Regional Transportation and Supply Chain Modeling for Large-Scale Emergencies

Maged Dessouky, Fernando Ordóñez, Zhihong Shen, and Hongzhong Jia

7.1 Introduction
Large-scale emergencies (or major emergencies) are defined as any event or occurrence that overwhelms local emergency responders, that severely impacts the operation of normal life, and that has the potential to cause substantial casualties and property damage. Examples are natural disasters (earthquake, hurricane, flooding, etc.) and terrorist attacks, like the one that took place on September 11, 2001. To manage the risks and consequences of large-scale emergencies and to mitigate their impacts to the population, wide-scale distribution plans of medical supplies must be developed.

Careful and systematic preplanning, as well as efficient and professional execution in responding to a large-scale emergency, can save many lives. Rational policies and procedures applied to emergency response could maximize the effectiveness of the scarce resources available in relation to the overwhelming demands. A key ingredient in an effective response to an emergency is the prompt availability of necessary supplies at emergency sites. It is challenging for practitioners to manage the huge volume of medical supplies to guarantee the readiness of the delivery and the freshness of the stock, as well as to deliver these massive supplies in a short time period to dispersed demand areas. Operations research models can play an important role in addressing and optimizing the logistical problems in this complex distribution process. Larson and colleagues (2005, 2006) conducted a detailed analysis based on well-known and recent large-scale emergencies. They emphasized the need for quantitative, model-oriented methods provided in the operations research field to evaluate and guide the operational strategies and actions in response to major emergencies.

In certain emergency settings, medication or antidotes must be applied within a specified time limit from the occurrence of the event to maximize their...
effectiveness and, in some situations, to save lives. Traditional pharmaceutical supply chains are not adequate to push a huge volume of medical supplies to the affected population within a short time frame. Usually the local caches are only enough to serve the first responders. The majority of the medical supplies used would be from the Strategic National Stockpile (SNS). The Strategic National Stockpile is part of the federal response to a bioterrorist attack on the United States and it is composed of a combination of vaccines, prophylaxis, and medical supplies aimed at minimizing the damage of an attack. This stockpile is separated into prepositioned push packages and a larger inventory managed by manufacturers (known as Vendor-Managed Inventories – or VMI – that correspond to about 80% of the SNS). The VMI provide additional flexibility and cost efficiency as they can deploy the specific medicine to combat a bio threat. Given the possible biological and chemical threats facing the nation and the limited budget to set up the SNS, the effective allocation of these resources in designing the SNS is key in managing the risk posed by these threats.

The SNS is administrated by the Centers for Disease Control and Prevention in the United States. For example, to address emergencies of infectious diseases, the SNS contains 300 million doses of smallpox vaccines and enough antibiotic to treat 10 million people for anthrax during a large-scale bioterror attack. We illustrate the distribution process in Figure 7.1. If there is a public health emergency severe enough to cause local supplies to run out, the SNS supplies can be accessed. Initial medical supplies from the push packages will be delivered to any state in the United States within 12 hours of federal and local authorization. This initial delivery can be followed by supplies from the VMI if needed. After the stockpile reaches the regional central depot, it is each state’s responsibility to have plans to receive and distribute SNS medicine and medical supplies to local communities as quickly as possible. The state investigates the efficient strategies and best practices in identifying potential points of dispensing (PODs) sites where the population in the nearby neighborhood comes to seek medical services, as well as in constructing suitable routes to distribute the supplies from the regional central depot to the selected PODs.

![Figure 7.1. Distribution process of large-scale emergency response.](image-url)
As the primary goal is to establish an effective and rapid distribution system for the medical supplies to respond to a large-scale emergency, it is natural to decompose the whole problem into three separate but interconnected key decisions, which can be integrated to provide a complete medical supply chain management solution to large-scale emergencies (see Figure 7.2). The three interconnected key decisions in the distribution process are:

- How to manage these medical supply inventories?
- Where to store them?
- How to distribute them?

A perishable inventory management model (Section 7.2), a facility location model (Section 7.3), and a routing model (Section 7.4) are formulated to address these three questions separately to respond to the challenges presented earlier. The models and solutions presented in this chapter are based on our previous work (Jia et al., 2007a, 2007b; Shen et al., 2009a, 2009b, 2011), and we refer the reader to these articles for in-depth explanations and additional results. The solutions to each problem can be coordinated into an overall efficient management strategy for the distribution process in the emergency response context. The minimum amount that must be kept in the inventory system is enforced by the SNS policy, which is the key constraint in the inventory model and also provides the capacity constraint to the location model as well as the total supply constraint to the routing model. The results of the location model are the selected PODs, which will feed into the routing model as the demand points to service. The goal is to establish efficient inventory management policies and to dispatch plans to help save lives through a timely delivery plan.

Figure 7.2. Problem decomposition for emergency response.
In a large-scale emergency setting, the initial prepositioned supplies will typically not be sufficient to cover all demands in the recommended times. The unmet demand in an emergency situation can directly result in loss of life, an impact that outweighs any operational costs. Another particular feature of large-scale emergencies is that the massive inventory in stock (medical supplies) is made up of perishable items. Therefore, the overriding objective in our work is to mitigate the impact of an emergency by minimizing unmet demand. It requires both an efficient inventory management scheme for the massive supplies to guarantee the readiness and freshness of the stockpile for continuous service and an effective plan to distribute the huge quantity of supplies as much as possible within the preferred timeline. We use a mix of optimization solvers and heuristics to solve the three models presented in this chapter. In this chapter, for the inventory management model, we developed a search procedure to find the optimal production policy. For the facility location model, we developed a standard locate-allocate heuristic to solve the problem while the routing model is optimally solved using the CPLEX commercial solver.

7.2 Inventory Management Problem
The readiness and freshness of the continuous massive supply from the VMI, managed by private vendors, is crucial to have the ability to satisfy the sudden demand that might arise in a large-scale emergency. It is imperative to enable massive stockpiles of medical supplies in an economical manner that would not put undue stress on vendors, who would otherwise require a higher compensation to participate. Consider, for instance, the requirement of the stockpile of Cipro, a common antibiotic used for anthrax infections and other illnesses. The federal government aims to disperse antibiotics to treat 10 million people in the case of a massive anthrax attack. This stockpile represents enough Cipro to meet regular market demand for several years. Current SNS policy allows firms to partially recoup costs by selling the stockpile in the private market when supplies are six months from expiration.

Since the size of the stockpile is large compared with the expected market demand, which leads to waste, an opportunity exists to implement a combined inventory policy that addresses both the regular market demand and the SNS minimum inventory in an efficient manner. From the manufacture’s perspective, a more sophisticated inventory holding policy that could enable using the stockpile to meet the regular market demand and production to maintain the stockpile inventory has the potential to reduce costs and thus reduce the price charged to the government. In this section, we explore such a combined inventory model and estimate its economic impact. For a general review of
inventory models for emergency management, the reader is referred to the paper by Ozguven and Ozbay (2014).

### 7.2.1 Inventory Model

We begin with a brief summary of the modified EMQ model proposed by Shen and colleagues (2011), which is used to obtain the optimal ordering policy for a perishable inventory system with a large minimum volume constraint. We assume a single fixed-life perishable item is produced, consumed, and stored for an infinite continuous time horizon. There are two types of demand. One is a known regular market demand with a constant rate of $D$ items per period. The production can start at any time at a constant maximum allowable rate of $P$, which is greater than $D$. The other demand is that the inventory system must maintain at every point in time at least $I_{\text{min}}$ of non-spoiled inventory.

Excess supplies could be disposed to some secondary market (e.g., overseas market) as long as it is before expiration of the item. The model assumes that inventory is consumed using the natural first-in-first-out (FIFO) policy and prices for both the primary and secondary markets are uniform. In this inventory model, the inventory cycle $T_{\text{inv}}$ is defined as the minimum length of time that an inventory pattern repeats and it is a given parameter. It is assumed that excess supplies above the $I_{\text{min}}$ level will be disposed (salvaged) once at the end of each $T_{\text{inv}}$ so that exactly $I_{\text{min}}$ items are present at the beginning of the next inventory cycle. To maintain a fresh inventory, every $T_{\text{inv}}$, the $I_{\text{min}}$ amount produced in the previous inventory cycle is consumed and another $I_{\text{min}}$ amount is produced. The production batch size $Q$ is the sole decision variable and we can determine all the relevant quantities to describe this model as a function of $Q$. For any given $Q$, we initially run a regular EMQ cycle (underlying regular EMQ cycle with cycle length $T = QD$), and we may need to make some adjustment near the end of the inventory cycle to satisfy the minimum inventory requirement (last production cycle).

Given different possible combinations of $T_{\text{inv}}$ and $I_{\text{min}}$, there is a total of five different possible scenarios. Figure 7.3 provides illustrative inventory plots for all five cases. The classification is based on three criteria: 1) if a last production cycle is needed; 2) where the inventory cycle ends relative to a regular underlying EMQ cycle; 3) where the last production cycle starts. Detailed explanation can be found in Shen and colleagues (2011). The total cost of the modified EMQ system is a non-continuous, non-differentiable function of $Q$.

Let $TRC(Q)$ be the total cost function. Shen and colleagues (2011) show that $TRC(Q)$ for each case among the five cases is a quadratic function of the order quantity $Q$ as well as boundary conditions for transitions between the different cases. To illustrate we present $TRC(Q)$ for case 1. The total costs within
Figure 7.3. Five cases for combined inventory policy: perishable inventory, $I_{\text{min}}$ minimal inventory requirement, $T_{\text{inv}}$ inventory cycle.
a single inventory cycle are composed of three parts: inventory holding cost $h$, fixed setup cost $A$, and unit purchasing cost $v$. After some simplifying steps, Shen and colleagues (2011) show that

$$TRC(Q) = A(N+1) + \left( \frac{Q}{D} - T_{inv} + N \frac{Q}{D} \right) Dv$$

$$+ \frac{1}{2} h \left[ (N+1) \frac{Q^2}{D} \left( 1 - \frac{D}{P} \right) - D \left( \frac{Q}{D} - T_{inv} + N \frac{Q}{D} \right)^2 \right]$$

where $N$ is the number of complete regular production cycles in one $T_{inv}$ cycle. For a given segment, $N$ is fixed so $TRC(Q)$ is a quadratic function of $Q$ for case 1. By proving that when $N$ is greater than a threshold value the total cost will monotonically increase as $N$ increases, the exact optimal $Q$ can be obtained with pseudo-polynomial complexity. Shen and colleagues (2011) show that this model only holds when $I_{min} \leq T_s$ and $T_{inv} \leq T/2$, where $T_s$ is the shelf life of the inventory.

### 7.2.2 Example

Computational experiments have compared the cost of maintaining a massive $I_{min}$ inventory in an integrated system to the cost of separate inventories for regular demand and for the SNS stockpile. Given the following problem parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{min}$</td>
<td>million pills</td>
<td>1200 (= 10 × 60 × 2)</td>
</tr>
<tr>
<td>$T_{inv}$</td>
<td>month</td>
<td>36</td>
</tr>
<tr>
<td>$Y$</td>
<td>month</td>
<td>36 (= 108−2 × $T_{inv}$)</td>
</tr>
<tr>
<td>$P_{gov}$</td>
<td>mil $/mil pill</td>
<td>0.95</td>
</tr>
<tr>
<td>$P_{market}$</td>
<td>mil $/mil pill</td>
<td>4.67</td>
</tr>
<tr>
<td>$v$</td>
<td>mil $/mil pill</td>
<td>0.2</td>
</tr>
<tr>
<td>$w$</td>
<td>mil $/mil pill</td>
<td>−0.3 (3 years)/−0.075(6 years)</td>
</tr>
<tr>
<td>$D$</td>
<td>mil pill/year</td>
<td>300</td>
</tr>
<tr>
<td>$P$</td>
<td>mil pill/year</td>
<td>600</td>
</tr>
<tr>
<td>$A$</td>
<td>mil $/time</td>
<td>2 (= 100 × $h$)</td>
</tr>
<tr>
<td>$h$</td>
<td>mil $/mil pill/year</td>
<td>0.02</td>
</tr>
</tbody>
</table>

we obtain that the integrated system has a total cost of $78$ million per year, which is less than the $111.45$ million per year obtained for a standard model. This standard model splits the production capacity of 600 million pills per year into two parts: an $I_{min}$ inventory renewed every $T_{inv}$ and a separate inventory for the regular demand. We use 400 million pills per year for the regular market.
demand with the regular EMQ model and the remaining 200 million pills per year constantly running to build the 1.2 billion pills $I_{\text{min}}$, which is completely refreshed every six years. The disposed $I_{\text{min}}$ inventory is sold at 7.5 cents per pill on the secondary market.

Shen and colleagues (2011) study the sensitivity of the optimal solution with respect to various parameters. For instance, we observe that the industry profits per year increase with the price per pill and also with the amount of time remaining in salvaged inventory (increasing $Y$).

The different slopes observed on the profit versus flexibility ($Y$) plot in Figure 7.4 have to do with the different impacts of increasing flexibility ($Y$) the amount of time remaining in the salvaged inventory. The flat slopes are due to cost savings in the regular EMQ part of the inventory for changes in $Y$, while the sharp profit improvement is due to the additional gain from the extra salvage amount.

### 7.3 Storage Problem

In the event of a large-scale emergency such as a terrorist attack (e.g., September 11) or a major natural disaster (e.g., hurricane), tremendous demands for medical supplies will occur at the incident site(s) in a short time period. Emergency preparedness requirements for such events need pharmaceutical caches to be prepositioned or staging areas to be determined so as to enable a rapid disbursement of supplies from the national stockpiles (CDC, 2005). An important consideration in selecting the locations of facilities (i.e., the pharmaceutical caches
or staging areas) is the coverage of the demand areas. A sufficient coverage of the demand areas by the facilities can ensure prompt medical services to the population, hence minimizing the loss of life caused by the emergencies.

Most facility location models, including the ones in a context of emergency services, consider providing a single facility (possibly including a few backup facilities) to cover a demand point (Church & ReVelle, 1974; Hogan & ReVelle, 1986; Huang et al., 2010; Marianov & ReVelle, 1996; Paluzzi, 2004; Schilling et al., 1979). Such emergency models, however, typically have not considered conditions where a tremendous demand and low frequency combine to create situations with insufficient supplies and large uncertain demands. These conditions, which may occur in large-scale emergencies, require a modification in the definition of facility coverage to allow for redundant facility placements and tiered facility services to ensure an acceptable form of coverage of all demand areas. Given the occurrence of a large-scale emergency, the resources of a number of facilities need to be applied to quell the impact of the emergency. This implies that, to service a demand point, a number (multiple quantity) of facilities should be used, which can be classified in terms of the distance/time to the demand point. A facility that is close to a demand point provides a better quality of coverage to that demand point than a facility located far from that demand point (Dessouky et al., 2006).

Jia and colleagues (2007a, 2007b) introduced a new facility location model for the medical supply distribution for large-scale emergencies. This model aims to maximize the population coverage, and addresses the uncertainty of demand and lack of supplies at each facility by requiring each demand point to receive a multiple facility quantity and quality of coverage. Specifically, the multiple facility quantity of coverage means that each demand point is covered by a multiple quantity of facilities that can be decided based on the attributes of the demand point such as the population density and the likelihood to be impacted by an emergency. The multiple facility quality of coverage means that the demand points are to be serviced by the facilities at different and tiered distance levels. We note that facility location problems for routine demand services with uncertainties have been previously investigated; for example, see Chen and Lin (1998) and Snyder and Daskin (2006). For a complete review of facility location problems with uncertainties, see Snyder (2006).

7.3.1 Facility Location Model

The facility deployment problem considers a given geographical territory and assumes that the requests for medical services are concentrated among various (finite) demand areas within the territory. To properly allocate the medical resources, the eligible facility sites need to be identified and the potential demand areas need to be categorized. Each demand area has distinct attributes
such as population density, economic importance, geographical feature, weather pattern, etc. Therefore, the likelihood for a certain type of large-scale emergency to occur in one area, as well as the corresponding impact level, are different from the other areas. As such, the quantity of facilities that needs to be allocated to a demand area should consider the attributes of the demand point and hence may be different from the number of facilities allocated to other areas. The selection of eligible facility sites must also consider a set of criteria that are suited for large-scale emergencies. For instance, the facilities should have easy access to more than one major road/highway. The sites should be secured and insusceptible from damages caused by the emergencies.

To formulate the covering facility location model for large-scale emergencies, we consider a set $I$ of demand points and a set $J$ of eligible facility sites. Indexed on these sets the following types of integer variables are defined, requiring each demand point to be serviced at quality levels $r \in \{1, \ldots, q\}$. Readers interested in details of the model are referred to Dessouky and colleagues (2006) and Jia and colleagues (2007a, 2007b).

$$x_j = \begin{cases} 1 & \text{if a facility is placed at eligible site } j; \\ 0 & \text{otherwise.} \end{cases}$$

$$u_{ir} = \begin{cases} 1 & \text{if demand point } i \text{ is covered at quality level } r; \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, the following input parameters are defined:

**Inputs:**

- $M_i = \text{the population of demand point } i$;
- $Q_i^r = \text{the minimum number of facilities that must be allocated to demand point } i \text{ so that } i \text{ can be considered as covered at quality level } r$;
- $P = \text{the maximal number of facilities that can be placed in } J$;
- $d_{ij} = \text{the distance from eligible facility site } j \text{ to demand point } i$;
- $D_i^r = \text{the distance requirement for demand point } i \text{ to be serviced at quality level } r$;
- $a_{ij}^r = \begin{cases} 1 & \text{if eligible facility site } j \text{ can cover demand point } i \text{ at quality level } r, \text{i.e. } d_{ij} \leq D_i^r \\ 0 & \text{otherwise} \end{cases}$
- $c^r = \text{the importance weighting factor of the facilities that have quality level } r$.

In what follows, we present a maximal covering model that is suitable to the location problems of large-scale emergencies. In this model, each demand point is considered covered only if it can be serviced by a multiple quantity of facilities that are located at different quality levels (distances). The model
has an objective of maximizing the demands covered by sufficient quantity of facilities at different quality levels.

\[
\nu(\text{MCLP}) = \max \sum_r c^r \left( \sum_i M_i u^r_i \right) = \sum_r \sum_i c^r M_i u^r_i \quad (7.1)
\]

Subject to:

\[
\sum_{j \in J} x_j \leq P \quad (7.2)
\]

\[
x_j = \{0, 1\} \quad \forall j \in J \quad (7.3)
\]

\[
\sum_{j \in J} a^r_j x_j \geq Q^r_i u^r_i \quad \forall i \in I, r = 1, \ldots, q \quad (7.4)
\]

\[
u^r_i = \{0, 1\} \quad \forall i \in I, r = 1, \ldots, q \quad (7.5)
\]

In the objective function, the weight parameter, \(c^r\), is used to prioritize the importance of the facilities at each different quality level. For any demand point, in general, a high-quality-level facility coverage is more crucial than a low-quality-level facility coverage. Therefore, we define \(c^1 \geq c^2 \geq \cdots \geq c^q\). Constraints (7.2) and (7.3) are used to represent that there are at most \(P\) facilities to be located in a set \(J\) of possible locations. Constraint (7.4) states that demand point \(i\) is considered as covered at quality level \(r\) only if there are more than a required quantity \(Q^r_i\) of facilities located within the corresponding distance constraint servicing it. Note that the number of quality levels \(q\) and the quantity of the facilities at each quality level \(Q^r_i\) for each demand point can be determined by considering the population, the importance, and the emergency occurrence likelihood at each demand point. In general, a demand point that is of a greater economic/political importance, has more population, and is more likely to receive an emergency should be covered by more facilities located at different quality levels. Also note that we consider that a facility of quality \(r\) is also of quality \(r + 1, \ldots, q\) since it is closer than required for larger quality of coverage. Thus we must satisfy the requirements: \(Q^r_i \leq Q^{r+1}_i\) for all qualities \(r = 1, \ldots, q - 1\). A facility that services a demand point at a high quality level (e.g., the first quality level) is also considered able to service this demand point at lower quality levels. Finally, constraint (7.5) enforces the integrality of variables \(u^r_i\). This problem is similar to the maximal covering location problem (MCLP) considered by Church and ReVelle (1974), Meggido and colleagues (1983), and Galvão and ReVelle (1996), but with different conditions.
to consider a point covered. Note that any solution $x$ that locates $P$ facilities is a feasible solution to the problem; simply set all $u_i = 0$ to satisfy constraint (7.4). Variables $u$ are auxiliary variables that quantify the number of points that are covered for any location $x$ of up to $P$ facilities.

For different large-scale emergencies, the demand for medical supplies at any point $i$ may vary, and it depends on various factors such as the impact of the emergency and the likelihood for the emergency to affect a demand point. In the model, we use the population parameter, $M_i$, to represent the demand for medical supplies at point $i$.

### 7.3.2 Illustrative Example

The emergency instance we consider is an anthrax emergency in Los Angeles County, which has a total population of 9.5 million. Anthrax is an acute infectious disease caused by the spore-forming bacterium. There are several types of anthrax infections (cutaneous, inhalation, and gastrointestinal), and each requires vaccines and antibiotics to cure infected people and immunize the high-risk population. During an anthrax emergency, the Strategic National Stockpile (SNS) is mobilized for local emergency medical services. The problem is then to select a given number of eligible facility sites as staging areas for mass distribution of the medical supplies from the SNS to cover as much of the population as possible.

To formulate the facility location problem, we first identify the demand points. We use the population density pattern that is available for Los Angeles County (ESRI website 2005), and consider the centroid of each census tract as a demand point to represent the aggregated population in this tract. Thus we obtain 2,054 discrete demand points that have different population densities. In addition, we select the set of eligible facility sites by deviating the facility sites that were identified by the emergency management of Los Angeles County by a small distance. The county has identified a little more than 200 eligible facility sites. In this chapter, we use 200 as the number of eligible facility sites.

To ensure adequate facility service, we consider multiple quantity coverage requirements for the demand points. The quantity requirements for a single quality level are dependent on the population at each demand point and they are defined as follows:

1. $Q_i = 1$, if the population of demand point $i$ is less than 3,000;
2. $Q_i = 2$, if the population of demand point $i$ is between 3,000 and 6,000;
3. $Q_i = 3$, if the population of demand point $i$ is between 6,000 and 9,000;
4. $Q_i = 4$, if the population of demand point $i$ is greater than 9,000.
Furthermore, we use double-quality coverage to provide medical services for each demand point. The distance requirements for the first and second quality levels are defined respectively as 8 km and 16 km. We consider the quantity of facilities required to service a demand point at the second quality level double the quantity of facilities required at the first quality level. For example, if a demand point has a population of 5,392, then two facilities will be required at quality level 1 (within 8 km of the demand point), and four facilities will be required at quality level 2 (within 16 km of the demand point). In addition, for simplicity, we define the two different quality level coverages as having the same importance, i.e., $c_1 = c_2$.

We consider the location problem of incrementally opening the complete set of the eligible facility sites in Los Angeles County. In a large-scale emergency, it will most likely be impractical to open all the sites at the same time due to the limited availability of the emergency responders and medical supplies (i.e., they do not arrive all at one time). Thus, it is likely that the sites will have to be opened sequentially. In this set of experiments, we gradually increase the set of facilities by 20 each time. Thus the numbers of opened facilities after each time interval are: 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, and so on. Furthermore, we consider that once a facility is opened, it cannot be closed.

We use the LocAlloc heuristic (see Jia et al., 2007a) to solve the series of location problems. To evaluate the solutions, we also solve these problems by uniformly selecting facilities throughout Los Angeles County, which is a plausible solution strategy during an emergency. We refer to this solution as the uniform solution. Figure 7.5 depicts the coverage results obtained from the heuristic solutions and the uniform solutions, including the primary (first facility) coverage, first quality level coverage, and both first and second quality level coverage for the sequential location problems.

From these results, it can be seen that the heuristic is able to generate much better location solutions than the uniform selection approach for all the location problems (except for the last one, in which the results are equal). Based on the solutions obtained from the heuristic, a significantly larger percentage of the population can receive the primary (first facility) coverage, first quality level coverage, and both first and second quality level coverage.

In addition, it is shown from the figure that most of the population is covered by the first 100 opened facilities and the later opened facilities provide only limited marginal coverage. This implies that during the emergency, the resources for the first 100 facilities should be deployed as early as possible so that the impact of the emergency can be mitigated and the loss of life can be minimized.
The objective of the routing problem we consider in a large-scale emergency situation is to minimize the unsatisfied demand that can occur because resources and first responders are overwhelmed. In fact, unmet demand in an emergency situation can result in loss of life, an impact that outweighs other commonly used vehicle routing problem (VRP) objectives such as the travel time or number of vehicles used. We note some researchers have combined the location and routing problems (see, e.g., Oran et al., 2012). In this chapter, we treat these two problems separately since the combined problem becomes computationally tractable when considering all the uncertainties associated with a large-scale emergency.

Unmet demand in these emergency situations can be due to having insufficient resources or delays in reaching the affected population. In particular, the administration of medication to the population within a time frame makes an appreciable health difference in a large-scale emergency. It is therefore important to associate each demand with a time window, to address situations where late delivery directly leads to loss of life.

The unpredictable nature of large-scale emergencies leads to many stochastic elements in the problem, including stochastic demand and travel times. First, at a given demand point (e.g., a neighborhood block), the quantity of required emergency supplies (antidotes, protective equipment, medication, etc.) is often proportional to the unknown size of the population and/or numbers of casualties.

Figure 7.5. The coverage results with sequentially available resources.
Second, the casualties’ exposures, or demands among “worried well” are hard to accurately predict. In addition, in an emergency situation, the traffic condition can become highly uncertain, due to unpredictability in people’s behavior. At the same time, the emergency event itself may directly affect road conditions. For example, an earthquake may destroy a segment of the freeway.

When we consider the problem of routing inventory in response to large-scale emergencies, both routes and delivery quantities are important in creating an effective dispatching plan. These two aspects of the solution are related to the following uses of efficient inventory routing plans for a large-scale emergency: First, from a planning perspective, agencies can use efficient routes in mock trial runs of an emergency event to predesign the routing policies and provide training opportunities. Second, from an operational perspective, the models can be used to update the predesigned routes and promptly determine exact delivery quantities to dispensing or treatment sites to generate real-time, near-optimal dispatching solutions after an emergency situation has happened. In this chapter, we focus on the planning aspect of the routing problem, which appeared in Shen and colleagues (2009a). The recourse operational aspect appears in Shen and colleagues (2009b).

### 7.4.1 Vehicle Routing Model

As mentioned earlier, in the planning aspect, we aim to minimize the unmet demand, taking into account the uncertainty in demand and travel time. Such a problem is related to the routing problem with profits, since minimizing the unmet demand maximizes the collected profits when the profit of each node is its demand level. The route generation is also constrained by deadlines associated with every demand node. However, in our problem, the unmet demand is not only caused by the deadlines, but also potentially by the vehicle capacity and the insufficient supply at the depot.

In our problem setting, the large-scale emergency makes both the demand level and the traffic condition highly unpredictable; hence uncertainty exists both in demand and in travel time. The overwhelming demand with limited resources and the urgency of the timely delivery make the problem constrained by both the vehicle capacity (in demand dimension) and the service deadline (in temporal dimension). We formulate the route-planning stage as a stochastic programming problem where we quantify all unmet demand, which can be due to insufficient capacity, lack of supply, or late delivery. We use chance constraints to represent the uncertainty in demand and travel times.

In addition to minimizing the unmet demand, the planning problem considers a secondary objective to guide the construction of a complete route. Instead
of ignoring the nodes not selected by the routing problem with profits due to tight deadline, insufficient supply, or limited vehicle capacity, we include them at the end of each route according to this secondary objective. These complete routes that visit all demand nodes, even after the deadline or with an empty truck, provide a blueprint for recourse actions in the operational stage. The secondary objective is the arrival time at each node. We use a significantly small coefficient $k$ to weigh this secondary objective so it will not interfere with the route generation guided by the minimum unmet demand primary objective.

Next, we define the notation and present the mathematical model for the vehicle routing problem with uncertainty in demand and travel time used to minimize unmet demand while visiting all nodes. Readers interested in details of the model are referred to Dessouky and colleagues (2006) and Shen and colleagues (2009a, 2009b). We consider a set $K$ of vehicles and a set $D$ of demand nodes. We identify an additional node, node 0, as the supply node (depot) and let $C = D U \{\text{node 0}\}$ represent the set of all nodes. Indexed on sets $K$ and $C$, we define the following decision variables

$$X_{i,j,k} = \begin{cases} 1 & \text{if vehicle } k \text{ goes from node } i \text{ to node } j \\ 0 & \text{otherwise.} \end{cases}$$

$$S_{i,k} = \begin{cases} 1 & \text{if node } i \text{ can be serviced by vehicle } k; \\ 0 & \text{otherwise.} \end{cases}$$

$Y_{i,j,k}$ = amount of commodity traveling from node $i$ to node $j$ using vehicle $k$

$U_i$ = amount of unsatisfied demand of commodity at node $i$

$T_{i,k}$ = visit time at node $i$ of vehicle $k$

$\delta_{i,k}$ = delay incurred by vehicle $k$ in servicing $i$.

We also define the following input parameters:

Inputs:

$s$ = amount of supplies at the supply node (depot)

c$_k$ = load capacity of vehicle $k$

dl$_i$ = service deadline at demand node $i$.

$\tau_{i,j,k}$ = uncertain time required to traverse from node $i$ to node $j$ for vehicle $k$

$\zeta_i$ = uncertain amount of demand at node $i$.

$\alpha_D$ = confidence level for chance constraint that defines the unmet demand

$\alpha_T$ = confidence level for chance constraint that defines vehicle arrival time

We also let $M$ be a large constant used to express nonlinear relationships through linear constraints. With this notation we express the minimum unmet demand routing problem as the problem below. We notice that the first three
constraints in the formulation plus cycle canceling constraints enforced by the chance constraints define feasible vehicle routes. The fourth constraint defines the delays at nodes and measures arrival time only for the nodes visited by a vehicle. The next three constraints identify the nodes serviced by vehicle \( k \) and ensure no flow is left at \( i \) by \( k \) if the node is not serviced by \( k \). The eighth constraint enforces the initial supply of resources at the depot and the ninth constraint allows the transfer of supplies only where vehicles have moved. The last two constraints are chance constraints that require the solution be such that the demand satisfaction constraints and the definition of arrival time are satisfied with high probability.

\[
\begin{align*}
\min & \sum_{i \in D} U_i + \kappa \sum_{i \in D, k \in K} T_{i,k} \\
\text{s.t.} & \sum_{j \in C} X_{0,j,k} = \sum_{j \in C} X_{j,0,k} = 1 & k \in K \\
& \sum_{j \in D, k \in K} X_{i,j,k} = \sum_{j \in D, k \in K} X_{j,i,k} = 1 & i \in D \\
& \sum_{j \in D} X_{i,j,k} = \sum_{j \in D} X_{j,i,k} & i \in D, k \in K \\
& 0 \leq T_{i,k} \leq \min \left\{ M \sum_{j \in C} X_{i,j,k}, \delta_{i,k} + dl \sum_{j \in C} X_{i,j,k} \right\} & i \in D, k \in K \\
& \delta_{i,k} \leq M(1 - S_{i,k}) & i \in D, k \in K \\
& S_{i,k} \leq \sum_{j \in C} X_{i,j,k} & i \in D, k \in K \\
& MS_{i,k} \geq \left( \sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k} \right) & i \in D, k \in K \\
& S \geq \sum_{k \in K} \left( \sum_{j \in C} Y_{0,j,k} - \sum_{j \in C} Y_{j,0,k} \right) \\
& Y_{i,j,k} \leq c_{i,j}X_{i,j,k} & i, j \in C, k \in K \\
& X_{i,j,k}, S_{i,j} \in \{0,1\} & i, j \in C, k \in K \\
& U_i \geq 0, T_{r,k} \geq 0, T_{0,k} = 0, \delta_{i,k} \geq 0 & i \in D, k \in K \\
& P \left\{ \sum_{k \in K} \left( \sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k} \right) + U_i \geq \zeta_i \right\} \geq 1 - \alpha_D & i \in D \\
& P \left\{ T_{i,k} + \tau_{i,k} - T_{j,k} \leq M \left( 1 - X_{i,j,k} \right) \right\} \geq 1 - \alpha_T & i \in D, j \in C, k \in K
\end{align*}
\]
Since the chance constraints in this formulation have only one random parameter on each probability statement that appears by itself, if the uncertain parameters have known distributions, the chance constraints can be simplified by using the 1-\(\alpha\) critical level of the distribution. That is, given an univariate random variable \(Z\) with cumulative distribution \(F_Z\), the statement \(P\{Z \leq f(x)\} \geq 1-\alpha\) is equivalent to \(F_Z^{-1}(1-\alpha) \leq f(x)\). We can therefore simplify the chance constraints in the above formulation if we assume that travel times between nodes of the graph and the demand at a node follow lognormal distributions. We assume that \(\log(\zeta_i)\) follows \(N(\mu_{\zeta_i}, \sigma_{\zeta_i})\) and that \(\log(\tau_{ijk})\) follows a \(N(\mu_{\tau_{ijk}}, \sigma_{\tau_{ijk}})\). This means we can solve the problem with chance constraints using the following deterministic constraints, where \(\kappa_T\) and \(\kappa_D\) correspond to the inverse of the standard normal at the 1-\(\alpha\) critical level:

\[
\left( T_{i,k} + \exp\left(\mu_{\tau_{i,j,k}} + \kappa_T \sigma_{\tau_{i,j,k}}\right) - T_{j,k} \right) \leq M \left(1 - X_{i,j,k}\right) \quad i \in D, j \in C, k \in K
\]

\[
\sum_{k \in K} \left( \sum_{j \in C} Y_{j,i,k} - \sum_{j \in C} Y_{i,j,k} \right) + U_i \geq \exp\left(\mu_{\zeta_i} + \kappa_D \sigma_{\zeta_i}\right) \quad i \in D
\]

7.4.2 Example

Our computational experiments are primarily focused on comparing the alternatives between short-and-risky routes and long-and-safe routes as well as exploring how the chance constraints safety factors influence the quality of the routes. We test our model on 10 different problem scenarios. For each scenario, we uniformly generate: 50 demand nodes and one depot node in a 200 by 200 square domain; and a mean demand quantity for each demand node ranging from 5 to 15. We service this demand with a fleet of 10 incapacitated vehicles. We use the Euclidean distances between any two of the nodes and assume a symmetric complete graph topology. The mean travel time between any two nodes is proportional to the distance. In the deterministic model (DP), we use the mean value of the demand quantities and travel times as its parameters. In the chance constrained model (CCP), we use a lognormal distribution with the same mean value as was used in the DP model; the standard deviation is set to be proportional to the mean value for the demand (20% of the mean value) and inversely proportional to the mean value for the travel time (\(\sigma = \mu (UB - \mu)/100\)), \(UB\) is the upper bound for the inversely proportional transformation, whose value is dependent on the graph topology, which must be greater than the longest arc in the graph. We restrict our analysis to this type of travel time distribution because we
aim to compare the trade-offs between short-and-risky routes and long-and-safe routes. For the chance constrained model, we set the confidence level as 95%, thus setting the “safety factor” values $\kappa_D$ and $\kappa_T$ in the deterministic equivalent of the chance constraints to 1.65.

We conduct a simulation to compare the DP and CCP routes. For 30 test cases over each one of 10 problem scenarios, we solve the problem formulations using the same tabu heuristic. We then generate 100 instances of emergencies to evaluate how well these preplanned routes behave. We allow adjusting the supplies on the route after observing the realization of the demand and travel time uncertainty. Our results appear in Figure 7.6.

The results suggest that, when the deadline is very tight (between 40% to 60% of the base route length), the deterministic routes outperform the CCP routes. There is no benefit for utilizing a conservative routing strategy under tight deadlines since the longer, less-risky trips (arcs) in most generated instances are longer than the deadlines. In this case, the short-and-risky routes generated by the DP model at least have a chance in some of the generated instances of arriving on time, especially for the demand points in the beginning of the route. In the situation where there is limited supply (e.g., when the total supply is only 70% of the total demand), the DP and CCP routes yield roughly the same amount of unmet demand even under a more relaxed
Regional Transportation and Supply Chain Modelling

Because of the insufficient supplies at the depot, the extended deadline cannot provide a better coverage.

Under moderate deadlines and supply quantities (when the percentage of the total supply over the total demand and the percentage of the deadline over the base route length both are between 80% and 120%), we can conclude that CCP routes outperform DP routes 2% to 6% in terms of the percentage of the unmet demand over the total demand, according to our simulation experiments. As the deadline and the total supply are relaxed, the percentage of the unmet demand is decreasing for both DP and CCP routes; at the same time, the advantage of CCP routes over DP routes gradually increases. The simulation results on the percentage of the unmet demand from the CCP routes are approaching to its lower bound due to the shortage of the supply, with an increasing deadline. We believe that it is because of, under the relaxed enough deadline and supply quantities, its conservative nature that the CCP model shows its merits. It leads to similar number of nodes in different routes in the CCP model. In contrast, the DP routes are more prone to have an uneven number of nodes between different paths, leading to a higher chance of having unmet demand when they are tested over different realization instances.

Another experiment we conduct is to fix the deadline and the total supply at the depot and vary the confidence level, which has a one-to-one correspondence to the $\kappa_T$ and $\kappa_D$ values in the chance constraints. We use the same value for both $\kappa_T$ and $\kappa_D$. For one specific network topology and demand distribution (a problem scenario), we fix the deadline at 80% of the base route length and the total supply at 90% of the total demand. We generate different sets of CCP routes with different safety factors, and run the same simulation as described previously. We plot the percentage of the unmet demand over the total demand in Figure 7.7. For the specific problem scenario we are testing, the estimated safety factor is 0.75 by averaging the variance value of all the edges in the graph in a statistical sense. It is shown as a dot on the plot. The point marked by a square gives the best average unmet demand ratio, which is also the value we used for the CCP model for the comparison of Figure 7.6. It corresponds to the confidence level at 95% and the safety factor as 1.65.

Figure 7.7 shows that with an increasing safety factor starting from the deterministic point coverage is improved. However, if we are too conservative as further increasing the safety factor after the “best CCP point,” the quality of the result deteriorates. Under a given deadline, if the safety factor is too big such that all the short and risky arcs have been “stretched” to exceed the
deadline, then the route planning based on those “over-stretched” arcs is not very meaningful in providing any guidance since every node would miss its deadline. Hence it is crucial to plan using an “optimal” safety factor, which is neither too opportunistic nor too conservative. Robust optimization, which aims to optimize the worst-case scenario, also shares the same problem structure and computational complexity as the deterministic and the CCP formulations under a bounded uncertainty set or a given set of scenarios. Robust optimization seeks a solution that satisfies all possible constraint instances. It achieves this by enforcing that the strictest version of every constraint is satisfied by substituting the largest uncertainty parameter value in each constraint. This maximal value for the uncertainty parameter has a one-to-one correspondence with a safety factor value in the chance-constrained formulation and lies somewhere on the plot in Figure 7.7.

### 7.5 Conclusion

In this chapter, we decompose the overall transportation and distribution of supplies in response to a large-scale emergency into three interrelated
models: (1) Perishable Inventory Model; (2) Facility Location Model; and (3) Vehicle Routing Model. For the perishable inventory management system, a trivial extension of a regular EMQ model is adequate when the required minimum inventory is not significant compared with the amount consumed by the regular market demand rate within the shelf life since it can be timely and completely depleted and refreshed by the regular market demand. However, when we consider the VMI in the SNS for the large-scale emergency setting, the minimum inventory requirement is in a scale that is comparable with the total market consumption within the shelf life. A more sophisticated inventory management strategy is required to provide the fresh and massive stockpile throughout the time horizon in the system. Hence in this chapter, we modeled the perishable inventory management problem with a minimum inventory volume constraint as a modified economic manufacturing quantity (EMQ) model. We observed that at a given profitability level of the firm, there are trade-offs between the lower amount paid by the government to firms (by reducing the unit price $p_{gov}$) with the higher flexibility the government allows to firms (longer time before the expiration to salvage the pills). For the facility location problem, we show how the traditional models can be extended to incorporate redundancy in the system to provide better coverage. In experimental settings, we demonstrate how this modified facility location provides more redundant coverage than traditional models on a representative large-scale emergency scenario in Los Angeles County. For the routing problem, we proposed a chance constrained model to minimize the unmet demand for the planning stage. The influence of the safety factor for the planning stage model at different values is also illustrated by simulation results. We demonstrated that it is crucial to plan with an “optimal” safety factor.

Acknowledgment

This work was supported by the U.S. Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE). However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of Homeland Security.

REFERENCES


