

Contextual Areas

Designing Response Supply Chain Against Bioattacks

David Simchi-Levi,^a Nikolaos Trichakis,^b Peter Yun Zhang^c

^aDepartment of Civil and Environmental Engineering, Institute for Data, Systems, and Society, and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ^bOperations Research Center and Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ^cInstitute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contact: dslevi@mit.edu (DSL); ntrichakis@mit.edu,  <http://orcid.org/0000-0002-8324-9148> (NT); pyzhang@mit.edu,  <https://orcid.org/0000-0002-0422-834X> (PYZ)

Received: July 30, 2017

Revised: August 15, 2018; December 2, 2018

Accepted: March 8, 2019

Published Online in *Articles in Advance*:
August 23, 2019

Subject Classifications: inventory/production:
linear decision rules; military: logistics

Area of Review: Military and Homeland Security

<https://doi.org/10.1287/opre.2019.1862>

Copyright: © 2019 INFORMS

Abstract. We study a prescriptive model for end-to-end design of a supply chain for medical countermeasures (MCMs) to defend against bioattacks. We model the defender's MCMs inventory prepositioning and dispensing capacity installation decisions, attacker's move, and defender's adjustable shipment decisions so as to minimize inventory and life-loss costs subject to population survivability targets. We explicitly account for the strategic interaction between defender's and attacker's actions, assuming information transparency. We consider the affinely adjustable robust counterpart (AARC) to our problem, which enables us to deal with realistic networks comprising millions of nodes. We provide theoretical backing to the AARC performance by proving its optimality under certain conditions. We conduct a high-fidelity case study on the design of an MCMs supply chain with millions of nodes to guard against anthrax attacks in the United States. We calibrate our model using data from a wide variety of sources, including literature and field experiments. We produce policy insights that have been long sought after but elusive until now.

Funding: This work is supported by Ford Motor Company, The Boeing Company, and United Technologies Corporation.

Keywords: bioterrorism • supply chain design • robust optimization

1. Introduction

In this paper, we study the problem of designing a medical countermeasures (MCMs) supply chain in preparation against bioattacks, that is, the intentional release of pathogens or biotoxins against humans to cause serious illness and death. Bioattacks have been a rising first-order concern for many countries worldwide in the last 15 years; the United States alone, for instance, has poured approximately \$60 billion into biodefense preparedness since the 9/11 attacks (Kaiser 2011). Bioattacks are considered a major threat because a minute quantity of pathogens is sufficient to infect humans; furthermore, the appropriate MCMs need to be administered within a short time window to effectively reduce casualties. For example, if bioterrorists were to release *Bacillus anthracis*, that is, anthrax, over a large city, hundreds of thousands of people could be at risk, and MCMs would need to be administered to them within a few days to have the intended effect (Stroud et al. 2011, Nicholson et al. 2016). Distributing such large quantities of MCMs to the general public within such short time periods poses a considerable operational challenge, which is often compounded by attack-detection delays.

In order to be prepared to deliver adequate medication in a timely manner in response to a bioattack,

first, MCMs need to be stockpiled at appropriate locations. Some inventories can be stored in central locations to take advantage of deployment flexibility, pooling effects, and scale economies in holding costs, and others can be positioned closer to populous areas or even predisposed to the intended users in the form of home medical kits in order to reduce transportation time. Second, for end users to have prompt access to deployed MCMs, target areas need to have adequate dispensing capacity in place, for example, in the form of publicly accessible dispensing facilities or dedicated courier delivery services. Too little capacity would result in inefficient usage of available inventory and devastating delays; too much would result in wasted resources. Given candidate storage locations, where and how much inventory of MCMs should be prepositioned to cost-effectively defend against bioattacks? How much dispensing capacity should be installed at target areas? Our goal in this paper is to address these research questions and develop a decision-support framework that could guide policy design.

1.1. MCMs Supply Chain in the United States

The Centers for Disease Control and Prevention (CDC) maintains inventories of MCMs within the Strategic

National Stockpile (SNS), which is a critical component of the national health security programs managed by the CDC (Office of Public Health Preparedness and Response 2016a, b). The inventory storage and delivery infrastructure of the SNS forms a crucial line of defense for the nation's public health security.

For many bioattack threats, the SNS currently stores inventories of MCMs in central repositories, and its delivery infrastructure enables deployment of virtually any amount of stored MCMs to almost any major city in the United States within 12–36 hours at negligible shipping costs. But centrally stored inventories may not be responsive enough for events such as anthrax attacks. In recent years, the CDC has been increasingly aware of alternatives, namely the prepositioning of MCMs in locations closer to the population. Prepositioning has been raised as a first-order issue for the SNS, and the Office of Public Health Preparedness and Response (PHPR) policymakers have commissioned several studies to examine different strategies (Stroud et al. 2011, Nicholson et al. 2016). To date, several facets of the problem have been analyzed in isolation, including location and transportation options, cost of each system component, and health effects of delayed treatment.

However, a comprehensive cost–benefit analysis at the *system level* that *integrates* individual considerations into a practical quantitative decision-support tool has been elusive. Our work helps fill this gap, the importance of which is indeed highlighted in the following excerpt from a report prepared by the Board of Scientific Counselors for PHPR (Inglesby and Ellis 2012, p. 4):

Both analytic and simulation (experimental) modeling activities should be increased substantially. Modeling will allow [the Division of] SNS to make quantitatively-based decisions on how much inventory to hold and where to hold it. An end-to-end model capturing the flow of materials in the SNS, as well as costs and logistical and health measures, should begin at the SNS-managed inventory site and go all the way to the point of dispensing to the public. Using such models will reveal bottlenecks, provide cost estimates, and help SNS properly evaluate the costs and consequences of alternative Response Supply Chain configurations.

Here we focus on an end-to-end analytic model as described in the quote. The absence of such integrative models so far attests to the underlying technical challenges: (1) the multiplicity of decisions involved, both strategic and operational, (2) the subtle way prepositioning could influence bioattackers' actions, and (3) the scale of the problem. In particular, on the strategic level, policymakers need to decide how much inventory to store and how much capacity to install among thousands of locations in anticipation of an attack. On the operational level, in response to an

attack, appropriate shipment decisions need to be made to efficiently dispense MCMs based on the specific demand and supply conditions. The *static* inventory positioning and capacity installation decisions need to be made in conjunction with myriad *contingent* shipment policies. Furthermore, some information about planned responses is already available in the public domain, and prepositioning MCMs, particularly predisposing medical kits, is likely to grant to bioattackers even greater visibility to planned responses. Consequently, bioattackers, because they usually act in a preplanned way to inflict the greatest damage possible, are more likely to target underserved populated areas as opposed to ones that have access to abundant stockpiles. That attackers' actions are likely to be adversarial is well recognized by biodefense policymakers; see, for example, Stroud et al. (2011) and the intelligence and military operations research literature (Brown et al. 2006; Golany et al. 2009, 2012; Alderson et al. 2011; Kaplan 2012).

1.2. Our Methodology

We model the SNS as a network, inventory stockpiles as nodes, and shipment routes as edges. Bioattacks correspond to demand surges at some of the nodes. To model interactions between bioattackers' and policymakers' decisions, we adopt the defender–attacker–defender (DAD) model, which can be viewed as a *multistage* robust optimization (RO) model (Brown et al. 2006, Ben-Tal et al. 2009). Specifically, we consider prepositioning of MCMs as a first-stage static defender decision, bioattacks as second-stage adjustable attacker decisions, and shipments of MCMs as third-stage adjustable defender decisions. We calibrate the objective and uncertainty set of possible demand scenarios in ways that reflect policymakers' considerations. In particular, we deal with minimizing either inventory holding costs under survivability guarantees for all possible attack scenarios or worst-case inventory and health costs, in which each scenario comprises simultaneous attacks in multiple locations. To preclude overly pessimistic scenarios in which too many locations are attacked, our uncertainty sets include attack budget constraints, which are both standard and crucial in the DAD framework (Alderson et al. 2014).

Unfortunately, there are no scalable solution approaches that perform provably well even for the inventory prepositioning problem, let alone joint optimization of inventory and dispensing capacity decisions. In particular, DAD models and multistage RO problems of the type we study here are generally intractable (Brown et al. 2006, 2008; Ben-Tal et al. 2009; Alderson et al. 2014). There has been valuable effort in solving reasonably sized problems (Alderson et al. 2015), but they are orders of magnitude smaller

than what our problem entails. In the RO literature, a popular heuristic is the so-called affinely adjustable robust counterpart (AARC), which restricts adjustable decisions to be affine in the uncertainty variables (Ben-Tal et al. 2009). Although AARCs often lead to tractable formulations, they can perform rather poorly in high-dimensional problems (Bertsimas and Goyal 2012). Worse, all known performance guarantees for AARCs apply to problems with “symmetric” uncertainty sets. The aforementioned attack budget constraints that are necessary in our problem lead to sets that are the “least symmetric,” casting considerable doubt over the performance of AARCs in our setting (see our literature review in Section 1.4 and discussion in Section 3.2 for more details).

1.3. Our Contributions

First, for the inventory prepositioning problem, we provide an approximate solution approach with theoretical backing for its performance that can deal with network sizes on the order of millions of nodes. In particular, we consider the affinely adjustable robust counterpart, which reduces to a linear optimization problem (LP) by restricting the adjustable shipment policies to be affine in the demand shocks. We provide theoretical backing to its performance by proving that the AARC is, in fact, optimal under certain conditions. Our work contributes to the RO literature by proving optimality of the AARC in a new and important context (see our discussion in Sections 1.4 and 4). The numerical studies we conducted revealed that AARCs provide near-optimal performance for the inventory prepositioning problem under general settings as well.

Second, we consider the joint problem of inventory prepositioning and dispensing capacities optimization. We show how dispensing capacity, which is costly, can be captured as a first-stage decision and how the third-stage shipment decisions can be reformulated so as to reflect the underlying capacitated dispensing operations. We derive the AARC for this joint problem as an LP and conduct numerical studies to confirm that it still produces near-optimal solutions.

Third, successfully applying AARC to our problem speaks to the potential of its broader impact for other DAD problems. In particular, the AARC approach is applicable to trilevel programs that are nonlinear and with integer variables. In the sense of Alderson et al. (2014), the AARC is suitable for eight of the nine types of operator models (the only exception is a simulation-based operator model that does not have a closed-form description at all).

Fourth, we use our model in a thorough, large-scale case study of MCMs supply chain design for the SNS to defend against anthrax attacks. The tractability of

our approach enables us to perform the first study we are aware of in the literature that is at a realistic nationwide scale and deals with networks with millions of nodes. We calibrate our model using multiple literature sources that studied different facets of SNS design in isolation. Our integrated framework enables us to derive prescriptive recommendations to the CDC. For example, we find that if the CDC wanted to ensure 85% survivability for attacks that simultaneously affected (any) two states, each with at most two cities being attacked, with a detection time of under 48 hours, then the minimum required annual inventory and dispensing capacity budget would be about \$210 million; for a survivability target of 92%, the budget would increase to \$553 million. This increasing marginal cost phenomenon, along with other cost/policy implications, can be explored quantitatively within our model.

1.4. Literature Review

1.4.1. Resource Allocation for Military Applications. For many military defense problems such as ours, stochastic and priority-list approaches are inappropriate for capturing the strategic interaction between an intelligent defender and attacker (National Research Council 2008, Golany et al. 2009, Alderson et al. 2013). Therefore, it is desirable to use a sequential game approach, such as DAD. A main challenge for this approach is tractability. In the past decade, there has been concentrated and valuable effort in developing scalable solution approaches, especially for the interdiction type of problems (Brown et al. 2008; Atkinson 2009; Kaplan 2012; Alderson et al. 2013, 2015; Lazzaro 2016). Atkinson (2009) and Lazzaro (2016) show that the state-of-the-art algorithms can solve networks with thousands of nodes. Alderson et al. (2015) mention that for the significantly simpler bilevel games, problems with thousands of nodes can be readily solved by decomposition algorithms with commercial solvers.

We differ from the existing DAD papers in two ways. First, we deal with problem sizes on the order of *millions* of nodes and present for them scalable approximate solution approaches backed by performance guarantees. Such a drastic increase in problem size would render existing decomposition-based solution approaches insufficient. To deal with high dimensionality, we use the AARC heuristic from RO and contribute back to the theoretical RO literature by proving a novel performance guarantee of AARC (see related discussion). Second, we provide an integrated modeling of cost, health deprivation, and antibiotic efficacy over time for our application. This, in part, requires a precise analysis of time-expanded multicommodity network flow and how the AARC heuristic is analytically suited for it. We consider these

insights to be generalizable because multicommodity network flow models are considered essential primitives for DAD problems (Alderson et al. 2014).

Another set of military operations research papers (Berman and Gavious 2007, Zhuang and Bier 2007, Hao et al. 2009, Rios and Insua 2012) also study Stackelberg game problems. For example, Berman and Gavious (2007) study the defender–attacker problem as a min-max facility location problem. Zhuang and Bier (2007) study the difference between random and fully strategic attackers. Although they focus on equilibrium analysis, our paper formulates and solves a large-scale, high-fidelity prescriptive model to support decisions.

1.4.2. Biodefense. Existing biodefense literature does not focus on the interaction between defender and attacker. It generally assumes an abundance of inventories of MCMs and exogenous demand scenarios (Wein et al. 2003; Craft et al. 2005; Bravata et al. 2006; Lee et al. 2006a, b, 2009a, b; Hupert et al. 2009; King and Muckstadt 2009; King 2012). They study, for example, the optimal number and layout of points of dispense (Lee et al. 2006a, b, 2009a, b; Hupert et al. 2009), the effectiveness of predetermined exogenous inventory levels on the treatment of patients after bioattacks and influenza outbreaks (Bravata et al. 2006, King and Muckstadt 2009, King 2012), and a small number of response strategies for airborne anthrax attacks (Wein et al. 2003, Craft et al. 2005). Taking the supply of MCMs for granted and assuming attack scenarios to be known are limiting because, in practice, MCMs are shipped from different locations with different lead times (from hours to days) and shipment sizes, resulting in time-dependent flows, and terrorists' decisions are unknown in advance. In contrast, we take an integrated approach and study the entire supply chain within the DAD framework.

1.4.3. Inventory Prepositioning. Beyond the biodefense literature, other streams have studied prepositioning problems, including humanitarian logistics (Ergun et al. 2011, Caunhye et al. 2012, Çelik et al. 2012) and manufacturing (Simchi-Levi et al. 2014, 2015). The latter two papers provide a decision-support tool to evaluate performance under disruptive risks for a given inventory configuration and a given disruption scenario. Unlike the *descriptive* nature of these models, ours derives *prescriptive* suggestions that incorporate two stages of decision making.

Recent works by Uichanco (2015) and Simchi-Levi et al. (2018) study inventory prepositioning to hedge against capacity disruption and demand uncertainties. Uichanco (2015) provides a solution method that scales with the number of vertices in the polytope describing the network structure. Such a method

would lead to an intractable formulation in our case because a national biodefense network typically comprises millions of nodes and edges. In contrast, we provide a tractable model that can deal with problems of the size faced by U.S. policymakers. By incorporating dispensing capacity optimization, our model is more general too. In Simchi-Levi et al. (2018), the authors also focus on providing an exact solution method. Not surprisingly, this leads to an intractable formulation too, namely an LP with an exponential number of constraints. They propose a constraint-generation algorithm, which they find to work well in practice for bipartite graphs involving 200 nodes. Their constraint-generation algorithm still involves the solution of mixed-integer linear programs in its subroutines. Such approaches could not possibly be employed to tackle networks with millions of nodes such as ours. In contrast, our focus is to provide tractable and scalable (polynomial-time) algorithms, which require the solution only of linear optimization problems. Although our approach is not exact for general network topologies, we back it with optimality guarantees (as discussed).

1.4.4. Robust Network Flow. Atamtürk and Zhang (2007) study the computational complexity of two-stage robust network flow under budgeted uncertainty sets and find it to be intractable in general. For a tree structure, they use a dynamic programming (DP) approach to show that it can be solved in polynomial time. Unfortunately, their DP approach is not suitable for our problem. Because they are developed within a constructive proof to show that robust tree network flow problems are polynomially solvable, the DP approach breaks down and cannot be applied for general, non-tree network flow problems. Furthermore, it is also tailored to deal only with uncapacitated network flow problems, and it is unclear if it can be applied to tackle dispensing capacity issues. From a practical standpoint, because the SNS network need not be a tree, and because dispensing capacity is a major concern (Stroud et al. 2011), the DP approach by Atamtürk and Zhang (2007) cannot be applied to tackle the SNS design problem we study in our work. Atamtürk and Zhang also provided a cutting-plane algorithm that can be applied to solve general (capacitated) networks, but this exact-solution algorithm has no polynomial-time guarantee and is not practical for large instances, for example, the ones we solve in Section 6. Our goal is different. Instead of focusing on computational complexity, we study a tractable solution method for general networks involving a rich set of decisions (inventory, capacity, shipment, and dispensing over time) and provide strong theoretical and numerical backing for its performance. Our solution method

enables us to apply it to a practically relevant policy design problem, demonstrated in a case study involving millions of nodes and considering important practical decisions and constraints.

1.4.5. Affinely Adjustable Robust Optimization. An introduction to this topic is included in Ben-Tal et al. (2009). There have only been a handful of results that show AARC heuristics to be optimal under special circumstances (Bertsimas et al. 2010, 2011; Iancu et al. 2013; Ardestani-Jaafari and Delage 2016). Our work contributes to this stream by providing an optimality proof under a new context. Bertsimas et al. (2010) studies one-dimensional problems, for example, inventory problems that involve a single stocking level, and in the work of Ardestani-Jaafari and Delage (2016), the authors deal with newsvendor-type problems under demand uncertainty. We differ from these papers by studying multidimensional problems, that is, shipment decisions in a supply chain network. The papers by Iancu et al. (2013) and Bertsimas et al. (2011) deal with families of problems, and for AARC optimality (or “good” performance), they require the uncertainty sets to be lattices or to possess a “symmetric” structure. In our model, in order to preclude overly conservative solutions, we rely on simplex-type budget constraints that invalidate both the lattice and the symmetric structures. Put differently, we prove optimality for a problem that is according to Bertsimas et al. (2011) in some sense the “least conducive” to the AARC optimality conditions set forth in the literature so far.

2. The Inventory Prepositioning Problem

We begin by considering the inventory prepositioning problem for MCMs, assuming for now that sufficient dispensing capacity is installed. For exposition purposes, we use the SNS to exemplify our discussion, but our model and solution heuristic generalize to any network. The SNS network is designed to protect the public susceptible to bioattacks in a set of geographic locations, for example, densely populated towns or neighborhoods, which we shall refer to as *demand* locations. The demand locations are split into administrative divisions according to a hierarchical structure, for example, boroughs, municipalities, provinces, states, etc. There are $L + 1$ levels in the hierarchy, indexed by $\ell = 0, \dots, L$, whereby each division at the lowest (L th) level comprises precisely one demand location. Divisions at some intermediate ℓ th level comprise subsets of divisions at the lower ($\ell + 1$)th level in a nested fashion; the highest (0th) level includes a single division, that is, federal/national level. For a division i at the ℓ th level, let $\mathcal{P}(i)$ be its parent division at the $(\ell - 1)$ th level, $\mathcal{C}(i)$ the set of (children) divisions that it includes at the $(\ell + 1)$ th

level, and $\mathcal{D}(i)$ the set of demand locations that it includes; that is,

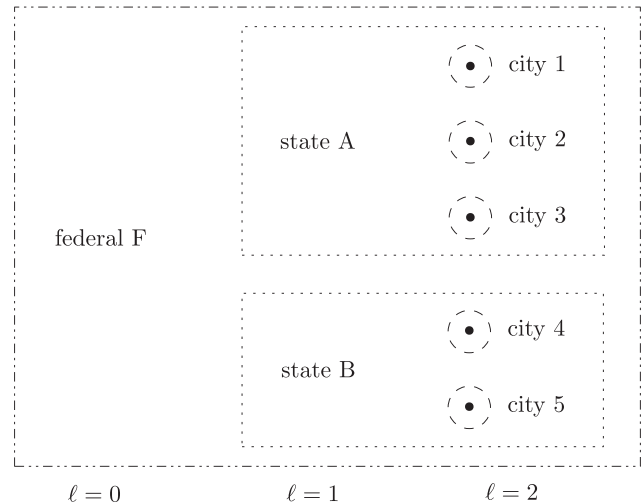
$$\mathcal{D}(i) = \left\{ \text{demand location } j : \underbrace{\mathcal{P}(\mathcal{P}(\dots \mathcal{P}(\mathcal{P}(j)) \dots))}_{L-\ell \text{ times}} = i \right\}.$$

Figure 1 includes an illustrative example in which the demand locations are cities, split under state and federal divisions.

To serve the demand locations, each administrative division at the $0, \dots, L - 1$ levels maintains a *stockpile* of appropriate medical countermeasures. In particular, in case of a bioattack, MCMs would typically need to be shipped from the—usually remotely located—stockpiles to the demand locations, where they are subsequently dispensed to the affected population.

Let the demand locations and stockpiles correspond to nodes in a directed graph (V, E) , which we refer to as *demand* and *stockpile* nodes, respectively. The subset of demand nodes is denoted with V_D . We index the root node (stockpile node at the highest level division) with zero. The edges in E connect stockpile nodes with demand nodes. Specifically, $(i, j) \in E$ if and only if inventory can be shipped from i to $j \in V_D$ once an attack occurs. Shipments have negligible costs and are not subject to capacity constraints for the purposes of the SNS network.¹ However, there is a fixed lead time τ_{ij} for shipping inventory from i to j for all $(i, j) \in E$. Note that even though a “hierarchical inventory network” such as the SNS resembles a tree structure and the demand and stockpile node sets are usually disjoint, our model

Figure 1. Five Demand Locations (Cities) Split into Three Administrative Divisions: Federal $\ell = 0$, State $\ell = 1$, and City $\ell = 2$



Note. Note that $\mathcal{P}(1) = \mathcal{P}(2) = \mathcal{P}(3) = A$ and $\mathcal{P}(4) = \mathcal{P}(5) = B$; $\mathcal{C}(A) = \mathcal{D}(A) = \{1, 2, 3\}$ and $\mathcal{C}(B) = \mathcal{D}(B) = \{4, 5\}$; $\mathcal{D}(F) = \{1, 2, 3, 4, 5\}$.

and solution approach are more general and can be applied to any network (V, E) , accommodating demand and stockpile on any node.

We focus on a particular bioattack threat, in anticipation of which a single MCM type is stored. Inventory of this MCM can be prepositioned at any node $i \in V$. If i is a demand node, the stored inventory corresponds to predisposed medical kits. Let $\mathbf{x} \in \mathbb{R}^{|V|}$ denote the amounts of stored inventories allocated at the nodes. The set of all feasible inventory allocations $X \subseteq \mathbb{R}^{|V|}$ is assumed to be a polytope and could include inventory-related constraints, for example, nonnegativity constraints and budget constraints subject to inventory setup cost and variable cost. A unit of inventory is adequate to treat precisely one individual. Associated with storing inventory at any node are purchasing, replenishment, and maintenance costs, etc. We refer to these costs simply as *holding* costs and denote the per-unit costs with $\mathbf{h} \in \mathbb{R}^{|V|}$. That is, storing x_i MCM units at node $i \in V$ costs $h_i x_i$.

When an attack takes place, a subset of demand nodes is affected, and parts of their populations are in need of treatment with the MCMs. We also say that division i is affected if any of its children demand nodes $\mathcal{D}(i)$ are affected. For example, if city 1 is attacked in Figure 1, we say that it is affected and so is state A. Let $\mathbf{d} \in \mathbb{R}^{|V_D|}$ be the (a priori unknown) vector with the realized number of affected individuals in each of the demand nodes. Once \mathbf{d} is revealed, inventory stored at the affected demand nodes is available immediately. In case of shortages, inventory from other stockpile nodes can be shipped to satisfy the demands as long as edges exist between them. For each edge $(i, j) \in E$, let f_{ij} denote the amount of inventory shipped from i to satisfy demand at node j . Note that shipment decisions are contingent on the realized demand, which we sometimes make explicit by writing them as $f_{ij}(\mathbf{d})$. Inventory shipped from i would be made available for treatment at node j only after τ_{ij} time units. Delays in treatment of the affected population could lead to lower probability of survival. We use $\rho_{ij} \in [0, 1]$ to denote the *survival probability* (or *survivability*) of an individual in node j if treated with inventory shipped from node i and $\bar{\rho} \in [0, 1]$ if left untreated. For example, suppose that node j is affected, and 50% of the affected individuals are treated with inventory prepositioned at that node, 30% are treated with inventory shipped from node i , and 20% are left untreated. The *average* survivability would then be $0.5\rho_{ij} + 0.3\rho_{ij} + 0.2\bar{\rho}$.

With respect to the possible attack scenarios, the CDC considers ones involving simultaneous attacks to multiple locations across the country (Inglesby and Ellis 2012). In view of this, we consider attack scenarios (or demand vectors) in which, for each administrative division i , at most Γ_i of its children divisions $\mathcal{C}(i)$ are

affected. We assume Γ_i to be nonnegative integers and refer to them as the *attack scale* parameters. For example, if the administration levels are states and cities, $\Gamma_i = 2$ for state i would mean that we consider attack scenarios with at most two cities in state i being affected.

In case demand location i is affected, the maximum number of individuals in need for treatment at that location is \hat{d}_i .² A further probabilistic characterization of the possible demand vectors \mathbf{d} appears to be, unfortunately, prohibitive for a variety of reasons. One relates to limited historical data, given that bioattacks have so far been rarely encountered in practice. A second, and more important, reason is the nature of terrorist attacks being *adversarial* and *endogenous* instead of purely random and exogenous. In particular, unlike natural disasters, for instance, bioattacks are premeditated and often carefully planned in order to maximize the damage inflicted. The choice of which locations to attack, then, which essentially drives the realization of \mathbf{d} , is likely to be influenced by the inventory decisions given that prepositioned inventory (or lack thereof) is partially visible to the public. Both issues of endogeneity and the adversarial choice of the attacked areas are indeed considered and acknowledged by policymakers (Stroud et al. 2011) and academics studying the DAD framework (Brown et al. 2008; Alderson et al. 2011, 2013, 2015; Kaplan 2012; Lazzaro 2016; Chan et al. 2018). In the next section, we consider a suitable model for \mathbf{d} that addresses these issues instead of postulating a probabilistic description.

In general, more (less) inventory in the system leads, on the one hand, to higher (lower) holding costs and, on the other hand, to fewer (more) potential casualties. Therefore, the CDC needs to balance inventory costs and the costs of life loss. We consider two possible ways to navigate this trade-off, leading to two problem formulations:

1. Policymakers explicitly quantify the cost of life lost to be b monetary units. The problem is then to select an inventory allocation so as to minimize holding costs plus worst-case life-loss costs under all possible attack scenarios. We refer to this formulation as *life-loss cost* (LLC).

2. Policymakers specify survivability targets. Let $1 - \epsilon_i$ be the target for average survivability in demand node i for some $\epsilon_i \in [0, 1]$, which we refer to as the *survivability target* parameters. The problem is then to select an inventory allocation so as to minimize holding costs while providing the survivability guarantees implied by the specified targets in all demand locations under all possible attack scenarios. We refer to this formulation as *life-loss guarantee* (LLG).

Our model is well suited for policy decision-making purposes. In particular, the attack scale parameters $\{\Gamma_i : i \in V\}$ enable policymakers to specify the severity

of attacks they want to hedge against, in terms of both *magnitude*, because higher values of Γ for lower-level nodes translate into more areas affected (e.g., cities), and *complexity*, because higher values of Γ for higher-level nodes translate into the affected areas being dispersed among more divisions (e.g., states). Furthermore, the life-loss cost b in the (LLC) formulation and the survivability target parameters $\{\epsilon_i : i \in V_D\}$ in the (LLG) formulation reflect the policymakers' aversion to casualties and insufficient coverage. Based on the policymakers' selections, the model prescribes appropriate inventory prepositioning strategies and elicits their minimum required costs, which allows for trade-off analyses.

3. Formulation and Solution Approach

For both (LLC) and (LLG), we formulate the inventory prepositioning problem as a DAD problem or multistage robust optimization problem. In particular, at the first stage, the defender decides on the prepositioned inventory allocation $\mathbf{x} \in X$. At the second stage, the attacker chooses an attack scenario or demand vector \mathbf{d} from the set of scenarios compatible with the specified attack-scale parameters, denoted by $U \subset \mathbb{R}^{|V_D|}$, so as to minimize expected survivors. We also refer to U as the *demand uncertainty set*. At the third stage, after the realized demand vector $\mathbf{d} \in U$ is revealed, the defender decides on the shipment strategy $\{f_{ij}(\mathbf{d}) : (i, j) \in E\}$. Formally, we model (LLC) as

$$\begin{aligned}
 \text{(LLC): } & \min_{\mathbf{x}, \mathbf{f}(\cdot), \mathbf{s}(\cdot)} \mathbf{h}^T \mathbf{x} \\
 & + \max_{\mathbf{d} \in U} b \sum_{i \in V_D} \left((1 - \bar{\rho}) s_i(\mathbf{d}) + \sum_{j: (j,i) \in E} (1 - \rho_{ji}) f_{ji}(\mathbf{d}) \right) \\
 \text{subject to } & x_i \geq \sum_{j: (i,j) \in E} f_{ij}(\mathbf{d}), \quad \forall i \in V, \forall \mathbf{d} \in U, \quad (1b) \\
 & s_i(\mathbf{d}) + \sum_{j: (j,i) \in E} f_{ji}(\mathbf{d}) = d_i, \\
 & \quad \forall i \in V_D, \forall \mathbf{d} \in U, \quad (1c) \\
 & s_i(\mathbf{d}) \geq 0, \quad \forall i \in V_D, \forall \mathbf{d} \in U, \quad (1d) \\
 & f_{ij}(\mathbf{d}) \geq 0, \quad \forall (i, j) \in E, \forall \mathbf{d} \in U, \quad (1e) \\
 & \mathbf{x} \in X. \quad (1f)
 \end{aligned}$$

The auxiliary variables $\{s_i(\cdot) : i \in V_D\}$ capture demand shortages, that is, the number of affected individuals left untreated at each node. Note that both the shipment decision variables $\mathbf{f}(\cdot)$ and the shortage variables $\mathbf{s}(\cdot)$ are adjustable, contingent on the realized demand vector \mathbf{d} . The objective is to minimize inventory holding costs plus worst-case life-loss costs. Constraint (1b) is a node capacity constraint: the total amount of inventory shipped from node i should be less than the amount stockpiled at i . Constraint (1c) defines the

demand shortage variable s_i for each node i , and d_i is the i th element of \mathbf{d} .

Similarly, we formulate (LLG) as follows:

$$\begin{aligned}
 \text{(LLG): } & \min_{\mathbf{x}, \mathbf{f}(\cdot), \mathbf{s}(\cdot)} \mathbf{h}^T \mathbf{x} \quad (2a) \\
 \text{subject to } & \bar{\rho} s_i(\mathbf{d}) + \sum_{j: (j,i) \in E} \rho_{ji} f_{ji}(\mathbf{d}) \geq (1 - \epsilon_i) d_i, \\
 & \quad \forall i \in V_D, \forall \mathbf{d} \in U, \quad (2b) \\
 & (1b), (1c), (1d), (1e), (1f). \quad (2c)
 \end{aligned}$$

Compared with (LLC), the (LLG) formulation has an objective of minimizing holding costs while ensuring that the average survivability at each node is higher than the target set as reflected by the added constraint (2b). Throughout this paper, we assume that $\rho_{ji} \geq \bar{\rho}$ for all $(j, i) \in E$ (i.e., treatment increases survivability) and $\bar{\rho} \leq 1 - \epsilon_i$ for all $i \in V_D$; otherwise, (LLG) reduces to a trivial problem with optimal cost of zero. Note that we impose a survivability target on a node level instead of in aggregation; this provides better control and could ensure equitable population protection across regions. We note later in Theorem 1 that (LLG) is a special case of (LLC).

3.1. Set of Attack Scenarios

We provide a formulation for the set of attack scenarios under consideration. In particular, recall that given some attack-scale parameters $\{\Gamma_i : i \in V\}$, a possible scenario involves demand nodes affected so that, for each administrative division i , no more than Γ_i of its children divisions $\mathcal{C}(i)$ are affected. Consider

$$\begin{aligned}
 U := \left\{ \mathbf{d} \in \mathbb{R}^{|V_D|} : \xi \in \mathbb{R}^{|V|}, d_i = \hat{d}_i \xi_i \quad \forall i \in V_D, 0 \leq \xi_i \right. \\
 \left. \leq \xi_{\mathcal{P}(i)} \quad \forall i \in V \setminus \{0\}, \xi_0 = 1, \sum_{j \in \mathcal{C}(i)} \xi_j \leq \Gamma_i \xi_i \quad \forall i \in V \right\},
 \end{aligned}$$

where ξ_i indicates whether location i is attacked and \hat{d}_i is the maximum number of people affected when location i is attacked. As we show, only the extreme points of U are relevant for the optimal solution of (LLC) and (LLG). Therefore, to argue that the polytope U represents the set of attack scenarios in which we are interested, it suffices to show that its extreme points precisely correspond to the attack scenarios under consideration. Indeed, as we also show, the auxiliary variables ξ are binary at the extreme points of U . The constraint $\xi_i \leq \xi_{\mathcal{P}(i)}$ enforces all parent divisions of an affected node to also be affected, that is, $\xi_j = 1$ for all j such that $i \in \mathcal{D}(j)$ and i is affected. Finally, constraint $\sum_{j \in \mathcal{C}(i)} \xi_j \leq \Gamma_i \xi_i$ ensures that, if i is affected, at most Γ_i of its children divisions are affected as we required. As a side note, it can be readily seen that the set U also includes all remaining attack

scenarios that are compatible with the attack-scale parameters but have a number of affected individuals $d_i \leq \hat{d}_i$. To facilitate exposition, we also let

$$\Xi := \left\{ \xi \in \mathbb{R}^{|V|} : 0 \leq \xi_i \leq \xi_{\mathcal{D}(i)} \quad \forall i \in V \setminus \{0\}, \xi_0 = 1, \right. \\ \left. \sum_{j \in \mathcal{L}(i)} \xi_j \leq \Gamma_i \xi_i \quad \forall i \in V \right\}.$$

We can then simply express $U = \{\mathbf{d} \in \mathbb{R}^{|V_D|} : d_i = \hat{d}_i \xi_i \forall i \in V_D, \xi \in \Xi\}$.

3.2. Solution Approach

By involving both static \mathbf{x} and recourse decisions $\mathbf{f}(\cdot)$ and $\mathbf{s}(\cdot)$, formulations (LLC) and (LLG) fall into the class of so-called multistage adjustable robust optimization (ARO) problems. Problems in this class are, in general, computationally intractable (Ben-Tal et al. 2009) because they require the optimization over functions or policies instead of vectors, and this makes them infinite-dimensional problems. Specifically, $\mathbf{f}(\cdot)$ and $\mathbf{s}(\cdot)$ are policies that could take different values contingent on the uncertain parameters' realization (namely the demand \mathbf{d}).

Most popular techniques to deal with ARO problems in the literature are heuristics and constraint-/column-generation methods. One popular heuristic, which we also adopt to tackle our problem, is to limit attention to policies restricted to depend affinely on the uncertain parameters, which are often referred to as *affine policies* (APs) as opposed to *fully adjustable policies* (FPs). This restriction often enables tractability (see, e.g., Ben-Tal et al. 2009). In our setting, the affinely adjustable robust counterpart of (LLG), for example, is

$$(\text{ALLG}): \min_{\mathbf{x}, \mathbf{F}, \mathbf{S}} \quad \mathbf{h}^T \mathbf{x} \quad (3a)$$

$$\text{subject to} \quad \bar{\rho}(\mathbf{S}_i^T \mathbf{d} + S_i^0) + \sum_{j:(j,i) \in E} \rho_{ji}(\mathbf{F}_{ji}^T \mathbf{d} + F_{ji}^0) \\ \geq (1 - \epsilon_i) d_i, \quad \forall i \in V_D, \forall \mathbf{d} \in U, \quad (3b)$$

$$x_i \geq \sum_{j:(i,j) \in E} (\mathbf{F}_{ij}^T \mathbf{d} + F_{ij}^0), \quad \forall i \in V, \forall \mathbf{d} \in U, \quad (3c)$$

$$(\mathbf{S}_i^T \mathbf{d} + S_i^0) + \sum_{j:(j,i) \in E} (\mathbf{F}_{ji}^T \mathbf{d} + F_{ji}^0) = d_i, \\ \forall i \in V_D, \forall \mathbf{d} \in U, \quad (3d)$$

$$\sum_{j:(j,i) \in E} (\mathbf{F}_{ji}^T \mathbf{d} + F_{ji}^0) \leq d_i, \quad \forall i \in V_D, \forall \mathbf{d} \in U, \quad (3e)$$

$$\mathbf{F}_{ji}^T \mathbf{d} + F_{ji}^0 \geq 0, \quad \forall (i,j) \in E, \forall \mathbf{d} \in U, \quad (3f)$$

$$\mathbf{x} \in X, \quad (3g)$$

where $\mathbf{F}_{ij} \in \mathbb{R}^{|V_D|}$, $F_{ij}^0 \in \mathbb{R}$ for all $(i,j) \in E$ and $\mathbf{S}_i \in \mathbb{R}^{|V_D|}$, $S_i^0 \in \mathbb{R}$ for all $i \in V_D$ are vectors of decision variables corresponding to the affine policies' coefficients. The affinely adjustable version of (LLC) is of similar form. We can reformulate (ALLC) and (ALLG) using standard robust optimization techniques (see corollary 1.3.5. in Ben-Tal et al. (2009)) to obtain linear optimization problems. Importantly, the resulting linear optimization problems are polynomial in size of the original inputs, enabling tractability and scalability.

On the flipside, tractability of the AP heuristic to deal with ARO problems often comes at the cost of suboptimal solutions. In fact, for some AROs, the suboptimality gap between the objective value under APs and the objective value of the original formulation under FPs can grow indefinitely with the dimension of the problem (Bertsimas and Goyal 2012). As we pointed out in the literature review, a handful of papers have recently identified conditions under which APs are indeed optimal. These conditions require the absence of simplex-type constraints in the uncertainty set. In our model, however, the simplex-type constraints $\sum_j \xi_j \leq \Gamma_i \xi_i$ are essential to preclude excessively conservative demand scenarios, in which an arbitrary number of nodes are affected. Consequently, given the current state of affairs in the robust optimization literature, the performance of AP heuristics (ALLC) and (ALLG) remains questionable. The next section is devoted to providing evidence that (ALLC) and (ALLG) are indeed likely to produce near-optimal solutions for the original (LLC) and (LLG) formulations.

4. Optimality of Affine Shipment Policies

We provide evidence that affine policies are near optimal for our problem formulations (LLC) and (LLG). First, and more important, we analytically show that under additional assumptions (on the survivability parameters and the network structure), APs are indeed optimal. Second, we conduct numerical studies illustrating that the suboptimality gap remains small for instances that violate the additional assumptions guaranteeing optimality.

4.1. Optimality Result

Consider the following assumptions.

Assumption 1. *The survival probabilities under treatment are all equal; that is, $\rho_{ij} = \rho \quad \forall (i,j) \in E$.*

Assumption 2. *Stockpiles serve all demand nodes in their division and only demand nodes in their division; that is, $(i,j) \in E \iff j \in \mathcal{D}(i)$.*

Assumption 1 requires that the difference in shipment times between stockpile and demand nodes bears no effect on the survivability of treated individuals.

Note that this assumption is violated for biotreats with incubation periods shorter than the shipment times, for example, for nerve agents that require treatment within minutes or hours after an attack (Stroud et al. 2011). However, it is satisfied for biotreats with longer incubation periods, for example, for anthrax attacks that are detected early (see also our discussion in Section 6). Assumption 2 requires that stockpiles, which are maintained by administrative divisions, serve affected locations only within their division. In the United States, this would mean that the stockpiles maintained by states are reserved for usage by their residents. In emergency situations, however, neighboring states, for example, could also provide assistance, and this assumption might not hold; we explore this further in the next section.

Let z_{LLC}^* , z_{LLG}^* , z_{ALLC}^* , and z_{ALLG}^* be the optimal values of formulations (LLC), (LLG), (ALLC), and (ALLG), respectively. We have the following result.

Theorem 1. *Under Assumptions 1 and 2, affine policies are optimal for (LLG); that is,*

$$z_{LLG}^* = z_{ALLG}^*.$$

Furthermore, if $\rho = 1$, affine policies are optimal for (LLC); that is,

$$z_{LLC}^* = z_{ALLC}^*.$$

To illustrate the applicability of our methodology beyond the scope of inventory prepositioning for MCMs, we prove our result in a more general model. In particular, for any node, we consider shipments to any of its children nodes, not just its demand (leaf) nodes. Furthermore, we allow demand to occur at any node in V , not just the (leaf) nodes in V_D . We formalize this generalization within the proof of Theorem 1, which can be found in Appendix A, and show that it subsumes problems (LLC) and (LLG) as special cases. We discuss alternative applications in Section 7.

4.2. Performance of APs for the General Case

To quantify the performance of the AP heuristic for more general cases, we conducted two numerical studies in which we sequentially relaxed the optimality-guaranteeing Assumptions 1 and 2. For brevity, we

present our studies for the (LLC) formulation; our studies on the (LLG) formulation yielded quantitatively similar results.

We measured the AP heuristic's performance via its suboptimality gap, defined as the relative difference between the heuristic's optimal cost z_{ALLC}^* and the true optimal cost z_{LLC}^* , that is, $(z_{ALLC}^* - z_{LLC}^*)/z_{LLC}^*$. Because the underlying problem is intractable, it is impossible to compute z_{LLC}^* for large instances. To quantify the gap, then, we rely on the method in Simchi-Levi et al. (2019) to generate a valid lower bound on z_{LLC}^* . Using this bound instead of the true optimal cost enables us to obtain an *upper bound* on the suboptimality gap. Despite such conservativeness, we found the performance of the AP heuristic to be strong across a wide range of parameter values: median suboptimality gaps were less than 1.5%, and importantly, they did not grow with the problem's scale.

In study 1, we relax Assumption 1 and allow the survival probabilities to vary. At a high level, we created 5,000 instances of (LLC). For each instance, we generated a random tree-style graph with varying numbers of nodes. Inventory cost, antibiotic efficacy, and attack-scale parameters were also randomly sampled; see Table 1 for the sampling ranges and Appendix B for more details.

In study 2, we relax Assumption 2 and introduce edges that violate the tree structure. In particular, we considered the same setup as in study 1, and then, for every pair of nodes that were not already connected and resided in adjacent levels in the graph, we added an edge between them with some probability p_{arc} so as to obtain "nontree" graphs. We sampled p_{arc} in $[0, 0.01]$, resulting in graphs that had up to approximately 200% more edges than the tree-style graphs in study 1.

Using a high-performance computing cluster, we ran the AP heuristic and the exact solution method in Simchi-Levi et al. (2019) for each instance. Each run was given a 4-hour time limit, 16 GB of memory, and two central processing units (2.1 GHz each). Across both studies 1 and 2, the AP heuristic successfully solved all 10,000 instances with an average run time of 6 seconds. Although the exact solution method managed to solve only 2% of the instances, it generated

Table 1. Parameter Setup for Study 1: b Is the Demand Loss Cost, and $|C_i|$ Is the Number of Children for Node i

Parameter	Sampling range
Approximate number of nodes	{100, 200, 500, 1,000, 2,000, 5,000, 10,000}
Number of levels	{2, 3, 4, 5}
Inventory cost	$[0, 0.2] \times b$
Efficacy	$[0, 1]$
Attack scale	$[0, 0.2] \times C_i $ for each node i
Demand (\hat{d}_i)	$[0, 1,000]$ for each node i

valid lower bounds for 91% of them with an average run time of 14,116 seconds. The generated lower bounds were then used for our comparisons.

4.2.1. Results. Table 2 provides statistics on the upper bounds of the suboptimality gaps we obtained by using the generated valid lower bounds on the optimal costs. Qualitatively, our analysis reveals that the suboptimality gap is small and does not have strong correlation with any of the parameters we varied, including, among others, the number of nodes and the attack severity. In particular, the studies suggest that the suboptimality gap does not grow with the network size. This is a key consideration given that we cannot tractably characterize the gap for even larger networks because of limitations of the exact solution method. More details are included in Appendix B.

5. Optimizing Dispensing Capacity

In this section, we relax the assumption of sufficient dispensing capacity being preinstalled and jointly optimize inventory prepositioning and dispensing capacity decisions. By *dispensing capacity*, we refer to the ability of the appropriate local authorities to distribute MCMs to the general public at demand nodes subsequent to a bioattack—technically, it corresponds to the delivery rate of MCMs from so-called receive, stage, and store facilities to end users. Such joint optimization is essential: too much inventory and not enough dispensing capacity would result in inventory buildup while the population suffers from dispensing delays; too little inventory and too much capacity would result in idle workers and underutilized resources.

Limited dispensing capacity could introduce further delays before the affected population has access to MCMs, which, in turn, could affect their efficacy. Unlike delays resulting from transportation lead times, which are independent of the quantity of MCMs shipped, delays resulting from limited dispensing capacity are dependent on the quantity of MCMs being dispensed. This creates the need to introduce a time dimension explicitly in our model.

We illustrate our approach using the (LLG) formulation. Subsequent to an attack, consider T discrete time periods, each with duration δ , indexed by $t = 1, 2, \dots, T$. Let u_j be the dispensing capacity in MCM units per time period at demand node j . In view of this capacity, a shipment from some node i ,

although it would still arrive at j after τ_{ij} time periods, might not be immediately available to the population in its entirety. To capture this, let f_{ijt} be the amount of MCMs shipped from i and dispensed at j at time t . To measure antibiotic efficacy, let ρ_t be the survival probability if treatment is received t time periods after the attack. We use δ small enough so that the probability of survival remains approximately constant and equal to ρ_t within time period $[(t-1)\delta, t\delta]$.

Let c_j be the cost per unit capacity installation (e.g., training of staff, preparation and maintenance of dispensing facilities), and let p_j be the cost per unit of MCMs dispensed after an attack. The joint inventory prepositioning and dispensing capacity optimization problem can be cast as

$$(LLGC): \min_{\mathbf{x}, \mathbf{u}, \mathbf{f}(\cdot), \mathbf{s}(\cdot)} \max_{\mathbf{d} \in U} \mathbf{h}^T \mathbf{x} + \mathbf{c}^T \mathbf{u} + \mathbf{p}^T \mathbf{f}(\mathbf{d}) \quad (4a)$$

subject to

$$x_i \geq \sum_{j:(i,j) \in E} \sum_{t=1}^T f_{ijt}(\mathbf{d}), \quad \forall i \in V, \forall \mathbf{d} \in U, \quad (4b)$$

$$s_j(\mathbf{d}) + \sum_{i:(i,j) \in E} \sum_{t=1}^T f_{ijt}(\mathbf{d}) = d_j, \quad \forall j \in V_D, \forall \mathbf{d} \in U, \quad (4c)$$

$$\bar{\rho} s_j(\mathbf{d}) + \sum_{i:(i,j) \in E} \sum_{t=1}^T \rho_t f_{ijt}(\mathbf{d}) \geq (1 - \epsilon_j) d_j, \quad \forall j \in V_D, \forall \mathbf{d} \in U, \quad (4d)$$

$$\sum_{i:(i,j) \in E} f_{ijt}(\mathbf{d}) \leq u_j, \quad \forall j \in V_D, \forall \mathbf{d} \in U, \forall t \in [T], \quad (4e)$$

$$f_{ijt}(\mathbf{d}) = 0, \quad \forall (i, j) \in E, \forall \mathbf{d} \in U, \forall t \in [\tau_{ij} - 1], \quad (4f)$$

$$f_{ij}(\mathbf{d}) \geq 0, \quad \forall (i, j) \in E, \forall \mathbf{d} \in U, \quad (4g)$$

$$s_i(\mathbf{d}) \geq 0, \quad \forall i \in V_D, \mathbf{d} \in U, \quad (4h)$$

$$\mathbf{x} \in X, \mathbf{u} \geq 0. \quad (4i)$$

Constraint (4e) ensures that no more than u_j MCM units are dispensed at demand node j across all origin nodes i at each time period t . Constraint (4f) ensures that no MCMs shipped from node i can be dispensed at node j before the associated shipment lead time τ_{ij} . All other constraints are similar in spirit to the ones we have in (LLG).

By restricting the flow and demand shortage variables to be affine in demand vector \mathbf{d} , the robust counterpart of (LLGC) remains a linear optimization problem. Thus, we are able to solve large problem instances that match the fidelity and scale of a national bio-defense network.

Table 2. Statistics of Upper Bounds on Suboptimality Gaps for AP Heuristic for Studies 1 and 2

Minimum	First quantile	Median	Mean	Third quantile	Maximum
0.0%	0.4%	1.4%	1.9%	3.0%	9.9%

Table 3. Input/Output for Our Case Study

Input	Output
Survivability target: ϵ_i	Total cost (inventory + capacity + dispensing)
Attack scale: Γ_i	Optimal inventory prepositioning strategy: \mathbf{x}^*
Detection time: τ_0	Optimal capacity installation: \mathbf{u}^*

Finally, note that introducing a time dimension here can be thought of as simply introducing additional edges in the set E . Therefore, beyond adding capacity variables and accounting for costs, the only essential way that (LLGC) differs from (LLG) is the inclusion of capacity constraints (4e). In Appendix B, we present a numerical study similar to the ones in Section 4.2, in which we quantify the AP heuristic's suboptimality gap for problems including such capacity variables, costs, and constraints. As before, we find performance to be strong with a median suboptimality gap less than 1%.

6. Strategic National Stockpile Supply Chain

In this section, we apply our work in a case study on the SNS design for aerosolized *Bacillus anthracis* (anthrax) attacks in the United States, which are of particular interest to public health experts owing to their relatively high probability of occurrence and potentially devastating impact (Stroud et al. 2011). Anthrax spores spread easily in the air to affect a large number of people through inhalation with high fatality rates if left untreated for even just a few days. More precisely, if t is the time between infection and treatment, then, according to the studies of Brookmeyer (2001), Brookmeyer et al. (2005), and Wilkening (2006, 2008), the survival probability $\rho(t)$ of the anthrax-infected population can be approximated, for $t \leq 200$ hours, by

$$\rho(t) = e^{-(0.004t)^2}. \quad (5)$$

Considering the time it would take to detect an anthrax attack and the time to ship and dispense MCMs from warehouses to individuals, which is likely to be on the order of days (Stroud et al. 2011), the CDC is exploring different options in prepositioning MCMs and coordinating the SNS network to achieve better antibiotic efficacy. We now discuss how to calibrate our model to tackle these problems.

6.1. Model Calibration

For the purposes of this study, we consider the (LLGC) formulation; that is, policymakers specify coverage targets, and the model optimizes the required inventory, capacity, and dispensing costs.

6.1.1. Network. There are 12 federally managed SNS warehouses for storage of MCMs, with their exact locations being classified for security reasons. It is known that the locations have been chosen such that MCMs can be transported from them to any state within 24–36 hours of a deployment call (Office of Public Health Preparedness and Response 2016b). In our model, it is appropriate, then, to model these warehouses as a single virtual federal stockpile node. In case of an attack, inventories can be shipped from the federal inventory node to state-managed warehouses, where additional inventory is usually stored. We therefore model the SNS stockpile network with one federal stockpile node and 52 state and special district nodes. Inventory from the state warehouses can then be forwarded to city, county, and local authorities for dispensing to the general public. Using Census data to obtain precise information about major cities or so-called Metropolitan Statistical Areas (MSAs) and counties, we include 377 MSA nodes, 3,221 county nodes, and 77,072 neighborhood nodes. In addition, we also include 77,072 household nodes, each denoting the aggregate demand in a neighborhood. The federal stockpile node serves all downstream state nodes. State nodes serve not only their downstream MSAs but also all MSA nodes in neighboring states. MSAs serve predominantly their downstream counties but also neighboring MSAs and counties, and so forth. In total, the network comprises 157,795 nodes and 335,010 edges.

6.1.2. Transportation Times. As we remarked earlier, the time to ship MCMs from the federal node to any state level node varies from 24 to 36 hours. From state level to MSA/county level, we assume the transportation time to be within 2–6 hours, for example, if

Table 4. Solution Time (Seconds)

Minimum	First quantile	Median	Mean	Third quantile	Maximum
192	719	798	671	839	954

Note. The resulting optimization problems for the case study instances have 1,100,000 variables and 115,000 constraints, approximately, after preprocessing using Gurobi.

Table 5. Statistics of Upper Bounds on AP Heuristic's Suboptimality Gaps for the Numerical Studies on the Joint Inventory and Dispensing Capacity Optimization Problem

Minimum	First quantile	Median	Mean	Third quantile	Maximum
0.00%	0.00%	0.00%	0.15%	0.00%	0.22%

shipped via ground transportation, depending on the geography of the state. According to field experiments, including, for example, a study in Philadelphia in 2005 (Agócs et al. 2007) and another one in Minneapolis in 2008 (Assistant Secretary for Preparedness and Response 2010), policymakers have estimated the shipment of MCMs from local authorities to dispensing points to take 10–12 hours.

6.1.3. Holding Costs. We consider the SNS design for one MCM type, for example, ciprofloxacin or doxycycline, to treat individuals affected by anthrax attacks in the United States. Inventory costs include the cost of purchasing, storage, management, replenishment, and shipment from manufacturer to warehouse. According to a commissioned paper by the Institute of Medicine (Guyton et al. 2011), the cost for anthrax antibiotics stored at regional warehouses amounts to \$2.10 per person, although home kits are more expensive at \$10 per person because of higher packaging and delivery costs. We estimate the federal, state, MSA, county, and neighborhood holding costs to be between \$1 and \$2 per person. In general, upstream holding costs are cheaper than downstream ones because of anticipated scale economies. Annual management and replenishment costs are 85% of purchase cost according to Guyton et al. (2011).

6.1.4. Demand. To estimate the demand volume in a neighborhood at each MSA in case of an attack, we assume that an airborne attack has the same spread radius in every geographic region, and we use a value of 3,000 square miles as in the field study on the Minneapolis MSA (Assistant Secretary for Preparedness and Response 2010). The affected population at node i can then be calculated as (population density at i) \times $\min\{3,000 \text{ sq miles, area of neighborhood}\}$. We obtained population density and area data on the neighborhood from statistics of the U.S. Census Bureau (2000).

6.1.5. Dispensing Modes and Costs. Two dispensing modes are currently available: (1) point of dispenses (PODs) and (2) door-to-door delivery by the U.S. Postal Service (USPS) (Wein et al. 2003). PODs are what the supply chain literature would consider a “pull” strategy, in which people commute to pre-specified locations in their neighborhoods to pick up MCMs for their families. The USPS option is a “push” strategy, in which the staff members of USPS deliver MCMs into the mailboxes of each household.

Dispensing capacity is shared among neighborhoods so that there are approximately two PODs per county.

Primary costs associated with the PODs setup are medical staff training costs and wages, supervisor training costs and wages, security wages, and public facility rental costs. Using the recommendation in Zaric et al. (2008), we estimate a labor wage of \$18.64 per hour, \$2,785 security cost per POD per day, and \$5,000 administrative overhead cost per POD per day; for each POD, the total number of staff is 300 and can support a dispensing service rate of 1,000 patients per hour. The resulting cost of operating a POD during the dispensing phase is \$3.39 per hour for a one person/hour increase in dispensing capacity. For capacity installation costs, we assume that a two-day drill is performed annually, resulting in \$162.72 per unit capacity per year.

The main costs associated with the USPS option are the wages for the delivery staff and the wages for the accompanying security officers. Based on the Minneapolis–St. Paul study (Guyton et al. 2011) with USPS delivery involving 179 staff, a wage of \$23.72 per person per hour, covering 205,000 households in 8.5 hours (estimate of 6.9 hours spent on transportation and 1.6 hours on material handling) for a population density of 1,000 households per square mile, we derive a population density–dependent USPS dispensing cost function: $\$(356/\sqrt{\text{density}} + 19)$ per hour for every one person/hour increase in delivery capacity. For capacity installation costs, we also assume a two-day drill every year. In addition to these costs, the USPS option also has an upper bound on how many workers can take up the delivery job because each delivery worker has to be accompanied by one security officer, and the total number of security

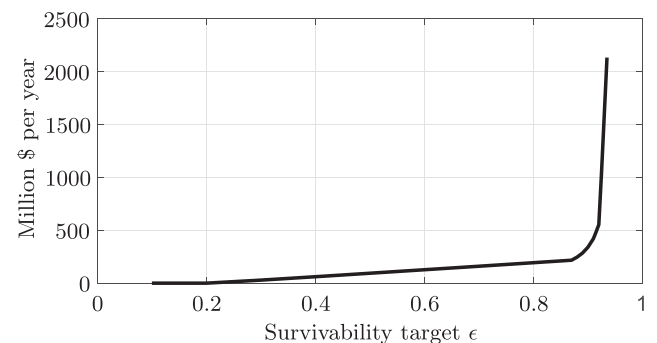
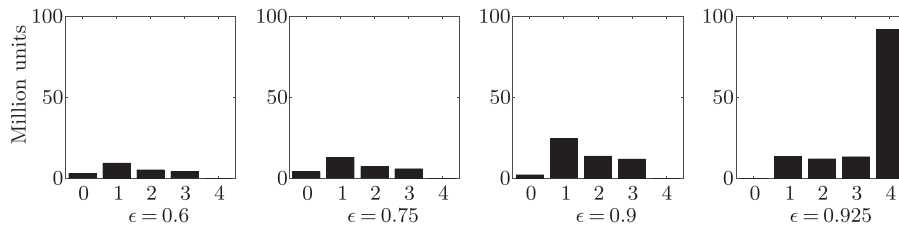
Figure 2. Minimum Required Annual Cost for Different Survivability Targets

Figure 3. Amount (y -Axis) and Location (x -Axis) of Prepositioned Inventory for Different Survivability Targets $\{0.6, 0.75, 0.9, 0.925\}$



Note. At the x -axis, level 0 is federal, level 1 is state, level 2 is MSA, level 3 is county, and level 4 is predisposed medical kits in households.

officers is likely a bottleneck during such an emergency situation. We assume that each county can support at most 10 delivery workers at any given time.

6.1.6. Survival Parameters. We choose a time step of $\delta = 8$ hours in accordance with the standard work shift length of 8 hours per group of staff for dispensing and delivery (Stroud et al. 2011). Then, based on the survival probability calculation in (5), for a detection time of 48 hours, we calculate $\{\rho_1, \rho_2, \rho_3, \rho_4, \dots\}$ to be $\{0.957, 0.944, 0.929, 0.912, \dots\}$. If a person inhales anthrax spores and is left untreated, the person's survival probability is estimated to be 0.2 (U.S. Food and Drug Administration 2015).

6.2. Results

Having calibrated the model, we compute the optimal costs for different policy parameters: survivability target, attack scale, detection time, and dispensing mode. Table 3 is a summary of the model inputs and outputs. Table 4 provides solution times and information about the size of the resulting optimization problems.

The default input parameters we consider are attack-scale parameters $\Gamma_i = 2, \forall i \in V$, and detection time $\tau_0 = 48$ hours.³ In the following, we perform several sensitivity analyses to illustrate the flexibility of our framework and how it can be used to guide policymaking.

6.2.1. Varying Survivability Target. Figure 2 depicts the minimum required annual cost as a function of the survivability target. For example, for an 80% survivability target, the annual budget needs to be about \$194 million dollars; if the survivability target is set to 92%, the budget would need to be increased to \$553 million, approximately. The curve in Figure 2 also illustrates that increasing the survivability target beyond 90% requires a rather steep cost increase. We can also interpret the curve as a Pareto frontier associated with the cost-survivability trade-off: the region left of the curve represents the achievable outcomes.

Figure 3 shows how much inventory and at which level it needs to be prepositioned for different survivability targets. It can be seen that as the target

increases, more inventory is required. Notably, for a target less than 90%, predisposed medical kits are redundant, whereas for a higher target, a rather large number of them is required.

Figure 4 shows county-level capacity decisions for different survivability targets. Although required capacity naturally increases with the desired target, interestingly, it drops at a steep rate beyond a certain point.

Put together, the analysis of varying survivability target illustrates a phase change that takes place around the target of 90%, approximately. To better understand this phenomenon, recall first that survivability targets translate into certain responsiveness requirements through relationship (5), for example, with higher targets calling for higher responsiveness. For targets lower than 90%, then, storing inventory upstream at federal/state levels provides acceptable responsiveness, and because it is so much more cost efficient per unit and also enjoys pooling benefits (in the sense that one kit could serve multiple downstream recipient households), it makes predisposed kits redundant. For targets higher than 90%, upstream storage no longer provides adequate responsiveness for the entire SNS network, and predisposed kits become necessary, which explains their introduction to the prepositioning strategy. The high rate at which these kits are introduced, then, compared with kits stored upstream, can be explained by the lack of pooling

Figure 4. Average Capacity Installation Per County (Number of People Served Per Hour) for Different Survivability Targets

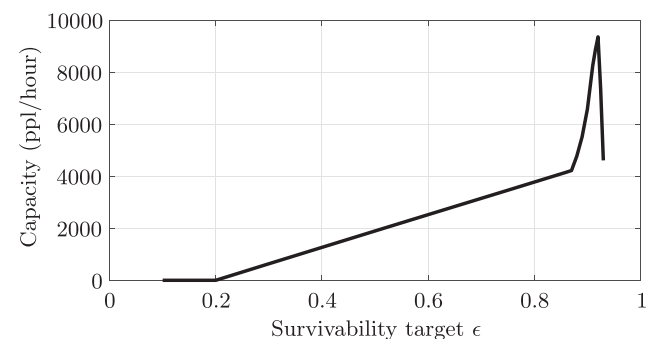
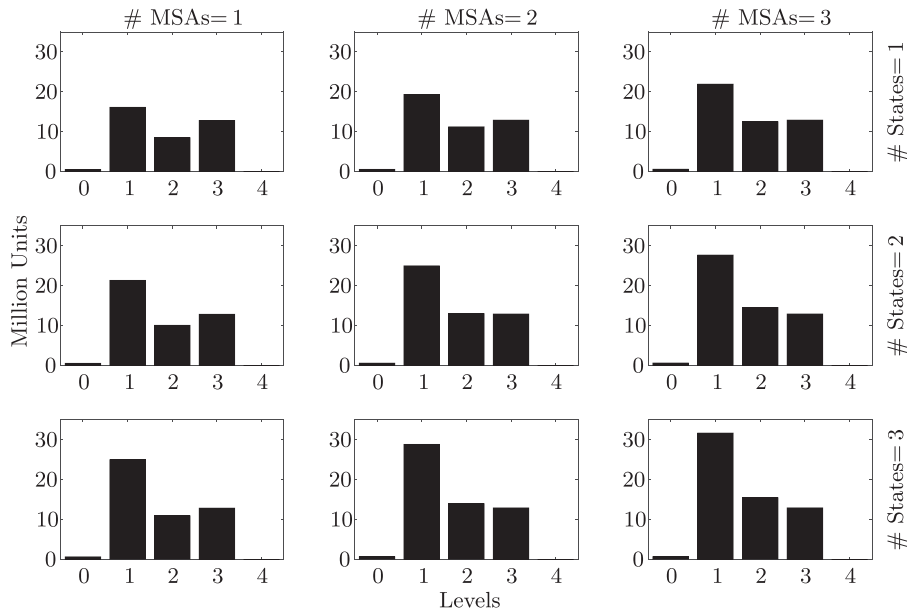


Figure 5. Amount (y-Axis) and Location (x-Axis) of Prepositioned Inventory for Different Attack-Scale Parameters

Note. At the x-axis, level 0 is federal, level 1 is state, level 2 is MSA, level 3 is county, and level 4 is predisposed medical kits in households.

benefits (one predisposed kit could only serve a single household). This, in turn, explains the observed steep increase in inventory costs and the decrease in required dispensing capacity.

6.2.2. Varying Attack Scale. We now explore different attack-scale parameters. Figure 5 reports the minimum required annual costs and optimal inventory amounts at each level, under nine different cases: $\{1,2,3\}$ states being attacked and each state having $\{1,2,3\}$ MSAs or counties affected. Recall that the first parameter reflects the geographic complexity of the attack and the second parameter reflects the magnitude of an attack within a state.

6.2.3. Varying Detection Time. Figure 6 shows the cost-survivability trade-off for different detection times τ_0 , highlighting the importance of timely response mechanisms. In particular, if the detection mechanism is able to identify an attack and start deployment within 60 hours, approximately \$788 million is needed for a 90% survivability target annually. For an enhanced detection mechanism capable of detecting attacks within 24 hours, the corresponding annual costs reduce to just more than \$210 million.

6.2.4. Choosing Between Different Dispensing Modes. Figure 7 illustrates the cost-survivability trade-offs for different dispensing modes. Using the USPS dispensing mode alone is generally more costly than using PODs because of the upper bound on USPS capacity and reliance on home medical kits. Overall,

we find that even though the USPS is empirically efficient in high-density neighborhoods, PODs should still remain as the primary dispensing channel because of their scalability.

7. Extensions and Concluding Remarks

In this paper, we considered a framework to tackle the problem of designing a cost-effective and responsive public health stockpile supply chain for protection against bioterrorism attacks. Our framework captured many of the key drivers facing the CDC in maintaining the SNS, for example, holding costs, pooling, capacity, and responsiveness. Our methodology can be extended to capture additional SNS design considerations or used in other supply chain applications. For example, an interesting direction for future research could be to allow for multiple types of MCMs for the same or different biothreats. In that

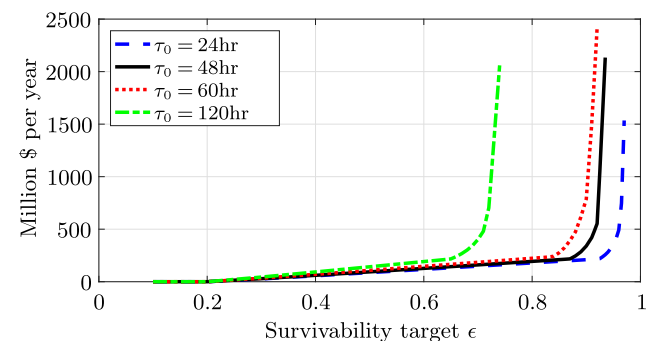
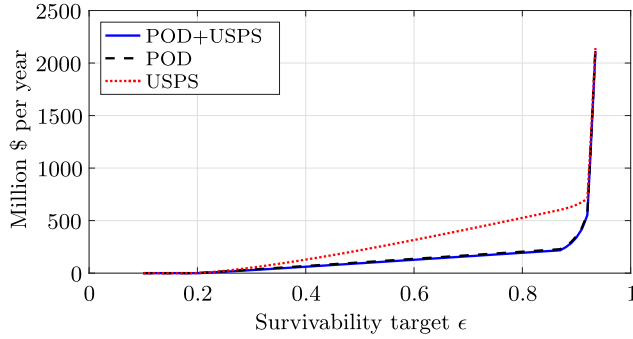
Figure 6. (Color online) Minimum Required Annual Costs for Different Survivability Targets and Detection Times

Figure 7. (Color online) Minimum Required Annual Costs for Different Survivability Targets and Dispensing Modes



case, our model could be used to guide the design of the CDC's overall portfolio and mix of MCMs in view of budgetary constraints.

Acknowledgments

The authors thank Area Editor Moshe Kress, the associate editor, and two anonymous reviewers for suggestions that substantially improved this article.

Appendix A. Proofs

Proof of Theorem 1

Under Assumption 2, our model can be reformulated to one in which, instead of inventory shipments over the edges of (V, E) , $\{f_e : e \in E\}$, we consider inventory shipments over the edges of the directed out-tree (V, T) with root node zero and $T := \{(i, j) : i \in V, j \in \mathcal{C}(i)\}$. This is because, under Assumption 2, any inventory shipment to node j occurs from some “parent node” i such that $j \in \mathcal{D}(i)$, that is, $\mathcal{P}^{L-\ell}(j) = i$ for some ℓ , where $\mathcal{P}^n(k) := \mathcal{P}(\mathcal{P}^{n-1}(k))$ for n positive integer, $k \in V$, and $\mathcal{P}^0(k) := k$. Thus, any f_{ij} can be thought of as a flow along the unique path $i \rightarrow \mathcal{P}^{L-\ell-1}(j) \rightarrow \dots \rightarrow \mathcal{P}(j) \rightarrow j$. It can then be readily seen that our model entails a path-based network flow formulation for inventory shipments in the graph (V, T) . In the rest of this proof, we consider the associated edge-based formulation and denote the shipments over the edges with $\mathbf{y} = \{y_t : t \in T\}$. Clearly, this is without loss because, for every feasible (\mathbf{f}, \mathbf{s}) in the path-based formulation, there exists \mathbf{y} such that (\mathbf{y}, \mathbf{s}) is feasible for the edge-based formulation and vice versa [for more details, see chapter 3.5 in Ahuja et al. (1993)]. Furthermore, we assume that demand occurs at every node; specifically, demand at node i is $d_i = \hat{d}_i \xi_i$ for all $i \in V$ and some $\xi \in \Xi$. Unmet demand at each node in $V \setminus V_D$ is treated in the same way as at nodes in V_D : either penalized at a cost rate b under the (LLC) or subject to a guarantee under the (LLG).

To prove optimality of affine policies, it suffices to show that for any static inventory allocation decision $\mathbf{x} \in X$, there exist policies that are affine in the uncertain demand for the adjustable decisions and achieve the same worst-case cost under fully adjustable policies. Thus, we henceforth consider the static inventory allocation decision \mathbf{x} as fixed.

We first deal with (LLC). At the end of the proof, we argue how (LLG) can be cast as a special case of (LLC). Using the edge-based formulation, the more general demand model, and a fixed inventory allocation $\mathbf{x} \in X$, it can be readily seen

that when $\rho = 1$, the (LLC) problem is equivalent to (in the sense that they have the same optimal set)⁴

$$\begin{aligned} \min_{\mathbf{y}(\cdot), \mathbf{s}(\cdot)} \max_{\xi \in \Xi} \quad & \mathbf{1}^T \mathbf{s}(\xi) \\ \text{subject to} \quad & s_i(\xi) + y_{\mathcal{P}(i)}(\xi) + x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij}(\xi) + \hat{d}_i \xi_i, \quad \forall i \in V, \\ & \forall \xi \in \Xi \\ & y_{\mathcal{P}(i)}(\xi) + x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij}(\xi), \quad \forall i \in V, \forall \xi \in \Xi \\ & \mathbf{y}(\cdot), \mathbf{s}(\cdot) \geq 0, \end{aligned}$$

where $\mathbf{1}$ is the vector of all ones. Let z_F be the optimal value of this fully adjustable formulation. Correspondingly, let z_A be the optimal value of its affinely adjustable counterpart, that is, when we restrict $\mathbf{y}(\cdot)$ and $\mathbf{s}(\cdot)$ to be affine in ξ . It suffices then to show that $z_F = z_A$.

We now introduce some useful notation:

- For some index set I , let $\xi_I := \{\xi_i : i \in I\}$ and $\text{proj}_I \Xi := \{\xi_I : \xi \in \Xi\}$.
- Let V^l be the set of nodes at the l th level, that is, $V^l := \{j \in V : \mathcal{P}^l(j) = 0\}$.
- Let A_i be the set of ancestor nodes of i , that is, $A_i := \{i, \mathcal{P}(i), \mathcal{P}^2(i), \dots, 0\}$.
- We frequently look into subgraphs of (V, T) , specifically out-trees rooted at some node $i \in V^l$, denoted by (V_i, T_i) , where $V_i := \{j \in V : \mathcal{P}^k(j) = i \text{ for some } k = 0, 1, 2, \dots, L-l\}$ and $T_i := \{(k, l) : k \in V_i, l \in \mathcal{C}(k)\}$.
- Let $\mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i})$ be the set of feasible policies for (V_i, T_i) , with inventory at node i being x_i , the inventory on $V_i \setminus \{i\}$ being \mathbf{x}_{-i} , and the uncertain parameter being ξ'_i at node i ; that is,

$$\begin{aligned} \mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i}) := \{ \{ \mathbf{y}(\cdot), \mathbf{s}(\cdot) \} : \\ & x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij}(\xi), \quad \forall \xi \in \Xi_i, \\ & s_i(\xi) + x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij}(\xi) + \hat{d}_i \xi'_i, \quad \forall \xi \in \Xi_i, \\ & y_{\mathcal{P}(k)}(\xi) + x_k \geq \sum_{j \in \mathcal{C}(k)} y_{kj}(\xi), \quad \forall k \in V_i \setminus \{i\}, \forall \xi \in \Xi_i, \\ & s_k(\xi) + y_{\mathcal{P}(k)}(\xi) + x_k \geq \sum_{j \in \mathcal{C}(k)} y_{kj}(\xi) + \hat{d}_k \xi_k, \\ & \quad \forall k \in V_i \setminus \{i\}, \forall \xi \in \Xi_i, \\ & \mathbf{y}(\xi), \mathbf{s}(\xi) \geq 0, \quad \forall \xi \in \Xi_i \}, \end{aligned}$$

where $\Xi_i(\xi'_i) := \{\xi \mid \xi \in \Xi, \xi_i = \xi'_i\}$. We use shorthand notation Ξ_i when there is no ambiguity in ξ'_i .

- Let $\Theta_i(\xi'_i, \mathbf{s})$ be the worst-case demand loss in (V_i, T_i) under $\xi_i = \xi'_i$, let (feasible) resource allocation policy $\{\mathbf{y}(\cdot), \mathbf{s}(\cdot)\}$, $\Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i})$ be the worst-case demand loss in V_i under $\xi_i = \xi'_i$ with an optimal fully adjustable policy, and let $\mathcal{Q}_i^*(\xi'_i, x_i, \mathbf{x}_{-i})$ be the set of all optimal policies:

$$\begin{aligned} \Theta_i(\xi'_i, \mathbf{s}) &:= \max_{\xi \in \Xi_i(\xi'_i)} \sum_{k \in V_i} s_k(\xi), \\ \Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i}) &:= \min_{\mathbf{y}(\cdot), \mathbf{s}(\cdot)} \Theta_i(\xi'_i, \mathbf{s}) \text{ subject to} \\ & \quad \{ \mathbf{y}(\cdot), \mathbf{s}(\cdot) \} \in \mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i}), \\ \mathcal{Q}_i^*(\xi'_i, x_i, \mathbf{x}_{-i}) &:= \arg \min_{\mathbf{y}(\cdot), \mathbf{s}(\cdot)} \Theta_i(\xi'_i, \mathbf{s}) \text{ subject to} \\ & \quad \{ \mathbf{y}(\cdot), \mathbf{s}(\cdot) \} \in \mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i}). \end{aligned}$$

Similarly, we define the affinely adjustable counterpart $\Omega_i^A(\xi'_i, x_i, \mathbf{x}_{-i})$ as the worst-case demand loss in V_i , restricted to $\{\mathbf{y}(\cdot), \mathbf{s}(\cdot)\} \in \mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i})$ and affine.

Using our notation, $z_F = \Omega_0^F(1, x_0, \mathbf{x}_{-0})$ and $z_A = \Omega_0^A(1, x_0, \mathbf{x}_{-0})$. It thus suffices to show that

$$\Omega_0^F(1, x_0, \mathbf{x}_{-0}) = \Omega_0^A(1, x_0, \mathbf{x}_{-0}). \quad (\text{A.1})$$

We now present two useful intermediate results. The first provides a recursive expression for Ω_i^F , and the second shows that Ξ has binary vertices.

Proposition A.1. For all $i \in V$, $\xi'_i \in [0, 1]$, $\mathbf{x} \geq 0$,

$$\begin{aligned} \Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i}) \\ = \left(\hat{d}_i \xi'_i - x_i + \max_{\xi_{\ell(i)} \in \text{proj}_{\mathcal{Q}(i)} \Xi_i(\xi'_i)} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) \right)^+. \end{aligned}$$

Moreover, $\Omega_i^F(\xi_i, x_i, \mathbf{x}_{-i})$ is convex in ξ_i over $\xi_i \in [0, 1]$.

Proposition A.2. Ξ is a polytope with binary vertices, that is, $\xi \in \text{ext}(\Xi) \Rightarrow \xi \in \{0, 1\}^n$.

Proposition A.1 shows the intuitive fact that given a fixed inventory decision and under the optimal policy, the demand shortage in V_i should be the sum of shortages in $\{V_j\}_{j \in \mathcal{C}(i)}$ plus the demand on node i , subtracting away inventory x_i . Based on such a recursive definition, we can view the tree as a collection of hub-and-spoke clusters, each consisting of one node and several edges pointing away from this node (Figure A.1).

Now we prove (A.1) by induction. For a given tree (V_i, T_i) and uncertainty set Ξ_i , we limit our attention to $\{\mathbf{y}(\cdot), \mathbf{s}(\cdot)\}$ that satisfy the following properties:

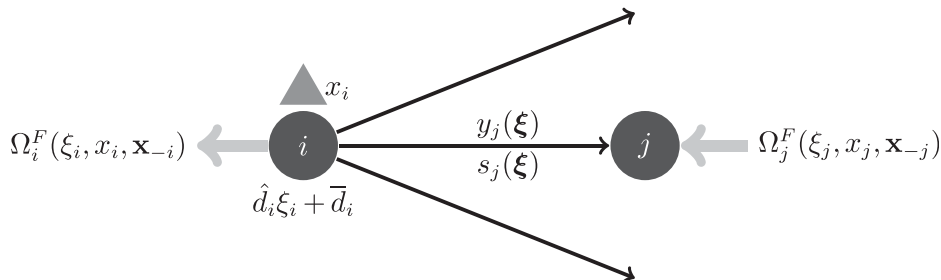
- i. $\forall m \in V_i, j \in \mathcal{C}(m), \xi \in \Xi_i: s_m(\xi) = s_m(\xi_m), y_{mj}(\xi) = y_{mj}(\xi_j)$; that is, the policies are univariate in ξ_m, ξ_j , respectively.
- ii. $\forall m \in V_i$: (a) $y_{\mathcal{P}(m)m}(0) = 0$, (b) $y_{jk}(0) = \Omega_k^A(0, x_k, \mathbf{x}_{-k}) = 0 \forall (j, k) \in T_m$, and (c) $s_j(0) = 0 \forall j \in V_m$; that is, these policies are linear.

Let

$$\begin{aligned} \bar{\mathcal{Q}}_i(\xi_i, x_i, \mathbf{x}_{-i}) &:= \{\{\mathbf{y}, \mathbf{s}\} \in \mathcal{Q}_i(\xi_i, x_i, \mathbf{x}_{-i}) : \{\mathbf{y}, \mathbf{s}\} \\ &\quad \text{satisfies (I)-(II)}\} \\ \bar{\mathcal{Q}}_i^*(\xi_i, x_i, \mathbf{x}_{-i}) &:= \bar{\mathcal{Q}}_i(\xi_i, x_i, \mathbf{x}_{-i}) \cap \mathcal{Q}_i^*(\xi_i, x_i, \mathbf{x}_{-i}). \end{aligned}$$

We are now ready to formalize the induction process.

Figure A.1. Recursive Representation of Ω_i^F



Induction Hypothesis

If, for some $l \in \{0, 1, \dots, L-1\}$, $\forall j \in V^{l+1}, \mathbf{x} \geq 0$, $\exists \{\bar{\mathbf{y}}^j(\cdot), \bar{\mathbf{s}}^j(\cdot)\}$ such that

$$(\text{Feasibility}) \forall \xi_j \in [0, 1], \{\bar{\mathbf{y}}^j(\cdot), \bar{\mathbf{s}}^j(\cdot)\} \in \bar{\mathcal{Q}}_j(\xi_j, x_j, \mathbf{x}_{-j}), \quad (\text{A.2})$$

$$\begin{aligned} (\text{Optimality}) \forall \xi_j \in \{0, 1\}, \Theta_j(\xi_j, \bar{\mathbf{s}}^j) &= \Omega_j^A(\xi_j, x_j, \mathbf{x}_{-j}) \\ &= \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}), \end{aligned} \quad (\text{A.3})$$

then $\forall i \in V^l, \mathbf{x} \geq 0$, $\exists \{\bar{\mathbf{y}}^i(\cdot), \bar{\mathbf{s}}^i(\cdot)\}$ such that

$$(\text{Feasibility}) \forall \xi_i \in [0, 1], \{\bar{\mathbf{y}}^i(\cdot), \bar{\mathbf{s}}^i(\cdot)\} \in \bar{\mathcal{Q}}_i(\xi_i, x_i, \mathbf{x}_{-i}), \quad (\text{A.4})$$

$$\begin{aligned} (\text{Optimality}) \forall \xi_i \in \{0, 1\}, \Theta_i(\xi_i, \bar{\mathbf{s}}^i) &= \Omega_i^A(\xi_i, x_i, \mathbf{x}_{-i}) \\ &= \Omega_i^F(\xi_i, x_i, \mathbf{x}_{-i}). \end{aligned} \quad (\text{A.5})$$

Base Case ($l = L-1$). Given some $\mathbf{x} \geq 0$, for each $j \in V^L$, we define $\bar{\mathbf{s}}_j^j(\xi) = \bar{\mathbf{s}}_j^j(\xi_j) = \xi_j(\hat{d}_j - x_j)^+$. This policy satisfies I and II by construction, and it is straightforward to check feasibility and optimality.

General Step

Now we prove the general induction step. Suppose that (A.2) and (A.3) hold for some $l \in \{0, 1, \dots, L-1\}$ and $\forall j \in V^{l+1}, \mathbf{x} \geq 0$. To construct $\{\bar{\mathbf{y}}^i(\cdot), \bar{\mathbf{s}}^i(\cdot)\}$, we define the following: because $\Omega_j^A(\xi_j, x_j, \mathbf{x}_{-j})$ is nondecreasing in ξ_j and non-negative, there necessarily exists $g_j \geq 0$ such that $\Omega_j^A(\xi_j, x_j, \mathbf{x}_{-j}) = g_j \xi_j$ for $\xi_j \in \{0, 1\}$. Let $\beta_i := \min\left(1, \frac{x_i}{\sum_{j=1}^{T_i} g_{(j)} + \hat{d}_i}\right)$ if $\sum_{j=1}^{T_i} g_{(j)} + \hat{d}_i > 0$ and $\beta_i := 1$ otherwise, where $g_{(j)}$ is the j th largest element of $\{g_j\}_{j \in \mathcal{C}(i)}$ (β_i is the surge demand coverage ratio).

We are now ready to construct affine $\{\bar{\mathbf{y}}^i, \bar{\mathbf{s}}^i\}$:

$$\{\bar{\mathbf{y}}^i, \bar{\mathbf{s}}^i\} := \begin{cases} \bar{y}_{ij}^i(\xi_j) = \beta_i g_j \xi_j, & \forall j \in \mathcal{C}(i), \\ \bar{s}_i^i(\xi_i) = (1 - \beta_i) \hat{d}_i \xi_i, \\ \bar{y}_{km}^i(\xi_m) = \bar{y}_{km}^{l+1}(\xi_m), & \forall (k, m) \in T_j, \forall j \in \mathcal{C}(i), \\ \bar{s}_k^i(\xi_k) = \bar{s}_k^{l+1}(1) \xi_k, & \forall k \in V_i \setminus \{i\}. \end{cases} \quad (\text{A.6})$$

Let $\bar{\mathcal{Q}}_{i1}^* := \bar{\mathcal{Q}}_{i1}^*(1, x_j + \bar{y}_{ij}^i(1), \mathbf{x}_{-j})$. By the induction hypothesis, $\bar{\mathcal{Q}}_{i1}^* \neq \emptyset$ because $x_j + \bar{y}_{ij}^i(\xi_j) \geq 0$ for $\xi_j \in [0, 1]$. We now check the feasibility and optimality of $\{\bar{\mathbf{y}}^i, \bar{\mathbf{s}}^i\}$.

For feasibility, note that properties I and II hold by construction. To show that $\{\bar{\mathbf{y}}^i, \bar{\mathbf{s}}^i\}$ satisfies the constraints in $\mathcal{Q}_i(\xi'_i, x_i, \mathbf{x}_{-i})$, note that the constraints involving node i are satisfied by construction of $\{\bar{\mathbf{y}}^i, \bar{\mathbf{s}}^i\}$. For nodes downstream from i , note that $\bar{y}_{km}^i(\xi_m) = \bar{y}_{km}^{l+1}(1) \xi_m$ and $\bar{s}_k^i(\xi_k) = \bar{s}_k^{l+1}(1) \xi_k$ for all $(k, m) \in T_j, j \in \mathcal{C}(i)$ and $k \in V_i \setminus \{i\}$. Because $\{\bar{\mathbf{y}}^0(\cdot), \bar{\mathbf{s}}^0(\cdot)\}$ and $\{\bar{\mathbf{y}}^{l+1}(\cdot), \bar{\mathbf{s}}^{l+1}(\cdot)\}$ satisfy the constraints of $\mathcal{Q}_j(0, x_j + \bar{y}_{ij}^i(0), \mathbf{x}_{-j})$

and $\mathcal{Q}_j(1, x_j + \bar{y}_{ij}^i(1), \mathbf{x}_{-j})$, respectively, it is straightforward to check the constraint satisfaction of $\{\mathbf{y}^i, \mathbf{s}^i\}$ for $\mathcal{Q}_i(\xi_i', x_i, \mathbf{x}_{-i})$.

For optimality, note that for $\xi_i' \in \{0, 1\}$, we have

$$\Theta_i(\xi_i', \bar{\mathbf{s}}^i) = \bar{s}_i^i(\xi_i') + \max_{\xi_i \in \Xi_i} \sum_{j \in \mathcal{C}(i)} \sum_{k \in V_j} \bar{s}_k^i(\xi_k) \quad (\text{A.7a})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \max_{\xi_j \in \Xi_j(\xi_j)} \sum_{k \in V_j} \bar{s}_k^j(1) \xi_k \quad (\text{A.7b})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \max_{\xi_j \in \Xi_j(\xi_j)} \sum_{k \in V_j} \bar{s}_k^j(1) \xi_k \quad (\text{A.7c})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \Omega_j^A(\xi_j, x_j + \bar{y}_{ij}^i(\xi_j), \mathbf{x}_{-j}) \quad (\text{A.7d})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j + \bar{y}_{ij}^i(\xi_j), \mathbf{x}_{-j}) \quad (\text{A.7e})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \left(\Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) - \bar{y}_{ij}^i(\xi_j) \right)^+ \quad (\text{A.7f})$$

$$= \bar{s}_i^i(\xi_i') + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \left(\sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) - \sum_{j \in \mathcal{C}(i)} \bar{y}_{ij}^i(\xi_j) \right) \quad (\text{A.7g})$$

$$= \left(\hat{d}_i \xi_i' - x_i + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) \right)^+ \quad (\text{A.7h})$$

$$= \left(\hat{d}_i \xi_i' - x_i + \max_{\substack{\xi_{\mathcal{C}(i)}: \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi_i' \\ \xi_j \in \{0, \xi_i'\}, j \in \mathcal{C}(i)}} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) \right)^+ \quad (\text{A.7i})$$

$$= \Omega_i^F(\xi_i', x_i, \mathbf{x}_{-i}) = \Omega_i^A(\xi_i', x_i, \mathbf{x}_{-i}). \quad (\text{A.7j})$$

Equalities (A.7a) and (A.7b) are straightforward applications of definitions of Θ_i and $\bar{\mathbf{s}}^i$.

To show equality (A.7c), we argue that it is sufficient to check $\xi_j = 1$ and $\xi_j = 0$ for $j \in \mathcal{C}(i)$. Suppose that $\Xi_j(\xi_j)$ has n extreme points ξ^1, \dots, ξ^n , and $W_m = \{k \in V_j : \xi_k^m = \xi_j\}$, $\bar{W}_m = V_j \setminus W_m$. Because, for each extreme point, ξ_k takes the value of either ξ_j or zero (cf. Proposition A.2), we can write

$$\max_{\xi_j \in \Xi_j(\xi_j)} \sum_{k \in V_j} \bar{s}_k^j(1) \xi_k = \max_{m \in \{1, \dots, n\}} \sum_{k \in V_j} \left\{ \bar{s}_k^j(1) \xi_k \mathbb{1}_{k \in W_m} \right\},$$

which is a convex function of ξ_j ($\mathbb{1}_{i \in I} = 1$ if $i \in I$ and zero otherwise). By this token, we can replace $\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i$ with $\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i \cap \{0, 1\}^{|\mathcal{C}(i)|}$ for the argument of the outer maximization because $\Xi_i(0)$ and $\Xi_i(1)$ have binary vertices.

To show equality (A.7d), note that, for $\xi_j = 1$, $\Omega_j^A(1, x_j + \bar{y}_{ij}^i(1), \mathbf{x}_{-j}) = \Theta_j(1, \mathbf{s}^j) = \bar{s}_j^j(1) + \max_{\xi_j \in \Xi_j(1)} \sum_{k \in V_j \setminus \{j\}} \bar{s}_k^j(1) \xi_k$, which is equal to $\max_{\xi_j \in \Xi_j(1)} \sum_{k \in V_j} \bar{s}_k^j(1) \xi_k$. For $\xi_j = 0$, $\Omega_j^A(0, x_j + \bar{y}_{ij}^i(0), \mathbf{x}_{-j}) = 0$. Equality (A.7e) holds by the induction hypothesis. Equality (A.7f) follows from the definition of Ω_j^F and (A.8). For equality (A.7g), by the definitions of β_i , we have $(\Omega_j^A(\xi_j, x_j, \mathbf{x}_{-j}) - \bar{y}_{ij}^i(\xi_j)) \geq 0$ for $j \in \mathcal{C}(i)$; therefore, we can rewrite (A.7f) into (A.7g). For the transition into (A.7h), it is straightforward to consider $\xi_i' = 0$ and $\xi_i' = 1$ separately.

Finally, (A.7i) holds by the convexity property of $\sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j})$, and (A.7j) holds by the recursive property, both shown in Proposition A.1. We have now completed the proof for the inductive step, thus showing $z_F = z_A$. \square

To complete the proof, we now turn our attention to the (LLG) formulation. By solving for $s_i(\mathbf{d})$ from (1c) and substituting in (2b), we can eliminate (1c) and rewrite (2b) as

$$\sum_{j:(i,j) \in E} f_{ji}(\mathbf{d}) \geq \frac{1 - \epsilon_i - \bar{\rho}}{\rho - \bar{\rho}} d_i, \quad \forall i \in V, \forall \mathbf{d} \in U.$$

At optimality, this constraint can be taken to be active without loss. (If it is not, we can scale down the associated flows into i so that it becomes active.) Thus, the (LLG) is equivalent to

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{f}(\cdot), \mathbf{s}(\cdot)} \quad & \mathbf{h}^T \mathbf{x} \\ \text{subject to} \quad & x_i \geq \sum_{j:(i,j) \in E} f_{ji}(\mathbf{d}), \quad \forall i \in V, \forall \mathbf{d} \in U', \\ & s_i(\mathbf{d}) + \sum_{j:(j,i) \in E} f_{ji}(\mathbf{d}) = d_i, \quad \forall i \in V, \forall \mathbf{d} \in U', \\ & s_i(\mathbf{d}) = 0, \quad \forall i \in V, \forall \mathbf{d} \in U', \\ & f_{ij}(\mathbf{d}) \geq 0, \quad \forall (i,j) \in E, \forall \mathbf{d} \in U', \\ & \mathbf{x} \in X, \end{aligned}$$

where $U' := \{\mathbf{d}' : d'_i = \frac{1 - \epsilon_i - \bar{\rho}}{\rho - \bar{\rho}} d_i \forall i \in V, \mathbf{d} \in U\}$. Then, this is, in turn, equivalent to

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{f}(\cdot), \mathbf{s}(\cdot)} \quad & \mathbf{h}^T \mathbf{x} + \max_{\mathbf{d} \in U'} b' \sum_{i \in V} (1 - \bar{\rho}) s_i(\mathbf{d}) \\ \text{subject to} \quad & x_i \geq \sum_{j:(i,j) \in E} f_{ji}(\mathbf{d}), \\ & \forall i \in V, \forall \mathbf{d} \in U', \\ & s_i(\mathbf{d}) + \sum_{j:(j,i) \in E} f_{ji}(\mathbf{d}) = d_i, \\ & \forall i \in V, \forall \mathbf{d} \in U', \\ & \sum_{j:(j,i) \in E} f_{ji}(\mathbf{d}) \leq d_i, \\ & \forall i \in V, \forall \mathbf{d} \in U', \\ & f_{ij}(\mathbf{d}) \geq 0, \quad \forall (i,j) \in E, \forall \mathbf{d} \in U', \\ & \mathbf{x} \in X, \end{aligned}$$

for b' large enough. However, this corresponds to an (LLC) instance that satisfies our assumptions and therefore admits an optimal affine adjustable policy. \square

Proof of Proposition A.1

Note that we can express $\Omega_i^F(\xi_i', x_i, \mathbf{x}_{-i})$ as $\max_{\xi_i \in \Xi_i(\xi_i')} r_i(\xi_i, x_i, \mathbf{x}_{-i})$, where

$$\begin{aligned} r_i(\xi_i, x_i, \mathbf{x}_{-i}) &:= \min_{\mathbf{y}, \mathbf{s}} \sum_{k \in V_i} s_k \\ \text{subject to} \quad & x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij}, \\ & s_i + x_i \geq \sum_{j \in \mathcal{C}(i)} y_{ij} + \hat{d}_i \xi_i, \\ & y_{\mathcal{Q}(k)k} + x_k \geq \sum_{j \in \mathcal{C}(k)} y_{kj}, \quad \forall k \in V_i \setminus \{i\}, \\ & y_{\mathcal{Q}(k)k} + s_k + x_k \geq \sum_{j \in \mathcal{C}(k)} y_{kj} + \hat{d}_k \xi_k, \\ & \quad \quad \quad \forall k \in V_i \setminus \{i\}, \\ & \mathbf{y}, \mathbf{s} \geq 0. \end{aligned}$$

It is straightforward to check that this optimization problem admits an optimal solution such that $s_i = (\hat{d}_i \xi'_i - x_i)^+$ and $\sum_{j \in \mathcal{C}(i)} y_{ij} = (x_i - \hat{d}_i \xi'_i)^+$. Also,

$$r_i(\xi, x_i + \Delta, \mathbf{x}_{-i}) = (r_i(\xi, x_i, \mathbf{x}_{-i}) - \Delta)^+, \quad \forall \Delta \geq 0. \quad (\text{A.8})$$

Another fact we use is that

$$\min_{\substack{\mathbf{x} \geq 0 \\ \mathbf{1}^T \mathbf{x} = K}} \sum_{i=1}^n (q_i - x_i)^+ = (\mathbf{1}^T \mathbf{q} - K)^+, \quad \forall \mathbf{q} \in \mathbb{R}^n, K \geq 0. \quad (\text{A.9})$$

Using all these properties, we have

$$\begin{aligned} \Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i}) &= (\hat{d}_i \xi'_i - x_i)^+ + \max_{\xi \in \Xi_i} \min_{\substack{\{y_{ij}\}_{j \in \mathcal{C}(i)} : y_{ij} \geq 0, \\ \sum_{j \in \mathcal{C}(i)} y_{ij} = (x_i - \hat{d}_i \xi'_i)^+}} \sum_{j \in \mathcal{C}(i)} r_j(\xi, x_j + y_{ij}, \mathbf{x}_{-j}) \\ &= (\hat{d}_i \xi'_i - x_i)^+ + \max_{\xi \in \Xi_i} \min_{\substack{\{y_{ij}\}_{j \in \mathcal{C}(i)} : y_{ij} \geq 0, \\ \sum_{j \in \mathcal{C}(i)} y_{ij} = (x_i - \hat{d}_i \xi'_i)^+}} \sum_{j \in \mathcal{C}(i)} (r_j(\xi, x_j, \mathbf{x}_{-j}) - y_{ij})^+ \end{aligned} \quad (\text{A.10a})$$

$$= (\hat{d}_i \xi'_i - x_i)^+ + \max_{\xi \in \Xi_i} \min_{\substack{\{y_{ij}\}_{j \in \mathcal{C}(i)} : y_{ij} \geq 0, \\ \sum_{j \in \mathcal{C}(i)} y_{ij} = (x_i - \hat{d}_i \xi'_i)^+}} \sum_{j \in \mathcal{C}(i)} (r_j(\xi, x_j, \mathbf{x}_{-j}) - y_{ij})^+ \quad (\text{A.10b})$$

$$= (\hat{d}_i \xi'_i - x_i)^+ + \left(\max_{\xi \in \Xi_i} \sum_{j \in \mathcal{C}(i)} r_j(\xi, x_j, \mathbf{x}_{-j}) - (x_i - \hat{d}_i \xi'_i)^+ \right)^+ \quad (\text{A.10c})$$

$$= (\hat{d}_i \xi'_i - x_i)^+ + \left(\max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i} \max_{\xi \in \Xi_i(\xi_j), j \in \mathcal{C}(i)} \sum_{j \in \mathcal{C}(i)} r_j(\xi, x_j, \mathbf{x}_{-j}) - (x_i - \hat{d}_i \xi'_i)^+ \right)^+ \quad (\text{A.10d})$$

$$= (\hat{d}_i \xi'_i - x_i)^+ + \left(\max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i} \sum_{j \in \mathcal{C}(i)} \max_{\xi \in \Xi_i(\xi_j)} r_j(\xi, x_j, \mathbf{x}_{-j}) - (x_i - \hat{d}_i \xi'_i)^+ \right)^+ \quad (\text{A.10e})$$

$$= (\hat{d}_i \xi'_i - x_i)^+ + \left(\max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) - (x_i - \hat{d}_i \xi'_i)^+ \right)^+ \quad (\text{A.10f})$$

$$= \left(\hat{d}_i \xi'_i - x_i + \max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) \right)^+. \quad (\text{A.10g})$$

Equality (A.10a) follows from the definition of r_i ; (A.10b) follows from (A.8); (A.10c) follows from (A.9); (A.10d) is an equivalent way of writing the maximization operator; exchanging the inner maximization and summation operators is allowed because r_j only depends on $\{\xi_k, k \in V_j\}$, leading to (A.10e); (A.10f) follows from the definition of r_j . Equality (A.10g) holds by considering the fact that $\Omega_j^F \geq 0$ for any $j \in \mathcal{C}(i)$ and checking two cases: one with $\hat{d}_i \xi'_i - x_i < 0$ and the other with $\hat{d}_i \xi'_i - x_i \geq 0$.

To show the convexity of $\Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i})$, we use an induction argument on the level of node i . First note that for $i \in V^L$, we have $\Omega_i^F(\xi'_i, x_i, \mathbf{x}_{-i}) = (\hat{d}_i \xi'_i - x_i)^+$, which is convex and nondecreasing in ξ'_i . For i in any other level of the

network, given the result we just proved (A.10g), it suffices to show that $\max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i(\xi'_i)} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j})$ is convex in ξ'_i . Because $\Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j})$ is convex in ξ_j (by the induction hypothesis), maximum is obtained at the extreme points. Given that $\text{proj}_{\mathcal{C}(i)} \Xi_i(\xi'_i) = \{\xi_{\mathcal{C}(i)} : \xi_j \in [0, \xi'_i], \mathbf{1}^T \xi_{\mathcal{C}(i)} \leq \Gamma_i \xi'_i\}$ and Ω_j^F is nondecreasing in ξ_j , we can equivalently write: $\max_{\xi_{\mathcal{C}(i)} \in \text{proj}_{\mathcal{C}(i)} \Xi_i(\xi'_i)} \sum_{j \in \mathcal{C}(i)} \Omega_j^F(\xi_j, x_j, \mathbf{x}_{-j}) + \sum_{j \in \mathcal{C}(i) \setminus S} \Omega_j^F(0, x_j, \mathbf{x}_{-j})$, which is convex, nondecreasing in ξ'_i . \square

Proof of Proposition A.2

For the sake of reaching a contradiction, suppose that there exists some extreme point $\xi \in \text{ext}(\Xi)$ such that its k th entry is fractional, $\xi_k \in (0, 1)$. Note that, by definition of Ξ , $\xi_0 = 1$, and thus, $k \neq 0$. Let $p = \mathcal{P}(k)$ be its parent; that is, $k \in \mathcal{C}(p)$. Then it must hold that $0 < \sum_{i \in \mathcal{C}(p)} \xi_i \leq \Gamma_p \xi_p$, and thus, $\Gamma_p > 0$ and $\xi_p > 0$.

Suppose that $\xi_p = 1$. Then the preceding condition becomes $0 < \sum_{i \in \mathcal{C}(p)} \xi_i \leq \Gamma_p$. Furthermore, $\xi_i \in [0, 1] \forall i \in \mathcal{C}(p)$. Now consider the $|\mathcal{C}(p)|$ -dimensional polytope $\{\gamma : \sum_{i \in \mathcal{C}(p)} \gamma_i \leq \Gamma_p, \gamma_i \in [0, 1] \forall i \in \mathcal{C}(p)\}$, which can be readily seen to have binary extreme points. Let these extreme points be $\{\gamma^j, j = 1, \dots, N\}$. Because $\xi_{\mathcal{C}(p)}$ belongs to this polytope, we can express it as

$$\xi_{\mathcal{C}(p)} = \sum_{j=1}^N \alpha_j \gamma^j, \quad \text{where } \sum_{j=1}^N \alpha_j = 1, \alpha_j \geq 0, \forall j \in \{1, \dots, N\}.$$

In particular, for $i \in \mathcal{C}(p)$ such that $\xi_i > 0$, we have

$$\sum_{j=1}^N \alpha_j \gamma_i^j = \xi_i \iff \sum_{j=1}^N \alpha_j \frac{1}{\xi_i} \gamma_i^j = 1. \quad (\text{A.11})$$

Now we use each γ^j to construct a vector $\xi^j \in \Xi$ and show that we can write ξ as a convex combination of $\{\xi^j, j = 1, \dots, N\}$ and $\xi \neq \xi^j, \forall j \in \{1, \dots, N\}$, which contradicts our assumptions.

We construct $\xi^j, j = 1, \dots, N$, as follows:

$$\begin{aligned} \xi_{V \setminus V_p}^j &= \xi_{V \setminus V_p}, \\ \xi_p^j &= \xi_p, \\ \xi_{\mathcal{C}(p)}^j &= \gamma^j, \\ \xi_{V_i \setminus \{i\}}^j &= \begin{cases} \frac{1}{\xi_i} \xi_{V_i \setminus \{i\}} & \text{if } \xi_i > 0 \text{ and } \xi_i^j = 1 \\ 0 & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{C}(p). \end{aligned}$$

It is straightforward to verify that for all $j = 1, \dots, N$, $\xi^j \in \Xi$ and $\xi^j \neq \xi$. We next argue that $\sum_{j=1}^N \alpha_j \xi^j = \xi$ by checking component-wise. This is straightforward for components $V \setminus V_p, p$, and $\mathcal{C}(p)$. For any $i \in V_i \setminus \{i\}$ for $i \in \mathcal{C}(p)$ such that $\xi_i = 0$, we have that $\xi_i^j = \xi_i = 0$ for all $j = 1, \dots, N$. Finally, for any $i \in V_i \setminus \{i\}$ for $i \in \mathcal{C}(p)$ such that $\xi_i > 0$, we have that

$$\begin{aligned} \sum_{j=1}^N \alpha_j \xi_i^j &= \sum_{j: \xi_i^j = 1} \alpha_j \xi_i / \xi_i + \sum_{j: \xi_i^j = 0} 0 = \sum_{j=1}^N \alpha_j \xi_i / \xi_i \xi_i^j \\ &= \sum_{j=1}^N (\alpha_j \xi_i^j / \xi_i) \xi_i = \sum_{j=1}^N (\alpha_j \gamma_i^j / \xi_i) \xi_i \stackrel{(22)}{=} \xi_i. \end{aligned}$$

Consequently, ξ cannot be an extreme point if $\xi_p = 1$. Because $\xi_p > 0$, we must have $\xi_p \in (0, 1)$. We can propagate this

argument upstream to eventually show that $\xi_0 \in (0, 1)$, which contradicts that $\xi_0 = 1$ for all $\xi \in \Xi$, and the proof is complete. \square

Appendix B. Numerical Studies on AP Heuristic

B.1. Relaxing Assumptions 1 and 2

We provide details on how problem instances are generated for studies 1 and 2 in Section 4.2 alongside a quantitative analysis of our results. For each problem instance, we generate a tree-style graph as follows:

- (1) Uniformly sample the target total number of nodes, $n \in \{100, 200, 500, 1,000, 2,000, 5,000, 10,000\}$.
- (2) Uniformly sample the number of levels, $l + 1 \in \{2, 3, 4, 5\}$.
- (3) For each nonleaf level $k = 0, 1, \dots, l - 1$, sample a random integer f_k between 5 and 15.
- (4) Find F (e.g., by line search) such that $\text{round}(1 + f_0/F + f_0 f_1/F^2 + \dots + f_0 \dots f_{l-1}/F^{l-1}) = n$, where $\text{round}(\cdot)$ is the usual rounding operation.
- (5) Grow the tree from the root in a breadth-first way: for each node in level $k = 0, 1, \dots, l - 1$, assign to it c_i children, where c_i is randomly picked from $\{(\prod_{j=0}^k f_j/F^k - 1)^+, \prod_{j=0}^k f_j/F^k, \prod_{j=0}^k f_j/F^k + 1\}$.

With this procedure, we generate 5,000 trees. For study 2, for each of the 5,000 trees, we create a nontree counterpart by introducing additional edges into the graph. Specifically, for each pair of nodes that are in adjacent levels in the previously constructed tree, we assign them an edge with probability p_{arc} , which is randomly sampled from $[0, 0.01]$ for each instance. The total number of edges in the nontree counterpart comes out to have 0%–200% more edges than its base tree.

To generate cost, efficacy, demand, and attack-severity parameter values, we first normalize the per-unit demand loss cost $b = 100$. Then, for each instance, we do the following:

- (1) Generate (float-valued) inventory cost upper and lower bounds uniformly: $\underline{h} \in [0, 10]$ and $\bar{h} \in [10, 20]$ and $\delta h = \bar{h} - \underline{h}$.
- (2) For each node i , assign an inventory cost $h_i \in [\underline{h}, \bar{h}]$ to it. More specifically, h_i is uniformly sampled from $[\underline{h} + \frac{l(i)+0.5}{l+1} \times \delta h, \underline{h} + \frac{l(i)+1}{l+1} \times \delta h]$, where $l(i) \in \{0, 1, \dots, l\}$ is the level in which node i resides.
- (3) Uniformly sample the average efficacy $\bar{\rho} \in [0, 1]$. For each path p , generate $\rho_p \in [0.5, 1] \times \bar{\rho}$. Assign a flow-cost efficacy $100\rho_p$ to the path.

- (4) For each demand node i , the maximum demand is \hat{d}_i , which is uniformly sampled from $[0, 1,000]$. (This value is then fed into the optimization formulation $d_i(\xi_i) = \hat{d}_i \xi_i$.)

- (5) To determine the attack scale for each node i , first generate $\bar{\Gamma} \in [0, 0.2]$. Then, for each node i in this network instance, uniformly sample an integer Γ_i between $\text{round}(0.8\bar{\Gamma}c_i)$ and $\text{round}(1.2\bar{\Gamma}c_i)$.

Figures B.1–B.6 depict, either through scatterplots or standard box plots, the upper bounds on the AP heuristic suboptimality gaps we obtained for studies 1 and 2 for a varying number of nodes, tree depth, inventory cost parameters, attack-severity parameters, antibiotic efficacy, and tree-violating edges. It can be seen that suboptimality gaps are small and do not appear to have strong correlation with any of the parameter values tested, including the number of nodes and the attack severity.

B.2. Relaxing Infinite Capacity Assumption

To explore the AP heuristic's performance for the joint inventory and dispensing capacity optimization problem, (LLGC) in Section 5, we conducted numerical studies similar to studies 1 and 2.

Recall that (LLGC) differed from (LLG) by introducing capacity constraints (4e) for subsets of edges in the graph and the associated capacity variables and costs. We conducted two studies, in both of which we considered the same setup as in study 1. In the first of these two new studies, we generated these subsets of edges in a random fashion. In the second, we considered subsets of edges that had similar structure as in (4e)—specifically, sets of edges that all shared the same demand node. Both studies yielded similar results; we focus only on the latter in the remainder.

To introduce dispensing capacity, for each demand node, we partitioned its incoming edges to k equisized sets; k was sampled from $\{2, 4, 6, 8\}$. For each resulting set of edges, we introduced a capacity variable and a constraint as in (4e). The associated capacity cost coefficient was sampled as a fraction of the life-loss cost.

Table B.1 provides statistics of the upper bounds on the suboptimality gaps we obtained. Figure B.7 depicts the suboptimality gap bounds for varying numbers of nodes. As before, our analysis reveals that the suboptimality gap is small and does not grow with the size of the network.

Figure B.1. Upper Bounds on AP Suboptimality Gap for Varying Number of Nodes (Left: Tree; Right: Nontree)

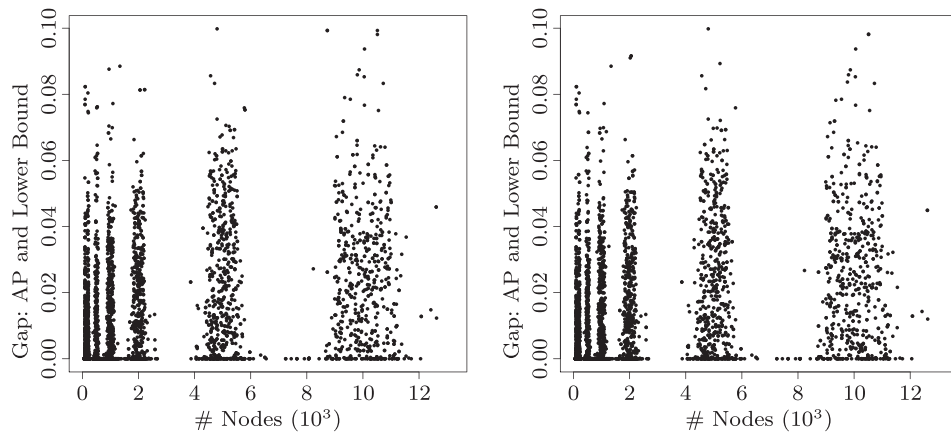


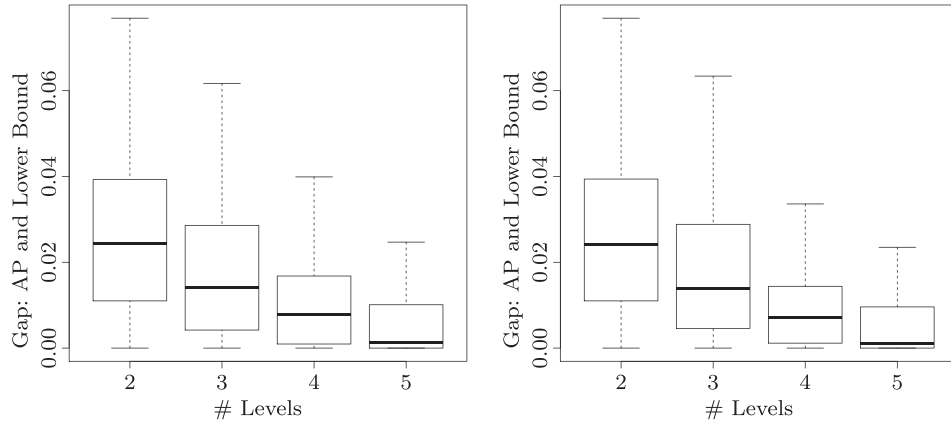
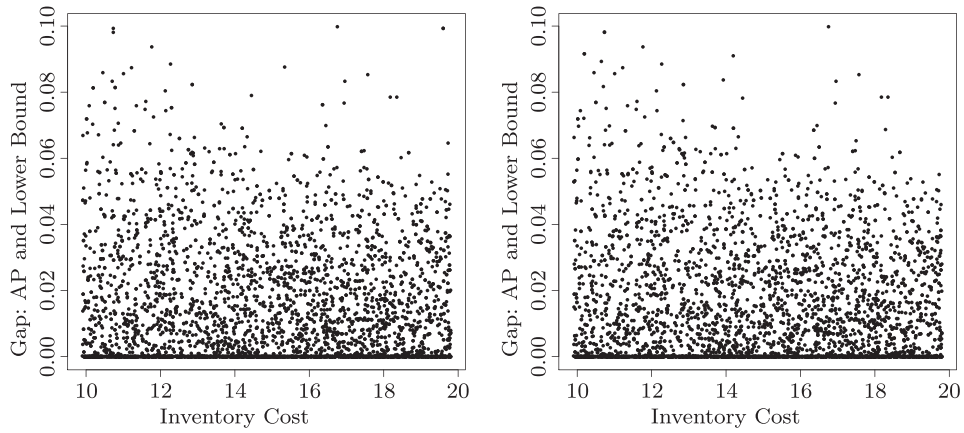
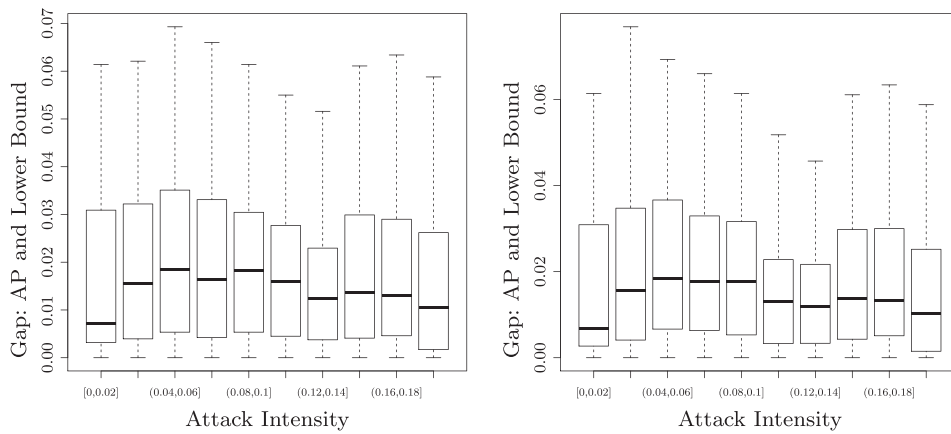
Figure B.2. Upper Bounds on AP Suboptimality Gap for Varying Number of Levels (Left: Tree; Right: Nontree)**Figure B.3.** Upper Bounds on AP Suboptimality Gap for Varying Inventory Cost, \bar{h} (Left: Tree; Right: Nontree)**Figure B.4.** Upper Bounds on AP Suboptimality Gap for Varying Attack Severity, $\bar{\Gamma}$ (Left: Tree; Right: Nontree)

Figure B.5. Upper Bounds on AP Suboptimality Gap for Varying Antibiotic Efficacy, $\bar{\rho}$ (Left: Tree; Right: Nontree)

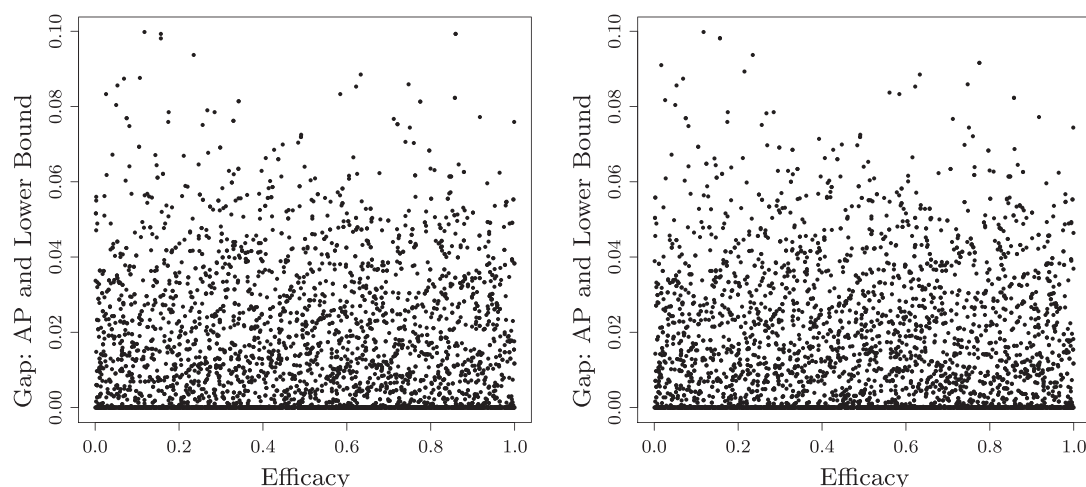


Figure B.6. Upper Bounds on AP Suboptimality Gap for Varying Network Complexity

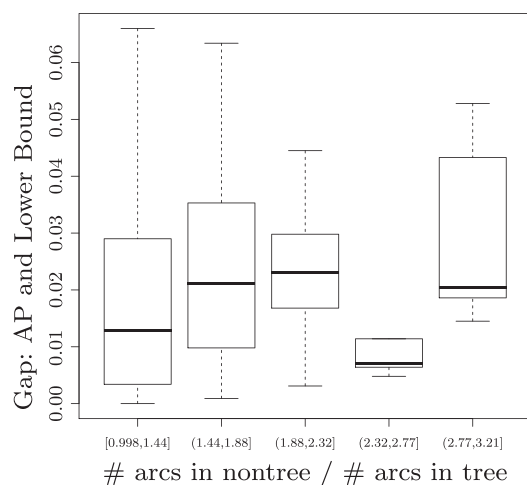
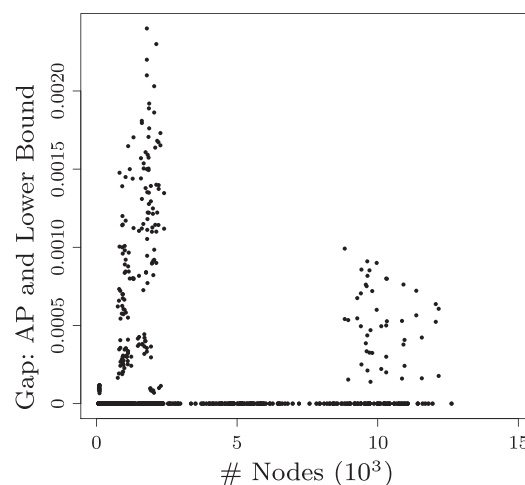


Figure B.7. Upper Bounds on AP Suboptimality Gap for Varying Number of Nodes, Capacitated Networks



Endnotes

¹ Unlike shipping MCMs, dispensing is often capacitated—a situation we address in Section 5.

² The maximum number of affected individuals for a bioattack is estimated by policymakers based on Census considerations, for example, population density; transmission and contagion characteristics, for example, the required quantity of spores to carry out an attack and their ability to remain aloft and travel further for aerosol attacks; and others. The estimation usually relies on experimental data; see, for example, our discussion in Section 6.

³ For a discussion of anticipated detection times in case of an anthrax attack, see Stroud et al. (2011).

⁴ Technically, this means that if (f, s) is optimal for (LLC), then there exists y such that (y, s) is optimal for the corresponding edge-based formulation and vice versa.

References

- Agócs M, Fitzgerald S, Alles S, Sale GJ, Spain V, Jasper E, Grace TL, Chernak E (2007) Field testing a head-of-household method to dispense antibiotics. *Biosecurity Bioterrorism* 5(3):255–267.
- Ahuja RK, Magnanti TL, Orlin JB (1993) *Network Flows: Theory, Algorithms, and Applications* (Prentice Hall, Englewood Cliffs, NJ).

- Alderson DL, Brown GG, Carlyle WM (2014) Assessing and improving operational resilience of critical infrastructures and other systems. Newman A, Leung J, eds. *Bridging Data and Decisions*, INFORMS TutORials (INFORMS, Catonsville, MD), 180–215.
- Alderson DL, Brown GG, Carlyle WM (2015) Operational models of infrastructure resilience. *Risk Anal.* 35(4):562–586.
- Alderson DL, Brown GG, Carlyle MW, Cox LA (2013) Sometimes there is no “most-vital” arc: Assessing and improving the operational resilience of systems. *Military Oper. Res.* 18(1):21–37.
- Alderson DL, Brown GG, Carlyle WM, Wood RK (2011) Solving defender-attacker-defender models for infrastructure defense. Wood RK, Dell RF, eds. *12th INFORMS Comput. Soc. Conf.* (INFORMS, Catonsville, MD), 28–49.
- Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Oper. Res.* 64(2):474–494.
- Assistant Secretary for Preparedness and Response (2010) National postal model for the delivery of medical countermeasures. Technical report, Department of Health and Human Services, Washington, DC).

- Atamtürk A, Zhang M (2007) Two-stage robust network flow and design under demand uncertainty. *Oper. Res.* 55(4):662–673.
- Atkinson MP (2009) Mathematical models of terror interdiction. Unpublished PhD thesis, Stanford University, CA.
- Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press, Princeton, NJ).
- Berman O, Gavious A (2007) Location of terror response facilities: A game between state and terrorist. *Eur. J. Oper. Res.* 177(2): 1113–1133.
- Bertsimas D, Goyal V (2012) On the power and limitations of affine policies in two-stage adaptive optimization. *Math. Programming* 134(2):491–531.
- Bertsimas D, Goyal V, Sun XA (2011) A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Math. Oper. Res.* 36(1):24–54.
- Bertsimas D, Iancu DA, Parrilo PA (2010) Optimality of affine policies in multistage robust optimization. *Math. Oper. Res.* 35(2): 363–394.
- Bravata DM, Zaric GS, Holty J-EC, Brandeau ML, Wilhelm ER, McDonald KM, Owens DK (2006) Reducing mortality from anthrax bioterrorism: Strategies for stockpiling and dispensing medical and pharmaceutical supplies. *Biosecurity Bioterrorism* 4(3):244–262.
- Brookmeyer R (2001) The statistical analysis of truncated data: Application to the Sverdlovsk anthrax outbreak. *Biostatistics* 2(2): 233–247.
- Brookmeyer R, Johnson E, Barry S (2005) Modelling the incubation period of anthrax. *Statist. Med.* 24(4):531–542.
- Brown GG, Carlyle WM, Wood RK (2008) Optimizing department of homeland security defense investments: Applying defender-attacker (-defender) optimization to terror risk assessment and mitigation. *Department of Homeland Security Bioterrorist Risk Assessment: A Call for Change* (National Academies Press, Washington, DC), 90–102.
- Brown G, Carlyle M, Salmerón J, Wood K (2006) Defending critical infrastructure. *Interfaces* 36(6):530–544.
- Caunhye AM, Nie X, Pokharel S (2012) Optimization models in emergency logistics: A literature review. *Socio-Econom. Planning Sci.* 46(1):4–13.
- Çelik M, Ergun Ö, Ben Johnson PK, Lorca Á, Pekgün P, Swann J (2012) Humanitarian logistics. Mirchandani PB, ed. *New Directions in Informatics, Optimization, Logistics, and Production*, INFORMS Tutorials in Operations Research (INFORMS, Catonsville, MD), 18–49.
- Chan TCY, Shen ZJM, Siddiq A (2018) Robust defibrillator deployment under cardiac arrest location uncertainty via row-and-column generation. *Oper. Res.* 66(2):358–379.
- Craft DL, Wein LM, Wilkins AH (2005) Analyzing bioterror response logistics: The case of anthrax. *Management Sci.* 51(5):679–694.
- Ergun Ö, Karakus G, Keskinocak P, Swann J, Villarreal M (2011) Operations research to improve disaster supply chain management. *Wiley Encyclopedia of Operations Research and Management Science* (John Wiley & Sons, Inc., Hoboken, NJ).
- Golany B, Kaplan EH, Marmur A, Rothblum UG (2009) Nature plays with dice? Terrorists do not: Allocating resources to counter strategic vs. probabilistic risks. *Eur. J. Oper. Res.* 192(1):198–208.
- Golany B, Kress M, Penn M, Rothblum G (2012) Network optimization models for resource allocation in developing military countermeasures. *Oper. Res.* 60(1):48–63.
- Guyton J, Kadlec R, Harjivan C, Farooqi S, Cavitt S, Buccina J (2011) A cost and speed analysis of strategies for prepositioning antibiotics for anthrax. Report, Institute of Medicine, Washington, DC.
- Hao M, Jin S, Zhuang J (2009) Robustness of optimal defensive resource allocations in the face of less than fully rational attackers. *Proc. 2009 Indust. Engrg. Res. Conf., Miami*, 886–891.
- Hupert N, Xiong W, King K, Castorena M, Hawkins C, Wu C, Muckstadt JA (2009) Uncertainty and operational considerations in mass prophylaxis workforce planning. *Disaster Medicine Public Health Preparedness* 3(S2):S121–S131.
- Iancu DA, Sharma M, Sviridenko M (2013) Supermodularity and affine policies in dynamic robust optimization. *Oper. Res.* 61(4): 941–956.
- Inglesby T, Ellis B (2012) Division of Strategic National Stockpile (DSNS) program review: A report from the Board of Scientific Counselors (BSC). Technical report, Center for Disease Control and Prevention, Atlanta.
- Kaiser J (2011) Taking stock of the biodefense boom. *Science* 333(6047): 1214–215.
- Kaplan EH (2012) OR forum—Intelligence operations research: The 2010 Philip McCord Morse lecture. *Oper. Res.* 60(6):1297–1309.
- King K (2012) Logistical models for planning and operating medical countermeasure distribution networks during public health emergencies. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- King K, Muckstadt JA (2009) Evaluating planned capacities for public health emergency supply chain models. Technical report, Cornell University, Ithaca, NY.
- Lazzaro G (2016) Tri-level optimization algorithms for solving defender-attacker-defender network models. Unpublished doctoral dissertation, Naval Postgraduate School, Monterey, CA.
- Lee EK, Chen C-H, Pietz F, Benecke B (2009a) Modeling and optimizing the public-health infrastructure for emergency response. *Interfaces* 39(5):476–490.
- Lee EK, Maheshwary S, Mason J, Glisson W (2006a) Decision support system for mass dispensing of medications for infectious disease outbreaks and bioterrorist attacks. *Ann. Oper. Res.* 148(1):25–53.
- Lee EK, Maheshwary S, Mason J, Glisson W (2006b) Large-scale dispensing for emergency response to bioterrorism and infectious-disease outbreak. *Interfaces* 36(6):591–607.
- Lee EK, Smalley HK, Zhang Y, Pietz F, Benecke B (2009b) Facility location and multi-modality mass dispensing strategies and emergency response for biodefence and infectious disease outbreaks. *Internat. J. Risk Assessment Management* 12(2–4):311–351.
- National Research Council (2008) *Department of Homeland Security Bioterrorism Risk Assessment: A Call for Change* (National Academies Press, Washington, DC).
- Nicholson A, Wollek S, Kahn B, Hermann J (2016) *The Nation's Medical Countermeasure Stockpile* (National Academies Press, Washington, DC).
- Office of Public Health Preparedness and Response (2016a) *Public Health Preparedness Capabilities: National Standards for State and Local Planning* (Center for Disease Control and Prevention, Atlanta, GA).
- Office of Public Health Preparedness and Response (2016b) *Strategic National Stockpile* (Center for Disease Control and Prevention, Atlanta, GA).
- Rios J, Insua DR (2012) Adversarial risk analysis for counterterrorism modeling. *Risk Anal.* 32(5):894–915.
- Simchi-Levi D, Schmidt W, Wei Y (2014) From superstorms to factory fires: Managing unpredictable supply chain disruptions. *Harvard Bus. Rev.* 92(1/2):96–101.
- Simchi-Levi D, Wang H, Wei Y (2018) Increasing supply chain robustness through process flexibility and inventory. *Production Oper. Management* 27:1476–1491.
- Simchi-Levi D, Wang H, Wei Y (2019) Constraint generation for two-stage robust network flow problem. *INFORMS J. Optim.* 1(1):49–70.
- Simchi-Levi D, Schmidt W, Wei Y, Zhang PY, Combs K, Ge Y, Gusikhin O, Sanders M, Zhang D (2015) Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces* 45(5):375–390.
- Stroud C, Viswanathan K, Powell T, Bass RR (2011) *Prepositioning Antibiotics for Anthrax* (Institute of Medicine, Washington, DC).

- Uichanco J (2015) A robust model for pre-positioning emergency relief items before a typhoon with an uncertain trajectory. Working paper, University of Michigan, Ann Arbor.
- U.S. Census Bureau (2000) *Population, Housing Units, Area, and Density: 2000* (U.S. Department of Commerce, Suitland, MD).
- U.S. Food and Drug Administration (2015) *Anthrax*. Accessed May 12, 2019, <https://www.fda.gov/vaccines-blood-biologics/vaccines/anthrax>.
- Wein LM, Craft DL, Kaplan EH (2003) Emergency response to an anthrax attack. *Proc. Natl. Acad. Sci. USA* 100(7):4346–4351.
- Wilkening DA (2006) Sverdlovsk revisited: Modeling human inhalation anthrax. *Proc. Natl. Acad. Sci. USA* 103(20):7589–7594.
- Wilkening DA (2008) Modeling the incubation period of inhalational anthrax. *Medical Decision Making* 28(4):593–605.
- Zaric GS, Bravata DM, Holty J-EC, McDonald KM, Brandeau ML (2008) Modeling the logistics of response to anthrax bioterrorism. *Medical Decision Making* 28(3):332–350.
- Zhuang J, Bier VM (2007) Balancing terrorism and natural disasters? Defensive strategy with endogenous attacker effort. *Oper. Res.* 55(5):976–991.

David Simchi-Levi is a professor of engineering systems at the Massachusetts Institute of Technology. His research currently focuses on developing and implementing robust and efficient techniques in supply chain and revenue management. He has published widely in professional journals on both practical and theoretical aspects of logistics and supply chain management.

Nikolaos Trichakis is the Zenon Zannetos (1955) Career Development Professor and an associate professor of operations management at the MIT Sloan School of Management. His research interests include optimization under uncertainty and data-driven optimization and analytics with application in healthcare, supply chain management, and finance.

Peter Yun Zhang is a PhD candidate at the Institute for Data, Systems, and Society at the Massachusetts Institute of Technology. His research interests include optimization and its applications in supply chain management and public policy.