i=1}^{N} \sum_{j=1}^{m_i} x_i t_{ij} m_j
\]

(1')

and go to Step 3. If an enclave does not exist, go to Step 6.

**Step 6**

For each service territory center \(i\), check if the summation of the difference of the \(x\) and \(y\)-coordinates between two successive iterations differs more than a small value \(\epsilon_1\) and the positions of all the centers between two successive iterations differs more than a small value \(\epsilon_2\). If yes: go to Step 3, otherwise stop.

Appropriately, Steps 1 through 6 describe a heuristic algorithm, since the centers of the districts are not known in advance. A FORTRAN code has been written for the implementation of the described algorithm (22). The code uses the IMSL library for the solution of the linear program 1–5 in Step 3.

After the districting problem has been solved for a given number of service areas, the performance of the emergency repair operations within each generated district is evaluated using the simulation procedure described in the next subsection.

**Simulation of Emergency Repair Operations**

The proposed simulation module simulates the emergency repair operations within any service territory designed by the districting model. After the calls have been generated, the ERTs are dispatched to the locations of the incidents. At this stage of the simulation module, each truck can serve only its own territory, and there is no interdispatching allowed between truck areas. Even though this might appear to be a constraint, it has the main benefit of the truck driver's familiarity with the underlying transportation network and failing equipment that results in reduced travel and repair times.

Two alternative dispatching policies are simulated by this module. The first policy uses a first-come, first-served dispatching rule, and the second policy uses a nearest-neighbor (NN) dispatching rule.

**CASE STUDY**

A case study related to the emergency repair operations of a large electric utility company is presented to demonstrate the applicability of the proposed methodology. The service area of the case study covers a geographic area of approximately 65 mi². Workload data for the area under consideration were obtained from the TCMS data base for a 1-year period and for the operations of the second shift (3:00 p.m. - 11:00 p.m.). Atoms having an area of 1 mi² were used as the unit of analysis.

Given the Manhattan-like structure of the underlying transportation network, the Manhattan metric was used to calculate the distances between points of interest in the study area. Travel speed data were obtained through personal interviews with ERT operators.

The proposed methodological framework was applied to evaluate the performance of the emergency repair operations for three numbers of ERTs and two dispatching policies.

Concerning the balancing of the workload and area, values of \(\alpha_1 = 0.10\) and \(\alpha_2 = 0.20\) were used. Complete balancing of area and workload is almost always impossible. Tighter balancing of the area frequently leads to enclaves and irregular shapes of the service territories.

The results of the case study are presented in graphical form in Figures 2 and 3 for two and three ERTs, respectively. Figure 4 shows the existing partition of the study area for \(N = 3\). Each square of the background grid is 1 mi². The numbers in the axes relate to the longitude and latitude values of the atoms as they are used by the company under study. When a single ERT was considered, the entire area constituted the service territory.

The actual percentage deviation from the average workload and area and the associated workload and area for each scenario and the existing partitioning are summarized in Table 1. The improvements compared with the existing partition can also be seen.

The results of the districting scenario were used as inputs to the simulation module (i.e., the polygons describing each service territory). Two alternative dispatching policies were simulated for the three districting scenarios and the current configuration. Four hundred days of operation were simulated for each scenario. The results of these simulations are shown in Figures 5 and 6.

**CONCLUDING REMARKS**

An integrated methodological framework for the optimum deployment of emergency repair fleets in the electric utility
industry has been presented. The proposed model takes into account the structural, operational, and technical characteristics of the emergency repair services of the electric utility industry and expands on work done in the area of emergency response operations.

A real-world case study was used to illustrate the applicability of the proposed methodology. The results of the case study suggest that the service restoration time relies heavily on the number of available ERTs. A decrease in service restoration time of about 64 percent was observed when the number of ERTs was increased from one to two. The decrease was substantial but not as significant when the number of ERTs was increased from two to three. Therefore, the marginal benefit of adding more servers diminishes after the second server. Most of the reduction observed in the service restoration time was due to the reduction of the dispatch time component.

Another interesting observation of the derived results is the difference in the performance of the ERTs between the two dispatching policies. In particular, under heavy workload (see Figures 5 and 6 for $N = 1$) the nearest-neighbor policy yields better performance of the system in terms of service restoration time than the first-come, first-served policy. This result is in agreement with recent analytical results dealing with the solution of the dynamic traveling repairman problem with rectangle service territories and uniform load distribution (14). Under light load (see Figures 5 and 6 for $N = 2$ and $N = 3$) the differences are not substantial.

The comparison of the existing partition with the proposed partition for $N = 3$ shows more balance in terms of area and workload partition. The generation of more homogeneous in terms of workload and area of responsibility truck areas has resulted in better system performance without increasing the number of required servers. The proposed partition results in a reduction of 2.5 min in the response time, which is attributed to the reduction of the average travel time due to the balancing of the areas and the consequent reduction of the dispatch time. Better improvements than the reported ones will be observed during periods of high demand since this is the time queues build up and balancing the workload has a profound impact on the performance of the system. This balancing also contributes to better working relationships between the truck operators and provides better response times especially under heavy load conditions.
With regards to the models implementation we can say that the proposed model provided the utility company with a tool for the efficient deployment of the fleet of service restoration trucks. Given the fact that the designation of service territories is a strategic decision for the utility companies that is established for a period of at least 5 years, computational requirements of the problem are not a burden for its use and implementation. Stated otherwise, designating the service restoration territories does not require real-time decision making and consequently the computational requirements of the districting model do not make the application of the model impractical. As a final note we would like to point out that during the final implementation of the proposed methodology, the results provided by the districting model were slightly modified with the help of the system operators to produce shapes of districts that reflect the real world operational conditions and constraints.

ACKNOWLEDGMENTS

This work has been supported by the Florida Power and Light Company. The authors would like to acknowledge with thanks the assistance of Jeff Mitchell, John Haupt, and Mannie Miranda.

REFERENCES