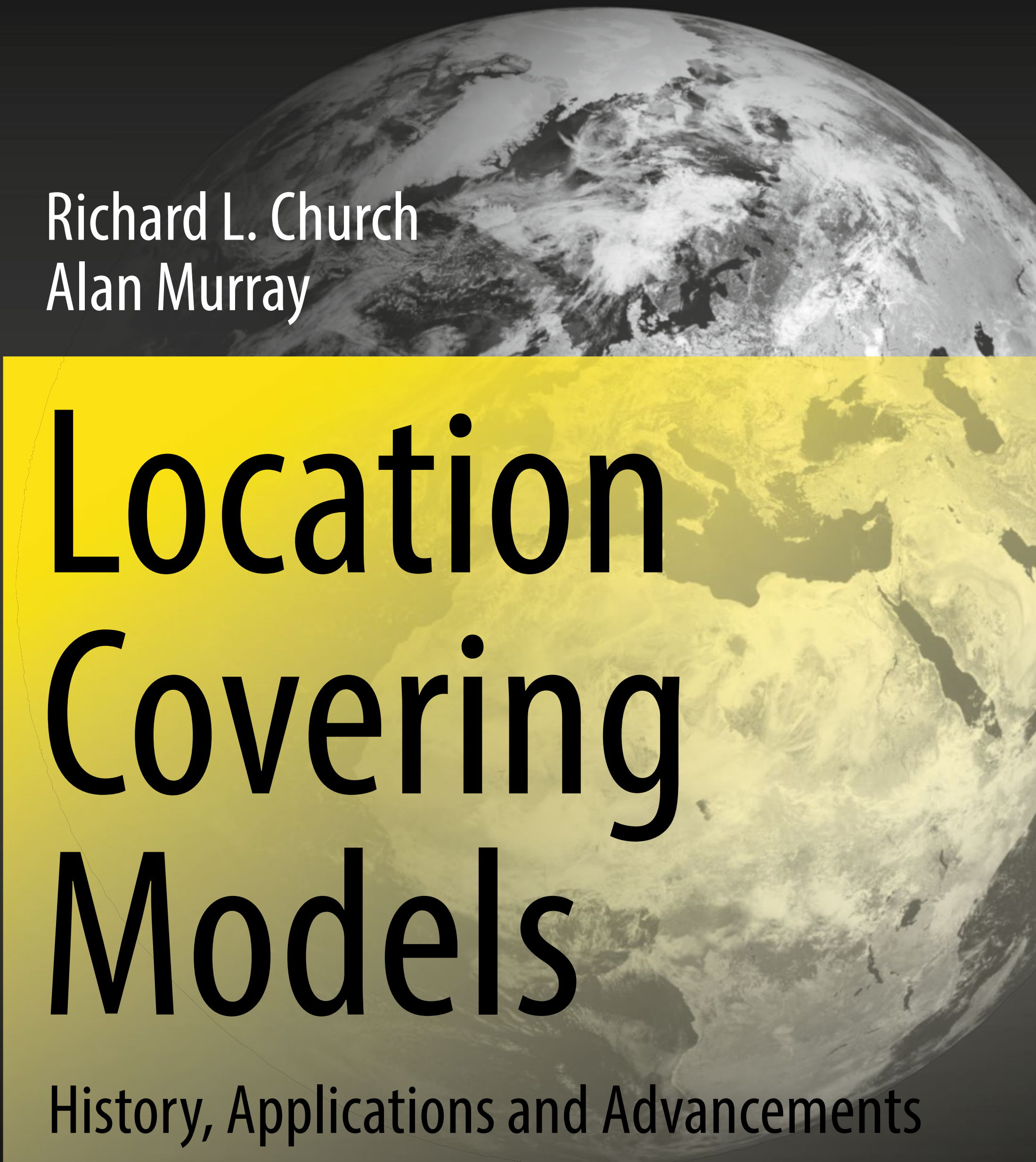
Advances in Spatial Science



Richard L. Church
Alan Murray

Location Covering Models

History, Applications and Advancements

Springer

Advances in Spatial Science

The Regional Science Series

Series editors

Manfred M. Fischer
Jean-Claude Thill
Jouke van Dijk
Hans Westlund

Advisory editors

Geoffrey J. D. Hewings
Peter Nijkamp
Folke Snickars

This series contains scientific studies focusing on spatial phenomena, utilising theoretical frameworks, analytical methods, and empirical procedures specifically designed for spatial analysis. Advances in Spatial Science brings together innovative spatial research utilising concepts, perspectives, and methods relevant to both basic science and policy making. The aim is to present advances in spatial science to an informed readership in universities, research organisations, and policy-making institutions throughout the world.

The type of material considered for publication in the series includes: Monographs of theoretical and applied research in spatial science; state-of-the-art volumes in areas of basic research; reports of innovative theories and methods in spatial science; tightly edited reports from specially organised research seminars. The series and the volumes published in it are indexed by Scopus.

More information about this series at http://www.springer.com/series/3302

Richard L. Church • Alan Murray

Location Covering Models

History, Applications and Advancements

Springer

Richard L. Church
Department of Geography
University of California
Santa Barbara, California, USA

Alan Murray
Department of Geography
University of California
Santa Barbara, California, USA

ISSN 1430-9602 ISSN 2197-9375 (electronic)
Advances in Spatial Science
ISBN 978-3-319-99845-9 ISBN 978-3-319-99846-6 (eBook)
https://doi.org/10.1007/978-3-319-99846-6

Library of Congress Control Number: 2018956344

© Springer International Publishing AG, part of Springer Nature 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This text is of our own design. We have attempted to be as complete as possible, but have had to make difficult choices about material to include in order to keep the length to a manageable size. We alone are responsible for any significant omissions.

We are deeply indebted to our wives, Carla and Patty, who have once again had to put up with us dedicating time to writing a text. We are truly grateful for their support, encouragement, and inspiration.

We also want to thank Barbara Fess of Springer for her patience and encouragement, as this endeavor took considerably longer than originally anticipated.

Finally, we wish to acknowledge Chuck ReVelle and Costis Toregas for their groundbreaking work in location set covering, an inspiring beginning to the subject of coverage and site selection. We also want to acknowledge Lou Hakimi and many others who help establish the field of location science. This work is dedicated to them.

Contents

Chapter 1
Location Modeling and Covering Metrics

1.1 Location Science

The field of location science is firmly rooted in several substantive developments, including the ground-breaking work of von Thunen (1826), Launhardt (1872), Weber (1909), Hotelling (1929), Hoover (1948, 1967), Christaller (1933), Lösch (1954), Weiszfeld (1937), Isard (1956), Moses (1958), Cooper (1963, 1964), Manne (1964), Hakimi (1964, 1965), Buffa et al. (1964) and Toregas et al. (1971). These authors may be considered founding fathers of location science, and they dealt with problems involving the competitive uses of land and land allocation, the location of industrial and communication facilities, the spatial arrangement of retail centers across a landscape, the location of competitors and competition through pricing, the layout of factory space, and the early use of computers in structuring and solving location problems. Since these early contributions, the field has expanded into new areas of application, new theoretical models, specialized solution approaches, and conceptual/technical forms of modeling location decisions and representing the spatial domain within Geographical Information Systems (GIS). Finally, as the field of location science has matured so too have the applications in both the public and private sectors.

The modern era of location science began with the development of quantitative models, in particular the optimal location of a production facility and associated transport investments (Launhardt 1872; Weber 1909) and explaining the allocation of land using economic and transport principles (von Thunen 1826). There are many branches of this field, and they are most often defined in terms of a unifying objective or construct. The central theme of this book involves the location of one or more facilities or objects in order to provide some type or level of coverage. Coverage is usually defined within the context of a performance standard. A number of examples of coverage standards are given in Table 1.1, but consider in particular the case of cellular phone service. A cell phone needs to be within a feasible communication distance of a service tower. This means that cell phone towers need to be placed so

© Springer International Publishing AG, part of Springer Nature 2018
R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_1

Table 1.1 Coverage standards examples

Standard	Context
3–50 km	Cellular antenna
5–6 min	Emergency 911 call
70 dB audibility	Outdoor warning or message
400 m	Reasonable walking distance for bus access
70 miles	Essential air service access for rural communities
800 m	Suitable access for rail/subway
Species presence	Nature reserve design
1 day ride (horseback)	Mail delivery
120 miles	Doppler radar moisture detection
Same day service	Express package delivery
6 h	Search and rescue
1500 m	Visibility distance of camera mounted on tower
60 min	Aeromedical response for trauma care
3 min	Cardiac arrest response
1000 m	Wildlife road crossing
150 feet	Street light intensity

that subscribers can access towers. Since cell phone towers cost money to build and maintain, the system must be designed to provide needed coverage while using the fewest towers possible. That is, cover as much regional demand for service as possible with minimal operational (tower) costs.

Although many of the concepts used in location covering models are simple and straightforward, any application may require additional elements that will likely increase problem complexity. For example, in the location of cell towers, the main objective may be to keep costs low by minimizing the number of towers that are sited. But, there are three complicating factors in the design of a network of cell towers that "cover" customers. First, even though cell phone transmission seems simple, it may be disrupted by intervening buildings, trees, terrain, weather and other barriers. A second issue is that signal strength degrades with distance. Both factors complicate service estimation associated with which areas are within "communication reach" of a possible tower site. A third issue is that the capacity of a tower to handle simultaneous calls is limited to the number of frequencies that have been assigned to that tower. Although it may be possible for a cell phone to communicate with a tower over a distance of twenty to thirty miles, cell towers are often located much closer to each other because of capacity limitations in handling calls and/or data transmission. A fourth issue is that when towers are sited closer together, the strength of the transmission signal is often decreased to reduce the possibility of interference between nearby towers using the same frequencies. Collectively, providing cell phone coverage involves the placement of towers in a way that most customers can make calls (i.e., people are within communication range of a cell tower when they need service), ensuring that the customer base relying on a given tower will not overwhelm the capacity of that tower, and assigning towers a set of frequencies that do no conflict with neighboring towers (see Erdemir et al. 2008).

Cellular tower placement is very representative of the complexities involved in coverage modeling. Coverage of a potential tower site is not trivial as it is predicated on spatial configuration and proximity. Service facilities may have a limited capacity. Beyond this, cellular tower siting is also a good example of defining and representing the problem domain. Most cell phone calls may be placed (or received) anywhere in a region, including along a highway connecting two populated areas, at a park or playground, on a boat floating in a lake, on a hiking trail, etc. This means that, conceptually, we want to potentially cover an entire region consisting of residential and commercial land parcels, schools, roadways, waterbodies, open space and the like. Potential cellular tower sites may consist of a finite set of points (e.g., on tall buildings, mountains and other high points on the landscape) or across continuous space that represent regions of feasible placement. Such detail means that the complexity of solving a coverage problem can be substantial.

The basic underlying motivation in this book is that we wish to locate one or more facilities (or objects/services) so as to provide service coverage of demand in an efficient manner. Covering is conceptually a very simple locational construct, however, there are many different ways in which to define and represent the provision of cover within a modeling and application domain. Further, most location covering models are NP-hard, and as such may be difficult to solve to optimality. A wide range of applications and model structures are addressed in this book. We present not only an introduction to this classic subject area of location science but provide as well an in-depth review of the main types of covering models and applications. Simply put, modeling coverage has expanded from its historical roots in the 1970s to a very large literature involving engineering, computer science, geography, management science, operations research, health sciences, as well as many other disciplines. The next section reviews standards-based covering constructs. Following this, site selection context and history is provided. The chapter then discusses a number of location design problems that have been addressed using a covering construct.

1.2 Standards Based Coverage

As suggested previously, location science is comprised of a number of formal constructs that represent general areas of application. These constructs are built around specific goals or objectives of interest. For example, the Classical Plant Location Problem involves siting a set of production facilities while minimizing the costs of product distribution, costs for building plants, and the costs of product manufacture (Manne 1964). The overarching objective is therefore one of maximizing system efficiency through the minimization of associated costs in production and distribution. This classic construct fits the needs of many private sector businesses and is considered to be a fundamental, classic model of location science. Whereas, efficiency, return on investment, and profit often characterize the goals of private sector companies, the public sector operates with wider sweeping goals of providing good service and addressing equity. In that vein, models like the Classical Plant Location Problem do not fully address the concerns of the public. Hakimi (1964) in a ground breaking article defined two location constructs: (1) minimize total weighted distance in serving a set of demand points, and (2) minimize the maximum distance in serving anyone. Whereas the first characterizes a goal for efficiency and accessibility, the second involves a concern for the individual that receives the poorest level of service. Hakimi (1964) described the problem of locating a switching center for a local phone network, where one would want to minimize the total length of wire in connecting each customer to the switching center. This is equivalent to minimizing the total weighted distance. However, when locating something like a police station or a hospital among a set of communities, Hakimi (1964) suggested that we should locate the station or hospital such that the maximum distance to any community is minimized. He called such positioning strategies medians and centers, respectively. Whereas the median location problem seeks efficiency, the center problem seeks fairness. This is viewed as fairness in the sense that the objective addresses those who experience the worst degree of access (in terms of furthest distance).

Although positioning services like hospitals and police stations by minimizing the maximum service distance is an attempt to be a fair as possible, attention to this alone may have a significant impact on all of the other demand to be served. In addition, this farthest or maximum service distance is often defined by one demand or community and may not be a measure of the true value of system service when considering everyone. For example, in urban areas, fire services are graded in their ability to respond to most fires within a desired response time, say 5 min. To meet this design requirement, fire stations need to be located so that most neighborhoods are within a mile and a half of a fire stations. This type of requirement, response within 1.5 miles, is a standards based service. A standards based service is often more meaningful than focusing only on those who receive the worst case of access, as defined by the maximum distance or time that anyone has to travel for service access. If the desired standard is service within 1.5 miles, how meaningful is it to focus on cases experiencing worst case access of 5 miles (or 15 min), as an example? Most would argue that although the worst case service level is of interest, what is important is that as many people as possible are provided a level of service that is meaningful, in this case service within 1.5 miles. For example, the quintessential problem might be to arrange fire stations in such a manner that all neighborhoods receive service within 1.5 miles (or 5 min response time). This problem is called the location set covering problem, and involves the search for a solution which "covers" or serves all within the standard while minimizing the resources required to provide coverage for all. This is the first true location covering problem and was first defined by Toregas et al. (1971). The underlying idea is that as long as neighborhoods and business districts are provided with "coverage", whether it is a 2 min or a 5 min response, service will be considered acceptable. But what if resources do not exist to cover or serve everyone with the desired standard? Then, we might turn our attention to providing as many people as possible with such a level of service. That is, maximize the demand that is provided coverage while spending no more than an allowable budget. This second construct is called the maximal covering location problem and was first posed by Church and ReVelle (1974). This does not mean that we are not interested in equity and do not have concern for those who are not well served, but that we want to provide what is deemed to be good service to as many as possible. There are ways we can address equity within a covering model, but the principal underlying element is that there is a standard that can be used to define if a suitable level of service has been provided and that the objective is to allocate resources in such a manner that everyone or as many as possible are provided such service coverage. Even though early covering models were suggested as a means for allocating and locating public services, their use and areas of application have expanded in both public and private sector contexts. Simply put, covering models address important standards-based problems, including siting cell phone towers, fire stations, surveillance cameras, and retail establishments.

Locating agents, stations, cameras, cell towers, etc. to cover demand is the central issue of this book. We address how coverage is modeled, the nuances of different problem settings, as well as how models have been expanded to address many different problems, even ones which do not appear to involve covering at first glance. One of the main themes in this book is that there can be a wide variety of problem domains. For example, is the region or object being covered defined by a continuous, bounded domain or is it defined by a finite set of discrete points? Are facilities, activities, or objects that are being located restricted to specific points or is their placement unrestricted within a continuous domain? Does demand for coverage vary spatially or is it homogeneously distributed? Is access in standards based service defined by a three dimensional view shed, are their limits to how sound travels that impacts audibility, or does travel only occur along road segments of a network? Are facilities logical objects, do they represent specific product types, or are they physical "bricks and mortar" entities? Are there a set of standards, defining perhaps good service, better service and best service rather than one single standard of coverage? Is the service extent defined in advance or is it to be determined endogenously? Can covering facilities be congested, altering whether a standard of service has been met? All of these nuances and more are addressed in the chapters included in this book.

Even though location science is a relatively new sub-discipline, principally expanding within the era of computers and operations research methods, standards of need involving distance or time limits, as well as other metrics, have been a part of the human fabric of civilization for centuries. The next section presents a historical perspective, noting elements of development where coverage standards or limits defined the very notion of how specific systems evolved or how specific problems were addressed. This historical take is meant to firmly establish the fact that desired service standards are at the very nature of things. Coverage is an important concept that has promoted safety, ensured service provision and relates to other functions in many societies. After this selective, but historical venture, the remaining sections of this chapter are devoted to describing a wide variety of problem application domains, all approached through the use of a location covering model.

1.3 Site Selection Context and History

It is likely that the first location problems were collectively solved by small groups of hunters and gatherers. They dealt with locating encampments and clearing needed trails. Important criteria considered was how easily an area could be cleared, whether resources could be efficiently accessed, and could sites be easily protected against intruders. There is no doubt that some location decisions were not very successful. For example, several of the California missions sited by Franciscan missionaries from Spain in the late 1700s and early 1800s had to be moved, including three because of flooding issues (Santa Cruz, San Gabriel, and Santa Clara), two because of unreliable water sources (San Diego and San Antonio), one after it was destroyed by an earthquake (Mission la Purisima), and another when the weather proved to be inhospitable (San Francisco). Mann (2005) describes issues with chosen locations for early cities in the Americas, from the Yucatan to Mississippian settlements. It is safe to assume that early failures heightened interest among societies and industries in making location decisions that promoted safety and security, enhanced livability and survivability, and eventually increased profits and productivity.

Military movements were examined in ReVelle and Rosing (2000), providing a post analysis of the Roman army deployment during the reign of Constantine as well as the British naval deployment in the early 1900s. The Legions of Rome had dwindled from 50 to 25 by the time Constantine came into power. The 25 legions were grouped into four field armies. The problem in deployment involved keeping a hedge on how far an army might be from the next uprising or enemy incursion. The British suffered a similar decline in naval resources and needed to consider a strategy in positioning only four battle fleets in order to protect their interests. ReVelle and Rosing (2000) demonstrated that the strategic positions and deployment of military forces could be structured using a covering framework, and further showed that the actual deployment solutions of Constantine and the British could have been improved.

The early western expansion of the United States involved strategic placement of military forts. These forts played a major role in providing protection and safe refuge to a surrounding area. For example, the City of Minneapolis, MN began with the development of Fort Snelling in 1819. Minneapolis is located at the only natural waterfall along the Mississippi River. This was a strategic location because the waterfall and flow volume provided energy for mills. In the latter 1800s virtually all wheat grown in the US was transported to Minneapolis for milling (most of it hauled by train). Figure 1.1 depicts the location of Texas forts in 1849. The systematic spacing of locations from north Texas to the Rio Grande River and the Gulf is evident, clearly suggesting the strategic focus of protection along the western front. Many of the settlement patterns of today in fact reflect a variety of site

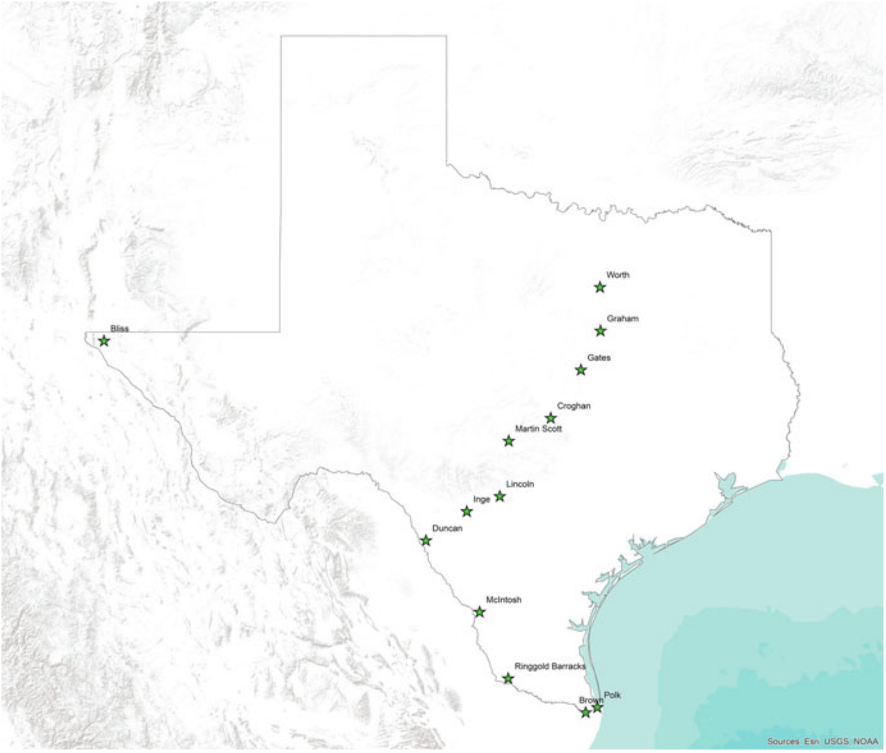


Fig. 1.1 Fort locations in Texas in 1849

selection objectives and metrics, include taking advantage of local resources (e.g., energy generation at St. Anthony Falls on the Mississippi River), minimizing costs of transportation (e.g., proximity along waterways), maximizing protection (e.g., the citadel at Quebec City, Quebec), and providing accessible services (e.g., Sutter's Fort in Sacramento, CA that supplied miners during the gold rush of the 1850s).

Embedded in this history of location is the fact that many site selections were periodic in position and reflected distance ranges of significance, a form of a covering or standards-based structure. For example, the 21 Spanish missions in California were placed approximately 30 miles apart along a 600 mile trail so that one could travel between one mission and the next in a "hard days ride". This periodicity can be found in other transportation/location examples, even during the era of the Pharaohs in Egypt (Church and Bell 1988). As another example, stage coach stations along the Butterfield stage route of the late 1850s were positioned approximately every 20 miles (a total of 139 along a 2800 mile route from St. Louis, Missouri to San Francisco). For most of the route, a stage coach would be no further than approximately 10 miles from a station (less than a 2 h journey). The route itself was chosen so that it would be passable in both winter and summer, by leaving St. Louis in a south westerly direction until it reached El Paso, Texas, crossing the

Rio Grande, and then heading principally in a westerly direction to Southern California, where it then turned north to San Francisco. The distance range between stations also reflected a pragmatic need to change teams of horses every twenty miles or so. If there was a breakdown or a horse was injured, one could reasonably make it to the next station by nightfall.

In the early 1930s Walter Christaller developed a theory of settlement location defined upon a theoretical triangular lattice of hamlets/villages. This lattice represented an unbounded plain of uniformly fertile soil and plentiful water. Distance reflected what a team of horses could travel in 2 h. It was reasoned that farmsteads would be no further than 2 h away from their closest village using a horse-drawn wagon. This allowed farmers to leave in the morning, travel to their market/service town, purchase supplies, and return in time for the main noontime meal. Thus, Christaller (1933) built into his theory the notion that settlements would be periodic in position. He also argued that retailers, in order to be successful, also needed a minimum number of customers or sufficient market size, and these market thresholds varied by the type of business. For example, we know that in our modern society people frequently purchase groceries and gasoline, whereas jewelry shopping is less frequent. To stay in business, a grocery store needs a smaller market than what might be needed for a jewelry store. To account for different market sizes necessary to support retailers of different products, Christaller (1933) reasoned that there would be a hierarchy of retail centers. This means that low level centers provide basic services, such as groceries and gasoline noted above, and higher level centers provide more specialized goods and services, like automobile dealerships and jewelry stores in the context of our previous examples.

Christaller (1933) used the “central place” hierarchical patterns to explain the location of settlements in Southern Germany. Others have applied central place theory to other farming regions like Iowa (Berry et al. 1962). Even though some of the underlying assumptions made in the development of central place theory rarely hold true, two of the concepts remain particularly important and relevant today. First, there is the fact that a retailer requires at least a minimum number of customers (called a market threshold) to remain in business. Second, it is true that customers are willing to go only so far for a particular type of product or service on a regular basis (called the range of a good). Accordingly, one can boil down Christaller’s theory as based upon two principal concepts of proximity: (1) the distance (or travel time) at which a given center would generate enough customers to be economically viable; and, (2) the distance (or travel time) beyond which customers will no longer be willing to travel to a facility on a regular basis. Storbeck (1988, 1990) showed that a coverage framework could be used to structure a model of central place theory based upon threshold and range concepts.

Even though Christaller and others like the Spanish missionaries used a periodic distance in facility placement that reflected basic needs (e.g. like being no further than a days’ ride on horseback), such factors are still relevant today. For example, a retail trade zone can be defined as the region that would be served by a given store location. Salvaneschi (2002) describes an approach used in branch store location: "... a market should be planned so that all its [retail trade zones] overlap by a certain percentage." He added that "... all stores should be impacted slightly; otherwise there is no guarantee that all customers will be properly served." When describing the McDonald's location strategy for Los Angeles, Salvaneschi (2002) states that "... the plan was to form a constellation of stores ... locating stores approximately 2–2 and a half miles away from each other." In fact, McDonald's specifies an address and not a territory for most of their franchisee participants. Of course, distances are not the only indicator of "reach" for a facility. Other factors or impedances may also be taken into account, like speed limit, level of congestion, etc. associated with deriving an effective distance. Nonetheless, the primary market of a store in densely populated areas is often based upon a predefined market size or proximity standard, just as distance and demand limits the effective coverage of a cell tower.

Beyond the notion that many facilities are located in periodic arrangements and reflect the effective reach of a facility or its services, like forts, market centers and stagecoach stations, one must understand the underlying standard for each problem, reflecting the desired range to be met. Examples of this abound. We wish to view as much of the forest as possible from a set of towers, so that fires can be detected. We want to place stations along a route so that horses can be relieved every few hours. Of course there are many others too. In the remaining sections we describe different areas of research and application that have involved the development and use of a covering model.

1.4 Surveillance, Sensors, and Warning Systems

A number of metrics can be used in determining the "effective service distance" of a facility. An interesting example is that developed by Agnetisa et al. (2009). Their problem involved the surveillance of a meandering river with the location of sensors (radar units) along the river. In monitoring traffic, or even an incursion along a river, they wanted to place sensors so that all segments of the river were within the field of view of at least one sensor. Each sensor had a maximum distance of view, or ability to detect an incursion. They represented this field of view as a disc with a given radius. Their intent was to locate a set of sensors (discs in this case) capable of monitoring all segments of the river while minimizing cost.

Surveillance resources are often positioned along a front (like a coastline), or to an area or region. A historical example of such a problem context is that the US Army deployed 11 observation towers along a 40 mile segment of the coastline of Delaware in order to detect enemy vessels during World War II in order to help protect key industries upstream on the Delaware River. Having towers located close together allowed observers at two neighboring towers to triangulate an accurate position for directing defensive fire. The distance of human observation has also played a major role in a number of other defense and security problems. Perhaps the most well-known is the use of lookout towers for early detection of forest fires. In 1910 there was a 3 million acre fire in Washington, Idaho, and Montana. It is said that the smoke from this fire reached Washington, D.C. In response to this devastating event, the US Forest Service issued a rule making local communities responsible for the costs of firefighting. This led to a high priority being placed on early detection and fire suppression. In order to detect forest fires, a vast network of fire lookout towers was built across the forested regions of the western United States. In Idaho alone nearly 1000 lookout towers were deployed. Such towers were often 30–100 feet or more in height and were usually placed on high ridges and peaks in order to maximize the size of the viewshed of each tower. An obvious objective was to place towers so that as much of the landscape was within view (covered) by the network of towers. Goodchild and Lee (1989) were the first to propose this as an optimization model using a covering framework.

Surveillance systems can be quite massive, as the previous example suggests. Another example is the Ground Observer Corps of the U.S. Air Force in the 1950s during the height of the cold war. They used hundreds of thousands of volunteers to man over 16,000 posts scattered along the gaps in radar coverage to detect a possible bomber attack from the Soviet Union. Many other examples of sensor placement are based upon a covering construct, and in some cases enhanced security is provided when multiple sensors can detect an intrusion at a given location. This means that it may be necessary to deploy facilities in such a way as to provide backup cover (Hogan and ReVelle 1986), multiple cover (Church and Gerrard 2003), and expected cover (Daskin 1983).

Although these historical examples show that visual surveillance has often taken considerable resources to implement and was deployed for a variety of reasons, visual surveillance is even more commonplace today. Visual surveillance is used to enhance security, prevent theft, and recreate crime scenes. Modern systems rely less on people and more on a variety of equipment, including satellites, cameras, motion sensors, image processing software, and face recognition algorithms. These systems are often connected with a communications network to a central monitoring or recording center. For example, with a CCTV (closed circuit television) network (i.e., video system), British police in London were able to piece together the activities of four bombers in a major terrorist attack in London, UK in 2005. But even more important is the design of surveillance systems so they "cover" an area at least cost (Murray et al. 2007). In a very recent monitoring project, Bao et al. (2015) use a maximal covering model to deploy camera-based watchtowers for fire detection in China. Monitoring can also involve other types of sensors, beyond cameras and radar. Examples include the design of monitoring networks for weather (e.g., Doppler radar), underground water pollutant plumes (Meyer and Brill 1988; Hudak and Loaiciga 1992), air pollutants (Hougland and Stephens 1976), and rain gages (Hogan 1990).

Another critical need in many communities is the ability to warn inhabitants of an impending emergency, like a tornado or an industrial accident. This can often involve a network of warning sirens. Sirens are often placed on top of buildings

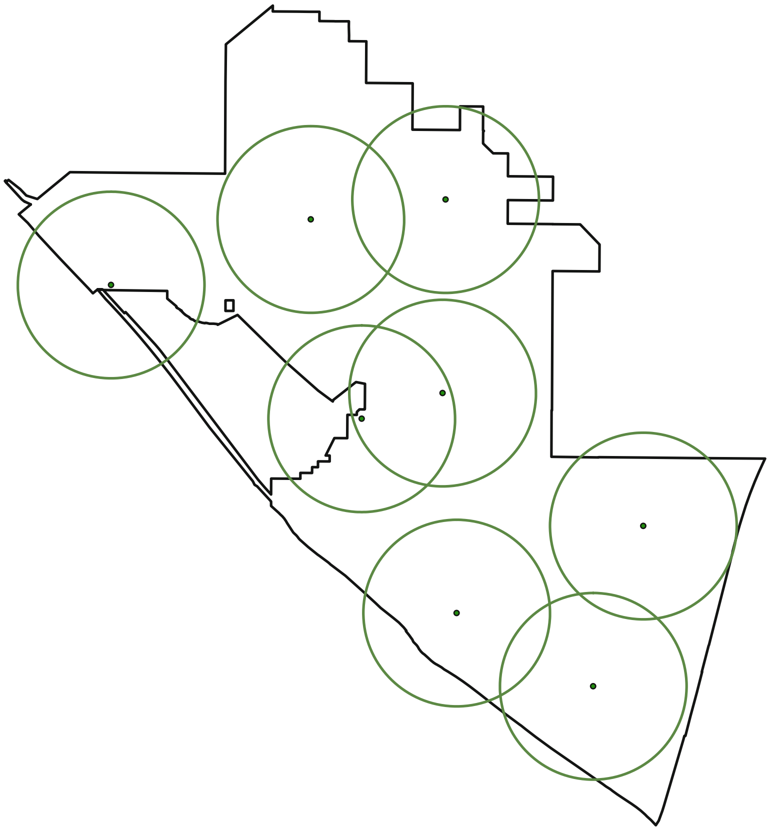


Fig. 1.2 Siren coverage for Huntington Beach, CA

and poles. Depending on their design, the distance at which they can be heard at a given decibel rating can be determined, as audibility decreases with distance from the siren. In this way a coverage pattern of a siren can be determined. In many cases coverage is circular, reflecting the distance which just meets a minimum decibel warning level. Some modern sirens are designed to follow the siren blast with a spoken announcement. For example, the City of San Francisco has recently replaced and updated an antiquated air raid siren system with a new Outdoor Public Warning System over the last 10 years. In all, the new system has 109 digital sirens and speakers. A number of these broadcast bilingual messages, some in English and Spanish and some in English and Cantonese. Figure 1.2 depicts the warning system of Huntington Beach, California, where the audible area for each siren is shown as circular. Modeling the efficient deployment of siren systems using a covering model started with the work of Current and O'Kelly (1992) and has been an area of considerable continuing research (Murray and Tong 2007; Murray et al. 2008; Wei and Murray 2014; Wei et al. 2014).

1.5 Emergency Response

Urban areas have experienced many significant fire catastrophes on a massive scale. In the United States, there were sizable city fire disasters in the latter 1800s and early 1900s. The Chicago fire of 1871 destroyed 18,000 buildings. The 1904 Baltimore fire consumed 1500 buildings in the central downtown. The great Atlanta fire of 1917 burned 1900 structures. The San Francisco earthquake and fire of 1906 wiped out 80% of the city. Much of what was not damaged by the earthquake was lost to fire. Such fires and destruction led to the development of better building standards, fire protection, and enhanced water availability through fire hydrant networks.

After a fire is ignited, it often takes only a few minutes to reach the point of flashover, which occurs when thermal conditions in a room are such that all combustible materials will ignite. This is when a small fire turns into a large fire, and begins to spread to other rooms. Besides heat, flashover is also associated with high levels of lethal gasses. Such events can easily occur within 4–6 min and makes the time to respond a critical issue. Once an alarm or call is made, a dispatcher must determine which crews are to respond. Notified crews often take a minute or so to assemble and take their places on their vehicles before travelling to a fire. Once on scene, it can also take time to assess the situation, set-up hoses and begin their rescue and attack. The response time model of Rand Corporation (Kolesar et al. 1975) estimates that it takes on average 3.1 min to travel to a fire when the travel distance is 1.5 miles. In general, it has been recommended that built up areas should have their closest station within this distance. For example, the City of Elk Grove, California established a maximal distance standard of 1.18 miles, where they estimate it will take no more than 5 min to respond (presumably including the time to assemble at the station, called the turn-out time, and the time to handle the call and send a dispatch request to a station) (Murray et al. 2012). Using such standards, cities today seek to locate fire stations in such a manner that as many structures and people are within their adopted time/distance response standards (Plane and Hendrick 1977; Schilling et al. 1980; Murray 2013). The majority of the yearly service costs are closely related to the number of stations used, as each requires a complement of fire fighters and equipment. In Los Angeles County alone the cost of fire service involves approximately 340 stations and costs more than $2 billion per year. Across the US, fire services are dispatched from more than 38,000 stations. Optimal deployment of stations in providing coverage can save money. In addition, coverage models can be used to optimally expand a system in growing communities as demonstrated in Murray et al. (2012).

Fire protection is but one element of public safety services. Another is the provision of police. Police departments are often organized by divisions or precincts, each a well-defined geographical area that includes a police station. For example, the Island of Manhattan is divided into 22 precincts. Stations usually contain holding cells, a booking area, changing rooms, offices, etc. Precincts are organized so that a police response from the station or defined patrol areas is kept within a maximal response time for the majority of emergency calls. Curtin et al. (2010) demonstrates how this deployment problem can be defined as a covering model, where they developed a decision support system to locate patrol activity centers (sectors) within police divisions in Dallas, Texas in order to improve response times to emergency calls.

Another major public service is that of emergency medical response. There are two general problems, one that involves responding to trauma cases and the second for other types of emergency calls. One of the first major applications of a coverage model for EMS occurred in Austin, Texas (Eaton et al. 1985). When locating EMS units, a major issue is calculating how often specific units are available to respond to a call and how often they are busy on another call. Daskin (1983) was one of the first to model the expected level of coverage for a system, relying on the use of a system-wide probability of busyness. This concept has been expanded in a number of ways, including the use of simulation, localized busyness estimates, queuing theory, and the development of a spatially defined queuing model called hypercube queuing. These approaches are explained in detail in Chap. 4.

ReVelle et al. (1976) were the first to propose a dual location problem where both hospitals and ambulances could be located so that their coordinated action could keep total service time within a standard. Their approach has been expanded to the location of trauma units and helicopters so that service can be provided within what has been called the golden hour of service (Branas et al. 2000). Issues of coordination, contingencies, and coordinating units (like ambulances and helicopters) are discussed in Chap. 3.

1.6 Nature Reserve Protection

Radar, lookout towers, and surveillance cameras all have physical limits to what can be detected. These limits of detection can be defined in terms of proximity, lighting, weather, presence of barriers, etc. Retail/service location strategies can involve time and distance limits as do public safety systems like EMS, police and fire response. But, there are other areas of application for covering models that are important and relevant in today's society. One of these is the protection of threatened and endangered species. From the 1980s, conservation biologists have been concerned with the location and selection of reserve sites. Such sites, which when taken together, are intended to protect as many species as possible (Csuti et al. 1997). With the exception of the model of Cocks and Baird (1989), many of the early attempts to develop site selection approaches relied on selection heuristics. For example, Kirkpatrick and Harwood (1983) suggested the following site selection strategy: select as the first site for the reserve that site which contains the greatest number of species of concern; for the second site in the reserve choose that site which contains the most species of concern that are not represented by the first site; for the third site pick the site which represents the most species not represented by the first two selected sites; continue this selection process until all species are protected. Many will recognize this as a greedy heuristic, which was first proposed for covering problems by Church (1974) and Chvatal (1979). But more important, the underlying concern was to select as few sites as possible while at the same time representing all species in the reserve system. Underhill (1994) was the first to recognize this to be a covering problem. Church et al. (1996) suggested that when resources did not exist to protect all species in a reserve system, then one could use a maximal covering framework. Since the mid-1990s, a relatively large number of covering models have been structured for the selection of a portfolio of sites as a biological preserve. These include stochastic elements such as uncertainty of species presence (Cocks and Baird 1989), site quality (Church et al. 2000), and threat of development and site degradation (O'Hanley et al. 2007). A review of such models can be found in Snyder and Haight (2016).

Many of the biological reserve site selection models do not consider the nearness of selected sites, but it might be important to not only represent species more than once within a reserve system, but also provide corridors of connectedness and nearby sites to support meta-populations (Williams 1998). This may mean that selection of one site for a reserve system may be contingent on selecting at least another nearby site as well. This type of contingent selection has been used in reserve design as well in other covering contexts (Williams et al. 2003). These topics are discussed at greater length in Chap. 3.

The simple reserve selection models, like that suggested by Underhill (1994) and Church et al. (1996), are logic-based covering models. This is because the intent focuses on what is contained within the site itself that is being covered, rather than those elements that are within some distance of the site, like the service area of a fire station. There are a number of other such logic-based covering models that have been defined in the literature. Each of these models involve the selection of a portfolio of elements such that altogether the portfolio contains all of the desired properties or as many of the desired properties as possible. We can think of the portfolio as covering a property when an item in the portfolio contains that property. For example, for reserve design, the portfolio represents the sites that have been selected and the items that are being represented are the species of concern. Other settings exist where the use of a logic-based approach for covering is applicable. For example, Klimberg et al. (1991) developed a model for scheduling inspections of possible violations of industrial facilities by the US FDA so that each possible violation would be inspected once in a 2-year planning horizon. In another logical design problem, Serra (2013) describes how to select a set of products with specific properties so that customers will find within this set a product matching many of their needs. This analysis technique is called TURF (Total Unduplicated Reach and Frequency analysis) and is approached as a covering problem. Cocking et al. (2009)

used a covering model to select dental color shade guides for teeth replacement, using real teeth as targets that dental shade guides should represent within a specific degree of accuracy. They found that a covering model identified a superior shade matching. Finally, covering models have been suggested as an aid in media selection for advertisers (Dwyer and Evans 1981).

1.7 Spatial Separation

While coverage has generally been viewed in positive terms thus far, there are situations where a standard may reflect an incompatibility, violation, danger, etc. That is, the notion of spatial proximity can indicate when facility siting, as an example, is unacceptable. From a modeling perspective, this can be viewed as a requirement for spatial separation. Moon and Chaudhry (1984) proposed a framework for models where the objective is to maximize the number of facilities being placed while maintaining that any two sited facilities are separated by at least a minimum prescribed distance. They called this the anti-cover location model. At first one might wonder whether such a problem construct is more theoretical rather than practical. One early suggested application of this approach involved the location of missile silos (keeping them apart would mean that a successful strike on one missile silo might not damage any nearby silos). Another possible application involved the location of hazardous materials, suggesting that they should be kept apart within a manufacturing site. Practically speaking, however, many applications for this covering construct exist. Recent applications of the anti-cover construct have evolved in many different ways and new innovative techniques have been developed to solve this type of problem. Recent applications of the anti-cover model have included habitat analysis (Downs et al. 2008; Church 2013; Church et al. 2015), examining liquor store market penetration (Grubesic et al. 2012), and assessing housing policy impacts (Grubesic and Murray 2008; Murray and Kim 2008) to name a few.

The definition of separation between facilities has been modified to suit the problem being addressed. An example of this involves the problem of scheduling harvesting units across a land holding (Murray and Church 1996). In one form of the problem harvest units are delineated in advance. Each harvest unit is usually less than a maximum allowable size and has a revenue value assigned to it based upon the inventory within the unit. Environmental restrictions often require that no two adjacent units be harvested in the same decade. Assume for simplicity that we are interested in one-time period, or decade, where we want to maximize revenue generated by determining which units to harvest such that no two harvested units are nearby to each other. Here the two units can be considered too close when they share a portion of their boundary, as an example. Other separation problems may allow spatial units to touch, but require that there be no overlap. For example, Grinde and Daniels (1999) place fabric cutting patterns on a roll of fabric using a covering construct. They have a set of patterns which are needed and are sewn together for a piece of clothing. Each pattern object can be oriented in only certain ways on the fabric so that when a set of pieces all are sewn together, it makes a presentable piece of clothing. The objective is to place as many sets of patterns on the fabric without overlap. Maximizing coverage of the fabric for this problem minimizes fabric waste.

Another problem that involves placing elements on a background (like fabric) involves automatic label placement on a map. Here, name labels, like the name of a river, must be placed along the object, but separated from other labels by a distance of separation. Such separation ensures legibility. Placing and sizing pivot irrigation systems is another example where spatial separation is important. The footprint of one pivot irrigation unit (map label, fabric cutout pattern, etc.) must not overlap with any others, but must be accomplished in such a manner that as much of the land is irrigated (or as many of the labels are placed on the map as possible, or the remaining fabric waste is minimized, etc.). All of these are examples of spatial separation requirements, where coverage models have been important. This topic is addressed in Chap. 5.

1.8 Expanded Locational Constructs

With the exception of the early research work on facility layout (see Buffa et al. 1964), most researchers have assumed that facility sites are points in a Cartesian plane (Launhardt 1872) or as points on a network (Hakimi 1964). Morgan and Slater (1980) suggested that a facility may be a structure on a network, like a path or a tree. They proposed several problems that involved finding a path of a network that maximized accessibility to all nodes of the network. Current (1981) expanded on this concept when he proposed the shortest covering path problem. He assumed that a path origin and a path destination were already known in advance. The shortest covering path is the shortest path on the network which starts at the origin and ends at the destination while traveling within a preset coverage distance of all network nodes. Current et al. (1985) proposed another construct called the maximal covering shortest path problem. These constructs have formed the basis for a wide variety of models formulated for transit system design (see for example: Curtin and Biba 2011; Murray 2003; Wu and Murray 2005; Matisziw et al. 2006; Laporte et al. 2011).

1.9 General Form of Coverage

For many applications, coverage can be defined as being provided or not. For example, is a neighborhood within the desired service range of a fire station or does a camera network provide complete area surveillance? There are circumstances, however, where coverage is not so easily defined. In fact, there may be degrees of coverage. For example, one may view service quality as a function of distance and quality level of the coverage provider. For EMS, units have been classified as Basic Life Support (BLS) and Advanced Life support (ALS), where coverage models allocate two vehicle types in order to provide everyone with at least one type of service level while maximizing the number of people provided ALS services (Eaton et al. 1985). Church and Roberts (1983) suggested that coverage should be defined over distance ranges to reflect defined standards like high, medium, and low service coverage. One of the reasons to suggest this construct is that even when a maximal time standard of 8 min is established for good EMS service, what about service within 8.10 min? Should that be counted as zero coverage, or an intermediate value of coverage (higher than zero and less than the desired, high-valued standard)? Church and Bell (1981) and Berman et al. (2003) have suggested that service values and coverage may degrade as a continuous function of increasing distance. Thus, the concept of coverage has been expanded to include problems where a discrete distance, time, visibility, or other metric is not so crisply defined with one simple cutoff. These issues are discussed in greater detail in Chap. 6.

1.10 Summary and Concluding Remarks

Location coverage problems can be found in a wide variety of activities and problem domains. These include public service delivery, transit system design, surveillance and monitoring systems, as well as biological reserve protection. Covering problems were addressed long before computational models were ever developed, and involved problems like protection (e.g., fort location), warning (e.g., fire lookout placement), administration, and retail location theory. Today, many of these problems are approached through the solution of a formal location covering model, using state of the art heuristics and algorithms. As outlined in Table 1.2, the goal of this book is to present representative examples of different types of covering models, problem domains, and to some extent solution approaches. This will involve probabilistic and stochastic coverage, logical coverage, quality of service coverage, and multiple-level coverage. Each chapter addresses a general area of coverage modeling or problem domain (e.g., discrete or continuous). We begin in Chap. 2 with the definition of the two classic models that underpin this subject of location science: the location set covering problem and the maximal covering location problem.

Table 1.2 Book coverage at a glance

Chapter	Topic	Major themes
1	Location modeling and covering metrics	• Location science • Standards based coverage • Site selection context and history • Surveillance, sensors, and warning systems • Emergency response • Nature reserve protection • Spatial separation • Expanded locational constructs • General forms of coverage
2	Classic beginnings	• Location set covering problem • Maximal covering location problem • Theoretical linkages • Fixed charges
3	Extended forms of coverage	• Multiple service • Existing service system • Site quality • Multiple objectives • Backup coverage • Coordinated systems • Hierarchical services • Multiple optima
4	Probabilistic coverage	• Reliable coverage • Expected coverage • Maximum reliable coverage • Queuing • Facility availability • Extensions
5	Anti-covering	• Separation context • Model construct • Mathematical structure • Relaxations and extensions • Inefficiency • Facets and more
6	Weighted benefit, variable radius, and gradual coverage	• Equity and implied value of service • Generalized maximal covering location problem • Expanded forms of generalized coverage • Endogenously determined coverage • Continuous endogenous coverage • Gradual coverage
7	Capture, capacities, and thresholds	• Maximum capture • Capturing/intercepting flow • Capacities • Thresholds • Franchise territory design
8	Continuous space coverage	• Problems • Formulations • Simplification and relaxation • Solution

(continued)

Table 1.2 (continued)

Chapter	Topic	Major themes
9	Disruption, protection, and resilience	• r-interdiction • Design sensitive to facility loss • Planning for greater levels of disruption • Issues when providing complete coverage • Coverage loss when supporting infrastructure is lost • Defensive coverage • Facility fortification • Applications
10	Coverage of network-based structures: paths, tours and trees	• Shortest covering path • Salesman and tour coverage • Maximal Covering Trees • Maximal coverage variants • Arc improvement • Applications
11	Grand challenges	• Big data • Developing better models and service assumptions • Problem transformation • Flexible and accessible location application software systems • GIS developers • Artificial intelligence and machine learning

References

Agnetisa A, Grandeb E, Mirchandani PB, Pacificic A (2009) Covering a line segment with variable radius discs. Comput Oper Res 36:1423–1436

Bao S, Xiao N, Lai Z, Zhang H, Kim C (2015) Optimizing watchtower locations for forest fire monitoring using location models. Fire Saf J 71:100–109

Berman O, Krass D, Drezner Z (2003) The gradual covering decay location problem on a network. Eur J Oper Res 151:474–480

Berry BJL, Barnum HG, Tennant RJ (1962) Comparative studies of central place systems. University of Chicago, Chicago, IL

Branas CC, ReVelle CS, MacKenzie EJ (2000) To the rescue: optimally locating trauma hospitals and helicopters. LDI Issue Brief 6(1):1–4

Buffa ES, Armour GC, Vollman TE (1964) Allocating facilities with CRAFT. Harv Bus Rev 42:136–158

Christaller W (1933) Central places in Southern Germany, Translated by CW Baskin (1966) Prentice-Hall, Englewood Cliffs, New Jersey

Church RL (1974) Synthesis of a class of public facility location models. PhD Dissertation, The Johns Hopkins University, Baltimore, MD

Church RL (2013) Identification and mapping of Map Habitat Cores (Chapter 9). In: Convis C, Craighead L (eds) Conservation planning from the bottom up: a practical guide to tools and techniques for the twenty-first century. ESRI Press

Church RL, Bell TL (1981) Incorporating preferences in location-allocation models. Geogr Perspect 8:22–34

Church R, ReVelle C (1974) The maximal covering location problem. Pap Reg Sci Assoc 32:101–118

Church R, Gerrard R, Hollander A, Stoms D (2000) Understanding the tradeoffs between site quality and species presence in reserve site selection. For Sci 46:157–167
Church RL, Bell TL (1988) An analysis of ancient Egyptian settlement patterns using location-allocation covering models. Ann Assoc Am Geogr 78:701–714
Church RL, Gerrard RA (2003) The multi-level location set covering model. Geogr Anal 35:277–289
Church RL, Roberts KL (1983) Generalized coverage models and public facility location. Pap Reg Sci 53(1):117–135
Church RL, Stoms DM, Davis FW (1996) Reserve selection as a maximal covering location problem. Biol Conserv 76:105–112
Church RL, Niblett MR, Gerrard RA (2015) Modeling the potential for critical habitat. In: Eiselt HA, Marianov V (eds) Applications of location analysis. Springer, Cham
Chvatal V (1979) A greedy heuristic for the set-covering problem. Math Oper Res 4:233–235
Cocking C, Cevirgen E, Helling S, Oswald M, Corcodel N, Ramelsberg P, Reinelt G, Hassel AJ (2009) Colour compatibility between teeth and dental shade guides in Quinquagenarians and Septuagenarians. J Oral Rehabil 36:848–855
Cocks KD, Baird IA (1989) Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia. Biol Conserv 49:113–130
Cooper L (1963) Location-allocation problems. Oper Res 11:331–343
Cooper L (1964) Heuristic methods for location-allocation problems. SIAM Rev 6:37–53
Csuti B, Polasky S, Williams PH, Pressey RL, Camm JD, Kershaw M, Kiester AR, Downs B, Hamilton R, Huso M, Sahr K (1997) A comparison of reserve selection algorithms. Biol Conserv 80:83–97
Current JR (1981) Multiobjective design of transportation networks. PhD dissertation, The Johns Hopkins University, Baltimore, MD
Current J, O'Kelly M (1992) Locating emergency warning sirens. Decis Sci 23:221–234
Current JR, ReVelle CR, Cohon JL (1985) The maximum covering/shortest path problem: a multiobjective network design and routing formulation. Eur J Oper Res 21:189–199
Curtin KM, Biba S (2011) The transit route arc-node service maximization problem. Eur J Oper Res 208:46–56
Curtin KM, Hayslett-McCall K, Qiu F (2010) Determining optimal police patrol areas with maximal covering and backup covering location models. Netw Spat Econ 10:125–145
Daskin MS (1983) A maximum expected covering location model: formulation, properties and heuristic solution. Transp Sci 17:48–70
Downs JA, Gates RJ, Murray AT (2008) Estimating carrying capacity for sandhill cranes using habitat suitability and spatial optimization models. Ecol Model 214:284–292
Dwyer FR, Evans JR (1981) A branch and bound algorithm for the list selection problem in direct mail advertising. Manag Sci 27:658–667
Eaton DJ, Daskin MS, Simmons D, Bulloch B, Jansma G (1985) Determining emergency medical service vehicle deployment in Austin, Texas. Interfaces 15:96–108
Erdemir ET, Batta R, Spielman S, Rogerson PA, Blatt A, Flanigan M (2008) Location coverage models with demand originating from nodes and paths: application to cellular network design. Eur J Oper Res 190:610–632
Goodchild MF, Lee J (1989) Coverage problems and visibility regions on topographic surfaces. Ann Oper Res 18:175–186
Grinde RB, Daniels K (1999) Solving an apparel trim problem by using a maximal cover problem approach. IIE Trans 31:763–769
Grubesic AH, Murray AT (2008) Sex offender residency and spatial equity. Appl Spat Anal Policy 1:175–192
Grubesic AH, Murray AT, Pridemore WA, Tabb LP, Liu Y, Wei R (2012) Alcohol beverage control, privatization and the geographic distribution of alcohol outlets. BMC Public Health 12:1–10

Hakimi SL (1964) Optimum locations of switching centers and the absolute centers and medians of a graph. Oper Res 12:450–459

Hakimi SL (1965) Optimum distribution of switching centers in a communication network and some related graph theoretic problems. Oper Res 13:462–475

Hogan K (1990) Reducing errors in rainfall estimates through rain gauge location. Geogr Anal 22:33–49

Hogan K, ReVelle C (1986) Concepts and applications of backup coverage. Manag Sci 31:1434–1444

Hoover EM (1948) The location of economic activity. McGraw-Hill, New York

Hoover EM (1967) Some programmed models of industrial location. Land Econ 43:303–311

Hotelling H (1929) Stability in competition. Econ J 39:41–57

Hougland ES, Stephens NT (1976) Air pollutant monitor siting by analytical techniques. J Air Pollut Control Assoc 26:51–53

Hudak PF, Loaiciga HA (1992) A location modeling approach for groundwater monitoring network augmentation. Water Resour Res 28:643–649

Isard W (1956) Location and space economy. MIT Press, Cambridge, MA

Kirkpatrick JB, Harwood CE (1983) Conservation of Tasmanian macrophytic wetland vegetation. Proc R Soc Tasmania 117:5–20

Klimberg R, ReVelle C, Cohon J (1991) A multiobjective approach to evaluating and planning the allocation of inspection resources. Eur J Oper Res 52:55–64

Kolesar P, Walker W, Hausner J (1975) Determining the relation between fire engine travel times and travel distances in New York City. Oper Res 23:614–627

Laporte G, Marin A, Mesa JA, Perea F (2011) Designing robust rapid transit networks with alternative routes. J Adv Transp 45:54–65

Launhardt W (1872) Kommercielle Tracirung der Verkehrswege. Arch Ingenieurverein

Lösch A (1954) The economics of location, translated by WH Waglorn and WF Stolper. Yale University Press, New Haven, CT

Mann CC (2005) 1491: new revelations of the Americas before Columbus. Alfred A. Knopf (a division of Random House Inc.), New York

Manne AS (1964) Plant location under economies-of-scale-decentralization and computation. Manag Sci 11:213–235

Matisziw TC, Murray AT, Kim C (2006) Strategic route extension in transit networks. Eur J Oper Res 171:661–673

Meyer PD, Brill ED Jr (1988) A method for locating wells in a groundwater monitoring network under conditions of uncertainty. Water Resour Res 24:1277–1282

Moon D, Chaudhry SS (1984) An analysis of network location problems with distance constraints. Manag Sci 30:290–307

Morgan CA, Slater PJ (1980) A linear algorithm for a core of a tree. J Algorithms 1:247–258

Moses LN (1958) Location and the theory of production. Q J Econ 72:259–272

Murray AT (2003) A coverage model for improving public transit system accessibility and expanding access. Ann Oper Res 123:143–156

Murray AT (2013) Optimising the spatial location of urban fire stations. Fire Saf J 62:64–71

Murray AT, Church RL (1996) Analyzing cliques for imposing adjacency restrictions in forest models. For Sci 42:166–175

Murray AT, Kim H (2008) Efficient identification of geographic restriction conditions in anti-covering location models using GIS. Lett Spat Resour Sci 1:159–169

Murray AT, Tong D (2007) Coverage optimization in continuous space facility siting. Int J Geogr Inf Sci 21:757–776

Murray AT, Kim K, Davis JW, Machiraju R, Parent R (2007) Coverage optimization to support security monitoring. Comput Environ Urban Syst 31:133–147

Murray AT, O'Kelly ME, Church RL (2008) Regional service coverage modeling. Comput Oper Res 35:339–355

Murray AT, Tong D, Grubesic TH (2012) Spatial optimization: expanding emergency services to address regional growth and development. In: Stimson R, Haynes K (eds) Applied geography. Edward Elgar Publishing, Cheltenham
O'Hanley JR, Church RL, Gilless JK (2007) Locating and protecting critical reserve sites to minimize expected and worst-case losses. Biol Conserv 134:130–141
Plane DR, Hendrick TE (1977) Mathematical programming and the location of fire companies for the Denver Fire Department. Oper Res 25:563–578
ReVelle C, Toregas C, Falkson L (1976) Applications of the location set-covering problem. Geogr Anal 8:65–76
ReVelle CS, Rosing KE (2000) Defendens imperium romanum: a classical problem in military strategy. Am Math Mon 107(7):585–594
Salvaneschi L (2002) Location, Location, Location: how to select the best site for your business. Oasis Press, Central Point, OR
Schilling DA, ReVelle C, Cohon J, Elzinga DJ (1980) Some models for fire protection locational decisions. Eur J Oper Res 5:1–7
Serra D (2013) Implementing TURF analysis through binary linear programming. Food Qual Prefer 28:382–388
Snyder SA, Haight RG (2016) Application of the maximal covering location problem to habitat reserve site selection: a review. Int Reg Sci Rev 39:28–47
Storbeck JE (1988) The spatial structuring of central places. Geogr Anal 20:93–110
Storbeck JE (1990) Classical central places as protected thresholds. Geogr Anal 22:4–21
Toregas C, ReVelle C, Swain R, Bergman L (1971) The location of emergency service facilities. Oper Res 19:1363–1373
Underhill LG (1994) Optimal and suboptimal reserve selection algorithms. Biol Conserv 70:85–87
Von Thunen JH (1826) Der Isolierte Staat in Beziehung auf landwrirtschaft und Nationalökonomie. Perthes, Hamburg
Weber A (1909) Theory of the location of industries. University of Chicago Press, Chicago, IL
Wei R, Murray AT (2014) Evaluating polygon overlay to support spatial optimization coverage modeling. Geogr Anal 46:209–229
Wei R, Murray AT, Batta R (2014) A bounding-based solution approach for the continuous arc covering problem. J Geogr Syst 16:161–182
Weiszfeld EV (1937) Sur le point pour lequal la somme des distances de *n* points donnes est minimum. Tohoku Math J 43:335–386
Williams JC (1998) Delineating protected wildlife corridors with multi-objective programming. Environ Model Assess 3:77–86
Williams JC, ReVelle CS, Bain DJ (2003) A decision model for selecting protected habitat areas within migratory flyways. Socio Econ Plan Sci 37:239–268
Wu C, Murray AT (2005) Optimizing public transit quality and system access: the multiple-route, maximal covering/shortest-path problem. Environ Plan B Plan Design 32:163–178

Chapter 2
Classic Beginnings

2.1 Introduction

The central theme of this book is the ability to identify the best location of one or more facilities or objects in order to provide some type or level of coverage. For example, let us suppose that we wish to place guards in an art gallery in such a manner that all areas of the gallery are within sight of one or more guards. In essence, we want the set of guard positions to "cover" or view the entire public area of the gallery. There can, of course, be many different configurations in which guards can view the entire gallery; however, it is of practical necessity to seek a pattern that deploys the fewest number of guards. As a second example, consider the case where we desire to provide fire protection to all neighborhoods of a city. In order to respond to a fire in a timely manner, we may set a standard that each neighborhood of the city should be no more than a mile and a half away from their nearest fire station (where fire trucks and crews can be housed to quickly respond when called). The fire service deployment problem can then be defined as finding the fewest fire stations (and their locations) so that each neighborhood is served or covered within a mile and a half of a station. Both the gallery guard positioning problem and the fire station location problem are examples of the Location Set Covering Problem, one of many covering problems that will be addressed in this book.

The definition of "cover" is usually defined within the context of a standard. Such a standard can be defined in terms of maximal distance (e.g., a mile and a half for a fire station), line-of sight (e.g., viewable by a security camera or guard), maximal service time (e.g., an 8 min response time by an available ambulance), the detection limit of a sensor (e.g., radar), audibility (e.g., loud enough to be heard), closer than a competitor (e.g., capture a percentage of customers by providing a closer bank branch), the amount of fabric uniquely covered by a cutout pattern, or by some other contextually meaningful metric (e.g., the presence of a target species at a biological reserve site).

© Springer International Publishing AG, part of Springer Nature 2018
R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_2

There are three principal forms of location covering models: those that seek complete coverage, those that maximize coverage subject to budgetary or resource constraints, and those that avoid coverage to the largest extent possible. For example, in terms of fire service deployment we may wish to identify the smallest number of stations and their locations while ensuring that each neighborhood is covered by service. This problem has been defined as the Location Set Covering Problem (LSCP) and seeks to cover all demand while using the fewest number of facilities. This facility location problem was first defined by Toregas et al. (1971) in order to locate emergency facilities. A second major development was based upon recognizing that resources needed to provide complete coverage may not be available. In such a case, the obvious question would be: "how much can be covered while expending no more than an allotted budget?" This, in essence, is the basis for the Maximal Covering Location Problem (Church and ReVelle 1974). A third approach represents an opposite tack, that of avoiding coverage, impact or detection. For example, we may wish to locate several facilities that need to be at least 50 miles from major population centers (e.g., a nuclear power plant) and perhaps 50 miles apart from each other. This last problem has been called the Minimum Impact Location Problem (Murray et al. 1998).

Since the early developments of the Location Set Covering Problem and the Maximal Covering Location Problem, and the latter developments of Minimizing Impact, models based upon some form of coverage have been extended and applied in a number of areas. Examples include remote sensing, watch tower location, surveillance camera placement, trauma care system design, biological reserve design, configuring FDA inspection teams, transit systems design, emergency shelter location, cell tower deployment, equipment storage location, cutting pattern layout of fabric, as well as many others. Since the ground breaking work of Berge (1957), Edmonds (1962), Toregas et al. (1971), and Church and ReVelle (1974), the science of "covering" has become a significant topic in many fields, including regional science, computer science, geography, operations research, industrial engineering, location science, ecology, and systems security. The remainder of this chapter presents two core problems that are central to the field of location covering.

2.2 Theory and Innovation

The field of location science spans a number of disciplines, including regional science, geography, engineering, business, computer science, and operations research. The beginning of this field is rooted in such classic works as Von Thünen (1826), Launhardt (1872), Weber (1909), Christaller (1933), Hotelling (1929), and Isard (1956) involving problems such as land use arrangements, industrial location, competition, and settlement patterns. This fundamental science was based upon mathematical constructs involving a blend of economics, behavior, and physics. For example, Launhardt (1872) attempted to identify the optimal location for a factory by finding the location which minimizes the costs of transport investment (e.g., a rail line) along with the costs of transporting needed raw materials to the factory and the cost of transporting the final product to the market. Such constructs, although easy to conceive, can often be hard to solve optimally due to issues of computational complexity. This fact alone hindered the development of this field until the invention of the modern computer. One might easily characterize the field of location science as having two distinct eras, one before digital computers were available and one since the development of the computer.

In the 1960s five different problem definitions emerged in the location literature: the p-median problem (Hakimi 1964, 1965), the p-center problem (Hakimi 1965), the simple plant location problem (Manne 1964), the plant layout problem (Armour and Buffa 1963), and the vertex cover problem (Fulkerson and Ryser 1961). Although these problems are rooted in a literature that preceded this modern era, there emerged an attempt to both structure and solve realistically sized problems in the 1960s using the resources of IBM's first solid state computer, the 7090 developed in 1959 and the subsequent 360 line of computers which became available in 1964. These key resources, along with computers of Control Data Corporation, Burroughs and Univac, played an instrumental role in the flourish of activity in this field. Although each of these problems is of interest in their own right, our discussion herein will be limited primarily to covering problems.

Berge (1957) was the first to define a cover problem involving a graph of nodes and edges.[1] Berge (1957) defined the cover problem as one of finding the smallest subset of nodes, B, such that every arc in the graph was adjacent to at least one node of B. Fulkerson and Ryser (1961) extended this concept by defining the minimal set representation problem. This problem involved sets and elements. Each set contains a list of elements that are members of that set, where specific elements can be members of more than one set. The minimal set representation problem can therefore be defined as follows:

Find the smallest number of elements in which each set contains at least one of the selected elements.

Define the following notation:

j = index of elements
i = index of sets

$$\delta_{ij} = \begin{cases} 1, & \text{if element } j \text{ is a member of set } i \\ 0, & \text{if otherwise} \end{cases}$$

[1]We have spent considerable time tracking references to various forms of "covering" in order to identify the original academic source. The trail leads to Berge (1957), Quine (1955) and Hall (1935). Hall (1935) defines a form of covering where a distinct representative of sets must be a group of members of which each set must contain one of these members. Quine (1955) was the first to discover how a set of truth functions could be reduced. Although Hall (1935) and Quine (1955) define the problems based upon representative sets and logic functions, Berge (1957) appears to be the first to define the cover problem within the context of a geometric shape such as a network structure.

$$x_j = \begin{cases} 1, \text{if element } j \text{ is selected} \\ 0, \text{ otherwise} \end{cases}$$

Fulkerson and Ryser (1961) formulated the minimal set representation problem as an integer-linear program as follows:

$$\textit{Minimize} \sum_{j=1}^{n} x_j \tag{2.1}$$

Subject to:

$$\sum_j \delta_{ij} x_j \geq 1 \quad \forall i \tag{2.2}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.3}$$

The objective (2.1) minimizes the number of times that elements are selected. Constraints (2.2) ensure that each set, i, is represented by at least one of the chosen elements, j. Constraints (2.3) impose binary integer requirements.

By definition, the goal is to select a minimum number of elements, so it is not necessary to choose an element more than once. In fact, selecting an element twice would be superfluous. Consequently, when requiring that each set is represented at least once, the decision variables, x_j, can be restricted to be binary, zero–one. For a given set i, constraints (2.2) ensure that at least one of the selected elements, j, will be a member of that set. Note that the coefficient of each variable in the constraint is either zero or one. Because of this, the left hand side of constraint (2.2) boils down to the sum of the element selection variables that are members of set i. For example, suppose that a set contained elements 1, 3, and 7. Then the constraint for this set would be equivalent to:

$$x_1 + x_3 + x_7 \geq 1 \tag{2.4}$$

Essentially, for constraint (2.4) to be satisfied, at least one of the three variables, x_1, x_3 and x_7, must be equal to one. Doing so ensures that this specific set is represented by at least one of its members, but the question is which ones should be selected if the number of selected elements is a minimum.

This formulation can be easily transformed to represent the edge adjacent covering problem of Berge (1957) by using i to represent edges, j to represent nodes, and by defining the δ_{ij} values as follows:

$$\delta_{ij} = \begin{cases} 1, \text{ if node } j \text{ is adjacent to edge } i \\ 0, \text{ if otherwise} \end{cases}$$

The model, (2.1)–(2.3), then captures the essence of the set covering problem as it involves minimizing the number of nodes selected such that each edge is incident to at least one selected node (see also Balinski 1965). Karp (1972) demonstrated that the problem of finding a solution which covers all edges with fewer than K vertices was NP-complete. In fact, the vertex cover problem was one of six fundamental problems initially used to define the NP-complete set (Garey and Johnson 1979). This initial set of problems included among others the clique problem, the Hamiltonian circuit problem, and the set partitioning problem. From a broader perspective, this means that this optimization problem, (2.1–2.3), is NP-hard.

Although early work in set representation and network edge covering problems focused on properties and formulations, Roth (1969) proposed a reductions algorithm for solving (2.1–2.3). This reductions process, when applied to the form given above, can effectively reduce many practical problems to a size that may be solved by hand or inspection.

One notable issue lacking in much of the initial work was a sense of application and use. Hakimi (1965) was perhaps the first to suggest a practical application involving the location of policemen on a highway network so that everyone was within a specified distance of a policeman. Unfortunately, his discussion was based upon the use of a Boolean function that proved to be less than fruitful as an approach for solving larger problems because it was thought to essentially require enumeration.

2.3 Location Set Covering Problem

In 1970, Constantine Toregas working under the direction of Charles ReVelle at Cornell University wrote a Master's thesis titled "A Covering Formulation for the Location of Public Service Facilities." This thesis was the first to structure a location model within the context of covering using an integer-linear programming formulation. Specifically, Toregas (1970) reasoned that emergency services, such as fire equipment and EMS vehicles, should be ideally located according to a response time standard. He suggested that there should be a maximum allowable response time in service provision in order to handle an event, such as a fire, before it gets out of hand. This would ensure a reasonable chance of saving lives and/or property. Simply put, Toregas (1970) described the following public service system design problem:

Minimize the number of facilities needed and locate them so that every demand area is covered within a predefined maximal service distance or time.

It was reasoned that the cost of building and equipping a facility, such as a fire station, would probably not vary much across a planning area as it involved the amortized costs of construction, site acquisition, equipment purchasing, etc. Since only the site acquisition costs would vary, this assumption would roughly hold in most locales. Thus, he argued that overall costs of providing service would indeed be a function of the number of facilities used. This was called the Location Set Covering Problem (LSCP).

2.3.1 Formulation

Toregas (1970) viewed the landscape of a region as a road network connecting a set of n nodes, which were defined as places of demand as well as potential facility locations. Consider the following notation:

i = index referencing nodes of the network as demand
j = index referencing nodes of the network as potential facility sites
S = maximal acceptable service distance or time standard
d_{ij} = shortest distance or travel time between nodes i and j
$N_i = \{j \mid d_{ij} \leq S\}$
$x_j = \begin{cases} 1, & \text{if a facility is located at node } j \\ 0, & \text{otherwise} \end{cases}$

Using this notation, Toregas (1970) defined the following model:

$$\text{LSCP}: \qquad \textit{Minimize} \sum_{j=1}^{n} x_j \tag{2.5}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq 1 \quad \forall i \tag{2.6}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.7}$$

The objective (2.5) involves minimizing the number of facilities being located. Constraints (2.6) ensure that for each demand node at least one facility is selected among the set N_i. This is the set of sites that are close enough to node i to provide service within the maximal service distance or time standard. Binary integer restrictions are imposed in constraints (2.7). Again, Toregas (1970) considered specifically the case where facilities were fire stations.

The LSCP is compact, involving only n constraints and n variables. In his Ph.D. dissertation at Cornell University, also under the supervision of Charles ReVelle, Toregas (1971) proposed two approaches to solve the LSCP, one based upon the use of integer-linear programming software and the other based upon a method called reductions.

What can be observed about the LSCP, (2.5–2.7), is that mathematically it is essentially the same as the minimal set representation problem structured in (2.1–2.3). The difference is context, where the LSCP explicitly involves spatial decisions that correspond to spatial service coverage. Accordingly, there are four associated points of contrast.

First, the variables x_j are spatially defined, where a specific node j represents a precise site on the surface of the earth. Second, although the sum in constraint (2.2) is over all j, as the coefficient δ_{ij} is used, the LSCP sums the x_j terms over the set $j \in N_i$.

Nevertheless, the constraints are equivalent as the only time when site j covers demand i is when $\delta_{ij} = 1$. More formally, this set may be stated as $N_i = \{j|\delta_{ij} = 1\}$, where the matrix $[\delta_{ij}]$ is called the coverage or set representation matrix. Third, the set N_i is derived and defined based on travel distance or time, and whether a demand location i can be served by a facility sited at j. This spatially defined set may be contiguous, non-contiguous, convex, non-convex, regularly or irregularly shaped, etc., depending on the context and associated criteria used in the assessment of attaining the desired standard of service. Fourth, the LSCP was conceived in an applied context for expressly helping to site emergency service facilities, such as fire stations.

From one perspective, mathematically the Toregas (1970, 1971) model is the same as that of Fulkerson and Ryser (1961), but from another it represents a substantial leap forward in location modeling. Whereas the vertex cover problem was more of a mathematical puzzle, the LSCP encapsulates the essence of design in efficiently allocating facilities across geographic space. In particular, the facilities of interest to Toregas (1970) were emergency services, such as fire protection resources or an ambulance/pre-hospital care. It is this significant real world application context which led to the modern foundation of location models using a covering objective. The applied setting involving geographic space often presents many challenges, ranging from representation to solution. It is precisely for these reasons, along with practical utility, that the LSCP has come to be so important and significant.

2.3.2 Planning Facility Deployment

Toregas et al. (1971) argued that the LSCP could be used for the purposes of planning by solving the model over a range of maximal service distance or time standards. By doing this, one can generate a tradeoff curve between the number of needed facilities in providing complete coverage as a function of the coverage standard, S. The tradeoff between the needed number of facilities and the maximal service standard will have a stair-step profile as the value of S is increased. When $S = 0$, presumably the number of needed facilities will equal, or nearly equal, the number of demand points in the problem. But as S is increased in value, the number of needed facilities will decline in number, until at some point, the value of S is large enough that all demand can be covered by a single facility capable of travel to all demand locations within the response standard.

In order to illustrate service response capabilities under different assumed standards, an example of the associated tradeoff is given in Fig. 2.1. In this case, an LSCP application instance was solved for each discrete value of S ranging from 1 to 29 by increments of 0.005 (a total of 5800 problems were therefore solved). At a value of 10, for example, it takes 9 facilities to cover all demand within this response standard. Alternatively, at a value of 12 only 7 facilities are required for coverage of all demand. There are several discrete increments in which the needed number of facilities decreases by more than one facility. Toregas et al. (1971) reasoned that this type of

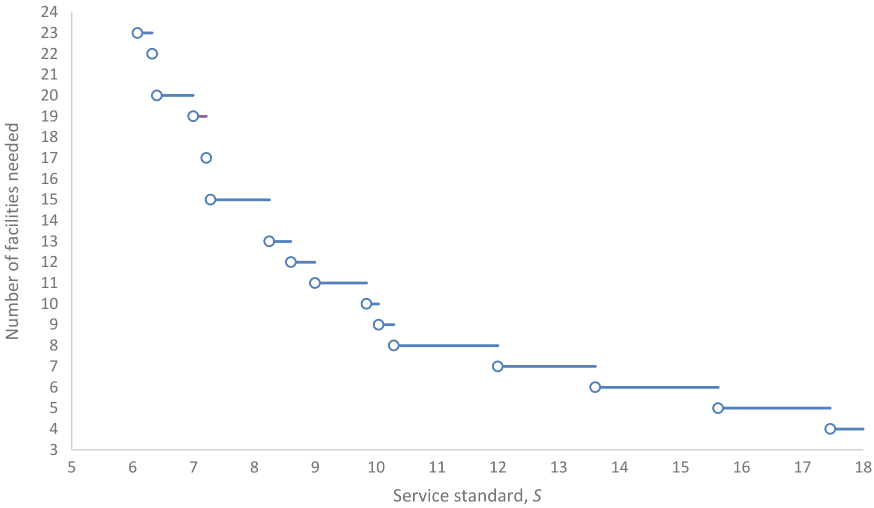


Fig. 2.1 Tradeoff curve between number of needed facilities and service standard, *S*

curve is useful in planning as one could identify the smallest maximal service distance for which the needed number of facilities equaled what was affordable. Worth noting is that one can think of these solutions as optimal for associated *p*-center problems.[2]

2.3.3 Applications

Even though the application focus of the LSCP helped to establish the use of a covering objective in location science, there were four other factors that helped to pique interest in this type of location model. The first factor was that the problem was simple to understand and it represented a major step towards making public safety services more efficient. The second factor was that Toregas (1971) and Toregas and ReVelle (1973) demonstrated that the use of a reductions algorithm in solving realistically sized LSCP instances was possible. Reductions approaches were first developed to reduce the size of logic and switching circuit problems (Quine 1955; Roth 1969). Toregas and ReVelle (1973) demonstrated that the principles used in reduction algorithms could be used to help solve the LSCP. The third factor was the availability of modest computing resources provided by the IBM 360 family of computers. Toregas (1971) solved 66 LSCP application instances and of these he was able to prove optimality for all but nine problems using the reductions process alone. For the remainder of the problems he used the reductions process followed by

[2]The *p*-center problem involves minimizing the furthest service distance that any demand must travel to their closest facility when locating exactly *p*-facilities. More discussion on this problem and its relationship to coverage problems is left for the chapters that follow.

the use of linear programming applied to a reduced form of the model given above. Having a simple solution process that was easy to program and could be supported by computer systems that were available in most large cities meant that the application frontier for this model was fertile. The fourth major factor was that after graduate school Constantine Toregas joined Public Technology Inc., an organization funded by member cities with the objective of solving municipal problems involving communication, public safety, environmental, and other emerging problems. Toregas, through Public Technology Inc., pushed the application of the LSCP and during the 1970s over 100 cities reorganized their fire protection plans based upon the use of the LSCP.

The application of the LSCP to fire station siting was the tip of the iceberg in terms of potential utility. It has been used or suggested for use in the location of ambulances (Berlin and Liebman 1974), the assignment of fire equipment to fire houses (Walker 1974), the selection of reserve sites for biological protection (Underhill 1994), the design of FDA inspection teams (Klimberg et al. 1991), dynamic repositioning of fire equipment (Kolesar and Walker 1974), accessibility of needed services for those with disabilities (Kwan et al. 2003), the location of warning sirens (Current and O'Kelly 1992), bus stop siting (Gleason 1975; Murray 2001), vehicle emissions testing stations (Swersey and Thakur 1995), weather radar (Agnetis et al. 2009), oil spill equipment location (Psaraftis and Ziogas 1985), security camera placement (Murray et al. 2007), art gallery guard placement (Chvatal 1975), and many others. Figure 2.2 shows LSCP results for the siting of emergency warning sirens in the coastal city highlighted in Chap. 1. The existing configuration of eight sirens is only capable of serving some 70% of the demand region. Use of the LSCP indicates that it would take 15 sirens to suitably serve the entire region in this case.

2.4 Maximal Covering Location Problem

The LSCP requires that each demand be covered or served within a maximal service distance or time standard. Recognizing that there may be situations where covering all demand within a strict standard, Toregas (1971) suggested that the maximal standard could be defined differently for each demand, e.g., S_i. For example, in the County of Santa Barbara it is recommended that ambulances be positioned in the region so that all urban areas are served within 10 min response time and that all rural areas be served within a maximal response time of 30 min. But, even when one defines such a standard, there are many circumstances when the resources required to provide complete coverage for urban and rural service exceeds what an agency may spend. This means that a service system configuration identified using the LSCP may be too expensive to implement. In such a circumstance it is reasonable to ask how much coverage can be provided with less investment. In particular, what can be achieved by deploying a fixed number of facilities? In a design sense, we can state this problem as:

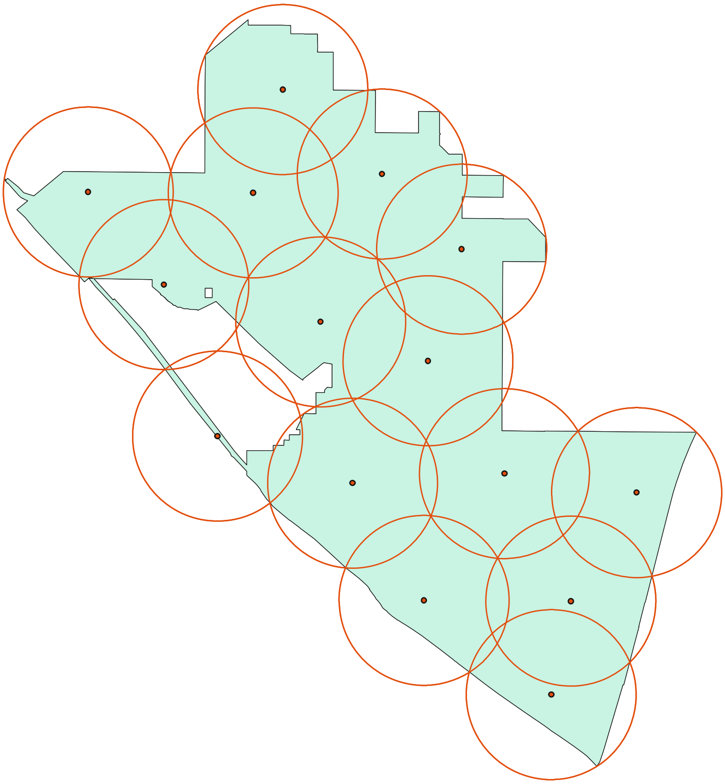


Fig. 2.2 LSCP solution for warning siren coverage in Huntington Beach, CA. This LSCP solution requires 15 siren locations to cover all of Huntington Beach

Maximize the amount of demand covered within a maximal service distance or time standard by locating a fixed number of facilities

This problem is called the Maximal Covering Location Problem (MCLP) and was defined originally in the Ph.D. dissertation of Church (1974) at The Johns Hopkins University as well as in Church and ReVelle (1974).[3]

Interestingly, at the same time White and Case (1974) also argued for more flexibility in determining a cover solution, stating that "... in a number of practical situations, it is not possible to provide the number of facilities required to cover *totally* all customers; rather, the number of facilities available for location is only sufficient to cover *partially* the set of customers." With this argument, White and Case (1974) defined the partial cover problem where the total number of customers (one customer per node) covered is maximized subject to the use of a fixed number

[3]Church and ReVelle in 1973 presented this at the North American Regional Science Council meetings in Atlanta, Georgia that was subsequently published as Church and ReVelle (1974).

of facilities. Both problem definitions are arguably important, as they expand the potential problem domain of location covering models. However, the issue of addressing an available budget makes them even more appealing.

2.4.1 Formulation

The following additional notation is introduced:

$p =$ number of facilities to be located
$a_i =$ service load or population at demand i
$y_i = \begin{cases} 1, & \text{if demand } i \text{ is covered within the service standard} \\ 0, & \text{otherwise} \end{cases}$

The formulation detailed in Church (1974) and Church and ReVelle (1974) can be stated as:

$$\text{MCLP}: \qquad \textit{Maximize} \sum_{i=1}^{n} a_i y_i \tag{2.8}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq y_i \quad \forall i \tag{2.9}$$

$$\sum_{j} x_j = p \tag{2.10}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.11}$$

$$y_i \in \{0, 1\} \quad \forall i \tag{2.12}$$

The objective, (2.8), seeks to maximize the total amount of demand covered by the placement of p facilities. Constraint (2.9) defines whether coverage to a given demand is provided or not. Essentially, if one or more facilities within the coverage set N_i is (are) selected, then the sum, $\sum_{j \in N_i} x_j$, is greater than equal to one in value. This allows the coverage decision variable, y_i, to be one, indicating that coverage has been provided to demand i. If no facilities have been selected among the coverage set, N_i, then the sum $\sum_{j \in N_i} x_j = 0$ and y_i is forced to be zero, indicating that demand i is not covered. The number of facilities to be located is restricted to equal p in constraint (2.10). Constraints (2.11) and (2.12) restrict decision variables to be binary, zero or one.

The MCLP contains $2n$ binary decision variables and $n + 1$ linear constraints. Formally, Church (1974) showed that this problem can be solved to optimality when relaxing the integer restrictions on the y_i variables, imposing only simple upper bounds, $y_i \leq 1$. This means that the MCLP has the same number of needed binary variables (e.g., x_j) as the LSCP for a given application.

Finally, if $a_i = 1$ for each demand i, then the MCLP would represent the partial covering problem of White and Case (1974). This implies then that the partial covering model is a special case of the MCLP.

From the outset, Church and ReVelle (1974) recognized that a heuristic approach would be necessary for solving large maximal covering applications. They suggested a procedure that was based upon “greedy” and “substitution” strategies. Since then, other strategies have been applied for solving large MCLP instances, including simulated annealing (Murray and Church 1996), genetic algorithms (Tong et al. 2009), heuristic concentration (ReVelle et al. 2008), vertex substitution (Gerrard et al. 1996), Lagrangean relaxation (Galvão and ReVelle 1996), and dual-based heuristics with branch and bound (Downs and Camm 1996), among others.

An important and interesting feature of maximizing coverage is that it may be approached as a minimization problem, detailed in Church (1974) and Church and ReVelle (1974). This alternate form is of value theoretically and computationally. Consider the following variables:

$$\bar{y}_i = 1 - y_i \tag{2.13}$$

As $\bar{y}_i \in \{0, 1\}$, then if $\bar{y}_i$ equals to one, the interpretation is that demand i is not covered. Alternatively, if then if $\bar{y}_i$ is equal to zero, the interpretation is that demand i is covered. If we solve the above (2.13) for y_i, then: $y_i = 1 - \bar{y}_i$. Substituting $1 - \bar{y}_i$ for y_i in the MCLP objective (2.8) yields:

$$\sum_{j=1}^{n} a_i \left(1 - \bar{y}_i\right) = \sum_{j=1}^{n} a_i - \sum_{j=1}^{n} a_i \bar{y}_i = A - \sum_{i=1}^{n} a_i y_i \tag{2.14}$$

where $A = \sum_{i=1}^{n} a_i$. Since A is a sum of constants (no decision variables), then it is not optimizable. Therefore, the MCLP objective function is equivalent to:

$$Minimize \quad \sum_{i}^{n} a_i \bar{y}_i \tag{2.15}$$

Similarly, we can also substitute $1 - \bar{y}_i$ for y_i in constraints (2.9) as follows:

$$\sum_{j \in N_i} x_j \geq 1 - \bar{y}_i \tag{2.16}$$

This is equivalent to:

$$\sum_{j \in N_i} x_j + \bar{y}_i \geq 1 \tag{2.17}$$

This condition forces the variable $\bar{y}_i$ to equal one when no sites have been chosen in the set N_i. In essence, when a demand i is not covered, $\bar{y}_i = 1$. Further, when demand i is covered, then $\bar{y}_i = 0$ because the constraint is satisfied without making $\bar{y}_i$ positive as well as combined with the orientation of objective (2.14). Collectively, this will force $\bar{y}_i$ to be zero whenever possible.

This means that the alternative formulation of the maximal covering location model can be summarized as:

$$\text{MCLP2 :} \qquad \textit{Minimize} \sum_{i=1}^{n} a_i \bar{y}_i \tag{2.18}$$

Subject to:

$$\sum_{j \in N_i} x_j + \bar{y}_i \geq 1 \quad \forall i \tag{2.19}$$

$$\sum_{j} x_j = p \tag{2.20}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.21}$$

$$\bar{y}_i \in \{0, 1\} \quad \forall i \tag{2.22}$$

Objective (2.18) involves minimizing the amount of demand not covered. Since the total demand is constant, maximizing coverage is equivalent to minimizing the demand not covered. This model takes the form of a constrained set covering problem. Constraints (2.19) specify coverage, extending constraints (2.6) in the LSCP to include the variable $\bar{y}_i$. Constraints (2.20) specify the number of facilities to be sited, p. Decision variables requirements are indicated in constraints (2.21) and (2.22). However, when solving this problem, it is only necessary to constrain the values of x_j to be binary integer. The integer conditions for $\bar{y}_i$ can be relaxed to simple non-negativity constraints, $\bar{y}_i \geq 0$, as these variables will never be greater than one in value in order to satisfy the covering constraint (2.19) combined with the orientation of the objective (2.18) and the fact that the x_j variables are integer in value.

Computationally, the MCLP2 is somewhat simpler to solve than the maximizing form, MCLP, as an upper bounded variable simplex algorithm is not needed when solving MCLP2 as a minimization problem. The solution results reported in Church and ReVelle (1974) were in fact based upon the MCLP2 formulation.

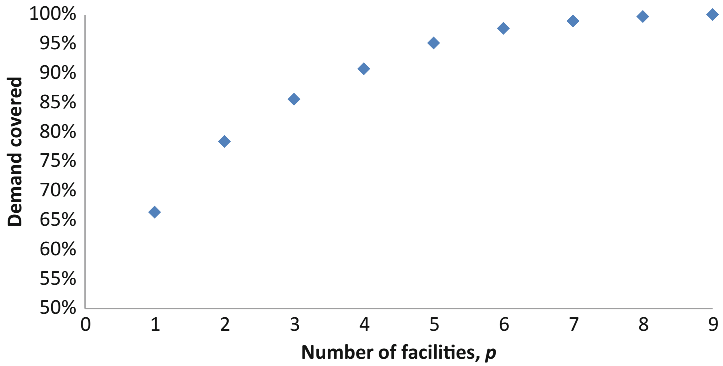


Fig. 2.3 Tradeoff curve between coverage and number of facilities sited

2.4.2 Planning Facility Deployment

Church and ReVelle (1974) suggested that the MCLP be solved for a given planning application by varying the number of facilities. Solving over a range of values for p, the number of facilities to site, makes it possible to generate a tradeoff between coverage provided and the investment in facilities. To illustrate this, Fig. 2.2 presents an associated tradeoff curve for a particular service standard. As is often the case, complete coverage requires a significant investment as compared to what is needed to provide 90% coverage, or even 95% coverage. For example, it requires nine facilities to cover all demand in this case within the standard. However, for five facilities coverage of 95% of the demand is possible. The strict elbow-shaped form of the coverage curve shown in Fig. 2.3 is quite common, indicating that in many cases, a high level of coverage can be provided at a substantial reduction to what it may take to provide complete coverage.

2.4.3 Applications

Models based upon the maximal covering concept have been applied in a number of fields. These include locating health clinics (Bennett et al. 1982), positioning ambulances (Saydam and McKnew 1985), locating fire stations (Murray 2013), selecting advertising venues from subscriber lists (Dwyer and Evans 1981), cluster analysis (Chung 1986), discriminate analysis (Chung 1986), selecting sites for nature reserves (Church et al. 1996), designing police patrol areas (Curtin et al. 2010), selecting color tabs for matching teeth to prosthetic teeth (Cocking et al. 2009), cutting pattern layout on fabric (Grinde and Daniels 1999), and locating humanitarian relief supplies (Balcik and Beamon 2008), just to name a few. Because the problem is relatively compact in size, many applications can be solved by a straight forward application of commercial software, like that of XPRESS, CPLEX,

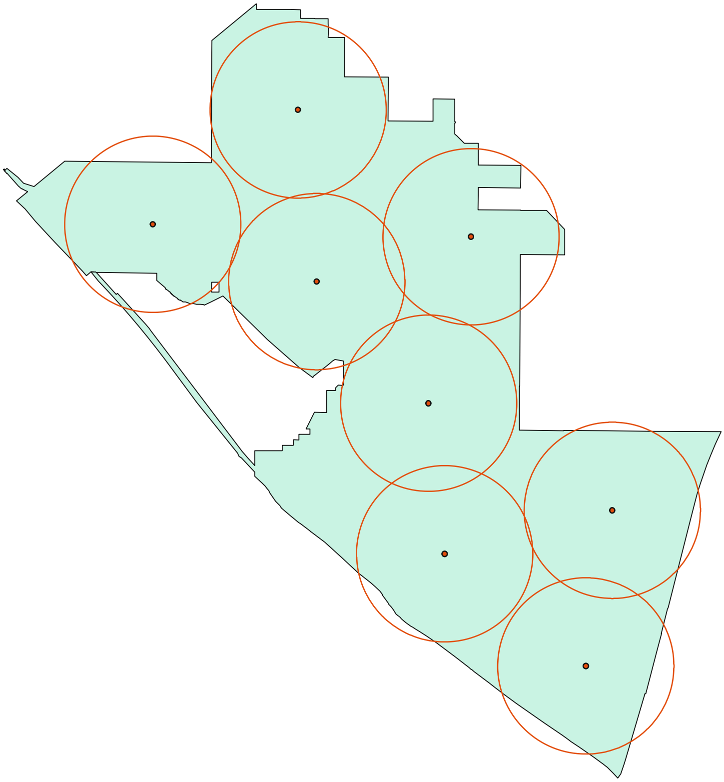


Fig. 2.4 Optimal configuration of eight warning sirens using the MCLP

and GUROBI. Further, some GIS (geographical information systems) software, like ArcGIS by Esri, contain heuristics that can solve the MCLP. Figure 2.4 shows MCLP results for the siting of emergency warning sirens discussed previously (Figs. 1.2 and 2.2). While the existing configuration of eight sirens serves approximately 70% of the demand region, application of the MCLP finds that an optimal configuration of eight sirens could actually serve over 85% of the regional demand in this case.

2.5 Theoretical Linkages

As was evident for the MCLP, alternative ways to mathematically structure a given problem often exist. Further, one formulation may prove more beneficial or desirable than alternatives, under particular conditions. Beyond this, there has been a more general interest in deriving and/or identifying a mathematical model that represents a unifying framework in facility location, and in particular be capable of structuring a range of important location modeling approaches.

Hakimi (1965) defined the following location problem:

Locate a fixed number of facilities such that the resulting sum of travel distances is minimized.

Four problem features were assumed: (1) each facility had the capability of handling all of the demand assigned to it; (2) each user is assigned to their closest located facility; (3) demand was concentrated at nodes of the network; and, (4) facilities could be located anywhere along the network (i.e., nodes and arcs). The motivating application for Hakimi (1965) was that of locating telephone switching centers. It was assumed that each customer would be connected by wire to their closest switching center. The objective was to minimize the total length of wire in serving all customers while locating p telephone switching centers, although it could apply to vehicular travel or any other appropriate form of connection as well. This has been called the p-median problem. Total service distance can be easily computed for any configuration of facilities along a network by first assigning each demand node to their closest facility, then multiplying the assignment distance by the demand to be served, and finally summing all demand weighted distance values. Hakimi (1964, 1965) proved that at least one optimal solution to a given p-median problem would utilize only nodal locations of the network. Since this groundbreaking work, virtually all models and solution techniques for the p-median problem have been based upon the assumption that facilities will be located only at nodes as such a restriction does not preclude an optimal solution from being identified. For example, ReVelle and Swain (1970) developed an integer-linear programming model for the p-median problem based upon this assumption. Accordingly, each node of the network was a place of demand as well as a potential facility site. Thus, a network of n nodes would have n demands and n sites. Consider the following notation:

$$z_{ij} = \begin{cases} 1, \text{if demand } i \text{ is assigned to facility } j \\ 0, \text{otherwise} \end{cases}$$

$$z_{jj} = \begin{cases} 1, \text{if node } j \text{ has been selected for a facility and assigns to itself} \\ 0, \text{otherwise} \end{cases}$$

ReVelle and Swain (1970) structured the following model:

$$p\text{-median}: \qquad Minimize \sum_i \sum_j a_i d_{ij} z_{ij} \tag{2.23}$$

Subject to:

$$\sum_j z_{ij} = 1 \quad \forall i \tag{2.24}$$

$$\sum_j z_{jj} = p \tag{2.25}$$

$$z_{ij} \le z_{jj} \quad \forall i, j \& i \ne j \tag{2.26}$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j \tag{2.27}$$

Objective (2.23) minimizes the total weighted distance of demand assignment. Each demand i is forced to assign to a facility in constraints (2.24). Constraint (2.25) establishes that exactly p facilities are to be sited. Constraints (2.26) ensure that demand i cannot assign to facility j (i.e., $z_{ij} = 1$) unless a facility at j has been sited (i.e., $z_{jj} = 1$). Integer restrictions are imposed in constraints (2.27).

A few observations regarding the p-median problem are in order. First, it has been widely applied and the focus of much academic research. Second, it is has been extended in various ways to address a range of nuances encountered in application. Third, whereas the LSCP has n variables and n constraints and the MCLP has $2n$ variables and $n + 1$ constraints (assuming for the moment that the number of demand locations is the same as the number of potential facility sites), the p-median problem formulation is comparatively quite large as it has n^2 variables and $n^2 + 1$ constraints. For example, a problem of 1000 nodes would entail one million variables and a million and one constraints. As a consequence, many researchers have devoted considerable effort to designing heuristic approaches to solving the p-median problem (Teitz and Bart 1968; Densham and Rushton 1992; Rolland et al. 1997; Rosing and ReVelle 1997; Mladenović and Hansen 1997).

There is, however, a practical and theoretical reason as to why this problem and model formulation is of importance. First, the LSCP and MCLP can be structured as a p-median problem, and it is possible to convert these covering models using a polynomial bounded algorithm. Second, all solution procedures for the p-median problem, even heuristics, can be directly applied to solve an equivalent LSCP or MCLP, after making an appropriate problem transformation. Church (1974) was the first to recognize how a covering problem could be converted into an equivalent p-median problem. For the case where each node is a demand location and a potential facility site, Church (1974) proposed the following distance (or travel time) transformation:

$$d'_{ij} = \begin{cases} 0, \text{if } d_{ij} \le S \\ 1, \text{otherwise} \end{cases} \tag{2.28}$$

If the above p-median problem formulation, (2.23)–(2.27), is solved using transformed distances, d'_{ij}, instead of the actual distances, d_{ij}, the objective function of minimizing total weighted assignment distance (i.e., $\sum_i \sum_j a_i d'_{ij} z_{ij}$ substituted in objective (2.23)) would count each demand i that is served within their standard, S, as zero times a_i, and each demand that is served beyond the standard as one times a_i. This means that the objective seeks to minimize the amount of demand that is assigned outside the service standard. As this is equivalent to maximizing the total demand served within the standard, the above distance transformation when used with a p-median problem represents a covering problem, precisely along the lines of the MCLP2 structured above. Of course, the MCLP can be used to solve the LSCP simply by starting at any value of p, the number of facilities to be sited, such that coverage of all demand is not possible, then increasing p and re-solving until all demand is covered. The case where all demand is covered and p is the smallest value possible represents an optimal solution for the LSCP.

Hillsman (1984) presented a transformation to formulate an LSCP as a uniform linear model, which is a general location construct that he proposed to solve a variety of discrete location problems. The transformation of Hillsman (1984) can be used in part to formulate an LSCP as a p-median problem. Such a transformation is a bit more nuanced as the p-median problem involves locating a fixed number of facilities whereas the LSCP deals with finding the fewest number of needed facilities. We can make such a transformation by first setting the value of p to be equal to the value of n, the original number of potential facility sites. For the sake of generality, we again assume that each node is a place of demand as well as a potential facility site. In addition, we need to add $n - 1$ pseudo nodes. Essentially, these pseudo nodes serve as places of pseudo demand and potential facility sites. These sites do not cover any real demand and cost nothing to use. We will refer to the total list of nodes in terms of $i, j = 1, 2, \ldots, n$ as the original nodes in the problem and $i, j = n + 1, \quad n + 2, \ldots, 2n - 1$ as the pseudo nodes. We can further define the demand at each node (pseudo or real) as one. The added “pseudo nodes” are included so that the total number of selected sites will always add to n, a property which will be explained below. Consider the following distance transformation:

$$\tilde{d}_{ij} = \begin{cases} 0, & \forall i,j \text{ when } d_{ij} \leq S, i \neq j, i \leq n \& j \leq n \\ 0, & \forall i,j \text{ when } i \geq n+1 \\ 1, & \forall i,j \text{ when } i = j, \ i \leq n \& j \leq n \\ M, & \forall i,j \text{ when } d_{ij} > S_i, i \leq n \& j \leq n \\ M, & \forall i,j \text{ when } \ i \leq n \& j \geq n+1 \end{cases} \tag{2.29}$$

where M is any number greater than n. The last condition of this transformation function creates a cost of serving any original demand by a facility placed at a pseudo node as the value of M. This is a high enough value that no original demand point will ever assign to a facility placed at a pseudo node (unless no feasible solution exists to the LSCP being solved). The cost of assignment for any pseudo demand is zero, whether it is to an original site or a pseudo site. This is established in the second condition of the transformation. Therefore, the added pseudo demands do not affect the objective function. If a pseudo site self assigns (i.e., $x_{jj} = 1$ when $j \geq n + 1$), then it means that it has been selected for one of the p facilities. The cost of selection (i.e., self-assignment) for a pseudo site is zero as well. Thus, such a selection does not affect the overall objective value. The rest of the transformation is explained by reviewing the effective distances for the original demand and facility nodes. If a real node (as opposed to a pseudo node) is selected for a facility, the self assignment distance is 1 in value (condition 3 of the transformation), which will be multiplied by the demand weight of 1, creating a total cost of assignment of 1. When an original demand assigns to a site that can cover that demand, the cost of assignment is zero (the first condition). Finally, if an original demand is forced to assign to a facility that is further than the coverage distance to a real site, the cost of assignment is M (condition 4). Assigning a real demand to a pseudo site will incur this same, prohibitive cost (condition 5). Since the cost of assigning any original demand beyond a coverage distance is so high compared to the cost of locating another real facility, a real facility will be located to cover each demand, reducing all assignment costs to zero except self-assignment. Thus, the assignment costs will ensure that all real demand points are covered by selected real sites. The only cost incurred in doing that is 1 for each selected real site (pseudo sites cost nothing). Thus, the objective will force the number of real sites selected to be as small as possible. Since pseudo sites cost nothing, enough of them will be selected to meet the p facility condition (that has been set arbitrarily large enough to have no impact on the problem). Consequently, when solving a p-median problem using the transformation (2.29), the resulting model [i.e., $\sum_i \sum_j a_i \tilde{d}_{ij} z_{ij}$ substituted in objective (2.23)] will minimize the number of real sites needed, while ensuring that each real demand node is covered. As a final note, for completeness transformation (2.29) is designed for the ReVelle and Swain (1970) formulation of the p-median problem, (2.23–2.27), with a complete $(2n - 1)$ by $(2n - 1)$ distance matrix, but by modifying the formulation it is possible to eliminate the representation of pseudo demands and all assignment variables to pseudo sites.

The significance of the two transformations, (2.28) and (2.29), is that any LSCP or MCLP can be solved as a p-median problem. Of course, this means that the alternative formulations now have an order of n^2 complexity. Since a wide variety of heuristics have been developed to solve the p-median problem, they can be employed without modification for solving these two covering problems, using the above transformations. This is precisely how the location-allocation routine found in ArcGIS allows one to solve covering and median problems. It performs all of the necessary transformations, upon the selection of a model option. However, it is important to note that the performance capabilities of this approach for some heuristics has not been tested or evaluated extensively, if at all. Thus, further testing is needed to fully vet the quality of obtained results using a specific heuristic in solving a covering problem through a transformation as a p-median problem. A final note is that Toregas et al. (1971) suggested that after deciding which level of deployment to pursue, p and S, for a given LSCP application, then it made sense to solve a distance constrained p-median problem. This involves modification of the p-median problem, and in particular constraints (2.24). The objective is to find the solution which minimizes total weighted distance, while locating p facilities and ensuring that each demand is within the S standard of their closest facility. This is commonly referred to as the p-median problem with maximum distance constraints.

2.6 Fixed Charges

One of the basic assumptions of Toregas et al. (1971) was that the cost of individual sites did not vary much, especially when compared to the cost of facility construction, the cost of equipment being housed, and the cost of crews needed. They argued that the number of facilities needed was a surrogate measure for cost, and by keeping the number of located facilities as low as possible meant that overall costs would be minimized as well. Such an assumption would be especially true if, for example, each fire station housed the same type and number of equipment, was no more expensive to build in one location as another, and site acquisition costs were approximately the same. But, there are applications where such an assumption would not necessarily be true. For example, consider the case where an early warning siren system is being located. The objective is to cover a political jurisdiction with siren towers such that all inhabitants can hear a siren when in use. Although the land costs for siren placement may be minimal, the site preparation costs for one site may be quite different from another. Sirens need to be connected to a protected communication network, provided a source of power, and have a solid geotechnical foundation. Utility connections could be considerable for one site and miniscule for another. The same is true for preparing a structural base. Consequently, site costs may be an important consideration in siren location. Another issue is almost always present: some facilities already exist. The cost of new facilities should be accounted for differently than existing facilities that have already been amortized. To address this issue, we can consider a cost sensitive form of the LSCP and MCLP.

In the case of the LSCP, this means that the objective should consider variable costs. This cost may be defined as follows:

c_j = cost of siting facility j

This means that the fixed cost LSCP (FC-LSCP) can be structured as:

$$\text{FC-LSCP}: \qquad \textit{Minimize} \sum_{j=1}^{n} c_j x_j \tag{2.30}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq 1 \quad \forall i \tag{2.31}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.32}$$

Objective (2.30) involves minimizing the total cost of siting facilities. Constraints (2.31) ensure that for each demand node at least one facility is selected among the set N_i. Binary integer restrictions are imposed in constraints (2.32). Worth noting is that FC-LSCP is mathematically equivalent to the set covering problem, extending the (2.1–2.3). Of course, the above mentioned points regarding the spatial significance of the FC-LSCP would apply in terms of its uniqueness and distinction from the set covering problem.

The MCLP also can be extended to account for fixed costs. Here a formulation is detailed based on MCLP2. The fixed cost MCLP2 (FC-MCLP2) is as follows:

$$\text{FC-MCLP2}: \qquad \textit{Minimize} \sum_{i=1}^{n} a_i \bar{y}_i \tag{2.33}$$

$$\textit{Minimize} \sum_{j=1}^{n} c_j x_j \tag{2.34}$$

Subject to:

$$\sum_{j \in N_i} x_j + \bar{y}_i \geq 1 \quad \forall i \tag{2.35}$$

$$\sum_{j} x_j = p \tag{2.36}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{2.37}$$

$$\bar{y}_i \in \{0, 1\} \quad \forall i \tag{2.38}$$

FC-MCLP2 is a multi-objective model. Objective (2.33) minimizes the amount of demand not covered. Objective (2.34) minimizes the total cost of siting facilities. Constraints (2.35) specify coverage of demand. Constraints (2.36) indicate the number of facilities to be sited. Technically speaking, this constraint is not necessary in the sense that objective (2.34) accounts for costs, so ultimately dictates the number of facilities that would be sited as a tradeoff of the importance to cover demand. Whether to include it partly a preference as this constraint represents another model parameter. Decision variable requirements are indicated in constraints (2.37) and (2.38).

Solution of problems involving multiple objectives can be challenging. Chapter 3 discusses this issue in more detail. Here we simply note that methods exist for dealing with multiple objectives in a manner that associated trade off solutions can be identified.

The FC-MCLP2 was first proposed by Church and Davis (1992). They show that site costs can make a difference in which solutions are non-inferior with respect to the tradeoff between site cost and coverage. To solve this problem, they used DUALOC, a dual ascent solution procedure (Erlenkotter 1978) designed to solve the classical fixed charge simple plant location problem. In casting the fixed charge maximal covering problem (they called this MCFix) as a simple plant location problem, Church and Davis (1992) used a transformation function like that given in (2.28).

Highlighted in this chapter is that there are often alternative ways to mathematically structure any given problem of interest. This was observed for the MCLP, as an example, but also that the MCLP and LSCP could be structured as a transformed p-median problem. Accordingly, fixed costs could readily be incorporated in the MCLP or even the p-median problem.

Both the FC-MCLP2 and FC-LSCP are simple, extended forms of MCLP and LSCP, respectively. They add a needed feature when specific site costs are particularly important. The majority of research involving covering problems has been based upon the assumption of uniform site costs. As stated earlier, site costs often represent only a small portion of the total system cost, and using the number of facilities (as in LSCP and MCLP) as a surrogate measure of overall cost is a reasonable assumption in many cases.[4]

2.7 Summary and Concluding Comments

The early roots in optimization models involving the notion of cover began in logic and set representation problems. Hakimi (1965) was one of the first to propose that the use of a coverage standard in allocating policemen across a network so that all parts of the network are within a prescribed distance of a policeman. Toregas et al. (1971) were the first to propose a mathematical formulation for a location model that involved covering a region with fire stations. This formulation, the LSCP, (2.5–2.7), along with that of the MCLP, (2.8–2.12), formed the basis for an important field of location science. Even though there are a large number of applications of these two models alone, there are also a number of elements that are not captured exactly by these two constructs. In addition, large LSCP and MCLP application instances might not be easy to solve. This has led to two major thrusts in covering model research: the development of different solution strategies, and the development of extended model forms.

[4]In Chap. 3 we present an interesting form of site cost proposed by Plane and Hendrick (1977).

References

Agnetis A, Grande E, Mirchandani PB, Pacifici A (2009) Covering a line segment with variable radius discs. Comput Oper Res 36(5):1423–1436

Armour GC, Buffa ES (1963) A heuristic algorithm and simulation approach to relative location of facilities. Manag Sci 9(2):294–309

Balcik B, Beamon BM (2008) Facility location in humanitarian relief. Int J Logist 11(2):101–121

Balinski ML (1965) Integer programming: methods, uses, computations. Manag Sci 12(3):253–313

Bennett VL, Eaton DJ, Church RL (1982) Selecting sites for rural health workers. Soc Sci Med 16 (1):63–72

Berge C (1957) Two theorems in graph theory. Proc Natl Acad Sci 43(9):842–844

Berlin GN, Liebman JC (1974) Mathematical analysis of emergency ambulance location. Socio Econ Plan Sci 8(6):323–328

Christaller W (1933) Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen. Gustav Fischer, Jena

Chung C-H (1986) Recent applications of the maximal covering location planning (M.C.L.P.) model. J Oper Res Soc 37:735–746

Church R, ReVelle C (1974) The maximal covering location model. Pap Reg Sci Assoc 32:101–118

Church RL (1974) Synthesis of a class of public facility location models, PhD dissertation, The Johns Hopkins University, Baltimore, MD

Church RL, Davis RR (1992) The fixed charge maximal covering location problem. Pap Reg Sci 71 (3):199–215

Church RL, Stoms DM, Davis FW (1996) Reserve selection as a maximal covering location problem. Biol Conserv 76(2):105–112

Chvatal V (1975) A combinatorial theory in plane geometry. J Comb Theory B 18:39–41

Cocking C, Cevirgen E, Helling S, Oswald M, Corcodel N, Rammelsberg P, Reinelt G, Hassel AJ (2009) Colour compatibility between teeth and dental shade guides in Quinquagenarians and Septuagenarians. J Oral Rehabil 36:848–855

Current J, O'Kelly M (1992) Locating emergency warning sirens. Decis Sci 23(1):221–234

Curtin KM, Hayslett-McCall K, Qiu F (2010) Determining optimal police patrol areas with maximal covering and backup covering location models. Netw Spat Econ 10(1):125–145

Densham PJ, Rushton G (1992) A more efficient heuristic for solving large p-median problems. Pap Reg Sci 71(3):307–329

Downs BT, Camm JD (1996) An exact algorithm for the maximal covering problem. Nav Res Logist 43(3):435–461

Dwyer FR, Evans JR (1981) A branch and bound algorithm for the list selection problem in direct mail advertising. Manag Sci 27(6):658–667

Edmonds J (1962) Covers and packings in a family of sets. Bull Am Math Soc 68(5):494–499

Erlenkotter D (1978) A dual-based procedure for uncapacitated facility location. Oper Res 26:992–1009

Fulkerson DR, Ryser HJ (1961) Widths and heights of (0, 1)-matrices. Rand Corporation, Santa Monica, CA

Galvão RD, ReVelle C (1996) A Lagrangean heuristic for the maximal covering location problem. Eur J Oper Res 88(1):114–123

Garey MR, Johnson DS (1979) Computers and Intractability: a guide to theory of NP-completeness. W.H. Freeman, New York

Gerrard RA, Stoms DA, Church RL, Davis FW (1996) Using GIS models for reserve site selection. Trans GIS 1(2):45–60

Gleason JM (1975) A set covering approach to bus stop location. Omega 3(5):605–608

Grinde RB, Daniels K (1999) Solving an apparel trim placement problem using a maximum cover problem approach. IIE Trans 31(8):763–769

Hakimi SL (1964) Optimum locations of switching centers and the absolute centers and medians of a graph. Oper Res 12(3):450–459
Hakimi SL (1965) Optimum distribution of switching centers in a communication network and some related graph theoretic problems. Oper Res 13(3):462–475
Hall P (1935) On representatives of subsets. J Lond Math Soc 1(1):26–30
Hillsman EL (1984) The *p*-median structure as a unified linear model for location-allocation analysis. Environ Plan A 16:305–318
Hotelling H (1929) Stability in competition. Econ J 39(153):41–57
Isard W (1956) Location and space-economy. MIT Press, Cambridge, MA
Karp RM (1972) Reducibility among combinatorial problems. In: Miller RE, Thatcher JW (eds) Complexity of computer computations. Springer, Boston, MA, pp 85–103
Klimberg R, ReVelle C, Cohon J (1991) A multiobjective approach to evaluating and planning the allocation of inspection resources. Eur J Oper Res 52(1):55–64
Kolesar P, Walker WE (1974) An algorithm for the dynamic relocation of fire companies. Oper Res 22(2):249–274
Kwan MP, Murray AT, O'Kelly ME, Tiefelsdorf M (2003) Recent advances in accessibility research: representation, methodology and applications. J Geogr Syst 5(1):129–138
Launhardt W (1872) Kommercielle Tracirung der Verkehrswege. Arch Ingenieurverein
Manne AS (1964) Plant location under economies-of-scale—decentralization and computation. Manag Sci 11(2):213–235
Mladenović N, Hansen P (1997) Variable neighborhood search. Comput Oper Res 24(11):1097–1100
Murray AT (2001) Strategic analysis of public transport coverage. Socio Econ Plan Sci 35:175–188
Murray AT (2013) Optimising the spatial location of urban fire stations. Fire Saf J 62:64–71
Murray AT, Church RL (1996) Applying simulated annealing to location-planning models. J Heuristics 2(1):31–53
Murray AT, Church RL, Gerrard RA, Tsui WS (1998) Impact models for siting undesirable facilities. Pap Reg Sci 77:19–36
Murray AT, Kim K, Davis JW, Machiraju R, Parent R (2007) Coverage optimization to support security monitoring. Comput Environ Urban Syst 31(2):133–147
Plane DR, Hendrick TE (1977) Mathematical programming and the location of fire companies for the Denver fire department. Oper Res 25(4):563–578
Psaraftis HN, Ziogas BO (1985) A tactical decision algorithm for the optimal dispatching of oil spill cleanup equipment. Manag Sci 31(12):1475–1491
Quine WV (1955) A way to simplify truth functions. Am Math Mon 62(9):627–631
ReVelle C, Scholssberg M, Williams J (2008) Solving the maximal covering location problem with heuristic concentration. Comput Oper Res 35(2):427–435
ReVelle CS, Swain RW (1970) Central facilities location. Geogr Anal 2(1):30–42
Rolland E, Schilling DA, Current JR (1997) An efficient tabu search procedure for the *p*-median problem. Eur J Oper Res 96(2):329–342
Rosing KE, ReVelle CS (1997) Heuristic concentration: two stage solution construction. Eur J Oper Res 97(1):75–86
Roth R (1969) Computer solutions to minimum-cover problems. Oper Res 17:455–465
Saydam C, McKnew M (1985) Applications and implementation a separable programming approach to expected coverage: an application to ambulance location. Decis Sci 16(4):381–398
Swersey AJ, Thakur LS (1995) An integer programming model for locating vehicle emissions testing stations. Manag Sci 41(3):496–512
Teitz MB, Bart P (1968) Heuristic methods for estimating the generalized vertex median of a weighted graph. Oper Res 16(5):955–961
Tong D, Murray A, Xiao N (2009) Heuristics in spatial analysis: a genetic algorithm for coverage maximization. Ann Assoc Am Geogr 99(4):698–711
Toregas C (1970) A covering formulation for the location of public facilities. Master's Thesis, Cornell University, Ithaca, NY

Toregas C (1971) Location under maximal travel time constraints. Ph.D. dissertation, Cornell University, Ithaca, NY

Toregas C, Revelle C (1973) Binary logic solutions to a class of location problem. Geogr Anal 5 (2):145–155

Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency services. Oper Res 19:1363–1373

Underhill LG (1994) Optimal and suboptimal reserve selection algorithms. Biol Conserv 70(1):85–87

Von Thünen JH (1826) Uer lsolierte Staat in Beziehung auf' Landwirtschaft un!d NutionaZokonomi, e, Part I, Perthes, Hamburg

Walker W (1974) Using the set-covering problem to assign fire companies to fire houses. Oper Res 22(2):275–277

Weber A (1909) Uber den Standort der Industrien, Tubingen, Translated as Alfred Weber's Theory of the Location of Industries (1929) by C. J. Friedrich, Chicago

White JA, Case KE (1974) On covering problems and the central facilities location problem. Geogr Anal 6:281–293

Chapter 3
Extended Forms of Coverage

3.1 Introduction

Industrial and public sector planning is often complex. Example contexts include making a city safer or designing a manufacturing supply chain. Models can assist in the design of such systems by tracking the major components and generating top performing solutions. Some of the best planning models are simple, easy to understand, and easily applied and solved. At a minimum, a model must capture the main essence of the problem. Two examples of such planning models are the LSCP (location set covering problem) and MCLP (maximal covering location problem) described in Chap. 2. Because they are so simple and powerful as planning models, they have been used in many different ways, from biological reserve design (Underhill 1994) to advertising (Dwyer and Evans 1981) to prosthetic teeth coloring (Cocking et al. 2009) to health clinic location (Bennett et al. 1982). There are, however, a number of elements and conditions that are not captured exactly using these two basic constructs of complete coverage or maximal coverage. Because of this, researchers have proposed a wide variety of extended model forms to better reflect specific elements of a given problem. In this chapter we present some of the work, including multi-service, hierarchical, and multi-factor issues, in coverage modeling.

Even though the concept of "standards-based" coverage is applicable in many areas of planning, its strict form often needs refinement. For example, Chap. 2 discussed the use of the LSCP for deployment of fire stations in an urban region, where the objective was to determine the fewest number of needed fire stations and their locations such that each demand area is served by a station within the maximal service distance or time standard. The assumption here is that each demand area needs the services provided by a station. But, what if one of the demand areas experiences a high incidence of service calls, making residential, school and office buildings significantly more at risk in the event of a fire? It may be that the workload associated with this area is large enough that it makes sense to provide increased

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_3

protection, perhaps by two or more stations. The point is that coverage needs may not always be the same across a region.

As another example, consider a coordinated system of hospitals and clinics, where clinic service should be provided within a short coverage distance and hospital services should be provided within a farther coverage distance. If hospitals can provide both clinic and hospital services, then the objective would be to maximize both clinic and hospital service coverage while locating a limited number of hospitals and a larger number of clinics. Thus, we would seek an efficient, coordinated hierarchical arrangement of both clinics and hospitals. A detailed description of coverage model extensions follows.

3.2 Multiple Service

One model extension addresses whether coverage by a single facility is sufficient. As discussed previously, there may well be situations where one facility may not be capable of providing the quality of service necessary. Figure 3.1 shows a situation where a demand area is required to be served by three sited facilities. Toregas (1970)

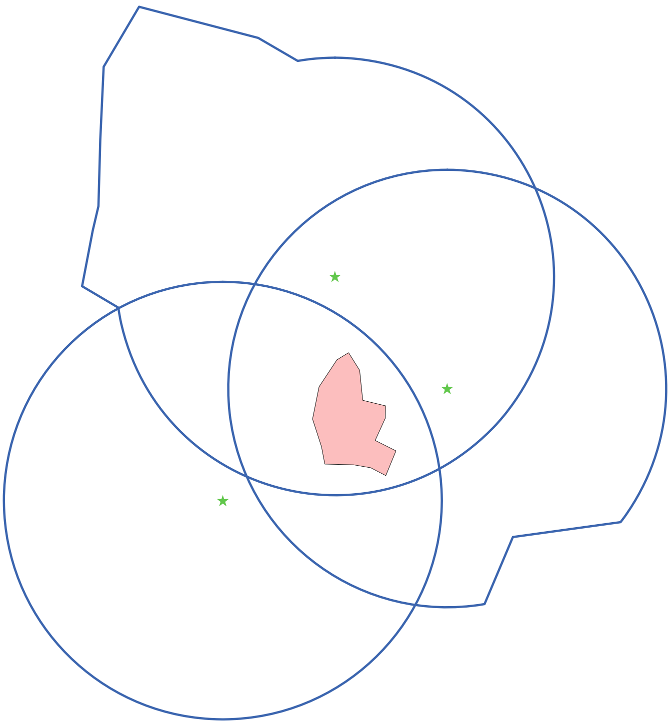


Fig. 3.1 Multiple service requirement for indicated demand area

recognized the potential need for multiple coverage, stating "… the value of the right hand side reflects the requirement that each node be serviced by at least one center which lies within a maximum of S distance units away. However, a situation may be envisaged in which certain nodes may require multiple service." It indeed makes sense that not all demand areas can be covered by the presence of just one facility. The following planning problem arises:

Minimize the number of facilities needed, and locate them in such a manner as to provide coverage to each demand by a desired number of facilities.

This is a relatively simple extension of the LSCP and has been called the Multi-Level Location Set Covering Problem, but here we will call it a Multi-Service LSCP (MS-LSCP) to help distinguish the features of this model from others in this chapter. To formulate this problem mathematically consider the following notation:

i = index of demand areas
j = index of potential facility sites
d_{ij} = distance or travel time between demand i and facility j
S_i = maximal service distance or time standard for demand area i
$N_i = \{j | d_{ij} \leq S_i\}$
l_i = number of facilities required for serving demand i

$$x_j = \begin{cases} 1, & \text{if site } j \text{ is selected for a facility} \\ 0, & \text{otherwise} \end{cases}$$

The mathematical model is as follows:

$$\text{MS-LSCP :} \qquad Minimize \sum_j x_j \tag{3.1}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq l_i \quad \forall i \tag{3.2}$$

$$x_\mathrm{j} \in \{0, 1\} \quad \forall j \tag{3.3}$$

Objective (3.1) follows that of the LSCP, where the number of sites selected for facilities is being minimized. Constraints (3.2) ensure that the number of facilities that serve a given demand equal or exceed l_i. Constraints (3.3) impose binary integer restrictions on decision variables.

Of course, if $l_i = 1$ for each demand i, then the above formulation is equivalent to that of the LSCP given in Chap. 2. Thus, the MS-LSCP is a general form of the LSCP. Toregas (1970) presented this model as being potentially useful for planning emergency services, where some areas of greater demand may need several facilities capable of making a quick response.

Specialized solution procedures have been developed for the MS-LSCP (see for example Hall and Hochbaum 1992). Of particular note is that just as reductions rules can play an important role in solving the (single service) LSCP, it is also possible to extend such rules to the multi-service case (Church and Gerrard 2003).

A multi-service counterpart for the MCLP is also possible. In fact, it has been proposed and structured somewhat indirectly through work on α service coverage (see Hogan and ReVelle 1986a, b). If we assume that facilities are sometimes busy and unable to respond in some instances (e.g., a service response team, like an ambulance, has already been dispatched to another emergency), then coverage of this request for service can only be provided if another facility is located within the response standard. Hogan and ReVelle (1986a, b) suggested that one could estimate how many facilities, l_i, are needed within the coverage region of a demand i in order to provide a desired confidence that demand i will have an available ambulance within the coverage standard α percent (or probability) of the time. The details of such probabilistic coverage are left for Chap. 4.

Structuring the multi-service MCLP (MS-MCLP) requires the following additional decision variables:

$$y_i = \begin{cases} 1, & \text{if demand } i \text{ is served by the required number of facilities} \\ 0, & \text{otherwise} \end{cases}$$

With this notation, the formulation is as follows:

$$\text{MS-MCLP}: \qquad \textit{Maximize} \ \sum_i a_i y_i \tag{3.4}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq l_i y_i \quad \forall i \tag{3.5}$$

$$\sum_j x_j = p \tag{3.6}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{3.7}$$

$$y_i \in \{0, 1\} \quad \forall i \tag{3.8}$$

Objective (3.4) seeks to maximize the amount of demand that is provided the specified level of coverage. Constraints (3.5) relate facilities sited to the coverage requirements. A demand i is counted as being covered only when it is served by at least l_i facilities. Constraint (3.6) specifies that p facilities are to be sited. Constraints (3.7) and (3.8) impose binary restrictions on decision variables.

MS-MCLP is a general form of the MCLP because coverage is considered provided only when enough locally placed facilities cover a specific demand. As was the case for the MS-LSCP, if $l_i = 1$ for all demand i, then the MS-MCLP would be equivalent to the MCLP.

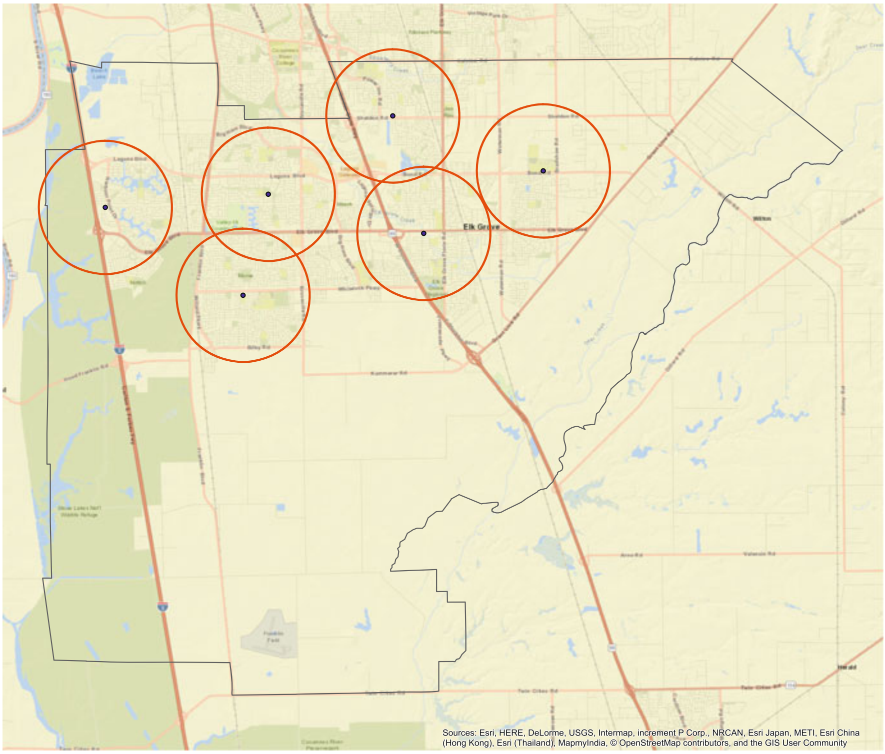


Fig. 3.2 Existing fire stations in an urban area

3.3 Existing Service System

An important element that has not yet been addressed when a model such as the LSCP is used in planning and design is when some facilities already exist in a service system. The existing fire response system and associated service areas for a city is shown in Fig. 3.2 to illustrate this. As some services already exist, the primary objective is to extend, enhance or replace current services with a new configuration of facilities. Plane and Hendrick (1977) dealt with this issue in reconfiguring fire stations in Denver, Colorado. They first identified critical areas containing hospitals and other important demand requiring multiple coverage. To do so, they structured a model similar to the MS-LSCP, (3.1–3.3). Since Denver already had a number of fire stations, it made little sense to them to simply abandon the current system for a new configuration. Therefore they accounted for the existing locations explicitly in the model. After all, the cost of acquiring a new site and building a new station must be more expensive than keeping an already established station in most cases. Thus, they replaced objective (3.1) with the following objective:

$$Minimize \ \ (1+\varepsilon)\sum_{j\notin EF} x_j + \sum_{j\in EF} x_j \tag{3.9}$$

where EF represents the set of existing facilities, and ε is the additional cost of building a new facility as compared to keeping an existing one.

This objective accounts for new stations being built and existing stations being retained. Since the cost of establishing a new station is ε higher than the cost of keeping an existing station, an existing station would be preferred over new sites as long as the existing site demonstrably added to coverage. Thus, Plane and Hendrick (1977) suggested a simple way in which to address the problem of designing a new system when an existing system already exists in a region. Worth noting is that the value of ε can be viewed from two different perspectives. First, it can be considered a true cost factor through a formal cost model analysis. When needed, this cost factor can vary according to the location of each potential site, e.g., ε_j. Second, it can be viewed as a general factor, where the model can be solved for a range of values of ε, thereby enabling solution alternatives to be identified that relate the costs of new stations to existing stations. If solutions are not sensitive to the value of ε, then it is not necessary in many cases to expend the effort in developing a detailed cost model.

Worth mentioning is that annual staffing and maintenance costs over the life of the facility service system may also be important. Murray (2013) discusses this issue in the context of fire service coverage through an extension of the MCLP. Existing facilities may prove to be spatially inefficient, so systematic analysis is essential for evaluating both short- and long-term costs.

3.4 Site Quality

An assumption implied in the models detailed in Chap. 2 is that quality of service provided by facility sites are indistinguishable, beyond the spatial coverage they provide. There are clearly situations where service quality may vary, and accordingly this may impact service in important ways. One example of this is in nature reserve design, where the goal is to establish a system to protect or preserve identified species of flora and/or fauna. A reserve likely ensures diversity and sustainability, but also can address concerns for vulnerability and viability, particularly in cases where a species is threatened or endangered.

A central problem in biological conservation is to determine a good site protection/preservation strategy. Protection can be expensive and so there has always been an attempt to keep site selection to a minimum. Margules (1986) described, and Margules et al. (1988), Pressey et al. (1993) and others have proposed heuristics for, nature reserve design seeking a minimum number of sites that ensure representation of each endangered/threatened species of interest. Underhill (1994) demonstrated that this design problem is equivalent to the LSCP. If the index i refers to each species of concern, then it is possible to derive the set of reserve sites j that contain individual species, N_i. This notation means that the formulation of the LSCP detailed in Chap. 2 may be structured and readily applied to address nature reserve design. Church et al. (1996) established the utility of the MCLP for reserve design (see also Camm et al. 1996). Subsequent research has reaffirmed the significance of both the LSCP and MCLP in supporting conservation efforts.

While representation is one consideration, facility service quality is also an important issue to take into account. For example, whether a species at a site is in decline, or that the site is of high quality with respect to a given species are important considerations. All sites are not necessarily the same in terms of site quality with respect to species representation. Church et al. (2000) suggested that such issues should be incorporated in the site selection process. In particular, they classified sites in terms of high, medium or low quality for a specific species. Of course, other categories of site quality or grading can be used, but the important issue is that the quality of protection can be integrated into the design problem.

The following additional notion is used to generalize the concepts of site quality:

k = index of site quality options (e.g., low, medium, high)
a_i^k = coverage value of demand i at quality k
N_i^k = set of facility sites j that can serve demand i at quality k
$$y_i^k = \begin{cases} 1, & \text{if demand } i \text{ is covered at quality } k \\ 0, & \text{otherwise} \end{cases}$$

Using this notation, a formulation of the site quality MCLP (SQ-MCLP) is as follows:

$$\text{SQ-MCLP}: \qquad \textit{Maximize} \sum_i \sum_k a_i^k y_i^k \tag{3.10}$$

Subject to:

$$\sum_{j \in N_i^k} x_j \geq y_i^k \quad \forall i, k \tag{3.11}$$

$$\sum_k y_i^k \leq 1 \quad \forall i \tag{3.12}$$

$$\sum_j x_j = p \tag{3.13}$$

$$y_i^k \in \{0, 1\} \quad \forall i, k \tag{3.14}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{3.15}$$

The objective, (3.10), maximizes the demand served across the various quality options. In the context of reserve design, this might correspond to the sum of protection value provided for each species, as an example. Constraints (3.11) bound coverage of demand by quality option through siting decisions. Constraints (3.12) limit coverage by quality type, allowing at most one to be tracked. Constraint (2.15) restricts the number of sites selected to equal p. Binary integer restrictions are imposed in constraints (3.14) and (3.15). The objective will ensure that each demand will be counted according to the highest level of coverage quality possible, assuming that higher quality corresponds to higher value.

As noted above, Church et al. (2000) provide an application to biological conservation, and show through analysis in California that using site quality can make a difference in the overall protection capabilities of a reserve system. Site quality has also been of interest in the location of public facilities. Bennett et al. (1982) used the SQ-MCLP to locate health clinics and health service providers in Colombia. They sought to replicate the concerns of the Colombia Health Ministry. First, the maximal travel distance standard was defined as 7.5 km, a distance that could be walked in about an hour. Further, it was specified that people traveling to a clinic should not have to ascend more than 30 m or descend more than 100 m in elevation to visit a clinic. Since hospitals were in the valleys at lower elevations, health professions thought that patients should not travel much higher in elevation to get to a clinic, especially if they were sick. Thus, service coverage was defined based upon distance and elevation change. Health clinics were operated by modestly trained health promoters, or promotoras. These health promoters were trained to educate rural people in personal hygiene, give immunizations, treat wounds, make referrals, and promote healthy habits (e.g., boil water for cooking and drinking). Villages that were considered feasible for the location of a promotora/clinic were those towns that had no hospital as hospitals provided clinic services as well. It was also considered desirable to place clinics in villages that had a safe water supply as this would eliminate the need for the promotora to boil water for use in the clinic. Further, it was desirable for a village to have electricity as this would allow a clinic to have a refrigerator and keep medicines and vaccines that would spoil without refrigeration. These last two elements were considered important but not to the point at which a site would be entirely excluded. Thus, the problem was one in which rural population coverage was to be maximized while using as many site locations that had potable water and electricity. Bennett et al. (1982) compared solutions identified by using the MCLP with site restrictions, without site restrictions, and a planning solution generated by health ministry planners. Even though planners had suggested that electricity and potable water were necessary, their solution used numerous sites which did not have one or the other or both amenities. In fact, the MCLP solution with site restrictions provided a level of coverage that was inferior to what had been generated by the planners. In contrast, the SQ-MCLP solution without site restrictions outperformed the planner's solution both in coverage and in terms of the numbers of sites which were used that had site amenities. Grubesic et al. (2012) evaluated the Essential Air Service program using the LSCP.

Site quality was accounted for in terms of airport efficiency, providing the capacity to distinguish between sites beyond the coverage they provided.

3.5 Multiple Objectives

While classic covering models detailed in Chap. 2 capture the important features of many problem contexts, another way in which extension has been useful is through the inclusion of multiple objectives. Schilling et al. (1980) were one of the first attempts to consider the case where coverage benefit could be measured in different ways. For example, there are two major objectives in fire protection services: (1) protect property, and (2) protect people. The distribution of property values across a city can differ considerably from the distribution of people. Maximizing property coverage may produce a different pattern of fire stations than that generated when maximizing the coverage of people. Both objectives are important, and accordingly Schilling et al. (1980) suggested that the best solution could be generated by considering the two objectives simultaneously.

Addressing such an extension requires the use of the following additional notation:

c_{jl} = cost for potential facility j with respect to consideration l
a_{il} = observed demand i with respect to consideration l

Given these attributes, extended coverage model objectives are possible. An extension of the LSCP to account for multi-objectives, which we will refer to as MO-LSCP (multi-objective LSCP), would be:

$$Minimize \sum_j c_{jl} x_j \tag{3.16}$$

This would give an objective for each consideration $l \in L$, for a total of $|L|$ objectives, and they would replace the single objective of the LSCP, (2.1).

Objectives (3.16) may be combined in a weighted form for solution purposes as:

$$Minimize \sum_l w_l \sum_j c_{jl} x_j \tag{3.17}$$

where w_l is a weight associated with each consideration l.

An extension along these lines for the MCLP is also possible, denoted MO-MCLP (multi-objective MCLP), and can be structured as follows:

$$Maximize \sum_i a_{il} y_i \tag{3.18}$$

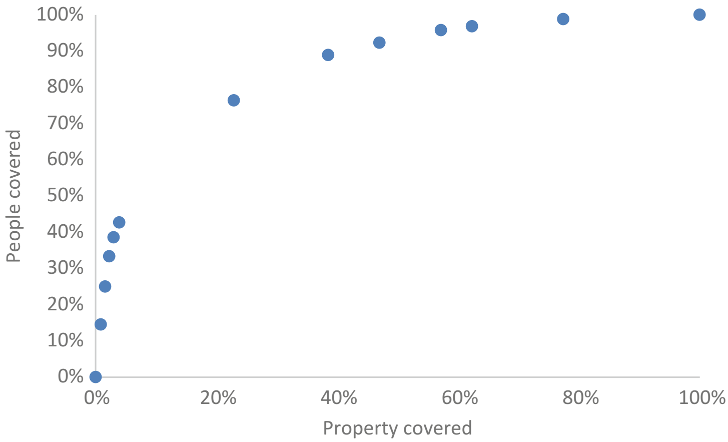


Fig. 3.3 Tradeoff curve summarizing considerations for different coverage objectives

As was the case for the MO-LSCP, this would give an objective for each consideration $l \in L$, for a total of $|L|$ objectives, and they would replace the single objective of the MCLP, (2.8).

Of course, objectives (3.18) may be combined in a weighted form for solution purposes as:

$$Maximize \sum_{l} w_l \sum_{i} a_{il} y_i \tag{3.19}$$

As noted above, Schilling et al. (1980) addressed two considerations, people and property. If $l = 1$ corresponds to people and $l = 2$ corresponds to property, as an example, then the first objective for (3.16) or (3.18) involves maximizing the coverage of people and the second objective involves maximizing the coverage of property values.

Integrated multi-objective problems, such as (3.17) and (3.19) for the MO-LSCP and MO-MCLP respectively, can then be solved for different combinations of values or weights (each of which must be non-negative) in order to find unique solutions. Often such solutions are summarized using a noninferior tradeoff curve. Technically, however, a weighting method approach, like (3.17) and (3.19), enable only supported noninferior solutions to be identified. In order to generate unsupported solutions as well, one must use a more specialized approach, e.g. a weighted Tchebychev (Steuer and Choo 1983) or the constraint method (Cohon 1978).

By varying weights and solving the weighted form of the problem, Schilling et al. (1980) were able to generate a tradeoff curve similar to that given in Fig. 3.3. Since the problem being solved is a discrete integer programming problem, the tradeoff is comprised of a set of discrete points.

Penalba (1980) in a follow-up study to Bennett et al. (1982) used a form of the MO-MCLP. One objective involved maximizing population coverage and the other objective involved maximizing the quality of the sites selected, where a site score was based upon the presence of electricity, water, etc. They generated a tradeoff of coverage versus site quality and found that the planner's solution was inferior to a number of solutions on the non-inferior tradeoff curve. Altogether, they concluded that it made sense to include an objective to encourage the selection of sites with valued amenities/qualities, which were not considered to be absolutely necessary.

Although much of the discussion in this book involves problems structured using a single objective, it should be noted that virtually every model could be cast within the framework of multiple objectives. One cannot always capture the range of perspectives and interests in planning and design problems with a single objective, whereas multiple objectives can be used to value different perspectives and metrics, generate tradeoffs, and aid in finding the best compromise solution. This is a subject in its own right, and the interested reader should consult Cohon (1978) and Steuer (1986) for general developments, and Solanki (1991), Current and Schilling (1994), Grubesic and Murray (2002), Wu and Murray (2005), Matisziw et al. (2006) and Akbari et al. (2017) for specific applications involving covering problems.

3.6 Backup Coverage

An important planning and design question is whether there is any flexibility in solution configuration for a coverage problem. In particular, the existence of alternate optima may be meaningful to management and decision making (Church 1974). For example, will modification of a solution configuration degrade system coverage? Or, is it possible to readjust a solution and maintain total coverage while increasing the number of demand areas that receive some degree of multiple coverage? If either of these situations is true, modeling extensions that enable their identification would be most useful.

With this in mind, the coverage model extensions detailed in this section have to do with the idea of multiple coverage beyond that provided by a single facility, without explicitly requiring specific levels of multiple coverage (as is the case with the MS-LSCP and MS- MCLP discussed in Sect. 3.2). Daskin and Stern (1981) posed the following problem:

Find the minimum number of facilities and their locations such that each demand is covered, while maximizing the number of backup coverage instances among demand areas.

One can consider achieved levels of multiple coverage along these lines as backup for a primary service facility. This problem definition calls for the use of the minimum number of facilities necessary to provide complete coverage as well as good backup coverage. Thus, what is sought is a solution to the LSCP, but also seeking out a solution configuration where the best backup cover is provided, if there is such flexibility. In essence, Daskin and Stern (1981) suggested that there could be multiple optima to a given LSCP and one should therefore engage in further effort to identify the best of these optima, measured by how many times demands are provided additional levels of coverage beyond the initial required single level of coverage. While Daskin and Stern (1981) refer to this as hierarchical objective set covering, we adopt a more consistent naming convention here.

Accordingly, we refer to the extension of the LSCP that addresses backup coverage as LSCP-B (location set covering problem with backup) as follows:

$$\text{LSCP-B}: \quad \textit{Minimize} \sum_j x_j \tag{3.20}$$

$$\textit{Maximize} \sum_i q_i \tag{3.21}$$

Subject to:

$$\sum_{j \in N_i} x_j - q_i \geq 1 \quad \forall i \tag{3.22}$$

$$x_\mathrm{j} \in \{0, 1\} \quad \forall j \tag{3.23}$$

$$q_i \geq 0 \quad \forall i \tag{3.24}$$

where q_i is a decision variable corresponding to the number of times that demand i is covered beyond the required initial coverage of one facility. The first objective, (3.20), of LSCP-B is equivalent to that of the LSCP, to minimize the total number of facilities. The second objective, (3.21), involves maximizing the number of times that demand is covered beyond a single facility. Constraints (3.22) require coverage of each demand. However, each constraint also includes explicit tracking of backup coverage using decision variable q_i. This is because q_i will be as high as possible given the orientation of objective (3.21), but is bounded by the number of facilities providing coverage to individual demand i, less the required level of one. Binary integer conditions are specified in constraints (3.23). Non-negativity requirements are imposed in constraints (3.24).

One may contrast the LSCP-B with the MS-LSCP. Of course, objective (3.21) is not included in the MS-LSCP. Beyond this, constraints (3.2) in the MS-LSCP are modified through a substitution of $l_i = (1 + q_i)$, which gives constraints (3.22).

Daskin and Stern (1981) actually structured their version of LSCP-B as a single objective model, adding together the two objectives, (3.20) and (3.21), using an importance weight, w, as follows:

$$\textit{Minimize} \ \ w \sum_j x_j - \sum_i q_i \tag{3.25}$$

Of course, minimizing the negative term in the second part of (3.25) is equivalent to maximizing the positive value of this term. The importance weight w must be large enough so that the first term, the number of located facilities, will be the lowest value possible. If this is the case, then it is equivalent to the LSCP. If it is desired that a LSCP-B solution is equal to that of the LSCP using the single objective, (3.25), then one valid bound on w for which a minimum number of facilities will be used is:

$$w \geq \sum_i q_i \tag{3.26}$$

Thus, condition (3.26) ensures that the smallest number of facilities possible in covering all demand will result.

Alternatively, one can also solve the above problem where $w > |I|$, as no facility can ever cover more than the total of all demands (either as required or as backup). Of course, there is an alternate way to ensure equivalence to the LSCP. A multi-objective solution approach for the LSCP-B can be used to obtain the full spectrum of Pareto tradeoff alternatives, among which would be one or more that represent an LSCP optimum.

A backup extension of the MCLP is also possible. Hogan and ReVelle (1986a, b) expanded on the concept of providing additional service coverage to demand. They argued that providing a second or multiple instances of service coverage should be biased towards those demands areas with higher service requirements. They also reasoned that it was more important to provide a second service facility coverage, or backup, to as much of the demand as possible, instead of maximizing the number of times demands are covered beyond a first instance. In essence, providing backup coverage to each demand should be accomplished before considering additional service coverage levels. This is clearly an extension of basic coverage notions, and further refines how additional service coverage may be viewed.

A formulation of the MCLP that is extended to account for backup (maximal covering location problem with backup, MCLP-B) is as follows:

$$\text{MCLP-B}: \quad Maximize \sum_i a_i y_i \tag{3.27}$$

$$Maximize \sum_i a_i q_i \tag{3.28}$$

Subject to:

$$\sum_{j \in N_i} x_j - y_i - q_i \geq 0 \quad \forall i \tag{3.29}$$

$$q_i - y_i \leq 0 \quad \forall i \tag{3.30}$$

$$\sum_j x_j = p \tag{3.31}$$

$$0 \leq y_i \leq 1 \quad \forall i \tag{3.32}$$

$$0 \leq q_i \leq 1 \quad \forall i \tag{3.33}$$

$$x_j = \{0, 1\} \quad \forall j \tag{3.34}$$

The first objective, (3.27), is the same as the MCLP, involving the maximization of the amount of demand provided coverage. The second objective, (3.28), seeks to maximize the total demand covered a second time. Constraint (3.29) tracks primary and secondary coverage of a demand, ensuring that it is not counted unless the appropriate number of facilities have been located to cover the demand. Constraint (3.30) establishes that secondary, or backup, coverage is counted only when first level coverage has already been tallied, $y_i = 1$. Constraints (3.29) and (3.30) keep each q_i at a zero value unless demand i is covered at least twice. Constraints (3.31) specify that p facilities are to be sited. Bounds on decision variables are required in constraints (3.32) and (3.33). Finally, binary integer conditions are imposed in constraints (3.34).

The MCLP-B largely corresponds to the BACOP2 approach of Hogan and ReVelle (1986a, b), though they allow facilities to co-locate, meaning that the x_i variable was defined as a general positive integer. Beyond this, their BACOP1 approach roughly corresponds to LSCP-B, but again allows for co-location. For consistency and comparative purposes, we have opted to not consider co-location here. However, the above models could readily be extended accordingly.

Storbeck (1982) was also concerned about finding those solutions that covered as much as possible a second or third time. He formulated a goal programming form of the MCLP that was referred to as goal-oriented location covering. Each covering condition was defined as a goal constraint, where the deviation below the goal represented leaving a demand uncovered and the deviation above the goal represented the provision of additional service coverage. By weighting shortfalls of the covering goals higher than providing additional service coverage, the Storbeck (1982) model could ensure that an MCLP solution was achieved. A negative weight for any deviations above a coverage goal means that additional service coverage was also valued. One can view the approach as an alternative to that of LSCP-B and MCLP-B.

3.7 Coordinated Systems

Early work on modeling service coverage was based upon the assumption that the facility, e.g. a fire station, provided coverage to those areas that were within a service distance or time standard of the facility. This assumption is supported when enough resources are located at each station to serve any calls for service within its coverage area. Unfortunately, resources are limited due to budget realities. There are two main issues in allocating resources. The first is concerned with the relative server busy-ness, and the second factor is associated with specific needs of the service. To explain the nature of these two factors, consider the differences between emergency medical response and fire response. Although both are important, the call volume for emergency medical services (EMS) is often ten times that of firefighting. Consequently, fire equipment is often idle much of the time whereas ambulances may be busy 50% or more of the time. In this sense one can observe that availability and busyness of servers may or may not be an issue, on average, when a call for service is received. Interestingly, many municipalities attempt to coordinate services to make better use of available resources. In particular, firefighters have been used to help with EMS, dispatched to calls when they are available as they are often close by and can respond quickly. Then, an ambulance arrives and takes over the service along with the transport of the patient to a hospital when that is necessary. When a patient transport occurs, it often takes the ambulance crew an hour to complete the service call and be available for handling another call. The firefighters, in contrast, make a quick response, help in stabilizing the patient, and return to the station once the ambulance crew arrives. Often, they are away from the station less than 10 min, and are still available to respond to a call for fire related services. The important point here is that these services are often coordinated. The remainder of this section details several coverage models that have been developed to locate and allocate different types of facilities or equipment. Chap. 4 focuses on problems involving availability and busyness issues.

3.7.1 Facility Type

The LSCP and MCLP are well suited for an application like that of fire station placement. This is due to the notion that a quick first response to all areas of a city can be made only when each neighborhood is covered by nearby service. But, service is more than just the presence of a station. The real service is provided by the crew and their equipment. The fact is that there are several types of equipment. The most common fire fighting vehicle is called a pumper or fire engine. Such vehicles have an array of hoses, some hand ladders, and other tools. They often have a small reservoir of water for immediate use in fighting a fire. The engine's pumps are strong enough to spray water at some distance. It is also designed to be connected to a hydrant or water tender for longer periods of use. Although there are many other types of equipment, one specialty vehicle is an aerial ladder truck, capable of being used for higher buildings and for above ground attack of a fire as well as rescue. Specialty trucks, like aerial apparatus, are not as common as the classic pumper and are not housed by themselves, but always in conjunction with one or more pumpers. Schilling et al. (1979) suggested that fire coverage could be defined by the equipment that was located/allocated. They proposed a model that would allocate both engines/pumpers and specialty equipment. They assumed that specialty equipment has broader response coverage standard than that of the engines. To have coverage, a demand area requires both engine and specialty equipment, though it may be housed at different stations.

The following notation will facilitate a generalization of basic coverage models that account for facility type:

t = index of facility type (entire T)
S^t = service standard associated with facility type t
$x'_{jt} = \begin{cases} 1, & \text{if facility of type } t \text{ located at site } j \\ 0, & \text{otherwise} \end{cases}$
$N'_{it} = \{j | d_{ij} \le S^t\}$

The gives rise to the following model, extending the LSCP to account for facility types (location set covering problem with facility types, LSCP-FT):

$$\text{LSCP-FT}: \qquad Minimize \sum_t \sum_j x'_{jt} \tag{3.35}$$

Subject to:

$$\sum_{j \in N_{it}} x'_{jt} \ge 1 \quad \forall i, t \tag{3.36}$$

$$x'_{jt} \in \{0, 1\} \quad \forall j, t \tag{3.37}$$

Objective (3.35) seeks a minimum total number of facilities. Constraints (3.36) ensure that each demand is covered by each facility type. Constraints (3.37) impose integer restrictions on decision variables.

Kolesar and Walker (1974) propose a model along the lines of the LSCP-FT where they substitute a fixed number of a new platform trucks for old aerial ladders. This was applied to assist the New York City Fire Department.

Another extension, of course, involves the MCLP. First some additional notation is presented:

p^t = number of facilities of type t to be located

With a budget or limit by facility type, the following formulation of the MCLP with different facility types (maximal covering location problem with facility types, MCLP-FT) is:

$$\text{MCLP-FT}: \qquad Maximize \sum_i a_i y_i \tag{3.38}$$

Subject to:

$$\sum_{j \in N'_{it}} x'_{jt} \ge y_i \quad \forall i, t \tag{3.39}$$

$$\sum_j x'_{jt} = p^t \quad \forall t \tag{3.40}$$

$$x'_{jt-1} \le x'_{jt} \quad \forall j, 1 < t < |T| \tag{3.41}$$

$$x'_{jt} \in \{0,1\} \quad \forall j,t \tag{3.42}$$

$$y_i \in \{0,1\} \quad \forall i \tag{3.43}$$

The objective of MCLP-FT is to maximize total demand covered within the stipulated response standard. Constraints (3.39) specify coverage of each demand by facility type. Notice that coverage is only possible if each facility type covers a demand. Constraints (3.40) limit the number of each facility type sited. Constraints (3.41) establish conditions on siting different types at a particular location, assuming that there is an existence necessity. For the example above, this might be that an engine must be sited before specialty equipment like a ladder truck. Binary integer restrictions are imposed in constraints (3.42) and (3.43).

The MCLP-FT is a general case of the model detailed in Schilling et al. (1979) referred to as TEAM (tandem equipment allocation model). They considered primary equipment and special equipment for fire response. A similar concern was also detailed in Brotcorne et al. (2003) for ambulance siting.

TEAM was conceived to allocate equipment to existing stations and applied to assist the Baltimore City Fire Department. TEAM represents a way for a fire department to reposition equipment among existing stations in order to increase coverage levels. As a part of that study, the fire department was also interested in identifying the response coverage that could be provided by optimizing station location and equipment allocation simultaneously. Thus, what is not considered in LSCP-FT and MCLP-FT is coordination with a facility. Schilling et al. (1979) proposed FLEET (Facility Location, Equipment Emplacement Technique) to address facility coordination.

Such an extension of the LSCP with coordinated facility types (location set covering problem with coordinated facility types, LSCP-CFT) is as follows:

$$\text{LSCP-CFT}: \qquad \textit{Minimize} \sum_j x_j \tag{3.44}$$

Subject to:

$$\sum_{j \in N'_{it}} x'_{jt} \ge 1 \quad \forall i,t \tag{3.45}$$

$$x'_{jt} \le x_j \quad \forall j,t \tag{3.46}$$

$$x'_{jt} \in \{0,1\} \quad \forall j,t \tag{3.47}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{3.48}$$

Objective (3.44) seeks a minimum total number of facilities. Constraints (3.45) ensure that each demand is covered by each facility type. Constraints (3.46) relate allocation of service types to a facility. Constraints (3.47) and (3.48) impose binary integer restrictions on decision variables.

An extension of the MCLP with coordinated facility types (maximal covering location problem with coordinated facility types, MCLP-CFT) is:

$$\text{MCLP-CFT}: \qquad \textit{Maximize} \sum_i a_i y_i \tag{3.49}$$

Subject to:

$$\sum_{j \in N'_{it}} x'_{jt} \geq y_i \quad \forall i, t \tag{3.50}$$

$$\sum_j x'_{jt} = p^t \quad \forall t \tag{3.51}$$

$$\sum_j x_j = p \tag{3.52}$$

$$x'_{jt} \leq x_j \quad \forall j, t \tag{3.53}$$

$$x'_{jt} \in \{0, 1\} \quad \forall j, t \tag{3.54}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{3.55}$$

$$y_i \in \{0, 1\} \quad \forall i \tag{3.56}$$

The MCLP-CFT objective, (3.49), seeks to maximize demand covered. Constraints (3.50) define coverage of each demand by facility type. Coverage is provided only when each facility type suitably serves a demand. The number of facility types is specified in constraints (3.51), while the number of stations is specified in constraints (3.52). Constraints (3.53) require that a station be sited before a facility type be allocated to the station. Finally, binary integer restrictions are specified in constraints (3.54), (3.55) and (3.56).

The MCLP-CFT is a general case of FLEET detailed in Schilling et al. (1979). The facility types are an engine and specialty equipment, for example. They may only be allocated to a station if it has been sited, which is structured in constraints (3.53).

3.7.2 Assisting Facility Types

Another form of coordinated coverage was developed for eradication of Dracunculiasis, more commonly called the Guinea worm. It is a water-borne parasitic disease that is endemic to a number of African and Asian countries. The disease can incapacitate a victim through severe pain or even paralysis. Because this disease can have such an impact on communities, people, and their economies, the United Nations embarked on an eradication program. Controlling this disease can be brought about by a combination of three main measures: water-supply management, health education, and health service intervention (Osleeb and Mclaferty 1992). A model was proposed by Osleeb and Mclaferty (1992) to allocate wells and schools in support of eradication efforts. Their view was that providing a safe-water well nearby to a community would help in eradication. If a school is located nearby to a village, then people would receive education about its spread and how to treat their water by filtering. They surmised that the best result would be afforded those villagers that were close to schools and safe water wells.

Additional notation is:

t' = index of facility types, including combinations

$a'_{it'}$ = amount of demand i served by a facility of type t'

$y'_{it'} = \begin{cases} 1, \text{if demand } i \text{ served by one or more facility types } t' \\ 0, \text{otherwise} \end{cases}$

Ω_t = set of facility types capable of assisting with services provided by a type t facility

In essence then, a facility of type t is capable of providing specific sorts of service, but in a coordinated system other facility types may be able to also provide such services under certain conditions, in addition to the specialized services that they contribute to a system. One can therefore think of this as an assisting system of facility types. According, the following model is an extension of the MCLP that accounts for assisting facility types (maximal covering location problem with assisting facility types, MCLP-AFT):

$$\text{MCLP-AFT}: \qquad Maximize \sum_i \sum_t a'_{it} y'_{it} \tag{3.57}$$

Subject to:

$$\sum_{j \in N'_{it}} x'_{jt} \geq \sum_{t' \in \Omega_t} y'_{it'} \quad \forall i, t \tag{3.58}$$

$$\sum_j x'_{jt} = p^t \quad \forall t \tag{3.59}$$

$$\sum_{t'} y'_{it'} \leq 1 \quad \forall i \tag{3.60}$$

$$x'_{jt} \in \{0, 1\} \quad \forall j, t \tag{3.61}$$

$$y'_{it'} \in \{0, 1\} \quad \forall i, t' \tag{3.62}$$

The objective, (3.57), of the MCLP-AFT is to maximize coverage by facility types. Constraints (3.58) relate facility siting decisions by type to coverage provided. Based on which associated facility types can provide assistance to another facility type, the set Ω_t indicates the assistance relationships that would constitute coverage (including combinations). Constraints (3.59) define the number of facilities of each type to be sited. Constraints (3.60) limit coverage to only one coordinated facility type. Constraints (3.61) and (3.62) specify binary integer requirements.

In the case of Osleeb and Mclaferty (1992) they considered facility types to be a well ($t = t' = 1$), school ($t = t' = 2$) and the combined well and school ($t' = 3$) scenario. The coverage set, N_{i1}, is the set of neighboring villages to i which can serve i with safe water if a well is placed at one of those villages in the set. The set N_{i2} represents the set of neighboring villages that can serve i for education should one of them be selected for a school. The objective maximizes the total risk reduction afforded the population across all villages. Risks can be reduced by providing well coverage, providing education coverage, or providing both well and school coverage. Of course, MCLP-AFT is a generalization of the original model detailed in Osleeb and Mclaferty (1992).

3.7.3 Coordinated Access

Since the 1970's and the introduction of legislation in the U.S. to establish standards in pre-hospital care services, special attention has been turned to the provision of trauma services. The overarching goal in the U.S. is to provide 90% of the population with trauma services within 60 min. It is important to respond to trauma cases immediately, transport them to a trauma center, marshal specialists, and begin surgery within what is termed the golden hour (60 min) in order to provide a high probability of patient survival. One of the models that has been formulated to address trauma services is a model called TRAMAH (Trauma Resource Allocation Model for Ambulances and Hospitals) (Branas et al. 2000). This model involves a different form of coordination where it is necessary to track exactly which pairs of facilities are being located in order to identify whether service coverage is provided. We term this coordinated access. Trauma centers are special additions to a hospital, and involve unique patient resources for victims of trauma, including teams of specialists that are on call and can provide aid immediately. While a trauma patient is brought to such a center, specialists are called and they come to the center as well, so when the patient arrives, a team of specialists is ready to address the patient's critical needs.

Since trauma centers are subsidized operations and are expensive, the objective is to establish as few of them as possible with the plan to ensure as much coverage as possible. Typically, such centers are co-located with a hospital, so one of the decisions is to choose which hospitals are to be selected for trauma services. Ground-based ambulances can be used to bring trauma victims to a hospital as long as they are not too far from a hospital. Outside of this ground-based travel range, aero-medical transport can be used (most often this is a helicopter). To provide cost effective trauma services requires designating a few hospitals to serve for trauma care (TC) and locating a fixed number of dispatch depots for helicopters (AD), so that altogether trauma services are within reach for as many people as possible. This is the basic aim of the TRAMAH approach. This model can be formulated as an extension of the MCLP that includes coordinated access (maximal covering location problem with coordinated access, MCLP-CA) using the following added/refined notation:

$J =$ the set of eligible TC locations
$K =$ the set of eligible AD locations
$N_i =$ the set of trauma care sites that are within the response time standard using ground transport
$M_i =$ the set of pairs (k,j) where an AD at k and a TC at j will provide coverage by air within the response time standard

$$y_i = \begin{cases} 1, & \text{if demand } i \text{ is covered by air or ground service} \\ 0, & \text{otherwise} \end{cases}$$

$$v_i = \begin{cases} 1, & \text{if demand } i \text{ is covered by ground transport} \\ 0, & \text{otherwise} \end{cases}$$

$$u_i = \begin{cases} 1, & \text{if demand } i \text{ is covered air service} \\ 0, & \text{otherwise} \end{cases}$$

$$x_j^{TC} = \begin{cases} 1, & \text{if location } j \text{ is selected for a trauma center} \\ 0, & \text{otherwise} \end{cases}$$

$$x_k^{AD} = \begin{cases} 1, & \text{if location } k \text{ is selected for an air transport depot} \\ 0, & \text{otherwise} \end{cases}$$

$$z_{kj} = \begin{cases} 1, & \text{if location } k \text{ is selected for AD and node } j \text{ for a TC} \\ 0, & \text{otherwise} \end{cases}$$

We can now formulate this model as follows:

$$\text{MCLP-CA :} \qquad \mathit{Maximize} \sum_i a_i y_i \tag{3.63}$$

Subject to:

$$\sum_{j \in N_i} x_j^{TC} \geq v_i \quad \forall i \tag{3.64}$$

$$\sum_{(j,k)\in M_i} z_{kj} \geq u_i \quad \forall i \tag{3.65}$$

$$\sum_{j\in J} x_j^{TC} = p^{TC} \tag{3.66}$$

$$\sum_{k\in K} x_k^{AD} = p^{AD} \tag{3.67}$$

$$y_i - v_i - u_i \leq 0 \quad \forall i \tag{3.68}$$

$$z_{kj} \leq x_k^{AD} \quad \forall k, j \tag{3.69}$$

$$z_{kj} \leq x_j^{TC} \quad \forall k, j \tag{3.70}$$

$$x_j^{TC} \in \{0,1\}, x_k^{AD} \in \{0,1\} \quad \forall j, k \tag{3.71}$$

$$y_i \leq 1 \quad \forall i \tag{3.72}$$

where all other variables are non-negative in value. Objective (3.63) of this model maximizes the number of people covered with trauma services by either ground transport or by air transport to an appropriately located trauma center. Constraint (3.64) dictates whether a given demand is within the response time standard of a trauma center based upon ground transport. Constraint (3.65) defines whether a trauma center and an air transport depot are each located so that trauma services are provided within the maximum response time standard for a given demand. Condition (3.68) allows coverage of a given demand i to be counted only when that demand is within ground based coverage of a trauma center or when it is within a combination of air base locations and trauma center locations which provided coverage or perhaps a combination of both air and ground services being available to provide coverage. Constraints (3.69) and (3.70) are used to define combinations of decisions involving both trauma center location and aerial depot location. The rest of the model represents the needed conditions on the decision variables. This MCLP-CA, when applied to a large planning area, can be quite difficult to solve, requiring the use of a heuristic (Branas and ReVelle 2001).

3.7.4 Conditional Coverage

Moon and Chaudhry (1984) were the first to suggest that facilities may be needed to cover each other. They proposed what they called a "double set covering" problem where facilities were required to cover all demand but also cover each other. This is yet a different form of coordination, which can involve the same type of facility, rather than "schools and wells" or "trauma centers and air transport depots." ReVelle et al. (1996) argue for this type of approach in the location of ambulances, where each ambulance is required to support coverage in a primary service area. In the event that a particular ambulance may be busy and another call is received in their service area, they argued that it would be desirable to have another nearby server that could move in and service that call. Moon and Chaudhry (1984) formulated this problem as follows, which we refer to as the LSCP-CC (location set covering problem with conditional coverage):

$$\text{LSCP-CC}: \qquad Minimize \sum_{j} x_j \tag{3.73}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq 1 \quad \forall j \tag{3.74}$$

$$\sum_{k \in Q_j} x_k \geq x_j \quad \forall j \tag{3.75}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{3.76}$$

where $Q_j = \{k| \ d_{jk} \leq T$ where $k \neq j\}$. The objective minimizes the number of facilities being deployed. Constraints (3.74) represent the classic covering constraints of the LSCP, which require that each demand is covered by at least one of the located facilities. The next constraint (3.75) requires that when site j is selected for a facility, then there must be at least one other facility located within T distance or time of site j. Note that the coverage set for site j, Q_j, includes all surrounding sites of j but excludes site j itself. This means that if a given site is chosen for a facility, then one of the surrounding sites must be chosen as well. Moon and Chaudhry (1984) found that when this model was solved as a linear programming problem by relaxing the integer conditions on the x_j variables, most of the decision variables were 0.50 in value and the sites that were chosen were pairs of nearby sites. Their attempts at creating a formulation which resulted in natural integer solutions were not very successful. This means that solving this type of model will often require a branch and bound algorithm. Chaudhry et al. (1987) explored the use of several greedy inspired heuristics in solving LSCP-CC. They found for modest sized problems that they could generate optimal solutions approximately 50% of the time, and a high percentage of the time they could find a solution which used no more than one additional site.

A MCLP-CC (maximal covering location problem with conditional coverage) follows from this. ReVelle et al. (1996) proposed two forms of what they termed the Maximal Conditional Covering Problem. Essentially, they assumed that all demand must be covered by at least one facility. This means that the number of facilities being located would be equal to or greater than $\bar{p}$, as defined earlier. Given a prespecified number of facilities $p \geq \bar{p}$, additionally, they wished to place facilities in such a manner that as many located facilities were provided a conditional backup, if at all possible.

3.8 Hierarchical Services

Facilities and their services can be related in different ways, for instance there could be a hierarchy of facilities. There are a number of examples of hierarchical services. The classic organization of schools is a good example, and includes elementary, middle (or junior high), and high schools. The lowest is the elementary school, often housing seven grade levels (kindergarten through sixth). The students of many elementary schools (7–10, depending on enrollments) are transitioned to one junior high. Similarly, students from multiple junior highs are transitioned to one high school. A similar hierarchical arrangement can be found in health services, namely clinics, hospitals, and regional medical centers. Medical services are often successively inclusive hierarchies, where the services provided at one level of the hierarchy are all provided at the next level. For example, clinic services are offered at a hospital and clinic and hospital services are offered at a regional medical center, where additional specialties are also provided (Engel 1968). Dökmeci (1973) was one of the first to formulate an optimization model to account for these hierarchical relationships.

3.8.1 *Successively Inclusive Services*

Moore and ReVelle (1982) proposed a model to locate successively inclusive services and applied it to a health services planning problem in Honduras. For their planning problem they located both hospitals and clinics, where hospitals offered services beyond that of clinics. Consider the following additional or modified notation:

c = an index to reference a clinic facility
h = an index to reference a hospital
p_k = the number of facilities of type k to be located, $k = c \text{ or } h$
l = an index to reference a service level, where $l = c$ or h
R^{kl} = the distance or time standard for a facility of type k providing service level l.

$$\bar{y}_i = \begin{cases} 1, & \text{if demand } i \text{ is not covered by at least one service type} \\ 0, & \text{otherwise} \end{cases}$$

$$x_j^k = \begin{cases} 1, & \text{if a facility of type } k \text{ is located at site } j \\ 0, & \text{otherwise} \end{cases}$$

S^{kl} = the distance standard for a facility of type k providing coverage for service l
$N_i^{kl} = \{j \in J | d_{ij} \leq S^{kl}\}$, the set of sites that if selected for a facility of type k will provide service level l for demand i

We will call the model of Moore and ReVelle (1982) the hierarchical coverage MCLP (maximal covering location problem with hierarchical coverage, MCLP-HC), and is formulated as:

$$\text{MCLP-HC}: \qquad \text{Minimize} \sum_{i \in I} a_i \bar{y}_i \tag{3.77}$$

Subject to:

$$\sum_{j \in N_i^{cc}} x_j^c + \sum_{j \in N_i^{hc}} x_j^h + \bar{y}_i \geq 1 \quad \forall i \tag{3.78}$$

$$\sum_{j \in N_i^{hh}} x_j^h + \bar{y}_i \geq 1 \quad \forall i \tag{3.79}$$

$$\sum_{j \in J} x_j^k = p_k \quad \text{for each } k = c, h \tag{3.80}$$

$$x_j^k \in \{0, 1\} \quad \forall j \text{ and each } k = c, h \tag{3.81}$$

$$\bar{y}_i \in \{0, 1\} \quad \forall i \tag{3.82}$$

This model seeks to minimize the amount of demand that lacks coverage at one or more levels of service, clinic or hospital. Constraints (3.78) establish that if a given demand i is not provided clinic coverage, the variable $\bar{y}_i$ will be forced to be 1 in value, indicating that demand i is left uncovered for one or more levels of service. Constraints (3.79) establish that if a given demand i is not provided hospital coverage, the variable $\bar{y}_i$ will be forced to be 1 in value, indicating that demand i is left uncovered for one or more levels of service. Altogether, constraints (3.78) and (3.79) force $\bar{y}_i$to be one if either service, clinic or hospital coverage, has not been provided to demand i. Constraints (3.80) specify that given number of facilities of each type (clinic or hospital) are to be located. The MCLP-HC represents a successively inclusive hierarchy as that defined by Moore and ReVelle (1982). It should be noted that the above model is equivalent to maximizing the population that is provided all levels of service.

Ratick et al. (2009) extended the MCLP-HC by replacing constraints (3.80) with a budget constraint. The budget constraint including fixed and variable costs for establishing facilities of each type in the hierarchy. The variable costs were a function of the potential size of population that could be served by a given facility within the hierarchy and included economies of scale. They applied their extended model to the Kohut district of Pakistan for a health services planning problem involving the positioning of clinics and hospitals.

3.8.2 Activities Between Levels

Church and Eaton (1987) described two different hierarchical arrangements that they found present in a medical services planning problem in Colombia, also involved clinics and hospitals. In this section we describe these two arrangements and give a general form of the problem that optimizes both types of service arrangements. In Columbia, medical services in rural areas are provided by health promoters (called promatoras) who manage clinics. There are also hospitals that provide both clinic and hospital services, just as in the MCLP-HC. The major difference is that the MCLP-HC does not account for any activities that may occur between levels. For example, if a patient goes to a clinic, they may be referred to a hospital, as the nature of their illness or injury is too severe to be handled at the clinic. Thus, a clinic needs to be within a reasonable distance of a hospital for patient referral from a clinic to a hospital to be accomplished in a timely manner. The Colombian system also had doctors from hospitals visiting clinics periodically. These periodic visits are difficult to accomplish unless the hospital is within a reasonable travel time to a clinic, so that a doctor may spend part of a day at a clinic and the rest of the day at the hospital, without taking too much time in transit between the clinic and the hospital.

Church and Eaton (1987) developed two models, one for hierarchical coverage with referral and one for hierarchical coverage with hospital services provided to clinics. Gerrard and Church (1994) presented a general form of this problem which optimizes both referral and hospital services to clinics while locating a hierarchical system of clinics and hospitals, which we present here. Consider the following additional/modified notation:

$$\bar{y}_i = \begin{cases} 1, & \text{if demand } i \text{ is not covered by a clinic} \\ 0, & \text{otherwise} \end{cases}$$

$$\bar{z}_i = \begin{cases} 1, & \text{if demand } i \text{ is not covered by a hospital} \\ 0, & \text{otherwise} \end{cases}$$

$$t_i = \begin{cases} 1, & \text{if demand } i \text{ is not covered by a clinic wich receives upward referral} \\ & \text{coverage from a hospital} \\ 0, & \text{otherwise} \end{cases}$$

$$r_j = \begin{cases} 1, & \text{if a clinic at } j \text{ is covered by hospital within the hospital to clinic} \\ & \text{referral distance} \\ 0, & \text{otherwise} \end{cases}$$

$a_i =$ the demand for clinic services at i
$\alpha_i =$ the fraction of demand at i that goes to a clinic and is then referred to a hospital for care
$b_i =$ the demand for hospital services at i

S^r = the maximum service distance for coverage of clinics by hospitals for referral services, either referral of a patient to a hospital of a visit of a doctor to a clinic
$M_j^r = \{k \in J | \, d_{kj} \leq S^r \}$

A coverage model along these lines to represent a generalization of hierarchical relationships is referred to as the MCLP-GH (generalized hierarchical maximal covering location problem) and is formulated as:

$$MCLP - GH: \quad \text{Minimize } \left[\sum_{i \in I} a_i \bar{y}_i, \ \sum_{i \in I} b_i \bar{z}_i, \ -\sum_{j \in J} r_j, \ \sum_{i \in I} \alpha_i a_i t_i\right] \tag{3.83}$$

Subject to:

$$\sum_{j \in N_i^{cc}} x_j^c + \sum_{j \in N_i^{hc}} x_j^h + \bar{y}_i \geq 1 \quad i \in I \tag{3.84}$$

$$\sum_{j \in N_i^{hh}} x_j^h + \bar{z}_i \geq 1 \quad i \in I \tag{3.85}$$

$$\sum_{j \in J} x_j^k = p_k \quad \text{for each } k = c, h \tag{3.86}$$

$$\sum_{k \in M_j} x_k^h \geq r_j \quad \forall j \tag{3.87}$$

$$x_j^c \geq r_j \quad \forall j \tag{3.88}$$

$$\sum_{j \in N_i^{cc}} r_j + \sum_{j \in N_i^{hc}} x_j^h + t_i \geq 1 \quad \forall i \tag{3.89}$$

$$x_j^k \in \{0, 1\} \quad \forall j \text{ and each } k = c, h \tag{3.90}$$

$$r_j \in \{0, 1\} \quad \forall j \tag{3.91}$$

$$\bar{y}_i \in \{0, 1\}, \ \bar{z}_i \in \{0, 1\} \ t_i \in \{0, 1\} \quad \forall i \tag{3.92}$$

The MCLP-GH has four objectives, listed as a vector. The first component of this multiobjective problem involves minimizing the demand that is not provided clinic coverage. The second component involves minimizing the demand that is not provided with hospital coverage. The third component maximizes the number of clinics that are covered with referral services by hospitals (referral of patients to hospitals, and doctor visits to clinics). The last component represents minimizing the number of referral patients from clinics that are not within a suitable distance of a hospital. Constraints (3.84) define whether a given demand is covered with clinic services by either hospitals or clinics. Constraints (3.85) define whether a given demand is covered with hospital services. Constraints (3.86) specify how many clinics and hospitals are to be located. Constraints (3.87) and (3.88) are used to define if a clinic has been located at a specific location j and it is within the referral coverage distance of a located hospital. If that happens then the variable r_j can be one in value. Otherwise it is forced to be zero. Constraints (3.89) are used to define whether a given demand has clinic service coverage and that clinic is within the referral coverage standard of a hospital. This occurs, when a clinic covers that demand and the clinic has referral coverage (i.e. $r_j = 1$) or that demand is covered by clinic services at a hospital. The reainder of the conditions specify the binary restrictions on the decision variables. Gerrard and Church (1994) give a detailed application of this general model to the Zarzal region of Colombia.

3.9 Multiple Optima

Church (1974) recognized that there could be many alternate optima for a given covering problem, especially simple constructs like the LSCP. He reasoned that one should search for those alternate optima that performed better than others in terms of criteria like access. Many of the models that have been formulated in this chapter help to search for those solutions that cover everything or a high percentage of demand while using sites that are of higher quality, provide higher levels of backup coverage, or even arranging facilities so that they cover each other. Brill (1979) suggested that one should seek alternative solutions to a given public planning problem because there could be a missing objective or an important issue that had not been articulated by planners and decision makers. Brill (1979) then described a method that can be applied to many of the problems, such as those detailed in this book, for identifying alternate, but different solutions that are close to optimal. Arthur et al. (1997) give a very simple way of generating alternate optima for a facility location problem in general, even though they specified the approach for finding alternate optima when using the MCLP in reserve site selection. Consider if you will that we have solved some problem, like the MCLP, to optimality. The optimal solution will be comprised of p sites, which we will designate as set F^1 and has an objective value of $\widehat{Z}^1$. We then solve a second problem with one additional constraint:

$$\sum_{j \in F^1} x_j \le p - 1 \tag{3.93}$$

This added constraint prevents the solution F^1 from being possible when solved as not all sites of F^1can be used in the subsequent solution to the model. Thus, it will be a different solution. Let this new solution have an objective value of $\widehat{Z}^2$ with the set of chosen sites F^2. If $\widehat{Z}^2 = \widehat{Z}^1$, then the new solution set F^2 is an alternate optimal solution. We can then solve another problem with an additional constraint like that of (3.93), except that it will prevent solution F^2 from being used in any subsequent solutions to the problem. This third problem then has two added constraints, one to prevent F^1 from being identified again and one to prevent F^2 from being identified again as a solution. Solving this third model will generates a new solution $\widehat{Z}^3$. We continue this process until we find a solution which has an objective that does not equal $\widehat{Z}^1$. All of the solutions generated except for the last one are alternate optimal solutions for the covering problem of interest.

3.10 Summary and Concluding Comments

In this chapter we have focused on the coordinated delivery of services. In its simplest form, we can think that coverage is needed not just once, but several times, as a fire prone area might need several fire stations within reach. From a biological perspective, it may take several sites that contain a specific species to provide support for a meta-population. To do this requires a simple but powerful modification to the LSCP and MCLP constructs. Planners have also been concerned with making changes to a system, from an old configuration to a new one, where the optimal solution is a combination of existing facilities and new sites.

While not discussed here, the hierarchical models are related to Central Place Theory. This concerns the arrangement of retail centers, from small to large, based upon the assumption of a successively inclusive hierarchical system of establishments. The ordering is dependent on the type of good. In such a system, low ordered goods (e.g. gasoline and groceries) are offered at low ordered centers as well as at all other centers, and higher ordered goods (e.g. Porsche car dealer) are offered at regional retail centers. Storbeck (1988, 1990) has suggested a covering model that replicates theoretical Central Place locations when applied to a triangular lattice. Altogether, there are many ways in which covering models have been extended so that they better capture specific nuances of possible applications. We have attempted to present the character as well as some of the details associated with this rich literature. Unfortunately we cannot provide a complete listing or review without excessively extending this book.

References

Akbari A, Pelot R, Eiselt HA (2017) A modular capacitated multi-objective model for locating maritime search and rescue vessels. Ann Oper Res 267:1–26. https://doi.org/10.1007/s10479-017-2593-1

Arthur JL, Hachey M, Sahr K, Huso M, Kiester AR (1997) Finding all optimal solutions to the reserve site selection problem: formulation and computational analysis. Environ Ecol Stat 4 (2):153–165

Bennett VL, Eaton DJ, Church RL (1982) Selecting sites for rural health workers. Soc Sci Med 16:63–72

Branas CC, MacKenzie EJ, ReVelle CS (2000) A trauma resource allocation model for ambulances and hospitals. Health Serv Res 35(2):489

Branas CC, Revelle CS (2001) An iterative switching heuristic to locate hospitals and helicopters. Socio Econ Plan Sci 35(1):11–30

Brill ED Jr (1979) The use of optimization models in public-sector planning. Manag Sci 25 (5):413–422

Brotcorne L, Laporte G, Semet F (2003) Ambulance location and relocation models. Eur J Oper Res 147(3):451–463

Camm JD, Polasky S, Solow A, Csuti B (1996) A note on optimal algorithms for reserve site selection. Biol Conserv 78:353–355

Chaudhry SS, Moon ID, McCormick ST (1987) Conditional covering: greedy heuristics and computational results. Comput Oper Res 14(1):11–18

Cocking C, Cevirgen E, Helling S, Oswald M, Corcodel N, Rammelsberg P, Hassel AJ (2009) Colour compatibility between teeth and dental shade guides in Quinquagenarians and Septuagenarians. J Oral Rehabil 36(11):848–855

Cohon JL (1978) Multiobjective programming and planning. Academic Press, New York

Church RL (1974) Synthesis of a class of public facility location models. PhD Dissertation, The Johns Hopkins University, Baltimore, MD

Church RL, Eaton DJ (1987) Hierarchical location analysis utilizing covering objectives. In: Ghosh A, Rushton G (eds) Spatial analysis and location-allocation models. Van Nostrand Reinhold, New York

Church RL, Gerrard RA (2003) The multi-level location set covering problem. Geogr Anal 35:278–289

Church R, Gerrard R, Hollander A, Stoms D (2000) Understanding the tradeoffs between site quality and species presence in reserve site selection. For Sci 46:157–167

Church RL, Stoms D, Davis F (1996) Reserve design as a maximal covering location problem. Biol Conserv 76:105–112

Current JR, Schilling DA (1994) The median tour and maximal covering tour problems: formulations and heuristics. Eur J Oper Res 73(1):114–126

Daskin MS, Stern EH (1981) A hierarchical objective set covering model for emergency medical service vehicle deployment. Transp Sci 15:137–152

Dökmeci VF (1973) An optimization model for a hierarchical spatial system. J Reg Sci 13 (3):439–451

Dwyer FR, Evans JR (1981) A branch and bound algorithm for the list selection problem in direct mail advertising. Manag Sci 27(6):658–667

Engel AD (1968) Perspectives in health planning. Athlone Press, London

Gerrard RA, Church RL (1994) A generalized approach to modeling the hierarchical maximal covering location problem with referral. Pap Reg Sci 73:425–453

Grubesic TH, Matisziw TC, Murray AT (2012) Assessing geographic coverage of the essential air service program. Socio Econ Plan Sci 46(2):124–135

Grubesic TH, Murray AT (2002) Constructing the divide: spatial disparities in broadband access. Pap Reg Sci 81(2):197–221

Hall NG, Hochbaum DS (1992) The multicovering problem. Eur J Oper Res 62(3):323–339

Hogan K, ReVelle C (1986a) Concepts and applications of backup coverage. Manag Sci 32:1434–1444

Hogan K, ReVelle C (1986b) Backup coverage concepts in the location of emergency services. Proc Pittsburgh Conf Model Simul 17:1423–1429

Kolesar P, Walker WE (1974) An algorithm for the dynamic relocation of fire companies. Oper Res 22(2):249–274

Margules CR (1986) Conservation evaluation in practice. In: Usher MB (ed) Wildlife conservation evaluation. Chapman and Hall, London, pp 297–314

Margules CR, Nicholls AO, Pressey RL (1988) Selecting networks of reserves to maximise biological diversity. Biol Conserv 43(1):63–76

Matisziw TC, Murray AT, Kim C (2006) Strategic route extension in transit networks. Eur J Oper Res 171(2):661–673

Moon ID, Chaudhry SS (1984) An analysis of network location problems with distance constraints. Manag Sci 30:290–307
Moore GC, ReVelle C (1982) The hierarchical service location problem. Manag Sci 28:775–780
Murray AT (2013) Optimising the spatial location of urban fire stations. Fire Saf J 62:64–71
Osleeb JP, Mclaferty S (1992) A weighted covering model to aid in Dracunculiasis eradication. Pap Reg Sci 71:243–257
Penalba JR (1980) Incorporating planning preferences into location models. Master's project report, University of Tennessee, Knoxville, TN
Plane DR, Hendrick TE (1977) Mathematical programming and the location of fire companies for the Denver Fire Department. Operat Res 25:563–578
Pressey RL, Humphries CJ, Margules CR, Vane-Wright RI, Williams PH (1993) Beyond opportunism: key principles for systematic reserve selection. Trends Ecol Evol 8(4):124–128
Ratick SJ, Osleeb JP, Hozumi D (2009) Application and extension of the Moore and ReVelle hierarchical maximal covering model. Socio Econ Plan Sci 43:92–101
ReVelle C, Toregas C, Falkson L (1996) Applications of the location set covering problem. Geogr Anal 8:65–76
ReVelle C, Schweitzer J, Snyder S (1996) The maximal conditional covering problem. INFOR 34:77–91
Schilling D, Elzinga DJ, Cohon J, Church R, ReVelle C (1979) The Team/Fleet models for simultaneous facility and equipment placement. Transp Sci 13:163–175
Schilling DA, ReVelle C, Cohon J, Elzinga DJ (1980) Some models for fire protection location decisions. Eur J Oper Res 5:1–7
Solanki R (1991) Generating the noninferior set in mixed integer biobjective linear programs: an application to a location problem. Comput Oper Res 18(1):1–15
Steuer RE (1986) Multiple criteria optimization. Theory, computation and applications. Wiley, New York
Steuer RE, Choo EU (1983) An interactive weighted Tchebycheff procedure for multiple objective programming. Math Program 26(3):326–344
Storbeck JE (1982) Slack, natural slack, and location covering. Socio Econ Plan Sci 16:99–105
Storbeck JE (1988) The spatial structuring of central places. Geogr Anal 20(2):93–110
Storbeck JE (1990) Classical central places as protected thresholds. Geogr Anal 22(1):4–21
Toregas C (1970) A covering formulation for the location of public facilities. M.S. Thesis, Cornell University, Ithaca, NY
Underhill LG (1994) Optimal and suboptimal reserve selection algorithms. Biol Conserv 70:85–87
Wu C, Murray AT (2005) Optimizing public transit quality and system access: the multiple-route, maximal covering/shortest-path problem. Environ Plan B Plan Des 32(2):163–178

Chapter 4
Probabilistic Coverage

4.1 Introduction

Much of the book is focused on facilities of various types, represented as points, nodes, lines, arcs, paths, tours, areas, etc., providing a wide range of services. The underlying assumption has generally been that facilities, or personnel at/from the facility, are available to serve when needed. This is not too surprising because a sizable number of location covering models are firmly rooted in the initial work Toregas et al. (1971) on the location set covering problem (LSCP), where they were specifically interested in public sector issues involving equity of access to service. In the LSCP facilities were viewed as available for service when needed. In particular, the application of the LSCP to site emergency services, like fire, ambulance and police response, helped to design and relocate such services so that they provide coverage to all. Little has changed, in fact, over the intervening years as emergency service contexts remain of great interest and coverage models have time and again been instrumental in helping to both understand existing service systems as well as develop management plans for emergency response while promoting fairness and equity in service access. Of course, there are many other areas of application for coverage models as well, but the emergency response context has continued to be both challenging and interesting as we better understand such systems and have better supporting data. The focus of this chapter involves the fact that facilities (or personnel) may not always be available when needed. That is, there is a non-zero probability that facility service coverage may not be provided even when every demand is within a desired maximal service standard of a facility. There are clearly many ways in which a facility would be unavailable for service. One situation is that personnel are already busy serving another demand. This is depicted in Fig. 4.1, where the fire engine has traveled from the fire station in response to a fire. However, while busy fighting this fire, another incident (vehicle crash and fire) has occurred across town. It is therefore not possible for a fire crew to respond immediately.

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_4

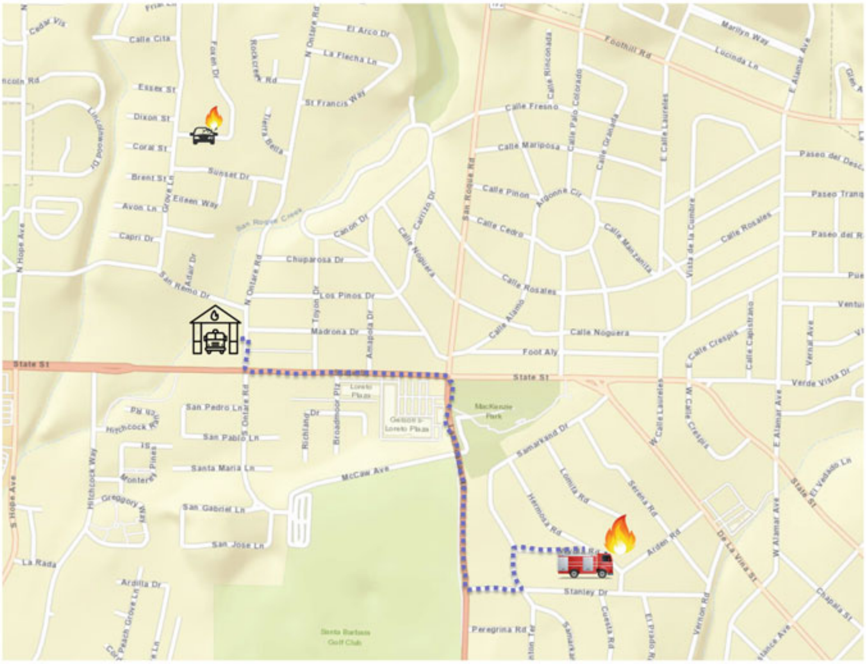


Fig. 4.1 Depiction of a facility busy serving demand

Another situation is that a facility may be unavailable due to a failure of some sort, such as equipment being broken, a power outage, a flood, or even an accident.

If one considers emergency services such as fire, ambulance/paramedic or police, as an example, then what is desired is speedy response to a call for service. As noted previously in this book, such response should be within a maximum service standard, S, in order to ensure safety, survivability, etc. As a result, the LSCP and MCLP detailed in Chap. 2 have been important modeling approaches in this area. However, there is an assumption that each facility is always available to provide service coverage. Such an assumption seems reasonable when considering fire stations and associated equipment, since they are usually idle and awaiting a call. Therefore, the probability of having two simultaneous nearby service calls is very small. Nevertheless, the time to serve calls can be substantial. Not taking this into account might result in the inability of a system of facilities to suitably respond to demand within the desired standard because nearby personnel are occupied with a previous service request. Accordingly, it might be critically important to take into account the probability or likelihood of facility unavailability, particularly given that the associated impacts on system service coverage may be significant.

Similar implications arise in the case of facility failure. Consider the case of cellular phone service. For a cell phone to work and be useful, it must be within the communication standard, S, of a service tower. However, proximity alone is not sufficient. The equipment at the service tower must be operable. While this equipment seems fairly reliable, it does fail, as power may be interrupted, infrastructure deteriorates, etc. As a result, system coverage is likely to be impacted at some point in time.

Several methods have been developed to take into account facility availability in deriving effective service coverage planning. These include the Probabilistic Location Set Covering Problem, the Maximal Expected Coverage Location Problem, the Maximum Availability Location Problem, and others. All of these models represent an attempt to address the availability of facilities in providing coverage. Simply put, response or service within the standard S may not always be possible, even though it is an underlying fundamental goal. Thus, from a probabilistic perspective one might expect that service of demand would most often be accomplished within the desired maximum service standard. Taking this a step further, we can establish a threshold under which "most often" can be stipulated. For example, we may want demand to be served within the S standard 85% of the time. Required standards along these lines are often codified in laws, regulations and guidelines. One example is that of the National Fire Protection Association in recommending that suppression resources be capable of arrival within 4 min travel time to 90% of the incidents (see Murray 2015). Many cities, counties and other local agencies have standards along these lines as well. Further, such standards are not limited to fire response. The California State Emergency Medical Service Authority, for example, promotes related standards for a range of EMS response categories, including Basic Life Support/CPR, defibrillation, Advanced Life Support, ambulance transport, etc. Irrespective of context, the stipulated threshold becomes another operational parameter. The literature often refers to this using the symbol α. In the above example, α would equal 85%.

4.2 Reliable Coverage

The first approach to take into account facility availability in coverage modeling was the Probabilistic Location Set Covering Problem (PLSCP) developed by Chapman and White (1974). The PLSCP essentially extends the LSCP by defining suitable coverage in terms of temporal availability as well as geographical proximity. That is, whether a demand location receives effective service is not solely dependent upon a site being within the service standard S of a facility, but also that this facility is available to serve. A critical question then is what is the likelihood of a facility being busy or otherwise unavailable? Suppose that we know this probability from historical records, quality guarantees associated with equipment, etc. If the likelihood of a facility being unavailable is q, then we can characterize the probability that at least one facility is within S and is not busy serving another demand. Further, ensuring that at least one facility is available for service to a given demand most of the time will likely require that that demand will have to be suitably covered by multiple facilities.

The notion that effective service should be at least some minimum threshold, α, in a probabilistic sense has been termed "α-reliable coverage". The reason for this is because the probability of effective coverage (facility availability within the service standard) must be at least α, where α is a fraction between zero and one. The PLSCP specifically seeks the minimum number and locations of facilities such that every demand receives α-reliable coverage.

Consider the following notation:

i = index representing demand for service
j = index representing potential facility sites
q = probability of a facility being unavailable or busy
d_{ij} = shortest distance or travel time from demand i to potential facility site j
S = the desired service standard (e.g. distance or travel time)
α = pre-specified level of reliability in service coverage

$$N_i = \{j | d_{ij} \leq S\}$$

$$x_j = \begin{cases} 1, & \text{if a facility is located at site } j \\ 0, & \text{otherwise} \end{cases}$$

The PLSCP is based upon the assumption that all facilities are independent and that a facility being unavailable is independent of the availability of any other facility in the system. As before, suppose that we are able to derive or that we know from experience the probability that any given facility is unavailable, q. With this probability, we can estimate for any given demand location the probability that a suitable or timely response within the standard S is possible, based upon the number of facilities capable of serving this demand within the standard. For example, if there is only one facility that can cover a given demand i then that demand would be served with a probability of $1 - q$. Suppose that there are three sited facilities, each within the service standard S for a specific demand i. The probability that all three facilities are unavailable (or busy) is $q \times q \times q$, or q^3. This would imply that the probability of coverage would therefore be $1 - q^3$. If this probability is greater than the establish reliability threshold, α, then we could consider demand i sufficiently served, taking into account the possibility that one or more facilities may be unavailable for service. It is worth noting that q may be viewed as a system-wide average unavailability measure, applying equally to each facility or potential facility location.

For any configuration of facilities, $\sum_{j \in N_i} x_j$ represents the number of facilities that have been located within the coverage standard of demand i. One can then derive the probability that demand i will receive service coverage when needed as:

$$1 - q^{\sum_{j \in N_i} x_j} \tag{4.1}$$

Therefore, equation (4.1) would reflect the probability of suitable coverage within the service standard *S*. The stipulation then is that this should be greater than or equal to α:

$$1 - q^{\sum_{j \in N_i} x_j} \geq \alpha_i \tag{4.2}$$

Note that α is defined here for each demand i, using α_i. This was the original mathematical specification in Chapman and White (1974). In application, however, they simply utilized α. This may have been for convenience, or perhaps due to the fact that justification for varying levels of reliability among demand may be problematic in public sector siting contexts.

Given that the probability of each facility being unavailable is the same, q, equation (4.2) can be rewritten in a linear form. Through algebraic manipulation we have:

$$1 - \alpha_i \geq q^{\sum_{j \in N_i} x_j} \tag{4.3}$$

Taking the log of each side of the inequality in equation (4.3) yields:

$$\log(1 - \alpha_i) \geq \log q^{\sum_{j \in N_i} x_j} \tag{4.4}$$

This then simplifies to:

$$\log(1 - \alpha_i) \geq \sum_{j \in N_i} x_j \log q \tag{4.5}$$

Finally, recognizing that log q is less than zero because q is assumed to be less than one, then simplification reverses the inequality in equation (4.5) to give:

$$\frac{\log(1 - \alpha_i)}{\log q} \leq \sum_{j \in N_i} x_j \tag{4.6}$$

Therefore, equation (4.6) denotes that the number of facilities that serve a demand area i (right hand side of inequality) must be greater than or equal to the quantity on the left hand side of the inequality in order to satisfy the stipulated level of reliability, α_i, given the probability of each facility being unavailable, q. Empirical evaluation may be useful in better clarifying what is happening here. Consider a reliability standard of $\alpha_i = 0.85$ and a probability of unavailability of $q = 0.25$. The interpretation of this is that demand i is expected to see response for service within the standard *S* at least 85% of the time. However, facilities may be busy some 25% of the time, serving other calls for service. Equation (4.6) in this case gives a left hand side value of 1.368. Thus, two or more facilities would clearly be necessary to provide reliable coverage of 85% or better.

Chapman and White (1974) structured the following model, incorporating this condition:

$$\text{PLSCP}: \quad \textit{Minimize} \quad \sum_j x_j \tag{4.7}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq \left\lceil \frac{\log(1-\alpha_i)}{\log q} \right\rceil \quad \forall i \tag{4.8}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.9}$$

The objective, (4.7), seeks the smallest number of facilities. This is precisely the same objective as the LSCP. Constraints (4.8) specify that suitable coverage must be provided to each demand at the indicated level α_i, where the operator function $\lceil \; \rceil$ represents the smallest integer greater than or equal to the enclosed value. Constraints (4.9) impose binary integer requirements on the decision variables.

The main constraint of the PLSCP is to ensure that the probability of an available facility being within the S standard for each demand i is greater than or equal to α_i. The right hand side quantity is nothing other than a coefficient, derived as a function of α_i (or α) and q, inputs that are known in advance. Assuming that each facility has the same unavailability probability (or busyness level) yields a problem where each demand location needs to be covered at least some number of times. Specifically, one can view constraints (4.8) as follows:

$$\sum_{j \in N_i} x_j \geq \gamma_i \quad \forall i \tag{4.10}$$

where $\gamma_i = \left\lceil \frac{\log(1-\alpha_i)}{\log q} \right\rceil$. That is, γ_i is the minimum number of facilities necessary to ensure coverage at the reliability level of α_i. Constraints (4.10) represent processed right hand side values of constraints (4.8), something that would be necessary to compute prior to solving a problem instance. One should recognize that the value of γ_i, strictly speaking, does not need to be the smallest integer value that exceeds the right hand side of constraint (4.8). Since the left hand side of equations (4.8) and (4.10) are sums of decision variable values, all required to be integer, the values of the left hand sides of conditions (4.8) and (4.10) will be integer in value. Thus, there is no additional requirement being imposed when we define each γ_i as being an integer-valued parameter. Additionally, there may be some benefit to integer valued model coefficients when applying solution techniques like branch and bound.

Chapman and White (1974) also discussed a version of the PLSCP where the unavailability of each facility varies. They did not, however, specify any technique to be used for the derivation of q or for the unavailability of specific facilities. Formalizing this is possible, but requires some modified notation:

q_i = probability of a facility being unavailable or busy in demand area i
γ'_i = minimum number of facilities necessary to ensure coverage at α_i given local busyness q_i

Based on our previous derivation, the minimum number of necessary facilities for each demand area is then calculated as:

$$\gamma'_i = \left\lceil \frac{\log(1-\alpha_i)}{\log q_i} \right\rceil \tag{4.11}$$

Constraints (4.8) therefore become the following when a local busyness probability is utilized:

$$\sum_{j \in N_i} x_j \geq \gamma'_i \quad \forall i \tag{4.12}$$

Ball and Lin (1993) suggest this basic constraint form as well based on assumptions of a Poisson distribution for facility unavailability. At any rate, constraints (4.12) suggest two forms of the PLSCP. One is reflected in (4.7)–(4.9) where a system-wide facility unavailability probability is assumed. The other version we can refer to as PLSCP′ and consists of objective (4.7) and constraints (4.9) and (4.12). PLSCP′ incorporates a local probability of a facility being unavailable. Irrespective of the model formulation, important issues are input parameters, namely α, α_i, q and/or q_i.

Upon close examination, one should recognize the equivalence of the PLSCP to the multi-level covering problem discussed in Toregas (1970) and Church and Gerrard (2003), and detailed in Chap. 3. The multi-level covering problem can be structured for both the LSCP and the MCLP, where there is a desire to cover demands with more than one facility. For the LSCP, the multi-level covering problem entails minimizing the number of facilities needed to cover each demand a desired number of times. What is important here is that there is a probabilistic derivation for the values of γ'_i and γ_i. Further, the resulting model is an equivalent deterministic model. Although, this type of problem is often solved by the use of integer linear programming software, it can be solved partially or even completely by the method of reductions [see Toregas and ReVelle (1973) for single level and Church and Gerrard (2003) for multi-level reduction algorithms].

4.3 Expected Coverage

Based on the discussion and models detailed in previous chapters, a natural extension of the PLSCP is to address the basic premise of the MCLP, where there are limited resources in siting facilities thereby preventing coverage of all demand within the desired service standard. Daskin (1982, 1983) incorporated the probability of a facility being unavailable, q, in creating a stochastic form of the MCLP. This model has been referred to as the Maximal Expected Covering Problem (MEXCLP), taking into account the probabilistic nature of facility availability.

Differing from the PLCSP, however, Daskin (1982, 1983) counted any level of coverage as providing some benefit, rather than requiring a minimum coverage reliability threshold. This was done by viewing availability as a binomial probability function. Consider the following additional notation:

k = index corresponding to the number of facilities sites
a_i = demand for service at location i
p = number of facilities to be sited

$$y_{ik} = \begin{cases} 1, \text{if demand } i \text{ is covered by at least } k \text{ facilities} \\ 0, \text{otherwise} \end{cases}$$

What is unique is that the y_{ik} variables are introduced to track the number of facilities capable of covering an individual demand i. Assume as before that each facility is unavailable some portion of the time, q. Thus, a demand site that is covered exactly once will be served $1 - q$ fraction of the time. We can represent this level of coverage by multiplying this fraction of coverage by the population at that site as the term $a_i(1 - q)y_{i1}$, where y_{i1} equals one in value if that demand is covered once. Technically, this is also equivalent to $a_i(1 - q)q^{k-1}y_{i1}$ as the term q^{k-1} equals 1 when $k = 1$. When two facilities cover demand i, then the probability of coverage is $1 - q^2$. The added fraction of coverage in increasing coverage from one facility to two facilities is the difference between $1 - q^2$ and $1 - q$, which is $(1 - q^2) - (1 - q) = q - q^2$. Thus, the added coverage is equivalent to $(1 - q)q^1$ or $(1 - q)q^{k-1}$ when $k = 2$. In general, the added probability of coverage for k facilities is $(1 - q)q^{k-1}$. Thus, the total increase in coverage for demand i when adding a kth facility that covers demand i is $a_i(1 - q)q^{k-1}$. Therefore, we can account for added levels of expected coverage. Based upon this, we can formulate the Maximal Expected Coverage Location Problem (MEXCLP) as follows:

$$\text{MEXCLP}: \quad Maximize \sum_i \sum_{k=1}^{p} a_i(1-q)q^{k-1}y_{ik} \tag{4.13}$$

Subject to:

$$\sum_{j\in N_i} x_j - \sum_{k=1}^{p} y_{ik} \geq 0 \quad \forall i \tag{4.14}$$

$$\sum_{j} x_j = p \tag{4.15}$$

$$x_j = \{0,1\} \quad \forall j \tag{4.16}$$

$$y_{ik} = \{0,1\} \quad \forall i,k \tag{4.17}$$

The objective (4.13) seeks to maximize expected coverage by taking into account the probability that each facility may be unavailable to serve a certain portion of the time. Constraints (4.14) account for the number of facilities that are capable of providing coverage to a particular demand. If demand i is covered only once then constraint (4.14) allows the sum of $\sum_{k=1}^{p} y_{ik}$ to be one in value. But since the additions to expected coverage in the objective function decrease as the number of units covering a demand increase, y_{i1} will then equal 1 in value. For this same reason, if $\sum_{k=1}^{p} y_{ik}$ equals two in value, both y_{i1} and y_{i2} will be one in value. In general, $y_{ik-1} \geq y_{ik}$. Consequently, the objective function combined with constraints (4.14) will accurately account for the expected coverage that is provided by the facilities that have been located. Constraint (4.15) limits the number of facilities to be located to exactly p facilities, which is a specific form of a budget constraint. Constraints (4.16) and (4.17) impose binary integer restrictions on the decision variables. Technically speaking, the integer restrictions on the y_{ik} variables are not necessary as long as these variables are constrained to be no larger than 1 in value. When all of the x_j are integer in value, so too will the y_{ik} variables be binary at optimality.

This formulation differs from the original version presented in Daskin (1982, 1983). It is quite possible in some circumstances that multiple facilities could be located/allocated to a site (called co-location). This may be necessary to ensure satisfactory response requirements. For example, one could house multiple ambulances at an EMS facility, or multiple response teams at a fire station. Accordingly, it may be necessary to allow the facility siting variable x_j to take on general, positive integer values in some situations. As the discussion in this chapter has focused on simply a facility, a decision was made to limit variables to binary. Extension along the lines originally presented in Daskin (1982, 1983) is a simple modification of the MEXCLP formulation presented here.

Suggested in the previous section is that the facility unavailability measure q may be known or unknown. If it is unknown, which would generally be expected for most service systems, then some sort of estimation is necessary. Daskin (1982) and

ReVelle and Hogan (1988) suggested the following basic system-wide busyness fraction estimate:

$$q = \bar{t} \sum_i a_i \Big/ T\, p \tag{4.18}$$

where $\bar{t}$ is the average time to complete service for a demand; and, T is the total availability possible (e.g., 24 h in a day) for a facility. With this, or any potential variation, it is possible to calculate the added amount of availability for the k^{th} added facility using the derivation above as:

$$\beta_k = (1 - q)q^{k-1} \tag{4.19}$$

With this term, we can re-write objective (4.13) as follows:

$$Maximize \quad \sum_i \sum_{k=1}^{p} a_i \beta_k y_{ik} \tag{4.20}$$

The added feature is that β_k may be defined in different ways, depending on the probabilistic function deemed most suitable for a particular application context. In this particular case, Daskin (1982, 1983) assumed a binomial function. Of course, any changes in attribute coefficients in this model (or others) may impact results, changing the configuration of facility sites and associated coverage in an optimal solution.

While objective (4.20), and objective (4.13), specifies one form of criteria to be optimized, other objectives may be possible. Perhaps there is a preference to move away from a system wide measure of unavailability. If so, then one might view that there could be a local area specific measure of busyness. ReVelle and Hogan (1988) suggested a local busyness fraction, q_{ik}, defined as a function of the number of facilities that could serve a particular demand. This local busyness measure is:

$$q_{ik} = \bar{t} \sum_{i' \in N_i} a_{i'} \Big/ T\, k \tag{4.21}$$

With this local measure, facility availability becomes more location specific, and can be summarized as follows:

$$\beta_{ik} = (1 - q_{ik})q_{ik}^{k-1} \tag{4.22}$$

This gives rise to an alternative objective, replacing objective (4.20) with:

$$\text{MEXCLP}': \quad \textit{Maximize} \; \sum_i \sum_{k=1}^{p} a_i \beta_{ik} y_{ik} \tag{4.23}$$

This objective combined with constraints (4.14)–(4.17) defines MEXCLP′, specifically accounting for and incorporating a local busyness measure. Such a model variant was, in fact, detailed in Sorensen and Church (2010).

While it is possible to present a formulation of the MEXCLP in a number of ways, there are two important distinctions to keep in mind. First, the unavailability or busyness of a facility may be viewed (and modeled) from a system-wide perspective, assuming that the probability of a facility being busy and unavailable for service is the same across a region, or it may be conceived that facility service loads are heterogeneous due to variations in demand across a region. Accordingly, there are two different objective functions of the MEXCLP that have been adopted, one generalized as (4.20) representing system-wide busyness and the other as (4.23) that takes into account a local busyness estimate, MEXCLP′. Second, facility siting may permit only one service entity at a location, or it may allow the co-location of facilities. Again, the formulation of the MEXCLP presented here assumes that co-location is not permitted. However, modification of this formulation is possible to allow for such a situation.

4.4 Maximal Reliable Coverage

The PLSCP and MEXCLP have been important formulations in the progression of developing capabilities that address issues of facilities being busy and unavailable when attempting to provide coverage. The PLSCP represents an attempt to ensure that a sufficient number of facilities provide coverage to each demand, doing so such that the probability of having a facility available when needed is at least α (or α_i if more specific detail is needed). The MEXCLP attempts to maximize the total amount of demand that is likely to receive service coverage without any delay. While a distinction can be made that the PLSCP extends the LSCP and the MEXCLP extends the MCLP, there are fundamental differences in the basic approaches taken. Accordingly, there is potential for a more direct extension of the PLSCP that is perhaps more consistent with the MCLP, but differs from the MEXCLP. Specifically, the MEXCLP does not require coverage of each demand be at least α (or α_i) as is the case with the PLSCP. Rather, the MEXCLP seeks to maximize the expected coverage provided.

To incorporate such a modeling feature ReVelle and Hogan (1989) introduced the Maximal Availability Location Problem (MALP). This model is characterized by an intent to maximize demand that is provided α-reliable coverage. Accomplishing this requires the use of γ_i, introduced in the discussion of the PLSCP. Recall that γ_i was defined as the minimum number of facilities that is necessary for demand i to have coverage reliability of at least α. Rather than require that all demand must be covered with this level of reliability, we can seek to maximize the demand that is provided α reliable coverage while locating p facilities. This of course could be considered to be an extension to the original maximal covering location problem. We can formulate the model of ReVelle and Hogan (1989) as follows:

$$\text{MALP}: \quad \textit{Maximize} \ \sum_i a_i y_{i\gamma_i} \tag{4.24}$$

Subject to:

$$\sum_{j \in N_i} x_j - \sum_{k=1}^{\gamma_i} y_{ik} \geq 0 \quad \forall i \tag{4.25}$$

$$\sum_j x_j = p \tag{4.26}$$

$$y_{ik-1} - y_{ik} \geq 0 \quad \forall i, k \tag{4.27}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.28}$$

$$y_{ik} = \{0, 1\} \quad \forall i, k \tag{4.29}$$

The objective, (4.24), maximizes the amount of demand that is reliably covered. Constraints (4.25) operate in a similar fashion as constraints (4.14) and help to define the level of coverage that is provided to a given demand i. What is different here is that constraints (4.14) in conjunction with the objective function (4.13) ensure that in the MEXCLP the variable y_{ik} was always less than or equal to the value of y_{ik-1}, without expressly forcing such a condition. Since the objective (4.24) counts only those demands that have received α-reliable coverage (i.e., when $y_{i\gamma_i} = 1$), without expressly forcing $y_{ik-1} \geq y_{ik}$, constraints (4.25) would allow $y_{i\gamma_i} = 1$ when there is only one facility that has been located that covers demand i. Therefore, constraints (4.27) are included to enforce this condition. That is, coverage at a higher level k cannot be counted unless it is provided at the preceding lower level, $k - 1$. Also note that coverage is counted up to and including the level γ_i with the use of y_{ik} variables and where the value of k is limited to be less than or equal to γ_i for each i. Constraint (4.26) limits the number of facilities sited and constraints (4.28) and (4.29) impose binary integer restrictions on the decision variables. Worth noting is the fact that constraints (4.27) can be added to the PLSCP model as well without any loss of generality. It is possible that these constraints may help to tighten the formulation and reduce solution times.

An important distinction with MALP, similar to MEXCLP, is that it can accommodate either system-wide or local busyness measures. Thus, there is a variant of

MALP using a system-wide busyness but also a version with a local busyness measure, which we can refer to as MALP′.[1]

One thing we have attempted to do in this book is highlight coverage model formulation nuances. In doing this, there have been many instances were alternative mathematical specifications are possible. Associated with MALP, the use of system-wide or local busyness is but one aspect of specification differences. It turns out that an alternative formulation is also possible, and there may be benefits to rely on such an alternative.

Consider the following additional notation:

$$y_i = \begin{cases} 1, \text{if demand } i \text{ covered at or above the reliability standard} \\ 0, \text{otherwise} \end{cases}$$

Given this variable, an alternative MALP formulation is the following:

$$\textit{Maximize} \sum_i a_i y_i \tag{4.30}$$

Subject to:

$$\sum_{j \in N_i} x_j - \gamma_i y_i \geq 0 \quad \forall i \tag{4.31}$$

$$\sum_j x_j = p \tag{4.32}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.33}$$

$$y_i = \{0, 1\} \quad \forall i \tag{4.34}$$

The objective, (4.30), maximizes the total amount of demand that is covered with α reliable coverage. Constraints (4.31) specify that the number of facilities that cover demand i must equal or exceed the value of γ_i in order for α reliable coverage to be counted for demand i. Constraint (4.32) limits the number of facilities to equal p. Constraints (4.33) and (4.34) impose binary integer restrictions on the decision variables.

Marianov and ReVelle (1996) as well as Sorensen and Church (2010) relied on this alternative formulation of MALP. There are a number of differences from (4.24) to (4.29). One of the benefits of this alternative is that it requires fewer variables and fewer constraints. This may be appealing for a number of reasons. It is more concise, and depending on the solver may actually lead to faster solution times. However, doing so likely reduces the “integer friendly” properties characteristic of the original

[1]The literature has often called these variants MALP I and MALP II, as introduced in ReVelle and Hogan (1989). Such a distinction is avoided in this chapter because the only difference is a change in model coefficients, with most of the variables and constraint structure remaining the same.

MALP formulation. More traditional approaches for solving integer programs, such as branch and bound without cuts as well as other modern advances for improving performance, may still encounter substantially more computational difficulty.

4.5 Queuing

An obvious concern in addressing a facility being unavailable or busy is a meaningful and accurate characterization of the associated probability of busyness. The models presented in this chapter up to this point are based on simple definitions of q, q_i and q_{ik}. These probabilities are either assumed to be known or are estimated in some fashion. Of course, many have questioned whether assumed or estimated probabilities are accurate, and the very existence of alternative measures and models suggests that enhancements have and continue to be necessary. Murray and Church (1992) examined system-wide and local busyness estimates, comparing measures relied upon in MALP to simulated busyness estimates associated with identified service system configurations. Their findings suggested that reliability estimates are not particularly accurate, and can negatively impact optimal facility configuration identification. Ultimately, the implications are that system performance may not conform to steady state expectations, which means that the reliability conditions within the model may not be achieved in actual application. As a result, these models may inadvertently identify a suboptimal solution, thinking it performs better.

One way that probability estimates have been improved is through the use of queuing models. For example, Batta et al. (1989) examined the MEXCLP and proposed an extension for adjusting expected coverage based on viewing service as a queuing system. Their intent was to address facility independence, varying busyness probabilities, and location influence on busyness probabilities. Goldberg et al. (1990), too, were interested in expected coverage, and relied on queuing and simulation to derive and update probabilities of facility response and service. Improved estimates of q_i values based on a queuing model were also used to enhance the derivation of the γ_i coefficients, the minimum number of needed facilities capable of covering demand i at or above the level α_i, in Marianov and ReVelle (1994) for improving the PLSCP and in Marianov and ReVelle (1996) for improving MALP.

The basic idea is that viewing the facility response process as a queuing system enables performance characteristics in steady state to be derived. In particular, assume that a system with k facilities observes a service arrival that is Poisson distributed at rate λ. Further, assume that service by a facility is exponentially distributed at rate μ. The rate diagram for this associated system is shown in Fig. 4.2 for this associated *M/M/k* queuing system. Such a system can be viewed as an *M/M/k* queuing system (k servers serving a system where arrival rates follow a Poisson process and where service times follow an exponential distribution). It is also often assumed that service calls are lost (no queue) when no facilities are available. For example, when a call for an EMS occurs and all ambulances are busy, then a rescue vehicle or some non-emergency patient transport company is

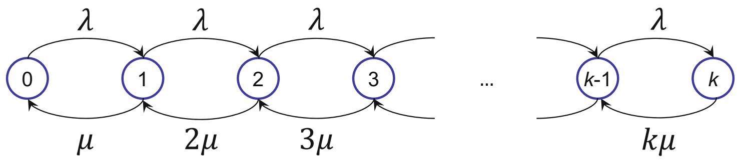


Fig. 4.2 Rate diagram for an assumed *M*/*M*/*k*/*k* queuing system

called to assist. Marianov and ReVelle (1996) used this queuing model to estimate within each local neighborhood of a demand how many servers would be necessary to satisfy the reliability constraint for that neighborhood. To do this, one needs to assume that the queue of one neighborhood is independent of the servers of other neighborhoods or vice versa. Unfortunately, this assumption has not been fully tested although questions have been raised as to the validity of this assumption (see for example Baron et al. 2009). This is essentially the same assumption made by ReVelle and Hogan (1989).

One of the underlying issues in locating facilities that can be congested is to ensure that each facility is not over committed in providing service. Marianov and Serra (1998) suggested a queuing based-form of the maximal covering problem, which explicitly addressed the capacity of individual servers. To formulate their approach we need to define additional notation:

λ_i = service (arrival) rate for demand area i
μ_j = service completion rate for facility j
c_j = coefficient adjusting for service capacity of facility j given reliability level α

$$z_{ij} = \begin{cases} 1, \text{if demand } i \text{ is covered and served by facility } j \\ 0, \text{otherwise} \end{cases}$$

With this notation, Marianov and Serra (1998) formulated a model that accounts for associated facility capacities. Doing this requires the use of allocation variables, z_{ij}, that tracks which demand is served by which facilities in order to ensure that, under steady state conditions, service to local areas of demand can be accomplished within desired α standards. This discrete model (maximal covering location problem with queuing, MCLP-Q) is as follows:

$$\text{MCLP-Q}: \quad Maximize \sum_i \sum_{j \in N_i} a_i z_{ij} \tag{4.35}$$

Subject to:

$$\sum_{j \in N_i} z_{ij} \leq 1 \quad \forall i \tag{4.36}$$

$$\sum_{j} x_j = p \tag{4.37}$$

$$z_{ij} \leq x_j \quad \forall i, j \in N_i \tag{4.38}$$

$$\sum_{i \in N'_j} \lambda_i z_{ij} \leq c_j \mu_j x_j \quad \forall j \tag{4.39}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.40}$$

$$z_{ij} = \{0, 1\} \quad \forall i, j \in N_i \tag{4.41}$$

The objective, (4.35), seeks to maximize the demand suitably covered. Constraints (4.36) indicate that each demand is allocated to at most one facility. Constraint (4.37) specifies the number of facilities to be sited. Constraints (4.38) restrict allocations to only those sites that are selected for a facility. Constraints (4.39) limit the arrival rates that are allocated from demands to a given site to be less than the capacity of the server at that site, based upon the service completion rate. The individual facility service capacities are based on a given reliability standard assuming queuing based steady state probabilities. Constraints (4.40) and (4.41) impose binary integer restrictions on decision variables.

While the use of queuing is an improvement in some ways, especially when considering the capacity of the servers as in MCLP-Q, many of the developed models rely on system-wide or local busyness measures. Murray and Church (1992) highlight the ways that this is problematic. Beyond this, there are a number of limitations associated with queuing approaches for supporting coverage modeling.

Often the demand/arrivals for an area i are based upon the assumption that service is provided by the set of facilities covering a demand, $\sum_{j \in N_i} x_j$. Unfortunately this set of facilities may also serve other demands outside the neighborhood N_i. Thus, an $M/M/\sum_{j \in N_i} x_i$ queuing system can overestimate service potential in a given demand area i because on average not all of the sited facilities will be available for potential service to demand i.

Further, questions have been raised about the degree to which a local estimate of busyness meets the assumptions under which the actual busyness estimate was made. For example, if you have a system that can cover all demand with one location and multiple units can be located at that site, then estimating the probability of busyness can be accurately accomplished either probabilistically, through the use of simulation, or by the use of a queuing model that meets the conditions of the system. But, when a system cannot be covered by one central location, then facilities will need to be scattered in order to cover all demands, which is more problematic. However, local busyness estimates can be somewhat accurate, if a local area does not share its servers with other demands outside the local coverage area or when units outside a local coverage neighborhood help to serve demands with that local neighborhood. But, the assumption that calls handled from outside servers to a local neighborhood N_i is balanced by servers within the neighborhood N_i handling calls that occur outside the neighborhood is likely to be violated. In fact, it is easy to overcommit the capacity of servers when service across the boundary of a neighborhood N_i occurs. Consequently, one of the foundational assumptions of the simple local busyness estimates is likely to be violated when call volumes are high and vary over a region.

4.6 Facility Availability

Daskin (1983) suggested that facilities could be unavailable for reasons other than service loads. For example, a facility may have reliability issues, needs regular maintenance (which requires downtime), or be subject to disruption due to infrastructure failure or flooding. Each of these examples involve aspects of a system that are unrelated to the demands that are being served, but where there may be an easily calculated probability of being unavailable for service. If all of the facilities have an equal probability of being unavailable (e.g., q) and each event of a facility losing service capabilities is independent of other service disruption events, then the models like MEXCLP and PLSCP can be used. The key element in the formulation of the MALP relies on the assumption that all facilities in the neighborhood of a demand have an average busyness given this number (i.e., q_{ik}), rather than specific facility availability estimates.

In general, busyness or unavailability of a facility to provide service may vary and not be the same (Lei et al. 2014), or that a facility's capability to provide service degrades as one moves further from that facility (Altınel et al. 2008). Both factors are issues in designing a wireless sensor network. Let's suppose that we are designing a wireless sensor network. There are two main choices in this design: (1) where to place each sensor, and (2) which type of sensor do we use at each chosen location. Let's say that this network is designed to detect an intruder or threat. Each sensor needs to be within communication of at least another sensor and altogether, the network of sensors must be able to communicate with each other (thus, the design of communication between sensors must represent a connected graph). For the purpose here, we will assume that the area in question does not hinder communication between neighboring sensors, given that enough sensors will be needed and are close enough together in order to adequately surveil an area. Furthermore, sensors are not perfect and there is a probability that they cannot detect an intruder. Also, the capability of a given sensor detecting an intruder declines with distance. Consider the following notation:

t = an index used to refer to sensor type, where $t = 1, 2, \ldots, T$
p_{ijt} = the probability of sensing point i with sensor type t located at point j
M_i = the maximum acceptable probability of missing an intruder at point i
c_{jt} = the cost of placing sensor type t at location j

$$x_{jt} = \begin{cases} 1, & \text{if sensor type } t \text{ is placed at location } j \\ 0, & \text{otherwise} \end{cases}$$

The probability of detection is commonly calculated as an exponential decay function, $e^{-\beta_t d}$, where the decay rate, β_t, is associated with the sensor type and d is the distance between the target point and the sensor. Using this function, we can calculate:

$$p_{ijt} = e^{-\beta_t d_{ij}} \tag{4.42}$$

This means that the probability of missing an intruder at point i can be calculated as:

$$\prod_{j \in J} \prod_{t=1}^{T} \left(1 - p_{ijt} x_{jt}\right) \tag{4.43}$$

With this function, we can now formulate the Effective Covering Problem (ECP) of Altınel et al. (2008) for multi-type sensor placement:

$$\text{ECP}: \quad \text{Minimize} \sum_{j \in J} \sum_{t=1}^{T} c_{jt} x_{ijt} \tag{4.44}$$

Subject to:

$$\sum_{t=1}^{T} x_{jt} \leq 1 \quad \forall j \tag{4.45}$$

$$\prod_{j \in J} \prod_{t=1}^{T} \left(1 - p_{ijt} x_{jt}\right) \leq M_i \quad \forall i \tag{4.46}$$

$$x_{jt} = \{0, 1\} \quad \forall j, \forall t \tag{4.47}$$

The objective, (4.44), involves minimizing the cost of the sensors that are used in the network. Note that the costs of sensors varies by type. Constraint (4.45) limits the location of only one sensor type at a given site j. Constraints (4.46) ensure that probability of missing an intruder at a given point i is less than the maximum acceptable probability of missing an intruder at point i. Constraints (4.47) restrict the decision variables to be binary in value. Unfortunately, this formulation is highly non-linear due to constraints (4.46), and not easily solvable by existing software. To address this formulation directly, one would need to design a heuristic.

There is, however, another approach to solving this problem and it involves a transformation. If we take the natural log of expression (4.46) we obtain:

$$\sum_{j}\sum_{t=1}^{T}\ln\left(1-p_{ijt}x_{jt}\right)\le\ln\left(M_i\right) \tag{4.48}$$

which is mathematically equivalent to constraint (4.46). Note that when $x_{jt} = 1$, the term associated with this variable in the summation on the left hand side of the inequality equals $\ln(1 - p_{ijt})$. Thus, when $x_{jt} = 1$ the term is equivalent to $\ln(1 - p_{ijt})x_{jt}$. Also note that when $x_{jt} = 0$, the associated term in the left hand side of the inequality equals $\ln(1 - p_{ijt} \times 0) = \ln(1)$. Since the $\ln(1)$ equals zero, this term is equivalent to $\ln(1 - p_{ijt})x_{ijt}$ when $x_{jt} = 0$. Thus, we can reformulate constraint (4.46) by taking the natural log of each side and using the properties when x_{jt} is binary in value to form the following equivalent constraint:

$$\sum_{j}\sum_{t=1}^{T}\ln\left(1-p_{ijt}\right)x_{jt}\le\ln\left(M_i\right) \tag{4.49}$$

The beauty of this constraint is that it is linear. Note that the probabilities p_{ijt} and M_i are less than one in value. The natural logs of these values will be negative. If we define $a_{ijt} = -\ln(1 - p_{ijt})$ and $b_i = -\ln(M_i)$, we can write constraints (4.49) in the following form:

$$\sum_{j}\sum_{t=1}^{T}a_{ijt}x_{jt}\ge b_i \tag{4.50}$$

Altogether, we can now formulate an equivalent form of the ECP as follows:

$$\text{Minimize}\sum_{j\in J}\sum_{t=1}^{T}c_{jt}x_{jt} \tag{4.51}$$

Subject to:

$$\sum_{t=1}^{T}x_{jt}\le 1\quad\forall j \tag{4.52}$$

$$\sum_{j}\sum_{t=1}^{T}a_{ijt}x_{jt}\ge b_i\forall i \tag{4.53}$$

$$x_{jt}=\{0,1\}\quad\forall j,\forall t \tag{4.54}$$

This model is now in the form of a classical multi-level covering problem, except that the contributions to covering a given area i are real values rather than integer values and the demand for coverage, b_i, is a real value rather than integer. Altogether, this is a general form of multi-level location set covering problem presented in Chap. 2. It can also be viewed as a generalization of PLSCP presented in Sect. 4.2.

We can also formulate a generalized form of MALP′, where facility availability varies geographically rather than being the same across a region. To do this, first recognize that we will consider only one type of facility being located, but where the probability of being available is site specific, p_j, or site and distance sensitive, p_{ij} (like in the sensor case). The site and distance sensitive form is appealing because not only can we handle a site specific availability of a facility, but we can also include the probability of reaching a given demand within a standard amount of time. Thus, we can define $a_{ij} = -\ln(1 - p_{ij})$ or $a_{ij} = -\ln(1 - p_j)$. With this, we can formulate a general form MALP′ (MALP′-G) as:

$$\text{MALP}'\text{-G}: \quad \text{Maximize} \sum_i a_i y_i \tag{4.55}$$

Subject to:

$$\sum_{j \in J} x_j = p \quad \forall j \tag{4.56}$$

$$\sum_{j \in N_i} \sum_{t=1}^{T} a_{ij} x_j \geq b_i y_i \forall i \tag{4.57}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.58}$$

$$y_i = \{0, 1\} \quad \forall i \tag{4.59}$$

where vehicle or facility availability varies geographically and where reaching a given demand from a specific site may also be probabilistic. The objective, (4.55), involves maximizing the demand that is provided α-reliable coverage. Constraint (4.56) represents the constraint on the number of facilities that are being located. Constraints (4.57) define whether α-reliable coverage has been provided to a specific demand i or not. Note that the value of $b_i = -\ln(M_i)$ where $M_i = 1 - \alpha_i$.

The problem becomes a bit more complicated if we want to develop a model to maximize expected coverage of demand, like that of Daskin (1983), when probabilities of facility unavailability vary geographically. MALP′-G is straight forward when we have a standard for whether a demand is covered or not, like an α-reliable covering standard, but if we wish to maximize expected coverage then determining and accounting for the amount of expected coverage to a given demand is considerably more difficult. The reason for this is that with the Daskin (1983) construct we could easily calculate the level of expected coverage for a demand based upon the number of facilities that covered that demand. But, if the probabilities of coverage vary in possibly covering a given demand, then the actual value of expected coverage varies on which sites are selected as well and not just on the number of selected site. This was the subject of recent work by Lei et al. (2014). To calculate expected coverage of a given demand, they first identified all of the different covering configurations of sites within the coverage area of that demand. For each covering configuration for that demand, they then determined the expected coverage that would be provided by that configuration given site independent availabilities. Their overall model then was designed to maximize expected coverage, where each demand is represented by the coverage provided by the best coverage configuration of that demand that exists among the selected facility locations. It is a relatively ingenious approach, but fraught with a significant degree of complexity, especially when there may exist many alternate covering configurations among those sites that can cover a given demand.

Rather than present the details of the maximal expected coverage model of Lei et al. (2014), we will present a far simpler, but approximate form of MEXCLP given varying probabilities of facility availability (site specific, p_j, or site and distance sensitive, p_{ij}). To do this, consider the following additional/modified notation:

α_k = expected coverage level k where $k = 1, 2, \ldots, K$, for example $\alpha_1 = 0.50$ or 50%.

$$y_{ik} = \begin{cases} 1, \text{if demand } i \text{ is covered at an expected coverage level } k \\ 0, \text{if otherwise} \end{cases}$$

The basic idea is that we discretize expected coverage levels into a small number, say 0.50, 0.60, 0.70, 0.80 and 0.90, or in this case $K = 5$. We can use the basic structure of constraint (4.53) to determine if a given level of coverage, say 50%, has been provided to a given demand. Using this discretized definition of expected coverage, we can define an approximate or discretized-level form of MEXCLP using site specific probabilities of availability:

$$\text{MEXCLP-GD}: \quad \text{Maximize} \sum_{j \in J} \sum_{k=1}^{K} \alpha_k a_i y_{ik} \tag{4.60}$$

Subject to:

$$\sum_{j \in J} x_j = p \quad \forall j \tag{4.61}$$

$$\sum_{k=1}^{K} y_{ik} \leq 1 \quad \forall i \tag{4.62}$$

$$\sum_{j \in N_i} \sum_{t=1}^{T} a_{ij} x_j \geq b_k y_{ik} \forall i \forall k \tag{4.63}$$

$$x_j = \{0, 1\} \quad \forall j \tag{4.64}$$

$$y_i = \{0, 1\} \quad \forall i \tag{4.65}$$

where $b_k = - \ln(1 - \alpha_k)$. The objective (4.60) maximizes the sum of expected coverages based upon what probabilistic level of coverage is attained (e.g., 60%) for each demand. Constraint (4.61) limits the number of facilities being located to equal p. Constraints (4.62) limit the levels of accounted coverage for each demand to be at most one. Constraints (4.63) are used to define whether a specific level k of expected coverage has been met for demand i. If so, then the constraint allows y_{ik} to equal one, denoting that expected coverage level k has been provided to demand i. Because only one level of coverage can be counted for a given demand, only the highest achieved level of expected coverage will be accounted for that demand in the objective function. The remainder of the constraints list the binary restrictions on the decision variables.

4.7 Extensions

There are a number of pathways in which the models presented in this chapter have been extended, often with the purpose of addressing issues experienced in applications. From the very beginnings of Location Science and the development of the classic LSCP and MCLP approaches, refinements have been proposed to address specific issues that arise in application. For example, Plane and Hendrick (1977) relocated fire stations, where the objective was to improve coverage as much as possible while keeping as many of the existing stations as possible. Often, there are many issues when taking a system and improving it by making selected changes as compared to designing something from scratch. That is, in application most problems are not "green field" in nature, but start with an existing system, which is either expanded, contracted, or reconfigured. A good example in analyzing an existing fire department and its fleet of rescue vehicles and fire trucks can be found in Pérez et al. (2016). Their motivation was to improve reliable coverage of emergency response taking into account different types of response as well as vehicle busyness in the city of Santiago, Chile. Another interesting application of covering to improve fire response can be found in the work of Aktaş et al. (2013). They analyzed the services of fire department locations in the City of Istanbul, Turkey to address many types of demand, including historic structures built during the Roman and Ottoman empires that have recently been the target of radical extremists.

Time is a major factor which has been virtually ignored in this chapter. The main reason for this is to keep our presentation as simple as possible. Demand for EMS can vary considerably over the hours of a week. Church et al. (2001) analyzed the temporal EMS demand for Kansas City (Missouri and Kansas) and presented a scheduling model that optimizes crew shifts and ambulance deployment over a week, so that periods of peak demand are well served while ensuring that periods of low demand are not overcommitted with resources. Repede and Bernardo (1994) have extended the MCLP to handle multiple time periods and Brotcorne et al. (2003) have developed a dynamic relocation model that repositions ambulances each time a call for service is made. Their model uses what they term the "double standard" where cover is maximized within one distance and required within a range that is more distant. The basic element is to reposition as few units as possible with the least overall distance of repositioning so that the double standard of coverage is met. This creates a system that is best positioned to answer the next call. Another example is that of Rajagopalan et al. (2008) in which they develop a multiperiod set covering model that they use to dynamically redeploy ambulances. Repositioning of emergency vehicles when some are busy has been a subject of interest since the classic work of Kolesar and Walker (1974) and has been found to help a system increase response coverage in real time.

Another type of extension is where facilities are co-located at a given site. Co-location helps to keep the number of stations or dispatching posts as small as possible. The idea is that co-location should be pursued whenever it is possible to accomplish without seriously reducing coverage or increasing response times. Several of the models that are presented in the previous section have been formulated with the option of allowing co-location whenever it is beneficial. Presumably, co-location may reduce the need for repositioning. It should be recognized that in less dense suburban areas, co-location may not be a viable option due to the fact that demand is spread over a larger area and facilities may need to be spread far apart, whereas in the dense urban cores co-location may be achieved without degrading coverage values.

It is also important to note that may emergency services have a goal of maximizing coverage and minimizing average response times. This means that models such as the p-median problem have also been used to deploy emergency resources. A good example of how that can be accomplished can be found in Weaver and Church (1985) where service is accounted for in terms of the closest vehicle as well as more distant vehicles to a demand. Thus, average response is a function of vehicle average availability. Because many coverage models are special cases of p-median equivalents (see Church and Weaver 1986), there exist many avenues in which coverage and average response can be handled together. There is also the possibility of using the probability chain approach developed by O'Hanley et al. (2013) to handle differing probabilities of availability in optimizing both coverage and average response distance.

4.8 Summary and Concluding Comments

The models described in this chapter represent a rather large literature devoted to emergency services and safety planning, including fire, EMS, and police as well as sensor network design. Many of these models feature covering as the primary objective. It is only natural that such models have been extended and tested in many ways. For example Borras and Pastor (2002) have analyzed several forms of probabilistic covering models. They found that the model of Ball and Lin (1993) almost always achieves the required reliability level. Their tests involved the use of a simulation model. They used their analysis to suggest a further extension of the Ball and Lin (1993) approach to achieve the same desired levels of reliability while using fewer vehicles. Sorensen and Church (2010) also used simulation to confirm whether local reliability is achieve within the context of a MALP′ approach. They also demonstrated that, in general, using local reliability estimates in an expected coverage model achieves higher levels of coverage than using local reliability estimates in MALP′. They suggested that an expanded MEXCLP approach better meets the operations standards of many ambulance systems. Altogether, simulation models like that of Erkut and Polat (1992) and Heller et al. (1989) are important tools in validating the results of optimization models in emergency management, but are useful in their own right.

The constructs that are reviewed in this chapter all involve simplifications of a highly complex system of congestion. The main reason for this is that the probabilistic nature of a complex design system is difficult to fully capture in a simpler deterministic model. However, such deterministic equivalents represent the only realistic approach to finding what are called "optimal" solutions. Unfortunately, few if any comparisons have been made between a heuristic solution process to a more realistic version of an actual system and the optimal solutions obtained from a deterministic equivalent model. It is time to demonstrate which pathway has been the most fruitful to date in addressing this important problem.

References

Aktaş E, Özaydın Ö, Bozkaya B, Ülengin F, Önsel Ş (2013) Optimizing fire station locations for the Istanbul metropolitan municipality. Interfaces 43(3):240–255

Altınel İK, Aras N, Güney E, Ersoy C (2008) Binary integer programming formulation and heuristics for differentiated coverage in heterogeneous sensor networks. Comput Netw 52 (12):2419–2431

Ball MO, Lin FL (1993) A reliability model applied to emergency service vehicle location. Oper Res 41 (1):18–36

Batta R, Dolan JM, Krishnamurthy NN (1989) The maximal expected covering location problem: revisited. Transp Sci 23(4):277–287

Baron O, Berman O, Kim S, Krass D (2009) Ensuring feasibility in location problems with stochastic demands and congestion. IIE Trans 41(5):467–481

Borras F, Pastor JT (2002) The ex-post evaluation of the minimum local reliability level: an enhanced probabilistic location set covering model. Ann Oper Res 111(1):51–74
Brotcorne L, Laporte G, Semet F (2003) Ambulance location and relocation models. Eur J Oper Res 147(3):451–463
Chapman S, White J (1974) Probabilistic formulation of the emergency service facilities location problems, presented at ORSA/TIMS National Meeting, San Juan, Puerto Rico (Reprint series no. 7407)
Church RL, Gerrard RA (2003) The multi-level location set covering model. Geogr Anal 35:277–289
Church RL, Weaver JR (1986) Theoretical links between median and coverage location problems. Ann Oper Res 6(1):1–19
Church R, Sorensen P, Corrigan W (2001) Manpower deployment in emergency services. Fire Technol 37(3):219–234
Daskin M (1982) Application of an expected covering model to emergency medical service system design. Decis Sci 13:416–439
Daskin M (1983) A maximum expected covering location model; formulation, properties, and heuristic solution. Transp Sci 17:48–70
Erkut E, Polat S (1992) A simulation model for an urban fire fighting system. Omega 20(4):535–542
Goldberg J, Dietrich R, Chen JM, Mitwasi MG, Valenzuela T, Criss E (1990) Validating and applying a model for locating emergency medical vehicles in Tuczon, AZ. Eur J Oper Res 49 (3):308–324
Heller M, Cohon JL, Revelle CS (1989) The use of simulation in validating a multiobjective EMS location model. Ann Oper Res 18(1):303–322
Kolesar P, Walker WE (1974) An algorithm for the dynamic relocation of fire companies. Oper Res 22(2):249–274
Lei TL, Tong D, Church RL (2014) Designing robust coverage systems: a maximal covering model with geographically varying failure probabilities. Ann Assoc Am Geogr 104(5):922–938
Marianov V, Revelle C (1994) The queuing probabilistic location set covering problem and some extensions. Socio-Econ Plan Sci 28(3):167–178
Marianov V, ReVelle C (1996) The queueing maximal availability location problem: a model for the siting of emergency vehicles. Eur J Oper Res 93(1):110–120
Marianov V, Serra D (1998) Probabilistic, maximal covering location—allocation models for congested systems. J Reg Sci 38(3):401–424
Murray AT (2015) Fire station siting. In: Eiselt HA, Marianov V (eds) Applications of location analysis. Springer, Berlin, pp 293–306
Murray AT, Church RL (1992) The reliability of α-reliability. Presented at ORSA/TIMS Joint National Meeting, San Francisco, California, USA, November 1–4, 1992
O'Hanley JR, Scaparra MP, García S (2013) Probability chains: a general linearization technique for modeling reliability in facility location and related problems. Eur J Oper Res 230(1):63–75
Pérez J, Maldonado S, Marianov V (2016) A reconfiguration of fire station and fleet locations for the Santiago Fire Department. Int J Prod Res 54(11):3170–3186
Plane DR, Hendrick TE (1977) Mathematical programming and the location of fire companies for the Denver fire department. Oper Res 25(4):563–578
Rajagopalan HK, Saydam C, Xiao J (2008) A multiperiod set covering location model for dynamic redeployment of ambulances. Comput Oper Res 35(3):814–826
Repede JF, Bernardo JJ (1994) Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky. Eur J Oper Res 75(3):567–581
ReVelle C, Hogan K (1988) A reliability-constrained siting model with local estimates of busy fractions. Environ Plan B: Plan Des 15(2):143–152
ReVelle C, Hogan K (1989) The maximum availability location problem. Transp Sci 23(3):192–200
Sorensen P, Church R (2010) Integrating expected coverage and local reliability for emergency medical services location problems. Socio-Econ Plan Sci 44(1):8–18
Toregas C (1970) A covering formulation for the location of public service facilities. Masters Thesis, Cornel University, Ithaca

Toregas C, Revelle C (1973) Binary logic solutions to a class of location problem. Geogr Anal 5 (2):145–155
Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19(6):1363–1373
Weaver JR, Church RL (1985) A median location model with nonclosest facility service. Transp Sci 19(1):58–74

Chapter 5
Anti-cover

5.1 Introduction

The anti-covering location problem (ACLP) is a well-recognized coverage-based dispersion model. Admittedly, reaching this conclusion requires a little work, but in fact this problem is related to the node packing, vertex packing, stable/independent set and r-separation problems, with considerable attention being devoted to each of these related problems (see Padberg 1973; Erkut 1990; Nemhauser and Sigismondi 1992; Murray 1995; Erkut et al. 1996; Murray and Kim 2008; Niblett 2014; Niblett and Church 2015). The name *anti-cover* can be attributed to Moon and Chaudhry (1984) who attempted to distinguish it from other well-known coverage problems. The name, therefore, reflects a sort of opposing goal compared to the set covering problem. The anti-covering location problem seeks to maximize the total weighted benefit of facilities sited in a region, doing so in a manner that ensures at least a minimum pre-specified distance or travel time between facilities and demand is maintained. If the benefit is the same for each potential facility location, then this is equivalent to maximizing the number of facilities that can be sited while maintaining minimum separation restrictions between all facilities and demand or between a sited facility and all other sited facilities. Of course, the goal of the location set covering problem detailed in Chap. 2 is to minimize the number of facilities needed for complete coverage of all demand, assuming the costs for selecting facilities is the same for every potential site. In this sense, then, the two problems have contrasting intents.

A host of practical planning contexts are recognized where the ACLP reflects goals and planning needs, including community impact assessment, service and trade area delineation, safety and security, environmental protection, military defense, forestry, water well design, habitat carrying capacity, etc. Few would argue the importance of siting nuclear facilities away from people and recreational spaces. A retail outlet would certainly not want another franchisee nearby that impinges on their consumer market area, possibly driving them out of business.

© Springer International Publishing AG, part of Springer Nature 2018
R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_5

Positioning two munitions depots next to each other would not be strategically sound as an accident, fire or attack at one would likely ignite the other. Excessive disturbance of a woodlands, forested area or watershed that harms flora and fauna is likely unsustainable. There are many other examples as well, all highlighting the significance of the ACLP across a range of application domains.

The literature associated with anti-covering location can be traced back to at least Berge (1957) who provides an early description of the stable set problem. Edmonds (1962) details the internally stable set or a packing problem, seeking the largest subset of vertices on a graph of vertices where no two vertices are joined by an edge. In this sense, the anti-covering location problem is nothing other than a node/vertex packing problem, or a stable/independent set problem, applied in a highly geographic context. This geographic context is highlighted in Moon and Chaudhry (1984), Murray (1995), Murray and Church (1996), Erkut et al. (1996) and Church and Murray (2009). The significance of the geographic context is that subsequent evaluation is necessary to transform place and distance into a network of nodes and arcs. With this, the node packing problem (or any other naming convention) arises. For convenience these inter-related problems are simply referred to as the anti-covering location problem.

5.2 Separation Context

The general problem of interest in this chapter involves the need to simultaneously site multiple facilities of the same type that provide some sort of service. Such a facility might be a power generation plant, like a nuclear reactor or coal fired operation. In this case the facility is often viewed as noxious, with many people and activities preferring not to be too close by. Moreover, if a facility is nearby, it is clearly undesirable to have other facilities nearby as well since such a possible concentration of facilities increases exposure and risk. Another example of a facility fitting the context here is one involving waste processing, like a dump, recycling center or transfer station. Some view these facilities, too, as being noxious or obnoxious. Alternatively, the facility might be a restaurant or outlet, one that is part of a regional or national chain. Of course, any outlet of the chain would not want their other restaurants too near as it would erode market share. The facility may also be part of military or defense services, such as a missile silo, fuel/munitions depot, etc. Siting these too close together could prove fatal, increasing system vulnerability to an attack or accident. As noted above, there are many, many more types of facilities possible where the common underlying goal is to locate many of these facilities to enhance service provision or access, but recognition that spatial separation is necessary for various reasons. Based on this, the anti-covering location problem can be stated as follows:

Select multiple facilities so as to maximize the total benefit of the facilities sited in a region while ensuring that there is a pre-specified minimum distance or travel time between facilities or between facilities and demand

Note that this problem definition includes two possibilities for enforcing separation: (1) keep each sited facility at least a minimum distance or travel time from all other facilities, or (2) keep each sited facility at least a minimum distance or travel time away from demands. It is also possible to consider the case where both types of separation standards must be met, that is maintain separations between facilities and between facilities and demand. Such a distinction has been highlighted in Francis et al. (1978) and Moon and Chaudhry (1984).

5.2.1 Separation to Avoid Concentration Around Demand

One interpretation of separation is that there is demand for a facility to provide service as long as there is no concentration in any one area (see Grubesic and Murray 2008; Church and Cohon 1976). To this end, one facility in or near a neighborhood is (reluctantly) deemed useful and needed in order to ensure access to the goods and services provided, but more than this would be considered excessive or even dangerous. There are many examples, many of which are NIMBY (not in my backyard) like. Perhaps most prominent are waste processing and recycling stations. Everyone producing waste needs the facility, implicitly and explicitly, but would prefer they be located in another nearby community. Certainly more than one would be considered unacceptable. Similar situations can be observed for rehabilitative and other social services. We need them, but only sparingly due to their localized impacts and/or negative externalities.

To illustrate this particular form of separation, consider Fig. 5.1. A neighborhood area is shown around one location, demand i. Of interest then in this situation is ensuring that no more than one facility be sited in the neighborhood. Imposing this in a model seeking to identify the best locations for facilities therefore requires stipulating conditions that prohibit multiple sited facilities in an area. This is generally done with constraints. The important point is that the constraints would be associated with each demand area, relating facility siting to the impacts on or in neighborhoods.

5.2.2 Separation Between Sited Facilities

In contrast to attention on demand, a second interpretation of separation is to focus solely on the facilities being located (see Zeller et al. 1980; Mealey et al. 1982; Downs et al. 2008; Ratick et al. 2008). Specifically, the intent is to ensure physical separation between sited facilities. This orientation too can be driven by the need to

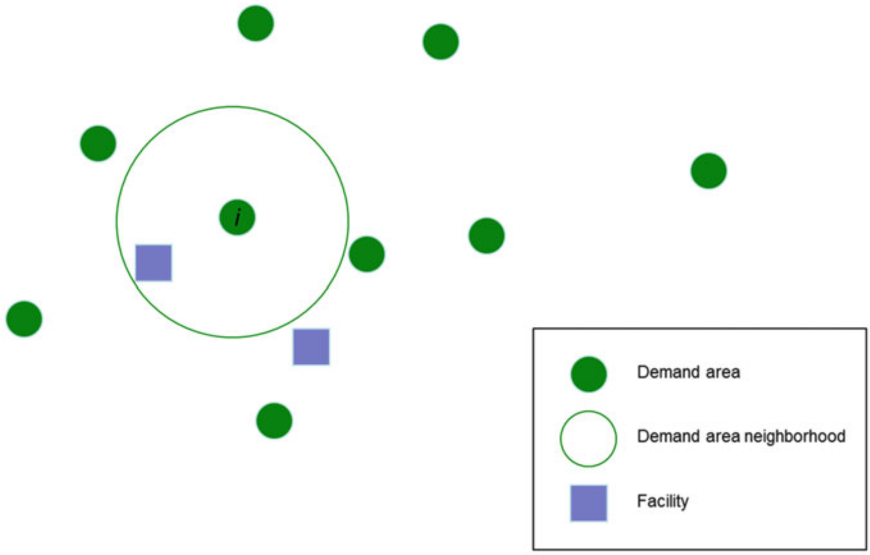


Fig. 5.1 Separation to avoid concentration around demand

minimize local area impacts. For example, environmental impacts associated with forest harvesting should be dispersed, with no two neighboring areas simultaneously scheduled for harvest (Thompson et al. 1973; Mealey et al. 1982; Murray 2007). Alternatively, for safety and security facilities may be prohibited from being too close to each other. Moon and Chaudhry (1984) discuss military installations separated in order to guard against simultaneous enemy/terrorist attack. Ratick et al. (2008) site backup facilities housing critical documents, data, emergency supplies, etc., providing protection in the event of a disaster at the main facility.

To illustrate this second situation, consider Fig. 5.2. An area is depicted around one of the sites, facility *j*, within which no other facility may be sited. In contrast with Fig. 5.1, the area of emphasis is now around the facility, not the demand. What is challenging in this case is that we do not know in advance which locations will be selected for facility placement. This means that restrictions must be structured on a conditional basis. That is, if a site is selected for facility placement, then no other facility may be sited that would be too close. Conditional restrictions are typically structured and imposed using constraints in the model. The important distinction in Fig. 5.2 is that the constraints are associated with potential facility sites, in contrast to constraints associated with demand sites in Fig. 5.1.

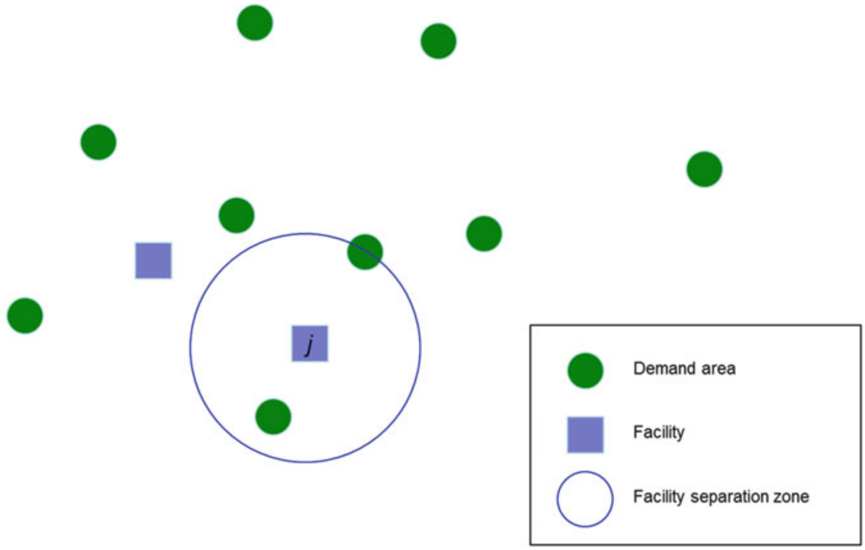


Fig. 5.2 Separation between facilities

5.3 Model Construct

An integer programming based formulation of the anti-covering location problem (as well as vertex packing, the r-separation problem, and other equivalent problems) is possible once spatial processing, analysis and evaluation is carried out. Without loss of generality, we limit the discussion to the case where separation between sited facilities is of concern (Fig. 5.2). Consider the following notation:

j (and k) = index of potential facility sites (entire set J)
α_j = benefit associated with locating a facility or allocating an activity at site j
d_{jk} = distance or travel time between potential facility sites j and k
M = large number
r = minimum required separation
$\Omega_j = \{k | d_{jk} \leq r\}$
$$X_j = \begin{cases} 1 & \text{if facility located at site } j \\ 0 & \text{otherwise} \end{cases}$$

This notation reflects geographic processing that is necessary, typically accomplished using GIS, in order to structure the ACLP. First, one must identify all potential facility sites that would be too close to a facility at site j, and they would be members of the set Ω_j. This is based on the separation requirement imposed, r, but also on the evaluation of proximity between potential facility sites, d_{jk}. Second, one must also derive the benefit associated with siting a facility at a potential site j, α_j.

This might be expected income or return on investment, but also could be safety, security, diversity, etc.

With this information and the above notation, a classic ACLP formulation can be specified (Moon and Chaudhry 1984, Erkut 1990):

$$\text{ACLP1}: \qquad \textit{Maximize} \ \sum_j \alpha_j X_j \tag{5.1}$$

Subject to:

$$MX_j + \sum_{k \in \Omega_j} X_k \leq M \quad \forall j \tag{5.2}$$

$$X_j = \{0, 1\} \quad \forall j \tag{5.3}$$

The objective, (5.1), is to maximize the total weighted benefit of facilities located. Spatial limitations on facility configuration are imposed in constraints (5.2), as locating a facility at site j implies that no nearby or conflicting locations may be selected. The core of this model is the set of restriction constraints (5.2). Note that if a given $X_j = 1$ then the value $MX_j = M$, which means that for all members k of set Ω_j, their associated X_k must equal zero in value. Constraints (5.3) impose binary integer restrictions on decision variables.

One can depict this decision making problem graphically, using a network of nodes and arcs. The nodes represent the potential facility sites and the arcs connect those sites that are deemed too close. Figure 5.3 shows a hypothetical case of 29 land parcels in an area, among which we seek to select as many parcels as possible as long as they do not share an edge. Think of this as a land use problem found in the forest industry where the parcels that we select will be harvested. Standards prevent us from harvesting any two neighboring parcels. Figure 5.4 depicts the network interpretation, where parcels are represented as nodes and arcs are between any two nodes (parcels) that would violate intended spatial separation if simultaneously selected for harvesting. In this case there are 29 nodes and 63 arcs. The intent is to select the greatest total weighted collection of nodes (parcels in this case) without any two selected nodes sharing an arc.

While the above proximity definition of separation is based on sharing an edge, other forms are possible as well, such as a minimum distance of separation. But, a prominent interpretation is adjacency, that two facility sites/units sharing a common edge or point are considered too close and therefore cannot be simultaneously selected like in the forestry example. Using the above model notation, we can introduce the restriction of not being able to share a boundary between two selected parcels as follows. Let $d_{jk} = 0$ if sites j and k are adjacent and 1 if not. If $r = 0$, then proximity based on adjacency would be imposed given this definition. The practical interpretation can be illustrated by considering Fig. 5.5 that depicts nine sites, and the network that would result is shown in Fig. 5.6. In this case, there are 16 arcs that reflect spatial separation restrictions in the interpreted network. Again, the intent is to

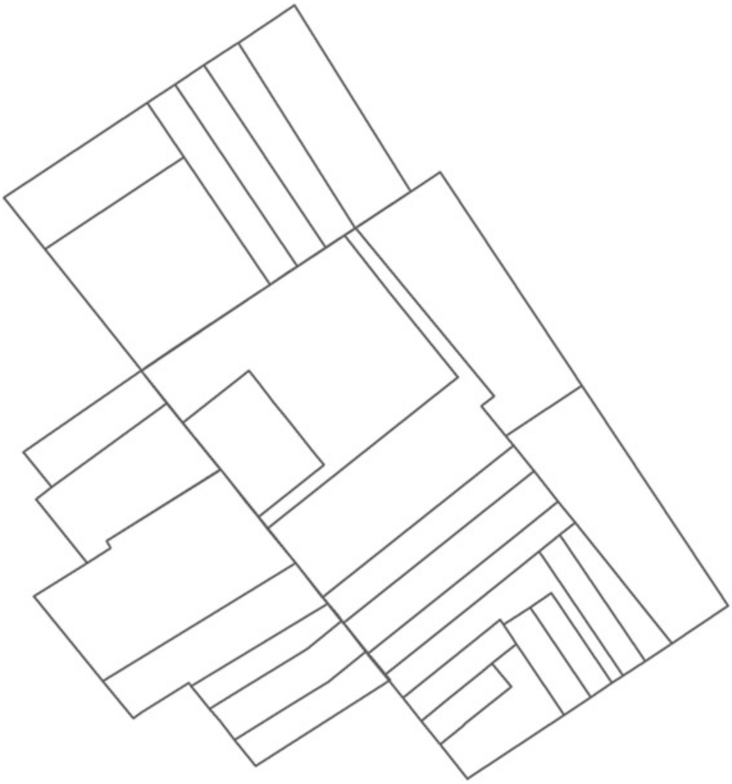


Fig. 5.3 Commercial parcels to select one or more facility sites

select the maximum total weighted collection of nodes without selected nodes sharing a common arc.

5.4 Mathematical Structure

There has been considerable interest in the structure of constraints (5.2) for imposing spatial limitations. Murray and Church (1997) and Church and Murray (2009) refer to (5.2) as a neighborhood adjacency constraint. The reason for this is somewhat intuitive as the set Ω_j represents those potential facilities in the spatial separation neighborhood of potential facility site j. Considerable interest in the structure of constraints (5.2) arises because the ACLP, (5.1–5.3), is recognized as being computationally challenging to solve. Moon and Chaudhry (1984), Nelson and Brodie (1990), Torres-Rojo and Brodie (1990), Yoshimoto and Brodie (1994), Murray (1995), Murray and Church (1995a, b, 1997) and Erkut et al. (1996) all discuss the difficulty of solving even small planning problem applications. This has everything

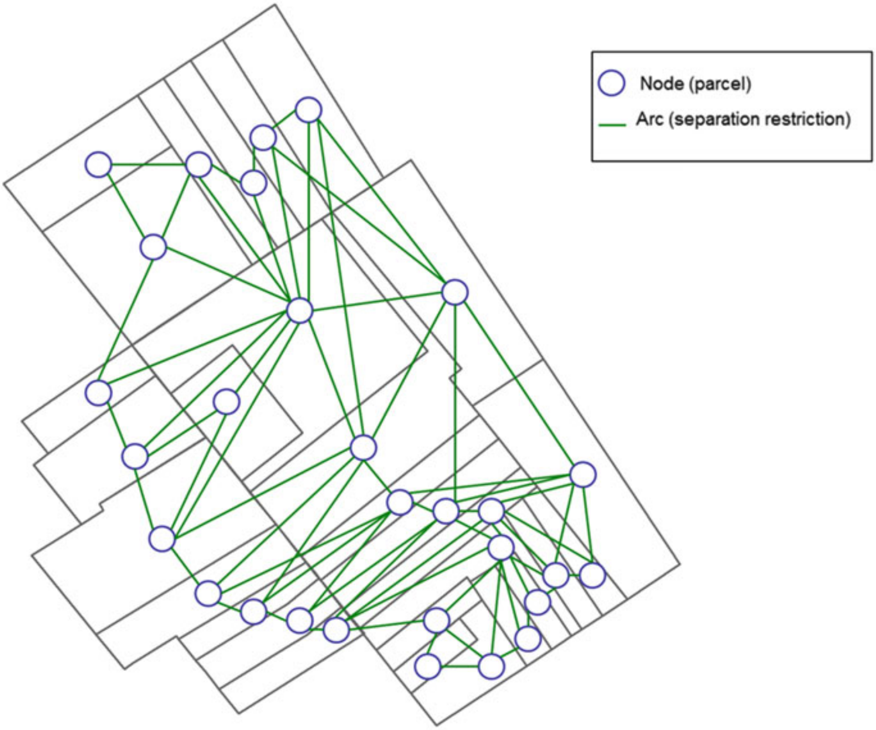


Fig. 5.4 Network representation of commercial parcels and separation restrictions

to do with the inherent mathematical structure associated with constraints (5.2) that impose spatial separation between sited facilities. Simply put, the structure of constraints (5.2) tends to require significant computational effort to solve to optimality, as this structure does not lend itself to natural integer-optimal solutions. The remainder of this section demonstrates that alternative formulations of the ACLP are possible. The significance is that such alternatives have very desirable mathematical properties that enhance their use, especially using commercial integer programming solvers. Thus, not only is it often possible to formulate a particular problem differently, but there may be good reason for considering such alternative models.

5.4.1 Cliques

An important concept in mathematical programming is a facet inducing construct known as a clique (Padberg 1973; Nemhauser and Trotter 1975; Nemhauser and Wolsey 1988). For our purposes here, members of a clique are mutually in conflict with each other. For our problem here, we want to ensure that selected facility sites

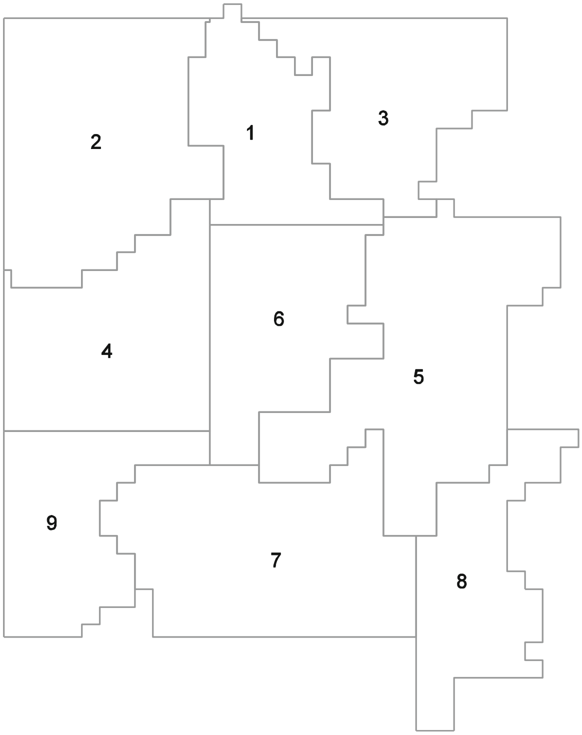


Fig. 5.5 Sites for activity selection

are spatially separated based on an established criteria or standard. Within this context, a clique is a set of two or more members, facility sites, where at most one member may be selected without violating some required facility separation. Relating this to the ACLP formulation above, the implication is that it is possible to identify groups or sets of potential facility sites that share mutual separation requirements. For example, suppose there are three potential facility sites, {1, 2, 3}, and they are all too close to each other, where the selection of one of these sites precludes the selection of any of the other two. More specifically, the decision variable for each potential facility site, X_1, X_2 and X_3, gives rise to the condition that at most one may be selected. Since the decisions are binary, a clique constraint for this set follows: $X_1 + X_2 + X_3 \leq 1$.

Let us suppose then that it is possible to identify all necessary clique constraints to impose all separation restrictions for a particular problem. This could be used to change the mathematical structure in the ACLP. Consider the following additional notation:

l = index of set of cliques (entire set L)
Φ_l = set of potential facility site members of clique l

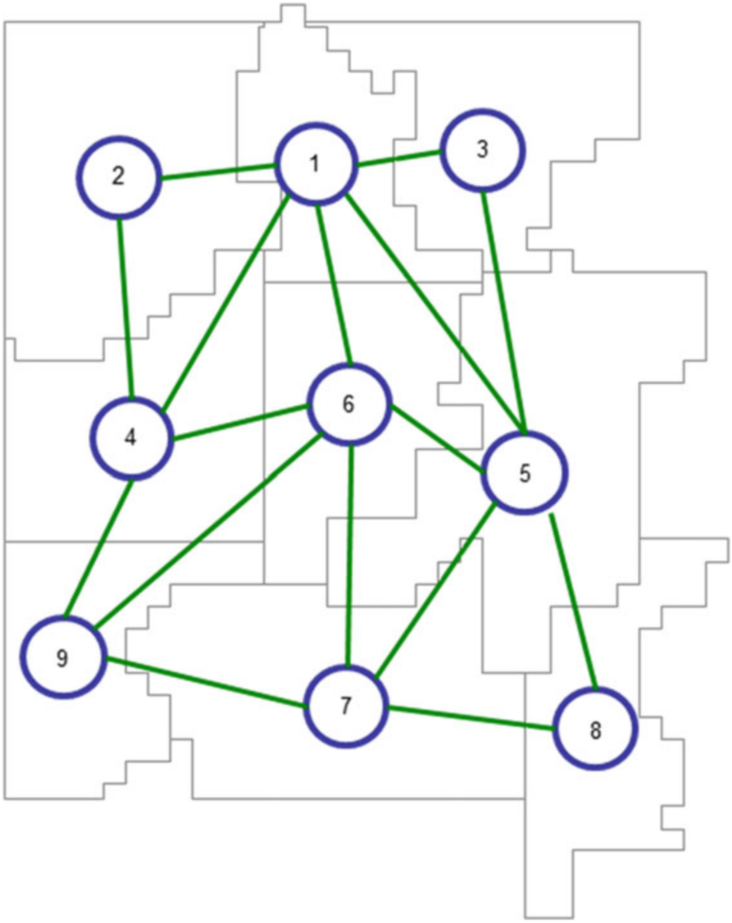


Fig. 5.6 Network resulting from adjacency separation

All that remains is to replace constraints (5.2) in the above formulation of the ACLP with clique constraints. The result is as follows:

$$\text{ACLP2 :} \qquad \textit{Maximize} \ \sum_j \alpha_j X_j \tag{5.4}$$

Subject to:

$$\sum_{k \in \Phi_l} X_k \leq 1 \quad \forall l \tag{5.5}$$

$$X_j = \{0, 1\} \quad \forall j \tag{5.6}$$

The objective, (5.4), remains to maximize the total weighted benefit of facilities located. The spatial separation between sited facilities is imposed in constraints (5.5) through the use of cliques. Constraints (5.6) impose binary integer restrictions on decision variables.

A special case of the clique is a pair of potential facilities that are too close to each other. Consider clique l consisting of potential sites 1 and 2, which means that $\Phi_l = \{1, 2\}$. The associated constraint is $X_1 + X_2 \leq 1$, a subset of the above clique involving potential facility sites 1, 2 and 3.

An entire set of pair based cliques can readily be specified using existing notation:

$$X_j + X_k \leq 1 \quad \forall j, k \in \Omega_j \tag{5.7}$$

This pairwise form is appealing because it is simple, and accounts for all separation conditions originally imposed using constraints (5.2). A potential negative is that the use of this constraint structure, pairs, will result in a relatively large number of constraints, which too has proven to increase computational difficulty when actual planning problems are solved. For superior mathematical structure, it is far more beneficial to have individual cliques with the largest possible number of members (Murray and Church 1997). That is, we would like $|\Phi_l|$ to be a big number, but ultimately the size of cliques is predicated on spatial separation requirements and associated structure that results.

Approaches to identify cliques can be found in Bron and Kerbosch (1973) and more recently Tomita et al. (2006) and Cazals and Karande (2008), among others. For use in the ACLP, the work of Jones et al. (1991), Murray and Church (1997), Goycoolea et al. (2005) and Murray and Kim (2008) offer practical approaches that have proven to be effective. The use of cliques in the ACLP or model extensions has proven to be beneficial in a number of ways, enabling problems to be solved faster as well as larger problem instances to be considered. A potential issue is that identifying a complete set of needed cliques may be computationally expensive. Theoretically, identifying a complete set of cliques involves enumerating all possible cliques. Given this, alternative constraint structures continue to be sought, offering the potential to provide good mathematical structure without excessive computational overhead to identify all cliques.

5.4.2 Hybrid

Recognizing that cliques, (5.5), may result in too many constraints or require too much time to identify, and that the structure of the neighborhood constraint, (5.2), has poor mathematical properties for successful solution using a commercial solver, Murray and Church (1997) explored an alternative constraint structure for imposing spatial separation (see also Erkut et al. 1996). They proposed a hybrid form that combined a maximal clique constraint associated with each potential facility site along with a modified neighborhood constraint structure. Consider the following additional notation:

$\widehat{L}$ = reduced set of cliques $\left(\widehat{L} \subset L\right)$
$\widehat{\Phi}_l$ = set of potential facility site members in clique l
$\widehat{\Omega}_j$ = reduced set of neighbors, $\left\{k \in \Omega_j \middle| k \& j \notin \widehat{\Phi}_l \text{ for any } l \in \widehat{L}\right\}$
$\widehat{\lambda}_j$ = smallest coefficient necessary to impose restriction associated with $\widehat{\Omega}_j$

Effectively, this reflects an attempt to avoid an excessive number and/or excessive search times in clique identification. This approach also adds important mathematical structure to the model formulation and improves the success in solving a problem using standard techniques. Structure is added by adding the best cliques (largest in size) to the original model (5.1)–(5.3). Then, whatever spatial separation conditions are not imposed in the identified sub-set of clique constraints will be imposed using neighborhood like constraints. The tradeoff here is that mathematical structure is being sacrificed somewhat in order to reduce pre-processing effort, while at the same time producing a tighter formulation. The ACLP model formulation with hybrid constraints, a combination of clique and neighborhood, is as follows:

$$\text{ACLP3}: \qquad Maximize \sum_j \alpha_j X_j \tag{5.8}$$

Subject to:

$$\sum_{k \in \widehat{\Phi}_l} X_k \leq 1 \quad \forall l \tag{5.9}$$

$$\widehat{\lambda}_j X_j + \sum_{k \in \widehat{\Omega}_j} X_k \leq \widehat{\lambda}_j \quad \forall j \left(\widehat{\Omega}_j = \varnothing\right) \tag{5.10}$$

$$X_j = \{0, 1\} \quad \forall j \tag{5.11}$$

The objective, (5.8), remains as before, to maximize the total weighted benefit of facilities located. The spatial separation between sited facilities is now imposed using two sets of constraints, (5.9) and (5.10). Constraints (5.9) are based on cliques, but do not likely constitute a complete set of cliques. To supplement constraints (5.9), whatever conditions that are not imposed by the subset of all clique constraints are enforced by the neighborhood constraints (5.10). Finally, constraints (5.11) impose binary integer restrictions on decision variables.

Murray and Church (1997) offered an approach for identifying hybrid constraints in forest planning. In the forestry context, the size of the largest cliques tends to be quite small. Murray and Kim (2008) developed a more general hybrid constraint set identification approach based on the use of GIS. Thus, a complete set of hybrid constraints is computationally efficient to identify, and their associated mathematical structure in the ACLP has proven to be effective for problem solution.

5.4.3 Theoretical Bounds

Up to this point, an argument has been made that constraint structure can impact the solvability of a model when applied to address a particular planning problem/situation. Why does this prove to be true? Taking as an example the ACLP, it is an integer programming (IP) problem due to the integer requirements on decision variables. In fact, it is a special case of an IP as it is has specific integer requirements that values be binary (0 or 1). The ACLP can also be considered a linear program (LP) with added restrictions on decision variables. The objective and major constraints are linear functions. It is of little surprise then that a common approach for solving an IP, and the ACLP, is based on LP. Specifically, LP with branch and bound is often used to solve an IP (Nemhauser and Wolsey 1988). Avoiding too much detail, this approach solves a series of LP relaxations of the IP (integer requirements temporarily ignored). The hope is that all decision variables in the LP turn out to satisfy all integer requirements. If not, which is likely, then the process of branching and bounding proceeds. At each branch linear constraints are added that serve as a cut, attempting to resolve each fractional decision variable in the associated LP relaxation. The process continues, branching off on a fractional decision variable present in a solution to a LP relaxation, until the optimal IP solution can be inferred.

With this in mind, we can now establish the significance of alternative formulations of the ACLP in a more formal manner. Assuming that the objective function of the ACLP is denoted using Z, the total weighted benefit, we can understand the relationships between the linear relaxations of the various ACLP formulations. Empirical evidence based on Moon and Chaudhry (1984), Erkut et al. (1996), Murray and Church (1997) and Murray and Kim (2008) suggests the following:

$$Z^* \leq Z^{LP}(\text{ACLP2}) \leq Z^{LP}(\text{ACLP3}) \leq Z^{LP}(\text{ACLP3}') \leq Z^{LP}(\text{ACLP1}) \qquad (5.12)$$

where Z^* is the optimal IP objective value, $Z^{LP}(\text{ACLP2})$ is the optimal LP solution to the relaxed model using clique constraints, (5.5), $Z^{LP}(\text{ACLP3})$ is the optimal LP solution to the relaxed model using hybrid constraints, (5.9) and (5.10), $Z^{LP}(\text{ACLP3}')$ is the optimal LP solution to the relaxed model using pair based clique constraints, (5.7), and $Z^{LP}(\text{ACLP1})$ is the optimal LP solution to the relaxed model using neighbor based constraints, (5.2).

An optimal solution to a relaxed LP provides a bound on the best possible solution meeting the IP conditions (Nemhauser and Wolsey 1988). In this case, the ACLP is a maximization problem, so the optimal IP solution, Z^*, will be less than or equal to any LP relaxation of the model formulation. Generally speaking, the tighter the LP bound, the more likely an all (or nearly all) integer solution will result. The implication in solving an IP using LP with branch and bound is that less computational effort will be required for solving an IP when the best possible LP relaxation is utilized. Thus, a facet inducing construct is one or more constraints that enhance model solvability by making the feasible LP region closer to the feasible IP region, thereby decreasing (or eliminating) branching and bounding.

Returning to relationships established in (5.12), it should be observed that Z^{LP}(ACLP1) provides the worst LP bound, explaining why in practice it has proven to be more difficult to solve an ACLP with neighborhood based spatial separation constraints. Similarly, the comparatively better Z^{LP}(ACLP2) bound supports why in practice the ACLP with clique based spatial separation constraints has proven very good for successful solution with the least amount of computational effort.

5.5 Relaxations and Extensions

One of the things that has been noted throughout the book, and in this chapter in particular, is that models can be formulated, reformulated, extended and modified in various ways. This is most certainly true of the ACLP, where its underlying mathematical structure is part of larger, more detailed planning model formulations. An example is forest harvest scheduling models where the decisions involve not only where but when as well (Thompson et al. 1973; Kirby et al. 1986; Vielma et al. 2007). This adds a temporal dimension to the ACLP as well as other supporting considerations, like limits on associated outputs in any given time period. Recognition of the mathematical relationship to the ACLP (and node packing) has ultimately resulted in improved methods for solving more detailed forest planning problems. There are many other examples as well, with the general point being that making the connection can be important in many ways, such as a more efficient formulation, a model that is easier to solve, enabling larger planning problems to be considered, etc.

While a detailed and lengthy review will be not be provided here, it is worthwhile to observe aspects of how the ACLP could be modified to support problem variants. The remainder of this section reviews a basic relaxation as well as an extension.

5.5.1 Relaxation

The suggestion raised in Hochbaum and Pathria (1997) is of interest in weighing the relative merits of spatial configuration against the economic returns associated with harvest activity. That is, certain restrictions between neighboring sites might be better to relax, or not imposed, if the overall economic return is significantly enhanced. The question of course is which restrictions and how to assess relative tradeoffs. To address this, Hochbaum and Pathria (1997) formulated what they termed the generalized independent set problem. This can be conceived of as a relaxation of the ACLP, or independent set problem, because it explicitly structures a way to allow sites in conflict to be selected simultaneously but accounting for benefits gained by doing so. Consider the following additional notation:

β_{jk} = penalty for not imposing separation between facilities sited at j and k

$$Y_{jk} = \begin{cases} 1, \text{if restriction between potential facility sites } j\&k \text{ is relaxed} \\ 0, \text{otherwise} \end{cases}$$

A penalty, β_{jk}, is introduced to quantify the significance or impact of relaxing the spatial conflict that exists between two selected facility sites j and k. Further, the variable Y_{jk} is used to track if this restriction has been relaxed or not. Using this notation, the generalized independent set problem (GISP) is formulated as follows:

$$\text{GISP}: \qquad Maximize \;\; Z_1 = \sum_j \alpha_j X_j \tag{5.13}$$

$$Minimize \;\; Z_2 = \sum_j \sum_{k \in \Omega_j} \beta_{jk} Y_{jk} \tag{5.14}$$

Subject to:

$$X_j + X_k - Y_{jk} \leq 1 \quad \forall j, k \in \Omega_j \tag{5.15}$$

$$X_j = \{0, 1\} \quad \forall j, \;\; Y_{jk} = \{0, 1\} \quad \forall j, k \in \Omega_j \tag{5.16}$$

The first objective, (5.13), remains the same as that of the ACLP, to maximize total benefit of selected facility sites. A second objective, (5.14), is introduced to track the total penalty incurred by relaxed spatial restrictions. Constraints (5.15) are similar to the pair based clique given in (5.7), but with an additional variable added to enable relaxing a specific separation restriction. Specifically, if $Y_{jk} = 1$, then both X_j and X_k can equal one and still satisfy the right hand side. That is, both could simultaneously be selected. Alternatively, if $Y_{jk} = 0$, then the original intention of the spatial separation condition is imposed where only one or the other site could be selected. Binary integer restrictions are imposed in constraints (5.16).

A visual interpretation of the GISP is offered in Fig. 5.7. The depicted network arcs correspond to the relationships associated with site 6 in order to simply the graphic. In contrast to the relationships Fig. 5.6, the dashed arcs shown in Fig. 5.7 may or may not be imposed. If not imposed, those dashed arcs would incur a penalty. For example, if the condition between sites 6 and 7 is relaxed, that would mean that $Y_{67} = 1$, with a penalty cost of β_{67} added to objective (5.14).

It should be noted that the original formulation presented in Hochbaum and Pathria (1997) integrated objectives (5.13) and (5.14) as a weighted combination (equivalent to the weighting method in multi-objective optimization, see Cohon 1978), but is presented here in a more general form. Though not recognized as related, the map labeling models detailed in Ribeiro and Lorena (2008), Cravo et al.

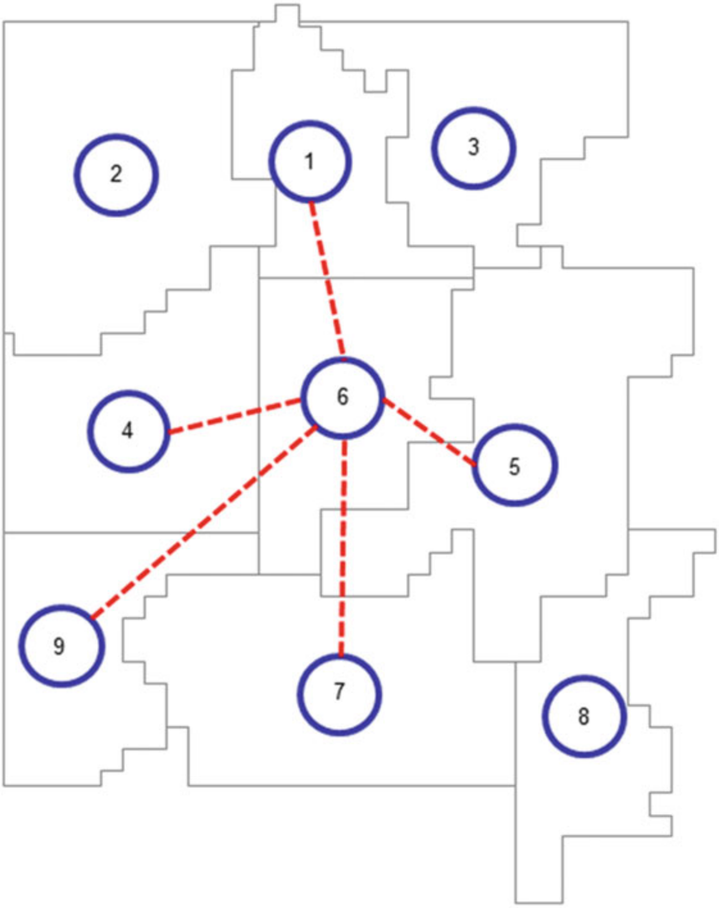


Fig. 5.7 Spatial conditions for site 6 in the generalized independent set problem

(2008) and Mauri et al. (2010) are effectively generalized independent set problems similar to that of Hochbaum and Pathria (1997).

The intent of the GISP is to expressly examine the impacts of restrictions on attainable benefits associated with the selection of facility sites. Clearly more total economic return is possible, larger objective (5.13) value, as more spatial restrictions are relaxed. However, this is at the expense of a greater total penalty, higher objective (5.14) value, associated with relaxing spatial restrictions. Thus the GISP is a multi-objective model with competing objectives, where gains are possible in objective (5.13) only by increasing objective (5.14). Conversely, decreases in objective (5.14) are only possible by reducing objective (5.13). Through the use of multi-objective solution techniques (see Cohon 1978), it is possible to generate and examine the tradeoff solutions associated with a particular planning problem, providing insights on the relative impacts of emphasizing whether spatial restrictions are to be imposed or not.

The significance of the GISP is that it is effectively a relaxation of the ACLP. Such a relaxation is accomplished through the addition of a variable that allows a constraint to not be imposed, but doing so incurs a penalty in the objective.

5.5.2 Extension

In addressing problems requiring the use of the ACLP, Wei and Murray (2012) noted that the positional inaccuracy of site boundaries creates various challenges, both in applying the ACLP and interpreting results. Effectively, there are limitations in using the ACLP as it based on the assumption that spatial information is accurate. To address this issue, Wei and Murray (2012) formulated what amounts to an extension of the ACLP, where boundary uncertainty can be explicitly considered in the model (see also Murray et al. 2014).

After detailed analysis of spatial information, Wei and Murray (2012) concluded that it is possible to characterize some relationships as known and others as uncertain. For example, even taking into account positional uncertainty of boundaries, some sites are clearly beyond intended separation standards. That is, $d_{jk} > r$ for sites j and k, where d_{jk} is the distance between sites j and k and r is the established separation standard (as defined previously). Additionally, some sites remain clearly within the separation standard under conditions of positional uncertainty. That is, $d_{jk} < r$ for sites j and k. However, there is the in between case where some sites may or may not be within the separation standard. In practice, all spatial information is subject to at least some degree of error or uncertainty. To account for these different cases, Wei and Murray (2012) introduced notation along these lines:

ρ_{jk} = probability that sites j and k would violate spatial separation requirement if both selected

Γ_j = set of a sites that would violate separation requirement if selected along with site j

Δ_j = set of sites that might violate separation requirement if selected along with site j

Given the probability measures, ρ_{jk}, it is possible to characterize the relationship between two sites j and k. If $\rho_{jk} = 1$ (or some other threshold value that is interpreted to be equivalent to one), then the two sites are deemed too close, even taking into account positional uncertainty. This means that they should not be allowed to simultaneously be selected as they would clearly be in violation of separation standards. Thus, site k would be in the set Γ_j as they are too close irrespective of spatial uncertainty issues. If $\rho_{jk} = 0$ (or some other threshold value that is interpreted to be equivalent to zero), then no separation constraints are needed as sites j and k have no chance of ever being in violation if both are simultaneously selected. The final case is when $0 < \rho_{jk} < 1$, where there is a chance that the two sites could be too close when positional uncertainty is considered. Of course, there is also a chance that the two sites would not be too close. In this case, site k would be in the set Δ_j, reflecting that possibility that the two sites may or may not be too close. With these sets and notation, the extension proposed in Wei and Murray (2012) referred to as the error ACLP (EACLP) can be formulated as follows:

$$\text{EACLP}: \quad \textit{Maximize}\ Z_1 = \sum_j \alpha_j X_j \tag{5.17}$$

$$\textit{Minimize}\ Z_2 = \sum_j \sum_{k \in \Delta_j} \rho_{jk} Y_{jk} \tag{5.18}$$

Subject to:

$$X_j + X_k \leq 1 \quad \forall j, k \in \Gamma_j \tag{5.19}$$

$$X_j + X_k - Y_{jk} \leq 1 \quad \forall j, k \in \Delta_j \tag{5.20}$$

$$X_j = \{0, 1\} \quad \forall j, \ \ Y_{jk} = \{0, 1\} \quad \forall j, k \in \Delta_j \tag{5.21}$$

The first objective, (5.17), is the same as that of the ACLP, to maximize total benefit of selected facility sites. A second objective, (5.18), is introduced to track the total probability of conditions that are considered uncertain. Constraints (5.19) impose certain spatial separation restrictions using the pair based clique in constraint (5.7). Constraints (5.20) represent the uncertain conditions where it may or may not be necessary to impose separation restrictions. An additional variable is added to enable relaxation of the restriction. Specifically, if $Y_{jk} = 1$, then the condition is not imposed. This means that both X_j and X_k can equal one. Alternatively, if $Y_{jk} = 0$, then the spatial separation condition is imposed. Binary integer restrictions are imposed in constraints (5.21).

To illustrate the uniqueness of the EACLP, consider the network shown in Fig. 5.8 for site 6. This may be contrasted with Figs. 5.6 and 5.7. The solid arcs represent those spatial separation conditions that must be imposed because they are certain, constraints (5.19). Alternatively, the dashed arcs represent those spatial separation conditions that may or may not exist when boundary uncertainty is taken into account, constraints (5.20). Unlike what is possible between sites 6 and 7 in Figs. 5.7 and 5.8 highlights that the EACLP requires separation between sites 6 and 7. However, other relationships may be viewed as uncertain, e.g., sites 6 and 2, sites 6 and 3 and sites 6 and 9, and if not imposed would incur the probability of violation, ρ_{62}, ρ_{63}, and ρ_{69}, respectively, in objective (5.18).

There is clearly a relationship between the ACLP, the GISP, and the EACLP to account for spatial error. Further, the EACLP is similar to the GISP. In fact, the two are equivalent if $\Gamma_j = \varnothing$ for all sites j and spatial separation restrictions are considered uncertain. Worth mentioning as well is that alternative types of extension are possible, such as the site based approach detailed in Wei and Murray (2015).

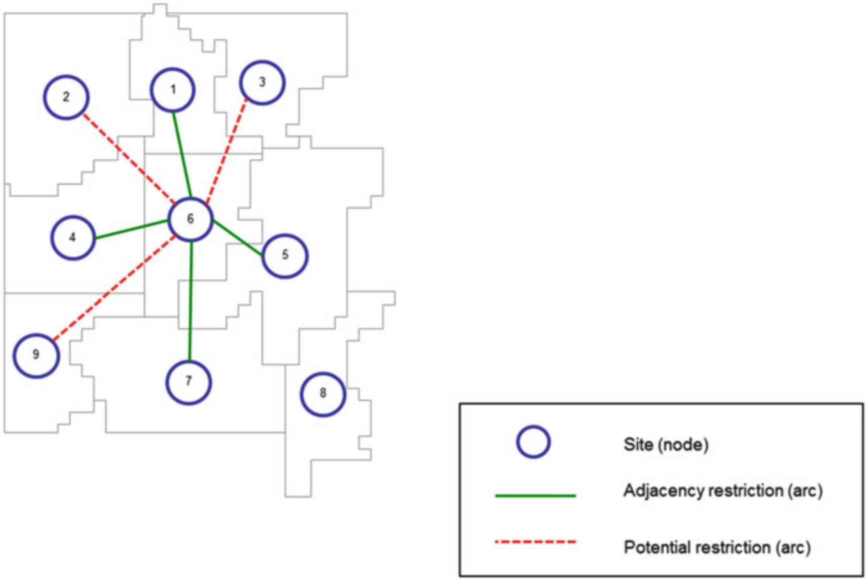


Fig. 5.8 Spatial conditions for site 6 in the uncertainty model of Wei and Murray (2012)

5.6 Inefficiency

The underlying premise of the ACLP is to site facilities in a manner that the greatest total benefit is derived, but in doing so ensure that service/separation ranges do not overlap. That is, the separation requirement between facilities and/or demand should not be violated. When the benefit for each site is the same (e.g., $\alpha_j = \alpha_{j'} \;\; j \neq j'$), this is equivalent to simply maximizing the total number of facilities sited in a region. Under these conditions then, the objective can be structured as:

$$Maximize \;\; Z = \sum_j X_j \tag{5.22}$$

The difference from the more general ACLP is that $\alpha_j = 1$ for each potential facility site in objective (5.22), reflecting the indistinguishability in benefits between potential facility sites. We can refer to the optimal solution to this ACLP special case as $Z^{\max}$. What can we say about a feasible, but sub-optimal solution? Well, the associated objective value will be less, or more precisely $Z < Z^{\max}$. But what if we were interested in knowing or identifying a feasible solution to the ACLP that is the most sub-optimal, yet there is no possibility of adding an additional facility to the region without violating spatial separation requirements. This is the problem examined in Niblett (2014) and Niblett and Church (2015). Identifying the worst case spatial configuration of facilities with respect to spatial restrictions provides important and meaningful context for an ACLP solution. This is particularly true when the use of the ACLP is oriented towards description and understanding of the spatial distribution of a good or service, something discussed in detail in Grubesic and Murray (2008), as an example.

Niblett (2014) and Niblett and Church (2015) refer to the worst case objective orientation as the disruptive ACLP (DACLP), because of the intent to configure facilities in such a manner as to disrupt the maximum potential that would be identified using the ACLP. The formulation is:

$$\text{DACLP}: \qquad \textit{Minimize} \ \sum_j \alpha_j X_j \tag{5.23}$$

Subject to:

$$MX_j + \sum_{k \in \Omega_j} X_k \leq M \qquad \forall j \tag{5.24}$$

$$X_j + \sum_{k \in \Omega_j} X_k \geq 1 \qquad \forall j \tag{5.25}$$

$$X_j = \{0, 1\} \quad \forall j \tag{5.26}$$

The objective, (5.23), now minimizes the total weighted benefit of facilities located. This is in contrast to that of the ACLP, where the objective is maximized. Spatial limitations on facility configurations are imposed in constraints (5.24). Constraints (5.25) require that the resulting solution cannot leave some site unselected, if the configuration allows that site to be selected while maintaining all separation restrictions. That is, the solution must eliminate all possibility of adding an additional facility without violating a separation constraint. Either a site j has a facility or one or more sites within the spatial restriction area of site j must be selected as a facility. Accordingly, as many facilities are sited as needed, ensuring what Niblett and Church (2015) refer to as a proper solution, while minimizing the benefits generated by the selected facilities. Constraints (5.3) impose binary integer restrictions on decision variables.

An optimal solution for the DACLP would produce a minimal objective value for the ACLP. For the unitary α_j values discussed above, then we can think of this objective as $Z^{\min}$ for a particular problem application. Accordingly, the solutions would maintain the relationship that $Z^{\min} \leq Z^{\max}$. Further, the difference would be very telling and informative in evaluating laws or restrictions that can be reflected in the use of the ACLP, or in the assessment of a franchise system as discussed in Chap. 7.

5.7 Facets and More

Throughout the chapter we have indicated that solving the ACLP has been of great interest because it remains a challenging problem to solve. As a result, there have been a wide range of exact and heuristic solution approaches proposed to solve the ACLP, and extensions of the ACLP.

On the exact side, Nemhauser and Trotter (1975) were among the first to report computational experience using cliques (5.5), along with other facet inducing constraints. This was done within a branch and bound process. Moon and Chaudhry (1984) examined solution using linear programming based on formulation ACLP1, (5.1–5.3), where neighborhood constraints are relied upon. Nemhauser and Sigismondi (1992) build upon the work of Nemhauser and Trotter (1975), detailing results based on a cutting plane approach. Erkut et al. (1996) and Murray and Church (1997) explored a number of different formulations, similar to those detailed above, reporting computational experience. A Lagrangian relaxation approach within branch and bound was detailed in Murray and Church (1996). More recently, the use of column generation is discussed in Warrier et al. (2005). Murray and Kim (2008) detailed an approach to identify effective hybrid constraints and examine their performance in solving the ACLP. Finally, a branch and cut scheme is presented in Giandomenico et al. (2013). Clearly there has been continued interest and improvement in solving the ACLP using exact methods that guarantee an optimal solution.

Similar interest and improved capabilities can be seen in heuristic solution development for the ACLP as well. Early work included greedy approaches by Chaudhry et al. (1986), but also more recently by Gamarnik and Goldberg (2010). Lagrangian relaxation was applied in Zoraster (1990). A genetic algorithm was proposed by Chaudhry (2006) and a greedy randomized adaptive search procedure (GRASP) was applied in Cravo et al. (2008). Recent work by Wei and Murray (2017) reports a multi-objective genetic algorithm for solving the EACLP, developed to provide an approximation of the non-inferior tradeoff curve in order to support planning and decision making.

5.8 Summary and Concluding Comments

The anti-covering location problem (ACLP), also referred to and/or related to the *r*-separation, node packing, vertex pack, maximum independent set and stable set problems, is an interesting and important planning problem in location coverage modeling. It has been considered and applied to a number of practical planning situations, including harvest scheduling, military defense, water well drilling, urban services, map labeling and others. Because of the broad application and appeal of this problem, there continues to be interest in the formulation, solution and use of this model as it has proven to be challenging to solve.

It should be of little surprise that the ACLP has been extended in various ways. For example, Zeller et al. (1980) structure a retail franchise design model that seeks to impose territorial exclusivity through the use of spatial separation constraints. Murray and Church (1999) developed a formulation based on integrating the MCLP with separation requirements (such an approach is also detailed in Berman and Huang 2008), discussing monitoring station and water well contexts. Williams (2008) presented a reserve design model with distance separation requirements where the intent is to ensure species survivability. Ratick et al. (2008) report a backup storage facility location model where facilities must be spaced apart for security reasons, accomplished using separation restrictions.

References

Berge C (1957) Two theorems in graph theory. Proc Natl Acad Sci 43(9):842–844

Berman O, Huang R (2008) The minimum weighted covering location problem with distance constraints. Comput Oper Res 35:356–372

Bron C, Kerbosch J (1973) Algorithm 457: finding all cliques of an undirected graph. Commun ACM 16:575–577

Cazals F, Karande C (2008) A note on the problem of reporting maximal cliques. Theor Comput Sci 407:564–568

Chaudhry SS (2006) A genetic algorithm approach to solving the anti-covering location problem. Expert Syst 23:251–257

Chaudhry SS, McCormick ST, Moon ID (1986) Locating independent facilities with maximum weight: greedy heuristics. Omega 14(5):383–389

Church RL, Cohon JL (1976). Multiobjective location analysis of regional energy facility siting problems. Brookhaven National Lab, United States Energy Research and Development Administration. Upton, NY (USA): US Government

Church RL, Murray AT (2009) Business site selection, location analysis and GIS. Wiley, New York

Cohon J (1978) Multi-objective programming and planning. Academic Press, New York

Cravo GL, Ribeiro GM, Nogueira Lorena LA (2008) A greedy randomized adaptive search procedure for the point-feature cartographic label placement. Comput Geosci 34:373–386

Downs JA, Gates RJ, Murray AT (2008) Estimating carrying capacity for sandhill cranes using habitat suitability and spatial optimization models. Ecol Model 214:284–292

Edmonds J (1962) Covers and packings in a family of sets. Bull Am Math Soc 68(5):494–499

Erkut E (1990) The discrete p-dispersion problem. Eur J Oper Res 40:48–60

Erkut E, ReVelle C, Ulkiisal Y (1996) Integer-friendly formulations for the r-separation problem. Eur J Oper Res 92:342–351

Francis RL, Lowe TJ, Ratliff HD (1978) Distance constraints for tree network multifacility location problems. Oper Res 26:570–596

Gamarnik D, Goldberg DA (2010) Randomized greedy algorithms for independent sets and matchings in regular graphs: exact results and finite girth corrections. Comb Prob Comput 19:61–85

Giandomenico M, Rossi F, Smriglio S (2013) Strong lift-and-project cutting planes for the stable set problem. Math Program 141(1–2):165–192

Goycoolea M, Murray AT, Barahona F, Epstein R, Weintraub A (2005) Harvest scheduling subject to maximum area restrictions: exploring exact approaches. Oper Res 53:490–500

Grubesic TH, Murray AT (2008) Sex offender residency and spatial equity. Appl Spat Anal Policy 1:175–192
Hochbaum DS, Pathria A (1997) Forest harvesting and minimum cuts: a new approach to handling spatial constraints. For Sci 43(4):544–554
Jones JG, Meneghin BJ, Kirby MW (1991) Formulating adjacency constraints in linear optimization models for scheduling projects in tactical planning. For Sci 37:1283–1297
Kirby M, Hager W, Wong P (1986) Simultaneous planning of wildland management and transportation alternatives. TIMS Stud Manag Sci 21:371–387
Mauri GR, Ribeiro GM, Lorena LA (2010) A new mathematical model and a Lagrangean decomposition for the point-feature cartographic label placement problem. Comput Oper Res 37(12):2164–2172
Mealey SP, Lipscomb JF, Johnson KN (1982) Solving the habitat dispersion problem in forest planning. Trans N Am Wildl Natur Resour Conf 47:142–153
Moon AD, Chaudhry S (1984) An analysis of network location problems with distance constraints. Manag Sci 30:290–307
Murray AT (1995) Modeling adjacency conditions in spatial optimization problems. PhD dissertation, UCSB, Santa Barbara, CA
Murray AT (2007) Spatial environmental concerns. In: Weintraub A, Romero C, Bjorndal T, Epstein R (eds) Handbook of operations research in natural resources. Springer, New York, pp 419–429
Murray AT, Church RL (1995a) Heuristic solution approaches to operational forest planning problems. OR Spektrum 17:193–203
Murray AT, Church RL (1995b) Measuring the efficacy of adjacency constraint structure in forest planning. Can J For Res 25:1416–1424
Murray AT, Church RL (1996) Constructing and selecting adjacency constraints. INFOR 34:232–248
Murray AT, Church RL (1997) Facets for node packing. Eur J Oper Res 101:598–608
Murray AT, Church RL (1999) Using proximity restriction for locating undesirable facilities. Stud Locat Anal 12:81–99
Murray AT, Kim H (2008) Efficient identification of geographic restriction conditions in anti-covering location models using GIS. Lett Spat Resour Sci 1:159–169
Murray AT, Wei R, Grubesic TH (2014) An approach for examining alternatives attributable to locational uncertainty. Environ Plan B Plan Des 41(1):93–109
Nelson JN, Brodie JB (1990) Comparison of a random search algorithm and mixed integer programming for solving area-based forest plans. Can J For Res 20:934–942
Nemhauser G, Sigismondi G (1992) A strong cutting plane/branch-and-bound algorithm for node packing. Oper Res 43:443–457
Nemhauser GL, Trotter LE (1975) Vertex packings: structural properties and algorithms. Math Program 8:232–248
Nemhauser GL, Wolsey LA (1988) Integer and combinatorial optimization. Wiley, New York
Niblett MR (2014) The anti-covering location problem: new modeling perspectives and solution approaches. PhD dissertation, UCSB, Santa Barbara, CA
Niblett MR, Church RL (2015) The disruptive anti-covering location problem. Eur J Oper Res 247 (3):764–773
Padberg MW (1973) On the facial structure of set packing polyhedral. Math Program 5:199–215
Ratick S, Meacham B, Aoyama Y (2008) Locating backup facilities to enhance supply chain disaster resilience. Growth Chang 39:642–666
Ribeiro GM, Lorena LAN (2008) Lagrangean relaxation with clusters for point-feature cartographic label placement problems. Comput Oper Res 35(7):2129–2140
Thompson EF, Halterman BG, Lyon TJ, Miller RL (1973) Integrating timber and wildlife management planning. For Chron 49(6):247–250
Tomita E, Tanaka A, Takahashi H (2006) The worst-case time complexity for generating all maximal cliques and computational experiments. Theor Comput Sci 363:28–42

Torres-Rojo JM, Brodie JD (1990) Adjacency constraints in harvest scheduling: an aggregation heuristic. Can J For Res 20:978–986

Vielma JP, Murray AT, Ryan DM, Weintraub A (2007) Improving computational capabilities for addressing volume constraints in forest harvest scheduling problems. Eur J Oper Res 176:1246–1264

Warrier D, Wilhelm WE, Warren JS, Hicks IV (2005) A branch-and-price approach for the maximum weight independent set problem. Networks 46(4):198–209

Wei R, Murray AT (2012) An integrated approach for addressing geographic uncertainty in spatial optimization. Int J Geogr Inf Sci 26:1231–1249

Wei R, Murray AT (2015) Spatial uncertainty in harvest scheduling. Ann Oper Res 232(1):275–289

Wei R, Murray AT (2017) Spatial uncertainty challenges in location modeling with dispersion requirements. In: Thill J-C (ed) Spatial analysis and location modeling in urban and regional systems. Springer, New York

Williams JC (2008) Optimal reserve site selection with distance requirements. Comput Oper Res 35 (2):488–498

Yoshimoto A, Brodie JD (1994) Comparative analysis of algorithms to generate adjacency constraints. Can J For Res 24:1277–1288

Zeller RE, Achabal DD, Brown LA (1980) Market penetration and locational conflict in franchise systems. Decis Sci 11:58–80

Zoraster S (1990) The solution of large 0–1 integer programming problems encountered in automated cartography. Oper Res 38(5):752–759

Chapter 6
Weighted Benefit, Variable Radius, and Gradual Coverage

6.1 Introduction

The previous chapters have primarily focused on application contexts and modeling approaches where predefined, discrete coverage metrics are appropriate. Examples of this include: a fire department adequately serves/covers those properties that are within 5 min of travel from a station, or a surveillance system monitors/covers the areas that can be viewed by one or more cameras. That is, coverage is defined as being achieved or not, a simple binary yes or no property. The fact that coverage is defined as being provided or not to an area or object conceived of as a demand for service makes many coverage problems relatively simple to construct, especially for problems that are discrete in nature. When both demand objects and potential facility sites are discrete locations and finite in number, it is possible to identify which sites are capable of covering specific demand objects. An important question, however, is whether coverage should be so crisply defined. For example, when demand for service requires five and a half minutes to respond to from the closest fire station, it may not be considered covered according to a desired 5 min service time standard. In reality, demand for service along these lines obviously receives some level of degraded response service, but just not complete coverage characteristics associated with an established service standard. This chapter therefore explores how coverage models have been extended to be more flexible by including multiple levels of coverage, or steps of coverage, as well as defining a range where coverage is gradually degraded or lost. The idea that service/coverage is degraded, lost or not provided is itself of potential concern, and raises issues of equity. Essentially, in a public setting, we should be concerned with treating those demands that are not covered as fairly as possible. How do we identify a facility configuration (solution) that is as equitable as possible? This, too, is a subject of this chapter. Finally, there are cases when the coverage capabilities at a given location can be a function of investment. That is, we might be able to expand what a facility can cover by enhancing or upgrading associated equipment. For example, a viewshed (or coverage range) of a fire lookout

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_6

tower might be extended by increasing its height. An emergency broadcast tower, as another example, might be outfitted with a superior transmitter providing a stronger signal, thereby increasing its range of reception. Such enhancements or upgrades likely are more costly, but do represent ways service capabilities may be altered. This chapter also addresses modeling where there may be options for increasing the coverage range of a facility along with making location siting decisions.

Before we delve into the nuances of expanding the range under which we define coverage, either in discrete steps or with some gradual decay function, a discussion of equity is given in the next section. Imbedded in this discussion is the general relationship of covering models and the *p*-median location problem, a topic first introduced in Chap. 2. This theoretical basis is important because most of the models defined in this chapter have equivalent *p*-median counterparts. From a theoretical perspective, many of the models are forms of a hybrid construct, defined between the strict decay values of service in the *p*-median problem and more discrete values of service implied when using a covering model.

6.2 Equity and Implied Value of Service

Since the early developments in covering models, there has been a concern for the definition of suitable service, especially in terms of measuring the value of being covered (Church and Bell 1981; Church and Roberts 1983) as well as concern for those not covered within what was termed the desired maximal service distance or time, *S*. Noted in Chap. 2, the original work of Church and ReVelle (1974) on the MCLP recognized that although a decision maker may be interested in covering as much demand as possible within a desired service distance *S*, there could be concern for those not served within this response standard. In reality, there may be a distance (or travel time) from facilities beyond which no demand should be. They called this a "mandatory closeness condition" and proposed a modified form of the MCLP described as follows: *maximize the demand that can be covered within a given maximal service standard S while at the same time ensuring demand will be served by a facility no farther than T away*. Here the mandatory coverage standard *T* is greater than the desired coverage standard *S*, with the intent to be as fair as possible to those not provided coverage within the desired coverage standard. There can be many reasons why complete coverage cannot be achieved, but often this is due to budget limitations). However, that does not imply that any demand should be completely ignored with respect to service response. Adding a secondary coverage standard that is mandatory provides a degree of equity that is not accounted for in the classic MCLP.

It was shown in Chap. 2 that the MCLP is equivalent to a special form of the *p*-median problem. The intent of the *p*-median problem is to locate *p*-facilities in order to minimize total weighted distance associated with serving all demand. Total weighted distance is defined as the sum of each demand weight multiplied by the assignment distance of that demand, where each demand is assigned to its

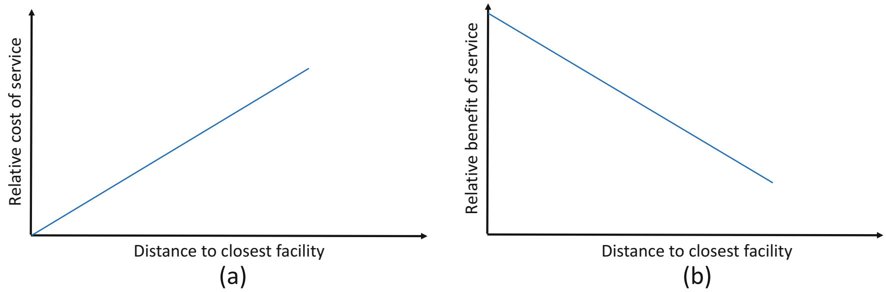


Fig. 6.1 Relative cost and relative benefit of service for each demand structured in the *p*-median problem as a function of distance from their closest facility

closest located facility. Based upon the relationship between the *p*-median problem and the MCLP, Church (1974) suggested a form of the *p*-median problem called the Fringe Sensitive Location Problem, subsequently applied in Bell and Church (1985) and Church et al. (1991). This problem can be defined as follows: *minimize the weighted distance/time of those not covered within a service standard S when locating p facilities*. The objective makes the system as accessible as possible to those not covered within the desired coverage distance *S*. Note, that this problem does not require a solution to maximize coverage, but high levels of coverage (if not maximal) are encouraged because one of the best ways in which to minimize the weighted distance/time of those not covered is to cover as much demand as possible. The basic idea is that one could alter facility locations that also achieve a maximal covering in order to make uncovered demand closer, providing some degree of equity.

While the *p*-median problem seeks to locate facilities and allocate demand such that the average distance to obtain service is minimized, there are no explicit costs accounted for other than distance (or travel time) from demand to a facility. That is, the service costs of the system do not explicitly account for anything beyond travel/accessibility. As a result, the relative cost of serving each demand is a linear function of distance. This is depicted in Fig. 6.1a. The interpretation is that the relative benefit starts at some positive level, β, and declines as a linear function of distance from a demand, as depicted in Fig. 6.1b. The *p*-median problem can therefore be structured and solved by either minimizing service costs (Fig. 6.1a) or maximizing relative benefit (Fig. 6.1b).

It is also possible to consider the MCLP in terms of costs and relative benefits. The covering construct is based upon the assumption that the relative value or benefit of service is some positive value, a constant, over the range between 0 and *S*. Beyond *S*, the value of benefit drops to zero. This is somewhat obvious, as the MCLP involves maximizing those served within the standard *S* without any distinction as to whether service is closer to a facility or further away. This relationship between benefit and proximity is depicted in Fig. 6.2b, where the relative benefit of coverage starts at the level β and continues at the level β until *S* is reached, where the relative benefit then drops to zero. Of course, when defining the MCLP in Chap. 2, there was

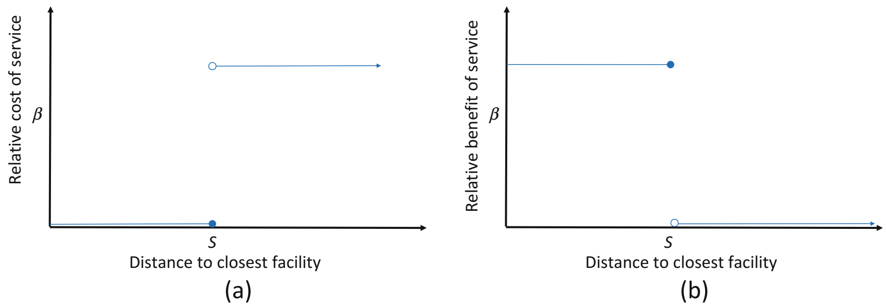


Fig. 6.2 Relative cost and benefit of service for each demand structured in the MCLP as a function of distance

no mention of the value or benefit of service. Nevertheless, since the only thing that counts is coverage within S, the implication is that the value or relative benefit of service beyond S must be zero. Alternatively, viewing the MCLP in terms of costs, Fig. 6.2a reflects that the relative cost of service is zero for a demand that is within S of a facility, but jumps up to some constant level beyond S. Just as the relative benefit was set at β within the range of 0 to S, the implied relative cost of service can be conceived to be β beyond S. Similar to the p-median problem, the MCLP can be structured and solved by maximizing the relative benefit of coverage, based upon the function summarized in Fig. 6.2b or equivalently by minimizing relative costs of service using the cost function given in Fig. 6.2a.

One of the important concepts here is that the MCLP can be solved as an equivalent p-median problem by using cost instead of an explicit distance function (Fig. 6.2a). This can be done through the use of a transformed shortest distance matrix, something discussed in Chap. 2. However, there is an even larger and more important issue here. Doing this implies that the value or benefit drops abruptly to zero at the desired maximum service standard S, raising the question of whether service benefit really drops to zero beyond S. Many may take issue with this property and argue that the relative benefit at the distance $S + \Delta$ may be lower than β but not zero (assuming the value of Δ is small).

The fact that coverage values can and do change based on proximity is the primary focus of this chapter. In the next section we present a generalized form of maximal covering which allows for a range of benefit values to be considered.

6.3 Generalized Maximal Covering Location Problem

There can be many application contexts for covering models where a strictly defined standard or metric accurately reflects whether coverage is provided (or not) by a facility serving demand. However, there also exists many cases where it might not be appropriate for such a crisp cutoff in the benefit function, like that depicted in

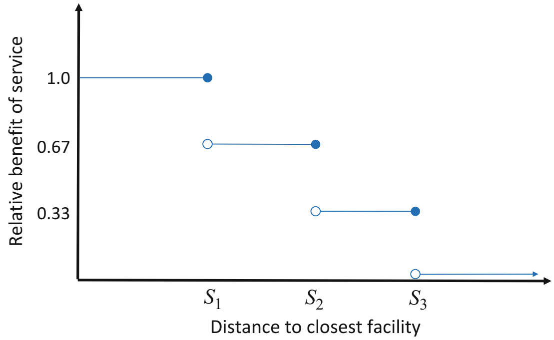


Fig. 6.3 A generalized form of relative benefits of service, where the relative benefits step down

Fig. 6.2. For example, consider a fire department relying on a response distance standard of 1.18 miles. This standard is actually used by the fire department serving the community of Elk Grove, CA (Murray 2013). A quick response to a fire is critical for saving lives and minimizing property losses. At this distance range, the expectation is that it is highly probable that response is possible within 5 min, even when accounting for the time taken to dispatch personnel. The classic view in fire planning involves maximizing what is covered within some response standard, while locating a fixed number of stations. When seeking a high level of coverage, it means only a minority will be outside the coverage range. Does that mean that the value of service drops to zero immediately beyond 1.18 miles, or even 1.5 miles? Most would agree that the answer is no. Clearly, a 6 min response will have benefit, but maybe not as high as what would be provided within the 5 min response time corresponding to the 1.18 mile standard. Thus, the actual benefit curve might look like the relative benefit of service function shown in Fig. 6.3. This curve represents three levels of relative benefits, constructed using three different service/coverage ranges. In this case, those served within the service standard of S_1 receive the relative benefit value of 1.0 while those served beyond the S_1 but within the distance of S_2 receive a benefit value of 0.67, with a relative reduction of a third compared to those served inside the first covering range. The benefit function drops even further for the third threshold (between S_2 and S_3) to 0.33. In this section, we will assume that the shape of the relative benefit curve can be represented as a series of steps with non-increasing values as distance (or travel time) increases, over a number of coverage ranges and eventually drops to zero. Such step functions can be broad, short, steep, regular, irregular, etc. The basic idea is that we can represent service coverage over a series of coverage ranges.

The MCLP can be generalized to distinguish the value of benefits received over a series of coverage ranges by expanding on the notation used in Chap. 2. Consider:

i = index referencing demand objects
j = index referencing potential facility locations
k = index of service standards
S_k = service standard threshold for range k

$b_i^k =$ relative benefit of coverage for demand i by a facility within the service standard range (S_{k-1}, S_k)

$d_{ij} =$ shortest distance or travel time between demand i and potential facility j

$N_i^k = \{ j \mid S_{k-1} < d_{ij} \leq S_k \}$, set of facilities j that are within the maximal service standard (or time) S_k, but beyond S_{k-1}, for demand i

$p =$ number of facilities to be located

$a_i =$ anticipated service load for demand i

$$x_j = \begin{cases} 1, & \text{if a facility is located at site } j \\ 0, & \text{otherwise} \end{cases}$$

$$y_i^k = \begin{cases} 1, & \text{if demand } i \text{ is covered within the standard } S_{k-1} \text{ and } S_k \\ 0, & \text{otherwise} \end{cases}$$

Given this notation, the idea is that coverage ranges can be used for each benefit step. The first coverage range, (S_0, S_1), defines service based on proximity from S_0 to and including S_1, assuming of course that $S_0 = 0$. The second distance range begins just beyond S_1 to and including S_2. The third and subsequent ranges are define in a similar manner. The mathematical formulation of this generalized maximal covering location problem (GMCLP) is as follows (Church and Roberts 1983; Berman and Krass 2002):

$$\text{GMCLP}: \qquad Maximize \sum_i \sum_k a_i b^k y_i^k \tag{6.1}$$

subject to:

$$\sum_{j \in N_i^k} x_j \geq y_i^k \quad \forall i, k \tag{6.2}$$

$$\sum_k y_i^k \leq 1 \quad \forall i, k \tag{6.3}$$

$$\sum_j x_j = p \tag{6.4}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{6.5}$$

$$y_i^k \in \{0, 1\} \quad \forall i, k \tag{6.6}$$

The GMCLP maximizes the total weighted relative benefits, objective (6.1), associated with the siting of p facilities. The relative benefit of coverage is evaluated for each demand and is based upon the coverage range of highest benefit. The provision of coverage for a specific demand i within a specific range k is defined in constraints (6.2). If one or more facilities are located within range k for demand i, then the sum of those locations will equal one or more and allow the variable y_i^k to be equal one in value. If there are no facilities that can cover demand i in distance range k, the term $\sum_{j \in N_i^k} x_j$ in (6.2) will equal zero, forcing y_i^k to be zero in value. Of course, it is possible that demand i might be covered within several distance ranges. However, constraints (6.3) prevent coverage for demand i to be counted no more than once over the different coverage ranges. If demand i is covered at two or more distance ranges, then the range with the highest relative benefit will be selected based on the objective function. Since the values are represented as a step function of decreasing benefit values, the highest and closest range will always be selected and accounted for. Constraint (6.4) specifies that p-facilities are to be located. Binary integer restrictions are indicated in Constraints (6.5) and (6.6).

A few additional comments are important at this point. As noted in Chap. 2, the LSCP (location set covering problem) reflects that situation where all demand must be suitably covered be a minimal number of sited facilities. Accordingly, a feasible solution necessarily satisfies a mandatory closeness condition for all demand. Thus, there is no need to consider a generalization along the lines of the GMCLP.

From an application perspective, the GMCLP has been solved using general purpose mixed integer-linear programming software, and heuristics have been explored as well (Berman and Krass 2002). Additionally, other formulations of this problem are possible that have a "tighter" structure when solving using a classic branch and bound algorithm.

6.4 Expanded Forms of Generalized Coverage

Austin (1974) suggested an approach to evaluate urban public facility locations that he terms an alternative to benefit cost analysis. This methodological framework was based upon explicitly evaluating the spatial distribution of impacts (both good and bad) associated with a public facility. Although the concepts that he introduced with respect to public facilities were explained within the context of urban areas, such concepts are more general in nature and can be applied regardless of the setting. Austin suggested four general forms of net impacts, and believed that most public facilities will fall into one of these types. One of the forms matches that given as the relative benefit of service function shown in Fig. 6.2 for maximal coverage. He described this form as being applicable to "... a park or a swimming pool or a governmental building, the effect of which (in net) remains fairly constant over the area it serves" (Austin 1974). To be accurate, the function of net benefits (or net impacts in his parlance) did not drop to zero at a service distance S, but rather went to zero at the boundary of the service area. Another form of net benefits is depicted in Fig. 6.4, described by Austin (1974) as a form in which "... accessibility to the facility is a prime factor in its impact on place utility." Church and Roberts (1983) characterize this type of function as one that accounts for a service that is desirable, but one that perhaps someone may not want next door (or too close). Consider a fire station, as an example. Living next to a fire station might afford great service should there be a fire, but on a day to day basis disruption and noise might be significant. This is true not only with respect to sirens signaling an emergency departure during day or night hours, but also risks associated with speedy departures that could put people and

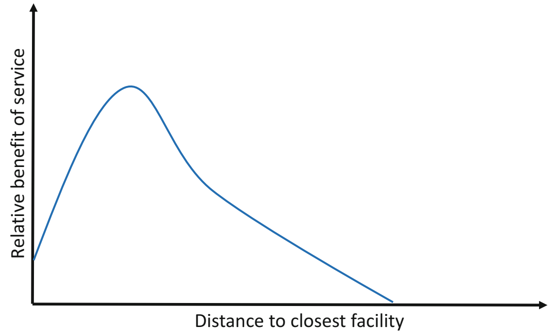


Fig. 6.4 A non-decreasing form of net benefits

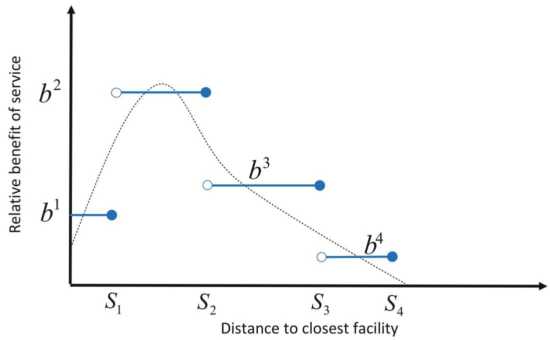


Fig. 6.5 Step based approximation of the nonlinear form of net benefits

children in harm's way. Thus, the net benefits for those living close to the fire station could be lower than the net benefits derived by those living some distance away.

The net benefit curve given in Fig. 6.4 can be approximated as a step function similar to that shown in Fig. 6.5. Here the relative benefit steps do not strictly decrease with distance as was the case in Fig. 6.3, but rather start at a lower value, b^1, then it jumps to b^2 at the standard S_1 where it remains until the distance S_2. Between standards S_2 and S_3 the relative net benefit values are constant but somewhat lower than b^1. Beyond the distance S_3, net benefits are considered to be zero. Of course we can better approximate this function with even more steps, but the three steps depicted in Fig. 6.5 show how such an approximation can be accomplished.

Unfortunately, the GMCLP cannot be utilized for addressing the case where relative benefit values do not strictly decrease with distance. The reason for this is quite simple. Assume that there are two service facilities that are close enough to be counted as providing some level of coverage for a given demand. Say one of these facilities is within S_1 and the other facility is greater than S_1 but is within S_2. The GMCLP will allocate the service provided as that of the second step (or a relative value of benefit of b^2, even when there is a closer facility that generates impacts that must be borne by that demand. In this case the true benefit that would be experienced by this demand would be b^1, not b^2. The reason for this is that the model selects the level of service that is provided to a demand based upon the larger contribution to the objective function. Thus, the GMCLP works for a problem with strictly decreasing relative benefit values as distance increases, and does not work appropriately in the case where there could be lower net benefits close to a facility compared to those for facilities further away.

In order to address this situation, a model form is needed that represents the relative benefit function as depicted in Fig. 6.5. Doing so is possible through an extension of the GMCLP. What is necessary is to add constraints that ensure that relative benefits are counted for the closest facility to a demand, rather than a more distant facility that might be higher in relative net benefits. This can be accomplished using constraints of the following form:

$$x_j + y_i^{k'} \leq 1 \quad \forall i, j \in N_i^k, k, k' > k \tag{6.7}$$

This represents a set of constraints for each demand i and each facility site j in order to account for cases where relative non-decreasing net benefit exists in coverage, ensuring that benefits are always attributed to the closest standard range k possible. Consider the first standard S_1 associated with demand i reflected in Fig. 6.5. Because this first coverage range has a relative benefit that is lower than the next range, constraints (6.7) would be necessary. The following would then be added to the model (Church and Roberts 1983):

$$x_j + y_i^2 \leq 1 \quad \forall i, j \in N_i^1 \tag{6.8}$$

where $k = 1$ and $k' = 2$ in this case. Each constraint prevents a second range step represented by y_i^2 from being assigned as the level of coverage, when a given site j that is within S_1 distance of demand i has been selected for a facility. Altogether, such constraints prevent assigning a higher valued coverage range (or step) from being allocated when any closer site in the first coverage range has been selected for a facility. Adding these constraints to the GMCLP results in a model that can solve problem instances where the relative benefit of coverage is reflected by the non-decreasing situation summarized in Fig. 6.5.

6.5 Endogenously Determined Coverage

Thus far, the coverage standard has been viewed as a constant, applying equally to each potential facility location. As noted previously, however, service capabilities may vary from facility to facility, and it may be beneficial to incorporate this as a part of the system design. One of the first modeling approaches to address this was reported in Goodchild and Lee (1989) involving the location of fire lookout towers used for viewing the landscape, where visibility is thought of as coverage. Their problem included both the placement of lookout towers as well as the height designation of each tower. The idea is that as the height of a tower is increased, the area of visibility also increases. Because towers can be costly to construct, building fewer higher towers may be less expensive than building more towers that are shorter in height. The same is generally true of emergency beacon emitters and other emergency communication facilities, where higher antennas are capable of communicating over obstacles, like buildings, and reach a broader area. Of course, this is precisely why antennas are often placed on high ridges. The distance of coverage could also be expanded with a higher strength broadcast signal, which may cost more than a lower strength signal. Altogether, there are many examples of where service coverage can vary, and possibly not be imposed in advance, but determined endogenously in the model. The problem can be thought of in two different ways: (1) each facility has a discrete set of alternate designs that represent different capabilities of coverage, and (2) the coverage capabilities at each facility can be represented as a continuous cost and capability function. In this section, we explore these two types of models: one to address the discrete case, and the other to handle the continuous case.

To formulate the discrete case, we need to introduce the following additional and/or modified notation:

o = index of service coverage options
c_o = cost of utilizing coverage option o
$N_{io} = \{j|$ site j using design option o to cover demand i $\}$

$$x_j^o = \begin{cases} 1, & \text{if a facility is placed at site } j \text{ with design option } o \\ 0, & \text{otherwise} \end{cases}$$

With notation used in earlier sections of this chapter along with the above modified notation, we can formulate two models along the lines of what is structured in Goodchild and Lee (1989), where different facility types/options are considered. For consistency purposes, we label these models the multi-radii location set covering problem (MR-LSCP) and the multi-radii maximal covering location problem (MR-MCLP).

$$\text{MR} - \text{LSCP}: \qquad \textit{Minimize} \sum_j \sum_o c_o x_{jo} \tag{6.9}$$

Subject to:

$$\sum_o \sum_{j \in N_{io}} x_{jo} \geq y \quad \forall i \tag{6.10}$$

$$\sum_o x_{jo} \leq 1 \quad \forall j \tag{6.11}$$

$$x_{jo} \in \{0, 1\} \quad \forall j, o \tag{6.12}$$

This model minimizes the total cost of site selection, objective (6.9), but does so taking into account coverage capabilities of different facility types. Constraints (6.10) specify that each demand must receive suitable coverage. Constraints (6.11) limit the number of options to at most one at each site. Binary integer restrictions are imposed in Constraints (6.12).

It should be clear from Chap. 2 that MR-LSCP represents a form of the location set covering problem, as the naming convention suggests, where the objective is to suitably cover all demand. What is unique here is that multiple facility types or options are considered. A different form of the problem can be easily formulated, where the objective is to maximize the demand that is covered while introducing a budget constraint on the cost of the selected facility sites along with choosing coverage capability options. Again, this type of extension was suggested in Goodchild and Lee (1989), with a heuristic proposed for its solution. Additional notation is the following:

$B =$ total budget in facility siting

The associated model is:

$$\text{MR} - \text{MCLP}: \qquad Maximize \ \sum_i a_i y_i \tag{6.13}$$

subject to:

$$\sum_o \sum_{j \in N_{io}} x_{jo} \geq y_i \quad \forall i \tag{6.14}$$

$$\sum_o x_{jo} \leq 1 \quad \forall j \tag{6.15}$$

$$\sum_o \sum_j c_o x_{jo} \leq B \tag{6.16}$$

$$x_{jo} \in \{0, 1\} \quad \forall j, o \tag{6.17}$$

$$y_i \in \{0, 1\} \quad \forall i \tag{6.18}$$

This model variant model maximizes demand covered, objective (6.13). Constraints (6.14) indicate whether coverage is provided to a demand. Constraints (6.15) limit facility selection to at most one type or option at a site. A total budget restriction is imposed in Constraint (6.16). Binary integer restrictions are imposed in Constraints (6.17) and (6.18).

Beyond consideration of multi-radii extension of the LSCP and MCLP, it should be evident that formulation in the generalized case is also possible. Accordingly, this model can be referred to as the multi-radii, generalized maximal covering location problem (MR-GMCLP):

$$\text{MR} - \text{GMCLP}: \qquad Maximize \sum_i \sum_k a_i b^k y_i^k \tag{6.19}$$

subject to:

$$\sum_o \sum_{j \in N_{io}} x_{jo} \geq y_i^k \quad \forall i, k \tag{6.20}$$

$$\sum_k y_i^k \leq 1 \quad \forall i, k \tag{6.21}$$

$$\sum_o x_{jo} \leq 1 \quad \forall j \tag{6.22}$$

$$\sum_o \sum_j c_o x_{jo} \leq B \tag{6.23}$$

$$x_{jo} \in \{0, 1\} \quad \forall j, o \tag{6.24}$$

$$y_i^k \in \{0, 1\} \quad \forall i, k \tag{6.25}$$

The MR-GMCLP maximizes the total weighted relative benefits, objective (6.19), associated with facility siting. Coverage of demand i within a specific range k is defined in constraints (6.20). Constraints (6.21) limit coverage of demand i to be counted no more than once over the different coverage ranges. Constraints (6.22) limit facility selection to at most one type or option at a site. A total budget restriction is imposed in constraint (6.23). Binary integer restrictions are indicated in constraints (6.24) and (6.25).

As with the GMCLP, the assumption for the MR-GMCLP is that relative benefit is decreasing for ranges further away, as depicted in Fig. 6.3. In situations where this is not the case, an extension that incorporates constraints (6.7) can be structured. Beyond this, the problems discussed in this section involve extending the coverage of a facility through strategic investment. These models can be extended to a multi-objective case where coverage is maximized and costs are minimized.

6.6 Continuous Endogenous Coverage

While the above models, MR-LSCP, MR-MCLP and MR-GMCLP, have treated discrete variable options of facility types (and their associated relative benefits) as an endogenous component of the decision making process through discrete mechanisms, it is possible to model the effective reach of a covering facility across a continuum of distances. Accordingly, this can be considered continuous endogeneity. Berman et al. (2009a, b) introduce a variant of this basic idea. Suppose we can extend a coverage range of a facility up to some distance $S_{\max}$, where the extension beyond some initial coverage distance S_{init} can be expressed as a function of Δ. In this case, Δ represents the added coverage distance of a facility, or $d_{ij} - S_{init}$. One can therefore think of a function, $\Gamma_j()$, capable of reflecting the cost of extending coverage capabilities of facility j. Consider then the following additional notation:

f_j = fixed cost to establish facility j
λ_j = cost of extending the coverage of facility j
$t_{ij} = \begin{cases} 1, & \text{if demand } i \text{ assigns to a facility at site } j \\ 0, & \text{otherwise} \end{cases}$
V_j = variable representing the cost to extend coverage of facility j

The variables t_{ij} exist whenever $d_{ij} \leq S_{\max}$. The cost of extension, λ_i, is derived based on the appropriately specified cost function $\Gamma_j()$ associated with a facility and distance variability. The cost of extending the coverage radius at a facility is a function of the maximum distance of any demand assigned to that facility. Note that this represents an additional cost of extending the coverage distance beyond an initial coverage distance S_{init}. If all assignments to a facility are less than this initial coverage distance, then the additional costs will be zero. However, there remains a fixed cost for establishing a facility. As was the case for multi-radii, it is possible to consider two basic variants of the continuous endogeneity, one based on set covering and the other on maximal covering.

The first to be detailed is the Continuous Endogeneity Set Covering Problem (CE-SCP). In this particular case the LSCP reference is avoided because of the inclusion of fixed cost terms. As noted in Chap. 2, the LSCP assumes all fixed costs are the same, so there is no need to include such detail in the model formulation. More generally, however, the set covering problem has incorporated fixed costs (see Murray 2005; Church and Murray 2009). The associated model is as follows:

$$\text{CE} - \text{SCP}: \qquad \textit{Minimize} \sum_j \left(V_j + f_j x_j\right) \tag{6.26}$$

subject to:

$$\sum_j t_{ij} = 1 \quad \forall i \tag{6.27}$$

$$V_j \geq \lambda_j t_{ij} \quad \forall i, j \mid d_{ij} > S_{init} \tag{6.28}$$

$$t_{ij} \leq x_j \quad \forall i, j \tag{6.29}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{6.30}$$

$$t_{ij} \in \{0, 1\} \quad \forall i, j \tag{6.31}$$

$$V_j \geq 0 \quad \forall j \tag{6.32}$$

This model minimizes the total fixed and variable costs of site selection, objective (6.26). Constraints (6.27) specify that each demand must be covered. Constraints (6.28) track the associated variable cost for facility coverage extension. Constraints (6.29) prevent assignment of coverage unless the facility is sited. Binary integer restrictions and non-negativity conditions are imposed in constraints (6.30), (6.31) and (6.32).

The idea here is that all facilities will cost at least f_j. As long as the demands that assign to a facility are at distances less than or equal to S_{init} there are no extra costs of coverage based upon a coverage distance extension. However, whenever a demand assigns to a facility that is further than S_{init} away from its assigned distance an extension in coverage distance is needed. The farthest distance of assignment over all demands that assign to a particular facility at j will define the additional coverage distance cost V_j. It is important to note that no variables t_{ij} are used in the model where the associated d_{ij} exceeds the maximum coverage radius of $S_{\max}$. This prevents the facility configuration using some infeasible coverage distance.

The second approach, based on maximal coverage, is referred to as the Continuous Endogeneity Maximal Covering Problem (CE-MCP). The associated model is as follows:

$$\text{CE} - \text{MCP}: \qquad \textit{Maximize} \sum_i a_i y_i \tag{6.33}$$

subject to:

$$\sum_j t_{ij} \geq y_i \quad \forall i \tag{6.34}$$

$$V_j \geq \lambda_j t_{ij} \quad \forall i,j \mid d_{ij} > S_{init} \tag{6.35}$$

$$t_{ij} \leq x_j \quad \forall i,j \tag{6.36}$$

$$\sum_j \left(V_j + f_j x_j\right) \leq B \tag{6.37}$$

$$x_j \in \{0,1\} \quad \forall j \tag{6.38}$$

$$y_i \in \{0,1\} \quad \forall i \tag{6.39}$$

$$t_{ij} \in \{0,1\} \quad \forall i,j \tag{6.40}$$

$$V_j \geq 0 \quad \forall j \tag{6.41}$$

This model maximizes total coverage provided, objective (6.33). Constraints (6.34) relate that service allocation to coverage. Constraints (6.35) track the associated variable cost for facility coverage extension. Constraints (6.36) prevent assignment of coverage unless the facility is sited. A total budget restriction is imposed in constraint (6.37). Binary integer restrictions and non-negativity conditions are imposed in constraints (6.38–6.41).

The two problems discussed in this section involve extending the coverage of a facility by greater investment. Both models can be extended to a multi-objective case where coverage is maximized and costs are minimized. Further, other extensions may be added, such as non-decreasing relative benefit.

6.7 Gradual Coverage

An overarching goal of this chapter has been to demonstrate the inherent flexibility of coverage modeling based on the use of service standards. That does not necessarily mean that such an approach is beyond reproach. Berman et al. (2003) have questioned the use of a coverage function with discrete steps such as that shown in Fig. 6.3. The basis for concern is along the following lines. Assuming that desired suitable coverage is S_1, as an example, does coverage drop or degrade in a discrete step exactly at S_1, as is the case for relative benefit in Fig. 6.3, or does the value of coverage decay gradually? Drezner et al. (2004) viewed the coverage function as an estimate of those being served. Service up to and including the distance of S_1 was thought to be complete, or rather everyone would be served or use their covering facility. But beyond the distance of S_1, patronage or attendance would gradually decay to zero, at a distance S_2. One might argue as to whether the coverage function represents the value of service, or the number of patrons that use the service, even at distances beyond a desired maximum service standard S_1. This is really a different interpretation of what coverage means and is generally applied, especially when it might be used in a public facility setting. That is, coverage typically represents whether something has been served within some standard or the relative benefit of that service.

This perspective of attendance or use, however, presents some problems of interpretation within the context of covering. In retail, for instance, one of the main problems is to estimate how many will shop at a given location and store type. There are models in the retail literature that are defined especially for this type of problem, as it is far more complicated than a simple coverage function. But, it is plausible that a coverage function could represent the attendance function at a store, even if we use an exponential decay function. Here, we will continue the description of gradual cover decay in terms of relative benefit, where the relative benefit of service declines as distance increases. The models of Berman et al. (2003), Drezner et al. (2004), and Berman et al. (2010) can be conceived along the lines of the relative benefit of service function shown in Fig. 6.6. The relative benefit of coverage is b^1 up to and including the service distance S_1. Beyond the distance of S_1 there is a steady decline until the distance of S_2, at which point it reaches a relative benefit value of zero. The decay function is depicted as linear, but other shapes can be easily modeled (Church and Bell 1981).

As noted in Chap. 2, it is possible to structure a coverage problem as a p-median problem equivalent. Doing so may have some associated benefits. In this particular case, the standard distance measure can be utilized for reflecting coverage decay. Consider the following modification of distance, d_{ij}:

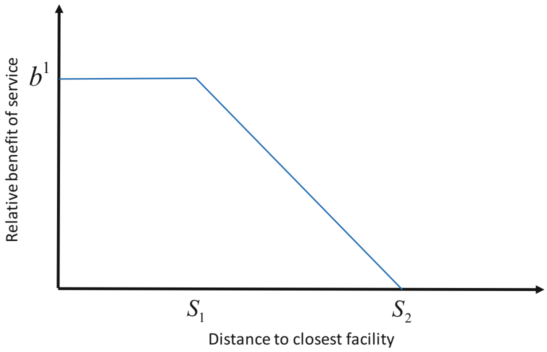


Fig. 6.6 Decay function in the gradual covering location problem

$$d'_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq S_1 \\ 1 - \left(d_{ij} - S_1\right)/(S_2 - S_1) & \text{if } S_1 < d_{ij} \leq S_2 \\ 0, & \text{if } d_{ij} > S_2 \end{cases}$$

$$z_{ij} = \begin{cases} 1, & \text{if demand } i \text{ assigns to facility } j \\ 0, & \text{otherwise} \end{cases}$$

With this notation, the following model may be structured:

$$\text{GC} - \text{MCLP}: \qquad Maximize \sum_i \sum_j a_i d'_{ij} z_{ij} \tag{6.42}$$

subject to:

$$\sum_j z_{ij} = 1 \quad \forall i \tag{6.43}$$

$$z_{ij} \leq x_j \quad \forall i, j \tag{6.44}$$

$$\sum_j x_j = p \tag{6.45}$$

$$x_j \in \{0, 1\} \quad \forall j \tag{6.46}$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j \tag{6.47}$$

This model maximizes the total demand covered, objective (6.42). Constraints (6.43) specify that each demand must be assigned to a facility, though the facility may not in fact provide suitable coverage. Constraints (6.44) prevent assignment of demand to a given facility unless the facility is sited. Constraint (6.45) requires p facilities to be sited. Binary integer restrictions are imposed in constraints (6.46) and (6.47).

This model, using the d'_{ij} matrix instead of the original d_{ij}, structures a general gradual covering decay problem. In fact, all of relative benefit of service functions specified in Figs. 6.2, 6.3, 6.4, 6.5, and 6.6 can be structured and solved in a similar fashion (see Church and Roberts 1983 for details). However, many functional forms of relative benefit may in fact require the use of closest assignment constraints. Such constraints are discussed in Church and Roberts (1983) and are expanded upon in Gerrard and Church (1996).

6.8 Summary and Concluding Comments

Most applications of location covering models involve the use of one maximal service distance or some other standard that is used to define whether coverage has been provided or not to a specific demand. In this chapter we have expanded on the notion that a demand is either covered or not, by introducing models that value coverage over several distance ranges, discounting the value of coverage as distance increases. For all practical purposes, the multiple range covering models described in Sects. 6.2 and 6.3 are quite practical and represent an approach that urban planners might find very appealing. We have also introduced models that endogenously determine a coverage distance where the coverage distance or extent of coverage at each facility can be expanded at a cost. One area that we did not discuss, and perhaps should have, is where the service benefit of coverage in some range could be negative (an issue discussed in general for public facility location problems by Austin 1974). Beyond this, there have been other extensions to some of the basic problems noted in this chapter, like the gradual covering location problem (sometimes called the gradual decay of service). They include a stochastic version (Drezner et al. 2010), a generalized ordered approach (Berman et al. 2009a, b), a probabilistic form (Berman et al. 2011), hub constructs (Setak and Karimi 2014; Peker and Kara 2015), and service quality (Eiselt and Marianov 2009). Space limitations prevent further model specification, but in general, extensions to address these and other concerns are possible.

References

Austin CM (1974) The evaluation of urban public facility location: an alternative to benefit-cost analysis. Geogr Anal 6(2):135–145

Bell TL, Church RL (1985) Location-allocation modeling in archaeological settlement pattern research: some preliminary applications. World Archaeol 16(3):354–371

Berman O, Krass D (2002) The generalized maximal covering location problem. Comput Oper Res 29(6):563–581

Berman O, Krass D, Drezner Z (2003) The gradual covering decay location problem on a network. Eur J Oper Res 151(3):474–480

Berman O, Drezner Z, Krass D, Wesolowsky GO (2009a) The variable radius covering problem. Eur J Oper Res 196(2):516–525

Berman O, Kalcsics J, Krass D, Nickel S (2009b) The ordered gradual covering location problem on a network. Discret Appl Math 157(18):3689–3707

Berman O, Drezner Z, Krass D (2010) Generalized coverage: new developments in covering location models. Comput Oper Res 37(10):1675–1687

Berman O, Krass D, Wang J (2011) The probabilistic gradual covering location problem on a network with discrete random demand weights. Comput Oper Res 38(11):1493–1500

Church R, ReVelle CR (1974) The maximal covering location problem. Pap Reg Sci 32(1):101–118

Church R, Current J, Storbeck J (1991) A bicriterion maximal covering location formulation which considers the satisfaction of uncovered demand. Decis Sci 22(1):38–52

Church RL (1974) Synthesis of a class of public facility location models. PhD Dissertation, The Johns Hopkins University, Baltimore, MD

Church RL, Bell T (1981) Incorporating preferences in location-allocation models. Geogr Perspect 48:22–34

Church RL, Murray AT (2009) Business site selection, location modeling, and GIS. Wiley, New York

Church RL, Roberts KL (1983) Generalized coverage models and public facility location. Pap Reg Sci 53(1):117–135

Drezner T, Drezner Z, Goldstein Z (2010) A stochastic gradual cover location problem. Nav Res Logist (NRL) 57(4):367–372

Drezner Z, Wesolowsky GO, Drezner T (2004) The gradual covering problem. Nav Res Logist (NRL) 51(6):841–855

Eiselt HA, Marianov V (2009) Gradual location set covering with service quality. Socio Econ Plan Sci 43(2):121–130

Gerrard RA, Church RL (1996) Closest assignment constraints and location models: properties and structure. Locat Sci 4(4):251–270

Goodchild MF, Lee J (1989) Coverage problems and visibility regions on topographic surfaces. Ann Oper Res 18(1):175–186

Murray AT (2005) Geography in coverage modeling: exploiting spatial structure to address complementary partial service of areas. Ann Assoc Am Geogr 95(4):761–772

Murray AT (2013) Optimising the spatial location of urban fire stations. Fire Saf J 62:64–71

Peker M, Kara BY (2015) The P-Hub maximal covering problem and extensions for gradual decay functions. Omega 54:158–172

Setak M, Karimi H (2014) Hub covering location problem under gradual decay function. J Sci Ind Res 73:145–148

Chapter 7
Capture, Capacities, and Thresholds

7.1 Introduction

There can be a number of different problem settings under which covering models can be defined and applied. Many of the models that are discussed in this chapter were originally inspired by issues associated with retail and competition. Although one at first blush may think of retail siting and coverage models as having little in common, except for something obvious like a pizza chain attempting to locate facilities so that it can deliver pizzas everywhere in a city within 30 min, there are surprisingly many retail elements that can be defined and addressed using coverage constructs.

Christaller (1933) attempted to address the question of why retail centers were arranged the way they were across a region. He reasoned that such centers needed a sustainable level of customers, called the threshold. Without a retail establishment attracting a threshold amount of customers, a business was not viable. Further, if a retail center had a large number of customers, much larger than the threshold, it was an invitation for other firms to enter that market area and establish their own facilities. Each of these elements can also be modeled using covering concepts, albeit with added components.

One of the first models to explicitly address a retail problem using a covering framework was the list selection problem in direct mail advertising (Dwyer and Evans 1981). Such a problem is characterized by a retail company wanting to rent various magazine subscriber lists in order to launch a mail-order catalogue, often for an upcoming holiday or event. The key issue is to determine which lists to rent in order to cover/reach as many of the potential customers as possible. The mathematical program developed by Dwyer and Evans (1981) has the exact same structure as the maximal covering location problem introduced and formulated in Chap. 2.

Another retail-based covering model is the maximum capture or sphere of influence location problem (ReVelle 1986). This model was inspired by the early work of Hotelling (1929) where two vendors compete for customers. The model

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_7

developed by ReVelle (1986) was based upon taking the perspective of one of the competitors in an attempt to capture as many of the other's customers while locating several retail facilities. Although this model is quite simple, it encompasses a key element in retail location, that of serving a region better than a competitor.

Finally, many settings suggest that facilities can handle only so many customers at the same time without being congested or losing service quality. That is, there is a limit or capacity to reasonably handle customers. In many applications, it is assumed that each facility can handle all of the customers it serves, but in other contexts this is simply not true. Consequently, the use of a capacity on each facility may be appropriate. This too, can be addressed in a covering model.

Collectively, the above examples have hopefully demonstrated that there are many possible applications for covering models in retail and service location. In the remainder of this chapter, we provide details on models that are designed to capture customers or an audience, models that maintain viability through the use of thresholds and capacity conditions, and models that carve up a region into franchise areas. In the next section, we discuss the problem of capturing customers in retail service provision.

7.2 Maximum Capture

Hotelling (1929) considered two competitors, each selling water from their own artesian wells along a linear market. In his highly constrained model of a game between the two competitors, he demonstrated that when each of the competitors maximized their own profits, an equilibrium would eventually be reached at which the linear market between the two water vendors would be equally divided with each customer visiting their closest market.[1] ReVelle (1986) envisioned a similar setting, but defined it on a network of nodes and arcs instead of a linear market. It was assumed that some of the nodes were places of potential retail location and other nodes were points of demand. Consider the situation that there are already one or more firms that offer the same product/service with the same price across this network. Presumably, the competing facilities have already divided the market with customers patronizing their closest facility.[2] With respect to this network of

[1]Hotelling (1929) suggested that if the two vendors could change location, they would eventually reach a price equilibrium while locating close together at the center of the linear market. Here they would attract an equal share of the customers. However, d'Aspremont et al. (1979) has since shown that a price equilibrium does not hold when the vendors are sufficiently close together.

[2]As in ReVelle (1986), we assume here that customers see no difference in price or offerings between competitors, so they patronize their closest facility. This may not be the case when making multi-purpose trips, like a person stopping off at a facility on their way to work, or on their way home.

demands, existing competing firms, and customer patronage patterns, ReVelle (1986) posed the following problem:

> If a retail operation plans to invest in one of more new facilities, where should the firm locate those facilities in order to maximize new customer gains?

When any new facility opens, the patronage patterns change. ReVelle (1986) suggested that, even if one lacks a reliable model of shopping behavior, a firm would still need some metric to measure the impact of locating any new facilities. He suggested the "maximum capture or maximum sphere of influence metric". Essentially, a demand node was considered to be captured from a competitor if a new facility was located closer to that demand than any existing competitor's facility. That is, a site which captures a large number of customers from existing retailers would be viewed as providing those customers with a closer, more accessible, facility. ReVelle (1986) noted that if the firm already has some facilities, customers at their existing facilities should not be counted as being captured if any of their newer facilities are closer to these existing customers. That is, the number being captured are customers from other firms because they are now provided with a closer facility. It should be noted that this problem is related to the Condorcet location problem, in which the objective is to locate a number of facilities in such a manner that a majority of people are better off (Hansen and Thisse 1981). Given this criteria, a solution to the maximum capture problem when 50% or more of the people (in terms of closeness) are captured meets the Condorcet property.

Consider the following notation:

I = set of demand nodes that are available for capture
J = set of eligible facility sites
i = index of demand nodes
j = index of facility sites
J_O = set of currently occupied facility sites
J_N = set of sites not currently occupied, but are eligible to have a new facility
d_{ij} = shortest distance or travel time between demand node i and facility site j
S_i = distance from node i to its closest existing facility

$$N_i = \left\{ j \in J_N \mid d_{ij} < S_i \right\}$$

$$\tilde{N}_i = \left\{ j \in J_N \mid d_{ij} = S_i \right\}$$

p = number of facilities to be located
a_i = demand representing potential customers at node i

$$x_j = \begin{cases} 1, & \text{if a facility is located at node } j \\ 0, & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1, & \text{if demand } i \text{ is now served by a new facility that is closer than } S_i \\ 0, & \text{otherwise} \end{cases}$$

$$z_i = \begin{cases} 1, & \text{if demand } i \text{ is now served by a new facility exactly at the distance } S_i \\ 0, & \text{otherwise} \end{cases}$$

The coverage sets N_i and $\tilde{N}_i$ are worth a little more discussion given their significance here. N_i is the set of available facility sites that are strictly closer to demand i than its current closest facility. Differing from this, $\tilde{N}_i$ is the set of sites j (occupied or unoccupied) that are equally distant to the closest current facility to demand i. Building on this discussion, two situations arise for capturing new customers in a region. Siting a new facility represents a case where it is now closer to demand than any existing facility of a competitor. The variables y_i track this situation. The second case occurs when a new facility is equally close to demand as an existing facility of a competitor. The variables z_i track this situation. We can now define the maximum capture location problem (MAXCAP) as:

$$\text{MAXCAP}: \quad \textit{Maximize} \sum_{i \in I} a_i y_i + \sum_{i \in I} 0.5 \ a_i z_i \tag{7.1}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq y_i \quad \forall i \in I \tag{7.2}$$

$$\sum_{j \in \tilde{N}_i} x_j \geq z_i \quad \forall i \in I \tag{7.3}$$

$$y_i + z_i \quad \leq \quad 1 \quad \forall i \in I \tag{7.4}$$

$$\sum_{j \in J_N} x_j = p \tag{7.5}$$

$$x_j \in \{0, 1\} \quad \forall j \in J_N \tag{7.6}$$

$$y_i \in \{0, 1\} \quad \forall i \in I \tag{7.7}$$

$$\mathrm{z_i} \in \{0, 1\} \quad \forall i \in I \tag{7.8}$$

MAXCAP is structured as originally formulated by ReVelle (1986). The objective, (7.1), seeks to maximize the total capture of new customers. The objective counts two types of customers as being captured. The first corresponds to capturing all customers that are now closer to one of the newly located facilities than any existing facilities. This is tracked in the first term of the objective. Whenever a demand node i is counted as closer than any existing facility, then the variable $y_i = 1$ and its population, a_i, is included in the amount being captured. The second type of capture involves those demands that are now equidistant to an existing facility and a newly sited facility. For those demands that now find their new closest facility at the same distance as an existing facility, it is assumed that the existing facility and new facility will share those customers equally. Thus, when a demand is equidistant between their closest new facility and their closest existing facility, the amount considered captured is one-half of the demand. This reduced level of capture is accounted for in the second term of the objective, where $z_i = 1$ means a new facility has been placed at the exact same distance from demand i as an existing facility. Constraints (7.2) allow variable y_i to equal one in value whenever one of more new facilities have been located among the sites j that are closer to demand i (that is, sites in the set N_i). Constraints (7.3) allow variable z_i to be one in value whenever a facility is located at the same distance to node i as an existing competitor's facility (that is, sites in the set $\tilde{N}_i$). To prevent any double counting, constraint (7.4) restricts capture to be counted in at most one way, complete capture or shared capture. Obviously, if a demand is provided both types of capture, the highest valued capture (i.e., complete) will be selected. Constraints (7.5) indicate that p new facilities are to be sited. Constraints (7.6), (7.7) and (7.8) specify binary restrictions on decision variables.

This formulation of MAXCAP does have two exceptions that require a more nuanced structure. Anytime there is a new facility that is closer than an existing facility, the above model considers that demand as being completely captured, regardless of the locations of other new facilities. This fits the definition of capture offered by ReVelle (1986). The above model also handles the case where each demand has a unique closest existing facility and where at most one new facility is located that is equidistant to that customer. In this case, the demand would be considered to be equally shared between the new and existing facilities, unless, of course, a different new facility is located even closer to capture all demand.

If a demand has two existing facilities the same distance away and the closest new facility is also located at that same distance, then the market needs to be shared three ways as the new facility would only capture a third of the demand not half. This is a nuance that can easily be added into the model objective. Following this logic, in the unlikely event that a demand has three existing closest facilities, and a new facility is located at that exact same distance, then the amount captured would be a fourth of the demand. This too, can easily be added to the objective by modifying captured fractions as appropriate to the second term of (7.1).

The problem becomes a bit more complex, however, when two new facilities are located equidistant from a demand that is the same distance away from its closest existing facility. The above model would consider the capture to be a half, but really the two new facilities and the existing facility would represent a split of demand into thirds, yielding a capture of two-thirds and not a half. Such a circumstance requires further model modification. Consider new variables:

$$w_i = \begin{cases} 1, & \text{if demand } i \text{ has a second new facility that is exactly at the distance } S_i \\ 0, & \text{otherwise} \end{cases}$$

These new variables will be one only when two new facilities have been located equidistant to demand i and where that distance equals S_i, the closest distance of an existing facility. We can modify the objective as follows:

$$Maximize \sum_{i \in I} a_i y_i + \sum_{i \in I} 0.5 \ a_i z_i + \sum_{i \in I} 0.166 \ a_i w_i \tag{7.9}$$

The basic idea is that a second, new, equidistant facility will add to the half already captured, raising the total to two-thirds, or an additional one-sixth (0.166). We also need to modify constraint (7.3) in the following manner:

$$\sum_{j \in \tilde{N}_i} x_j \geq z_i + w_\mathrm{i} \quad \forall i \in I \tag{7.10}$$

as well as introduce a new constraint:

$$y_i + w_i \leq 1 \quad \forall i \in I \tag{7.11}$$

These additions would also require constraints for enforcing the binary restrictions on new variables, w_i. The idea is that if two new facilities are located equidistant to a demand that is at that same distance to an existing facility, then the associated variable w_i would be equal to one. Constraints (7.10) will allow both w_i and z_i to be one in value when two equal distant facilities to demand i are located at the same distance as an existing facility serving this demand. If that is the case, the objective will count two-thirds of the demand as having been captured. If one new facility is located that is the same distance to demand i as the closest existing facility, then this constraint will allow only one of the two variables, w_i or z_i, to be one in value. In that case, the objective will force z_i to be one as it represents an added share of a half, and not a sixth. Finally, constraints (7.11) prevent w_i from being one in value if some other new facility is located even closer, where the demand is fully captured (i.e., $y_i = 1$).[3]

Chapter 6 introduced several models that were based upon expanding the definition of coverage. Rather than using a simple metric, say service within a maximal standard S, to define if coverage has been provided or not, Chap. 6 presented the GMCLP (general maximal covering location problem) that involved several service standards and valued the coverage provided differently for each standard. Chapter 6 also demonstrated how these models could be formulated as equivalent p-median problems, just as was done for the LSCP and MCLP in Chap. 2. ReVelle (1986) made the same connection for MAXCAP, providing a mechanism for using a

[3]This nuance is formulated differently than in ReVelle's original maximum capture paper, as the constraints proposed by ReVelle (1986) actually force demand nodes where there could be multiple shares of capture to have at least one new facility located at that distance, clearly an unintended feature.

transformed distance matrix in the p-median in order to structure MAXCAP.[4] The above model has been expanded to include the feature that some facilities may already exist and the extended form of MAXCAP optimizes capture while some of the existing facilities may be relocated, some may be kept, and others may be added (ReVelle and Serra 1991). Serra et al. (1999) and Drezner et al. (2002) suggest that each facility needs to have a minimum level of customers to be viable, a subject that is discussed later in this chapter. Other extensions have included a hierarchical set of facilities (Serra et al. 1992), uncertainty of demand (Serra et al. 1996), chance constraints on capture (Colomé et al. 2003), a weighted network (Eiselt and Laporte 1989), and a form of the problem involving shrinking a facility chain and ceding as few customers as possible in the process (ReVelle et al. 2007).

Given that we have formulated a model that involves a firm entering a market with an established competitor, a logical question might be: what if there were two firms, a leading firm (A) and a follower firm (B), where the leader firm moves into the market first and the follower firm makes site choices after the leader firm has made theirs? This is exactly the problem posed by Serra and ReVelle (1994). Although they recognized that this problem was a Stackelberg game, their model was that of the leader alone, which is a version of MAXCAP given above. They recognized shortcomings in their model, however, and proposed two heuristics that simulated the course of two firms making decisions and where firm A would site in response to the choices of firm B, providing a mechanism to examine resulting market share. The idea is that after firm A makes site choices of where to locate, Serra and ReVelle (1994) would then apply MAXCAP to see what firm B will do in response. Then, given firm B's selection, they would test possible moves by firm A one at a time associated with each facility in order to assess the likely response of firm B. The idea is to keep those moves that firm A makes where the move represents an improvement in the market share for firm A over the previous locational choice as evaluated after firm B has made their choices. The interest, of course, is to uncover the best strategy for firm A. This type of heuristic (moving facilities one at a time to seek out better locations) has proven relatively robust, but has not been tested in any complete way involving a game between two competitors. It is important to recognize that competitive games, like the above description involving two firms, is a complex problem that cannot be formulated as a single level optimization problem. Rather, it has been approached as a bi-level problem. Further discussion is beyond the scope of this chapter, and remains a promising area for future research, although several bi-level models are described in Chap. 9 for a different type of covering problem.

[4] Note that this equivalent form of the p-median problem cannot handle the case when more than one new facility is placed at the same distance of a competitors involving a given customer.

7.3 Capturing/Intercepting Flow

Most of this book involves cases where demand for service coverage is represented by a set of discrete points or encompassed by a continuous bounded area. However, these two situations do not necessarily address all types of demand, like that for a quick service restaurant, a gas station or a commercial vehicle enforcement facility (truck scale). Much of the potential demand for a restaurant may be reflected by discrete points, but there is another type of demand or target audience that cannot be represented by points or areas. This type of demand has been called "flow" or "traffic" based. Retail location experts are always interested in the amount of traffic, or flow, that streets experience when making site selection choices. Even though many firms might make location decisions based upon the population within a primary shopping zone of 5–7 miles, as an example, firms are also interested in those sites that experience high traffic volumes. That is, the quality of a potential facility site can be a function of both nearby demand and volume of traffic that happens to pass by. Traffic volume is not only of interest for retail site location but is also of value to transportation planners.

Traffic along a road segment over a day or a week can be can viewed as the sum of a number of trips being made that traverse that segment. In order to forecast those areas that will experience traffic congestion, and the need for road investment, better traffic flow control, etc., transportation planners estimate the travel demand that will occur between two locations. It is not uncommon for these locations to be transportation analysis zones, an official reporting unit by the US Census. The interest is in the traffic that originates in one area that is destined for another area. Given a network of nodes and arcs representing major streets/roads that connect two locations, daily traffic volumes indicate some level of interaction between an origin and a destination. Such interaction follows particular paths through the network. It is precisely this interaction that may be of value in facility siting.

Consider the following notation:

i = index of network nodes representing origins (entire set denoted I)
k = index of network nodes representing destinations (entire set denoted K)

$$Q = \{(i,k)|i \in I, k \in K\}$$

q = index of origin-destination pairs
t_q = volume of traffic that occurs between origin-destination pair q
j = index of potential facility location (entire set denoted J)
p = number of facilities to be sited
N_q = set of locations j that can intercept or capture flow of origin-destination pair q

$$\delta_{qj} = \begin{cases} 1, & \text{if } j \in N_q \\ 0, & \text{otherwise} \end{cases}$$

The set Q represents all origin-destination pairs, and through analysis we can determine which potential facility sites j are capable of capturing/intercepting a given origin-destination q pair. We call the set of sites that can capture origin-destination pair q set N_q. This information can then be used to specify an indicator coefficient, δ_{qj}, summarizing whether or not a potential facility j captures/intercepts the flow of origin-destination q pair.

Let's say that we wish to locate a quick service restaurant specializing in hamburgers and we want a site that has the largest number of people traveling passed it during the week. To do this we would have to identify the site j that has the largest traffic volume. This can be calculated as follows for each potential site j:

$$\sum_{q \in Q} \delta_{qj} t_q \tag{7.12}$$

The best site j would be the one with the highest traffic flow, determined by evaluating (7.12) in each case. But, if we wish to locate two or more hamburger joints having the highest number of people that pass by at least one of the restaurants each day, then the problem becomes a bit more complicated. For example, picking the site that experiences the second highest volume of traffic, using equation (7.12), may well identify a site with considerable flow in common with the highest volume site. In this case, the second joint will actually cannibalize sales from the first chosen site. As the goal is to find a site with the most total traffic flow, we want to keep cannibalization to a minimum. Consider the following additional notation:

$$x_j = \begin{cases} 1, & \text{if a facility located at site } j \\ 0, & \text{otherwise} \end{cases}$$

$$y_q = \begin{cases} 1, & \text{if a flow volume generated by origin-destinationn pair } q \text{ is captured} \\ 0, & \text{otherwise} \end{cases}$$

Hodgson (1990) structured a model to identify the best locations for capturing origin-destination flow along the lines described above. This problem is called the Flow Capture Location Model (FCLM) and formulated as follows:

$$\text{FCLM : } Maximize \sum_{q \in Q} t_q y_q \tag{7.13}$$

Subject to:

$$\sum_{j \in N_q} x_j \geq y_q \quad \forall q \tag{7.14}$$

$$\sum_{j \in J} x_j = p \tag{7.15}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.16}$$

$$y_q \in \{0, 1\} \quad q \in Q \tag{7.17}$$

The variable y_q is used to identify whether a given flow volume for origin-destination pair q passes by one or more of the selected facility sites. Even if a volume of flow generated by a given origin-destination pair passes by more than one of the selected facility locations, flow capture is counted only once. This means that the objective (7.13) represents the total traffic volume captured one or more times, so there will be no double counting of any flows. Thus, once captured there is no incentive to locate any facilities to intercept that same origin-destination flow again. Constraints (7.14) specify that flow q can be captured when at least one facility is located at a site capable of capture. When this occurs $y_q$will equal one in value. Otherwise the constraint will force y_q to be zero in value, indicating that flow q has not been captured. Constraint (7.15) specifies that exactly p facilities will be located. Constraints (7.16) and (7.17) impose binary restrictions on decision variables.

As discussed in previous chapters, only the location siting variables x_j technically need to be formally specified as binary when a general integer programming solver is used as the y_q variables will take only zero–one values as long as these variables are bounded to be no larger than 1 in value. One should also recognize the similarity of this model to that of the maximal covering location problem detailed in Chap. 2. In fact, it is exactly the same except the form has been defined so that it represents the problem of capture/intercepting flows instead of covering demand.

The FCLM can be used for a number of different applications, including the selection of billboards in an advertising campaign and the location of truck weighing stations as well as siting (or expanding or contracting) a chain of retail establishments. In each of these cases, the problem is to capture or intercept the maximum amount of flow. Interesting forms of this problem that have been developed and includes the capture of people who might evade a system of intercepting facilities like inspection stations (Marković et al. 2015), paths (Gutiérrez-Jarpa et al. 2010), alternative fuel stations (Hong and Kuby 2016), and a generalized form that addresses many possible nuances (Zeng et al. 2010).

7.4 Capacities

When solving for a configuration of facilities to cover an area (represented by points or other spatial objects), the workload of individual facilities is often overlooked. In many cases, this is a realistic approach. Consider, for example, a system of sirens designed to warn residents of potential danger (e.g., severe storms, wildfires, tornados, etc.). When locating such a system to cover an area, the main issue is to ensure spatial coverage so that all can hear a siren. Covering everything spatially as efficient as possible is the primary concern. The number of people served by an individual siren is not limited. The same can be true for other coverage based service systems, like sensors, cameras, beacons, etc. But, there are systems where the major concern is to seek high levels of coverage, where service at individual facilities can be congested; that is, there may be some limit on how many people can be served by an individual facility. This means that facilities may experience capacity limitations, and those capacities should not be exceeded at each located facility. To ensure that capacities will not be exceeded, explicit tracking of demand service by each facility is necessary. Further, ensuring that the total demand assigned to a given facility does not exceed its capacity is also required. The fact that we need to track or assign individual demands to facilities means that any adopted approach must rest on an allocation process. Accordingly, a covering location-allocation approach is necessary for addressing capacity issues, in contrast to many of the models in this book that account for coverage in a manner that is not explicit in terms demand service assignment. For example, the LSCP involves finding the smallest number of facilities needed to cover every demand, but there is no accounting for which facility is providing service, just that all demand is covered. In fact, some demand will be covered by many facilities while others will be uniquely covered by only one facility. Although we know that every demand is covered, it is not possible to know the amount of demand that is to be served at a facility without adding allocation variables. This is also true for the MCLP. To illustrate the complications that can arise, consider Fig. 7.1 showing a solution derived using the MCLP ($p = 13$). The facilities in this case serve households, and summarized in Fig. 7.1 is the total demand allocated to each sited facility. Of the 9116 households served by this solution, the workload for each facility varies considerably. This ranges from a low of 245 to a high of 1152. There is much variability from the average of 701, and as a result significant inequities across facilities may be experienced along with the potential to exceed facility capabilities (e.g., workload of 1152) as well as

Fig. 7.1 Facility sites identified using MCLP with closest assignment workloads indicated

unreasonably low demand for service (e.g., workload of 245) in some cases. Variability can be addressed through the imposition of capacities for each facility. The early efforts to address capacity issues in coverage modeling include the works of Chung et al. (1983), Smogy and Church (1985), Chung (1986), Current and Storbeck (1988), and Pirkul and Schilling (1989, 1991).

When introducing capacities into a coverage problem, an allocation process must be conceived. The way in which we might allocate demand is dependent upon our assumptions about the service being located as well as the amount of information concerning demand behavior. There are three classical ways to define how demand might be allocated amongst facilities: closest assignment (user optimal), equal assignment probability (based upon the principle of insufficient reason), and system convenience (system optimal). These three demand allocation/assignment approaches hinge on the definition of a system's operational characteristics. Let's say that a system attempts to provide service to a demand by responding to a call for service, like "I want a pizza in 30 minutes". If a system of pizza restaurants covers potential demand across a city within 30 min, and there are overlapping facility service areas, then a central call center can coordinate the response for service, allocating the order to a particular pizza restaurant for preparation, baking and delivery. Assuming that several outlets can satisfy the 30 min standard, then the call center can make the choice as to which of those outlets will do the task, likely picking the facility that has idle capacity. That is, the system makes the choice and not the individual customer with the intent of making the system operate as efficiently as possible. This is a system optimal assignment. If users, on the other hand, make the choice to call a particular facility for their pizza, or decide to travel to the facility to order and eat their pizza, then each individual user makes their choice based upon what maximizes their utility best, not the system. This represents a user optimal approach. A user optimal approach requires some type of choice formula, or utility function, which can be used to allocate demand to individual facilities. Finally, what if we do not know what each individual will do or how they will make their choice regarding facility assignment coverage. If we have no prior knowledge about allocation, then any choice set of alternatives for a customer will be probabilistic. That is, for three facilities covering a given individual demand, as an example, then the individual is assumed to patronize each facility one third of the time. This approach is based upon the work of Bernouli and Laplace (Dupont 1977/78) and is called the principle of insufficient reason. Bernouli, and later Laplace, suggested that if there was no reason to think that one alternative/event was preferred over another, then each choice or event will occur equally likely, hence the equal probabilities of choice in assigning demand to individual facilities in their coverage set. The remainder of this section discusses coverage models for each of these three approaches.

7.4.1 System Optimal Perspective

We begin with the system optimal approach, where the system decides which facility will serve a given demand and where all demands must be covered. Most of the

literature has adopted this form of allocation when adding capacity restrictions to covering facilities. Consider the following notation:
i = index of demand points/areas/objects (entire set denoted I)
j = index of potential facility sites (entire set denoted J)
S = desired maximal service standard (travel distance or time)
d_{ij} = shortest distance or travel time between demand i and potential facility j

$$N_i = \{j | d_{ij} \leq S\}$$

$$\Psi_j = \{i | d_{ij} \leq S\}$$

p = number of facilities to be located
a_i = amount of demand at i

$$x_j = \begin{cases} 1, & \text{if a facility is located at } j \\ 0, & \text{otherwise} \end{cases}$$

C_j = capacity of potential facility j
z_{ij} = fraction of demand i that is assigned to facility j

The introduction of the z_{ij} variables now enable allocation and tracking of service. Accordingly, a capacitated version of the LSCP is possible, as done in Current and Storbeck (1988). Here we formulate the system-optimal perspective of the capacitated location set covering problem—system optimal (CLSCP-SO) as:

$$\text{CLSCP-SO} : \textit{Minimize} \sum_{j \in J} x_j \tag{7.18}$$

Subject to:

$$\sum_{j \in N_i} z_{ij} = 1 \quad \forall i \in I \tag{7.19}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.20}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.21}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.22}$$

The CLSCP-SO involves locating just enough facilities and associated capacity such that all demand is served within the capacity limits of each facility, given the coverage capabilities of each facility. Demand allocation is done using variables z_{ij}, indicating the fraction of demand i served by facility j. The objective, (7.18), is equivalent to the original LSCP detailed in Chap. 2. Constraints (7.19) require that the sum of the fractional assignments of demand i add up to one, which means that 100% of that demand will be allocated within the coverage standard (to one or perhaps several facilities). Of course, all assignments must be within a maximal service standard of demand i as well. Constraints (7.20) require that the assigned demand to each facility j cannot exceed its established capacity, C_j. Observe that if a site j is not chosen for a facility, then $x_j = 0$ and the effective capacity on the right hand side of constraint (7.20) will equal zero (the product of zero times C_j). This will restrict the sum of assignments to that site to zero, effectively ensuring that allocations are made only to those sites that are selected for facilities. This model assigns demand at the convenience of the system, such that allocations all fit within the capacity, even though demand may be served by facilities that are not their closest or their preferred choice. Constraints (7.21) impose binary integer restrictions on facility siting variables. Constraints (7.22) indicate non-negativity conditions on assignment variables, which means that demand can be split and portions may be assigned to different facilities.

It should come as little surprise that the CLSCP-SO can be cast in a form that maximizes coverage. The system optimal form of the CMCLP (capacitated maximal covering location problem—system optimal, CMCLP-SO) was first detailed in Chung et al. (1983) and Smogy and Church (1985). The formulation is as follows:

$$\text{CMCLP-SO} : \textit{Maximize} \sum_{i \in I} \sum_{j \in N_i} a_i z_{ij} \tag{7.23}$$

Subject to:

$$\sum_{j \in J} z_{ij} = 1 \quad \forall i \in I \tag{7.24}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.25}$$

$$\sum_{j \in J} x_j = p \tag{7.26}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.27}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.28}$$

The objective, (7.23), is to maximize total demand covered within the desired maximum service standard. Constraints (7.24) ensure that each demand fully assigns for service. Note that allocation of demand to a facility may be beyond the stipulated service standard. The rationale is that demand beyond the standard may still seek out service, and as a result needs to be served. Constraints (7.25) impose capacities on facilities serving demand. Constraint (7.26) stipulates that p facilities are to be sited. Finally, binary restrictions are indicated for facility variables, constraints (7.27), and non-negativity is required on allocation variables, constraints (7.28).

An interesting distinction in the CMCLP-SO formulation is that there is no service standard limit in the assignment of demand. Demand may well be served beyond the maximum desired service standard, but the goal is in fact to maximize demand served within the standard. This goal is reflected in the objective, (7.23).

One may well be tempted to restrict service capacity to only those within the service standard. Such a constraint form would be as follows:

$$\sum_{i \in \Psi_j} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.29}$$

Unfortunately, the use of constraints (7.29) would ignore actual service demand within the system for those beyond the service standard, yet still obtaining service from the system. Such individuals remain likely and able to be served, just not within the desired standard.

The CMCLP-SO here is a restricted form of that developed by Smogy and Church (1985), but is equivalent to that of Chung et al. (1983) (see also Chung 1986). In Chung (1986) there is a suggestion to add the following Balinski type constraints:

$$z_{ij} \leq x_j \quad \forall i \in I, j \in N_i \tag{7.30}$$

The addition of constraints (7.30) in the CMCLP-SO helps to encourage integer solutions in linear programming based techniques, such as branch and bound. This would be considered a *tighter* LP form for this problem. The literature on the use of Balinski constraints in location-allocation models is quite extensive. Such constraints can improve the performance of general purpose software in solving many types of location-allocation models, especially those with similar structure to that of the p-median problem (mentioned in Chap. 2).

The extended model formulated in Smogy and Church (1985) was designed to allocate health personnel, but also ensure service was limited to a manageable number of patients. The approach made it possible to locate more than one facility (heath professional) at a given site through the use of an x_j variable that was allowed to be any positive integer in value.

7.4.2 User Optimal Perspective

When operating a set of facilities, like retail or many public services, people will decide exactly which facility they will visit. That is, rather than the system dispatching service to the customer/demand, like pizza delivery and or EMS response, the customer goes to the facility. In the former case, the system can decide which facility will respond to serve each demand, but in the latter case the user makes the decision as to which facility they will attend. In many cases that will boil down to the demand going to their closest facility. This is often observed for facilities like public libraries, post offices, etc. If we assume a closest assignment paradigm, we need to ensure that the system has enough capacity at each facility to serve all demand while assuming each demand will go to their closest facility. This is an allocation process called closest assignment.

For any facility j, all sites j' that are as close of closer to demand i can be identified. Formally, we define:

$$\Omega_{ij} = \left\{ j' \middle| d_{ij'} \leq d_{ij} \right\}$$

This then is the set of potential facility sites as close or closer to demand i, and includes facility j. Using this notation, we can define the capacitated LSCP with closest assignment (CLSCP-CA) as follows:

$$\text{CLSCP-CA} : \textit{Minimize} \sum_{j \in J} x_j \tag{7.31}$$

Subject to:

$$\sum_{j \in N_i} z_{ij} = 1 \quad \forall i \in I \tag{7.32}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.33}$$

$$z_{ij} \leq x_j \quad \forall i \in I, j \in N_i \tag{7.34}$$

$$\sum_{j' \in \Omega_{ij}} z_{ij'} \geq x_j \quad \forall i \in I, j \in N_i \tag{7.35}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.36}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.37}$$

The CLSCP-CA differs from the CLSCP-SO because of the use of closest assignment constraints (7.34) and (7.35). These constraints require that demand be allocated/served by its closest sited facility. Constraints (7.35) ensure that each demand wholly assigns to their closest facility, or in the case where a demand has two or more equidistant closest facilities, the sum of that demand's assignment across the set of equally close facilities must be one. This may not handle all of the issues of closest assignment raised in Gerrard and Church (1996) as they demonstrate how demand can be forced to equally share their assignment across a set of equally close facilities, in addition to forcing closest assignment.

As before, we can also formulate a closest assignment form of the user optimality perspective in addressing capacitated maximal coverage. Here the problem involves maximizing coverage (that is, service within some desired standard of service) while locating p facilities. An expectation is that demand would go to their closest facility, even when they are not covered. That is, just because a demand is not covered does not mean that the demand would not go to their closest facility when they needed service. Because of this, sufficient capacity must exist within the system to serve all demand, however some demand are served within the maximum coverage standard and some are not.

The closest assignment form of the capacitated MCLP (capacitated maximal covering location problem with closest assignment, CMCLP-CA) is as follows:

$$\text{CMCLP-CA}: \quad \textit{Maximize} \sum_{i \in I} \sum_{j \in N_i} a_i z_{ij} \tag{7.38}$$

Subject to:

$$\sum_{j \in J} z_{ij} = 1 \quad \forall i \in I \tag{7.39}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.40}$$

$$\sum_{j \in J} x_j = p \tag{7.41}$$

$$z_{ij} \leq x_j \quad \forall i \in I, j \in J \tag{7.42}$$

$$\sum_{j' \in \Omega_{ij}} z_{ij'} \geq x_j \quad \forall i \in I, j \in J \tag{7.43}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.44}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.45}$$

The objective, (7.38), is to maximize total demand covered within the desired maximum service standard. Constraints (7.39) require demand to be completely allocated to facilities. As was the case for the CMCLP-SO, allocation of demand to a facility may be beyond the stipulated service standard. Constraints (7.40) impose capacities on facilities serving demand. Constraint (7.41) stipulate that p facilities are to be sited. The major distinction of the CMCLP-CA is the inclusion of closest assignment constraints (7.42) and (7.43). Constraints (7.43) in particular require that each demand assign to their closest facility, or facilities in situations where they may be equidistant. Again, one caveat is that particular cases may arise where (7.42) and (7.43) insufficient (see Gerrard and Church 1996). Binary restrictions are indicated for facility variable, constraints (7.27), and non-negativity is required on allocation variables, constraints (7.28).

Little research has actually focused on user optimal situations in coverage modeling. A notable exception is the work of Gerrard (1996), with the CLSCP-CA and CMCLP-CA inspired by his research. It should also be pointed out that there are other forms of user assignment that can be used as well when crafting a user optimal location covering problem. Potential examples include the incorporation of assignments through the gravity model and utility based multinomial logit models. Such models, for all practical purposes, have yet to be developed.

7.4.3 Equal Fraction Perspective

The third perspective noted for allocation of demand is that of equal probability when there is insufficient information on demand preferences. The is referred to here as an equal assignment fraction approach because without additional information, one must assume that probabilistically there is an equal chance among potential choices. The idea is that if there are several facilities that can serve a given demand (within some market range), and if we lack information regarding preferences to make allocations to these facilities, then we are left with a situation where assignments are equally probable. That is, if a demand location is served within the range of three facilities, then assuming that one-third will go to one of the facilities, a third to the next and a final third to the last of the three facilities is the only reasonable conclusion. As mentioned above, the fractions of assignment will be equal based upon the principle of insufficient reason first raised by Bernoulli in the 1600s. It has also been called the principle of indifference and the equal distribution of ignorance (Dembski and Marks 2009). The model that we present below is inspired by the work of Balakrishnan and Storbeck (1991), although their work will be discussed later in the chapter.

The nature of capacitated location models, in general, requires a set of allocation variables, in addition to variables that represent the selection of specific sites for facilities. These two components represent the very heart of facility location-allocation modeling. The models in the previous two sections used allocation variables, z_{ij}, representing each possible assignment of a demand i to a potential facility j. Somewhat unique in the equal assignment case is that specific allocations of demand to specific facilities are not necessary. For instance, say demand i is served within the service standard range S by three different facilities. If we assume that demand i will be equally split between the three "covering" facilities, then we can determine and calculate the portions of demand i assigning to each individual facility by just knowing how many times demand i is covered. Thus, consider the following type of decision variable:

$$y_i^k = \begin{cases} 1, & \text{if demand } i \text{ is served (or covered) by exactly } k \text{ facilities} \\ 0, & \text{otherwise} \end{cases}$$

In the case of location set covering, we know each demand would be served at least once, but it is possible to be served by all sited facilities. Thus, the number of options is the set K, where $K = \{1, 2, 3, \ldots, k^{\max}\}$. Given this, the contribution of demand i for a given facility j providing service coverage is:

$$\sum_{k \in K} \left(\frac{1}{k}\right) a_i y_i^k \tag{7.46}$$

where the fraction of $^1/_k$ represents the portion of demand a_i that will be assigned to facility j given that there are exactly k facilities providing service to demand i. From this basis, we can now formulate an equal assignment form of the capacitated LSCP (capacitated location set covering problem with equal assignment, CLSCP-EA) as:

$$\text{CLSCP-EA}: \quad \textit{Minimize} \sum_{j \in J} x_j \tag{7.47}$$

Subject to:

$$\sum_{j \in N_i} x_j - \sum_{k \in K} k \ y_i^k = 0 \quad \forall i \in I \tag{7.48}$$

$$\sum_{k \in K} y_i^k = 1 \quad \forall i \in I \tag{7.49}$$

$$\sum_{i \in \Psi_j} \sum_{k \in K} \left(\frac{1}{k}\right) a_i y_i^k \leq C_j x_j \quad \forall j \in J \tag{7.50}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.51}$$

$$y_i^k \in \{0, 1\} \quad \forall i \in I, k \in K \tag{7.52}$$

Objective (7.47) is the same as the classic form of the LSCP, where the number of facilities providing service is minimized. Constraints (7.48) and (7.49) together define feasible values of y_i^k. Constraints (7.49) indicate that exactly one y_i^k variable will be equal to 1 for each demand i. Given this, constraints (7.48) account for the number of facilities that cover demand i exactly k times. With the definition of the set K involving only positive integers greater than zero, constraints (7.48) do not allow a given demand to be left uncovered. That is, at the bare minimum, demand i must be covered at least once. Constraints (7.50) ensure that the assigned demand to each facility does not exceed its capacity. Constraints (7.51) and (7.52) impose binary integer requirements.

An extension to address the option of equal assignment for the capacitated form of the MCLP (capacitated maximal covering location problem with equal assignment, CMCLP-EA) is as follows:

$$\text{CMCLP-EA}: \quad \textit{Minimize} \sum_{i \in I} a_i y_i^0 \tag{7.53}$$

Subject to:

$$\sum_{j \in N_i} x_j - \sum_{k=1}^{p} k \ y_i^k + y_i^0 = 0 \quad \forall i \in I \tag{7.54}$$

$$\sum_{k=0}^{p} y_i^k = 1 \quad \forall i \in I \tag{7.55}$$

$$\sum_{j \in J} x_j = p \tag{7.56}$$

$$\sum_{i \in I} \sum_{k=1}^{p} \left(\frac{1}{k}\right) a_i y_i^k \leq C_j x_j \quad \forall j \in J \tag{7.57}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.58}$$

$$y_i^k \in \{0, 1\} \quad \forall i \in I, k \in \{0, 1, 2, \ldots, p\} \tag{7.59}$$

The CMCLP-EA differs in subtle but important ways than the CLSCP-EA, but also the other CMCLP formulations. First, objective (7.47) now minimizes the total amount of demand not covered or served. This is equivalent to maximizing demand covered, a distinction highlighted in Chap. 2 for the MCLP. Second, constraints (7.54) have been extended to account for the case where no coverage is provided within the desired service standard. In particular, either one or more facilities are located so that a given demand i is covered or the associated y_i^0 variable is forced to equal 1. This option is also included in constraints (7.55). Third, there is constraint (7.56) specifying that exactly p facilities are to be sited. Constraints (7.57) establish capacity limits on sited facilities. In this case, however, only demand that is covered is allocated to the located facilities. This is in contrast to the assignment of all demand in the cases of CMCLP-SO and CMCLP-CA. Extension to account for the allocation of all demand in terms of capacity considerations is possible, but will require the use of assignment variables. We leave such a variant of the CMCLP-EA as a topic for future research. Binary integer requirements are imposed in constraints (7.58) and (7.59) for all decision variables.

7.5 Thresholds

The previous section was devoted to ensuring that located facilities would not be overwhelmed with demand when there are limits to what each facility can handle. We addressed three different methods of allocating demand to facilities and formulated coverage models, extensions of the LSCP and MCLP. Collectively, these capacitated models provide ways for helping to identify facility siting configurations that address a major issue in planning: functional capacity limits. But there is another issue that is important in planning a system, dealing with minimum service demand thresholds at each facility. As discussed in Chap. 1, Christaller (1933) developed a theory of central places, involving spatial patterns of service/retail centers across a region. He reasoned that the market area of each central place was limited to some distance that he called the "range", a distance beyond which people were unwilling to travel for shopping and services. The demand that was enclosed within the range distance represented the highest potential for a market center, or even an individual facility. He further suggested that each center needed some minimum amount of demand or customers in order to be profitable and stay in business. He called this the "threshold". Simply put, if the demand within the range of a site is not as large as the threshold for a given type of facility or service, then that site is not viable for that type of service. For all practical purposes, many activities require some threshold of service demand in order to be viable. This includes a restaurant, a dry cleaning business, a grocery store, etc. Location models, beyond those developed in central place theory, often ignore the need to address threshold issues. That does not mean that some approaches have not included constraints on facility thresholds. For example, the classic warehouse location model of Geoffrion and Bride (1978) includes capacity constraints as well as threshold constraints.

Perhaps the first covering model developed using threshold constraints was the regional solid waste planning model of Church (1980). This model was developed for the Tennessee Valley Authority (TVA), a federal agency. They wanted to carve up the region into solid waste planning areas or districts. Each district was to have a designated center, a place where solid waste recycling would be economically viable. The model that TVA employed divided the region of 201 counties into districts, by assigning counties to those locations selected as district centers. The location model sited centers and assigned counties to these centers based on optimizing three different objectives: minimize the weighted distances of possible waste haulage (to each designated center from the other counties in each district), maximize the amount of waste that could be economically transported to centers, and minimize the number of centers that would have an expected volume of waste that was less than 1000 tons per week. The first objective represents that of the p-median problem, the second objective reflects a maximal covering goal, and the third objective amounted to a desired threshold. Specifically, the TVA wanted as many viable recycling centers as possible, recognizing that a base level of activity (tonnage per week) was important. The threshold objective was a means for seeking this out within a multi-objective model. Solutions could be generated on the Pareto frontier based on all three objectives. The basic idea was twofold: one, covering models can often be paired with other location objectives resulting in a multi-objective location model; and two, it is possible to either require that all facilities in a solution meet a minimum threshold of business or service demand or the number of facilities can be minimized that do not have an expected demand that equals or exceeds a minimum threshold. For the remainder of this section, we will assume that thresholds should be enforced everywhere, rather than minimizing the number of facilities that do not meet a minimum threshold.

Three different approaches for demand assignment were explored in the previous section. They included system optimal, user optimal, and equal assignment. The former two approaches utilized capacity constraints (7.20) and (7.33) for the LSCP, and similarly imposed in the MCLP using constraints (7.25) and (7.40). Given this, it is easy to structure constraints that enforce minimum thresholds as:

$$\sum_{i \in I} a_i z_{ij} \geq T_j x_j \quad \forall j \in J \tag{7.60}$$

where T_j represents the minimum required threshold of service demand at facility j. One could readily add constraint (7.60) to the CLSCP-SO, CMCLP-SO, CLSCP-CA or CMCLP-CA models in order to address threshold requirements.

Capacity constraints (7.50) and (7.57) for the equal fraction approaches also suggest a straightforward structure for imposing thresholds:

$$\sum_{i \in I} \sum_{k=1}^{p} \left(\frac{1}{k}\right) a_i y_i^k \geq T_j x_j \quad \forall j \in J \tag{7.61}$$

Accordingly, the CLSCP-EA and CMCLP-EA could be extended by adding constraints (7.61) in order to address threshold requirements.

A variant of the CMCLP-EA involves substituting threshold constraints (7.61) for capacity constraints (7.57), yielding a form of the McTHRESH model of Balakrishnan and Storbeck (1991) developed for market location. Other covering based models that have involved threshold constraints include the works of Carreras and Serra (1999), Hong and Kuby (2016) and Drezner et al. (2002). Another interesting approach involving covering models and thresholds is that of Storbeck (1988, 1990) where theoretical central place patterns were analyzed. Storbeck (1988, 1990) was able to demonstrate that the central place patterns of Christaller (1933) could be generated using a protected threshold covering model applied to a uniform triangular lattice of demand and facility points.

When developing covering models for specific applications, both capacity and threshold constraints may be necessary. The issue at hand is that when such conditions are added to a covering problem, additional allocation variables or coverage service level variables are likely needed in order to track the exact demand loading on each sited facility. This often requires additional constraints as well. Altogether, the models are generally more difficult to solve using general purpose mixed integer programming software, possibly to the extent that exact solution is not possible. One of the reasons for this is that capacity and threshold constraints are a form of knapsack constraints. Knapsack-type constraints usually encourage fractional values among the allocation variables in a linear programming relaxation. This usually results in much larger branch and bound trees in the search for an optimal solution, and concomitantly longer solution times. As a result, development of specialized solution approaches has been necessary (see for example Carnes and Shmoys 2008).

7.6 Franchise Territory Design

Problems that deal with locating retail and service facilities, like a franchise operation (e.g., 7-Eleven, Carl's Jr., Dairy Queen, Subway, etc.), share a lot in common with Christaller's central place theory. In fact, several covering models have been developed to analyze central place patterns as well as locate franchise facilities. Although Christaller (1933) was principally interested in the location of villages, towns, and cities, where each of them had a variety of services and types of retail establishments, the same issues apply whether it is an individual facility being located or the emergence of a town or city.[5] We concentrate here on the issues of locating a set of facilities across an area, and in particular franchisee facilities. It is important to remember the concepts of threshold and range as they apply to a given type of business. The range represents the furthest extent to which a facility might capture or attract demand/customers and threshold represents the minimum amount of demand (or number of customers) such that an individual facility is viable/ profitable. One can think of the types of allocation (system optimal, user optimal, and equal assignment) as approaches to estimate the amount of demand that will be served by a facility. This will need to be as much or larger than the threshold for that type of business. In fact, we can structure models for multiple retail facility location using the constraints and variables already used. For example, the CMCLP-CA could be modified such that the capacity constraints, (7.40), are removed and threshold constraints (7.61) appended, giving us a retail chain facility location model. This would be an alternative to the McTHRESH model discussed above.

There is, however, an interesting problem that arises when retail establishments are franchisee owned and franchisor licensed. Wherever they are located, the franchisor would like to ensure that each facility maintains a threshold of demand. What is particularly interesting and important is that there are two perspectives, one of the franchisor who wants to locate and license as many viable franchisee facilities as possible, and the other perspective of the franchisee who hopes for a large market area and customer base with as few other facilities (or none) of the same franchisor in the region as possible. That is, the franchisor wants a lot of facilities in order to saturate a market area and the franchisee wants to see as few as possible. This problem was first raised by Current and Storbeck (1994), which they termed FVF (franchisor vs. franchisee).

[5]To be complete, Christaller (1933) suggests a hierarchy of centers and types of goods. For instance a village may have a grocery store, a hardware store, and a gas station, whereas a town will offer all of those goods as well as many other facilities, like clothing stores, automobile dealerships, and a hospital. Cities have an even larger set of services, including all of those offered in villages and towns, but in addition even "higher-ordered" goods and services. These include such businesses as medical specialists and high-end jewelry stores. It is important to note that Church and Bell (1990) have demonstrated that co-location of competitors can exist in stable market central place configurations, so that, depending on the nature of the business, competing firms will locate in the same area.

The FVF problem recognizes the tradeoff between the perspective of the franchisor wanting many facilities and the franchisee wanting as few as possible. Consider the following additional or modified notation:

$$\delta_{ij} = \begin{cases} 1, & \text{if demand } i \text{ is within the range of potential outlet location } j \\ 0, & \text{otherwise} \end{cases}$$

$$\theta_{ij} = \begin{cases} 1, & \text{if demand } i \text{ is within the threshold of potential outlet location } j \\ 0, & \text{otherwise} \end{cases}$$

$$x_j = \begin{cases} 1, & \text{if facility location } j \text{ is chosen for a franchisee outlet} \\ 0, & \text{otherwise} \end{cases}$$

$$r_i = \begin{cases} 1, & \text{if demand } i \text{ is not within the range of a facility} \\ 0, & \text{otherwise} \end{cases}$$

$$t_i = \begin{cases} 1, & \text{if demand } i \text{ is not within the threshold of a facility} \\ 0, & \text{otherwise} \end{cases}$$

e_i = number of additional facilities for which demand i is within the range
w_1, w_2, w_3 = importance weights for three different objective terms.

The Franchisor versus Franchisee (FVF) model can be structured as:

$$\text{FVF}: \quad \textit{Minimize} \quad -w_1 \sum_{j \in J} x_j + w_2 \sum_{i \in I} a_i t_i + w_3 \sum_{i \in I} a_i e_i \tag{7.62}$$

Subject to:

$$\sum_{j \in J} \delta_{ij} x_j - e_i + r_i = 1 \quad \forall i \in I \tag{7.63}$$

$$\sum_{j \in J} \theta_{ij} x_j + t_i = 1 \quad \forall i \in I \tag{7.64}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.65}$$

$$t_i \in \{0, 1\} \quad \forall i \in I \tag{7.66}$$

$$r_i \in \{0, 1\} \quad \forall i \in I \tag{7.67}$$

The objective, (7.62), of the FVF contains three terms, each weighted by an importance weight reflecting the relationship of one to the others. The first term is the franchisor's main goal of maximizing the number of outlets being located. The second term attempts to minimize the number of people who are not in the threshold of a facility. This, too, is an objective of the franchisor, who would like nothing better than having all demand served within the threshold of a facility. The third term of the objective minimizes the total demand that experiences multiple facilities outside the threshold distance but inside the range distance. This represents the objective of the franchisee, who wishes to keep competition within the market range to be at a minimum and potentially increase profits beyond the threshold of viability. Constraints (7.63) and (7.64) make use of coverage matrices, δ_{ij} and θ_{ij}, indicating demand within the range and threshold standards. They are derived in advance and represent spatial information and relationships structured in the model. Constraints (7.63) effectively track how many facilities are in the range of a demand. Similarly, constraints (7.64) account for whether a demand is part of the threshold demand of a facility. Notice that constraint (7.64) also ensures that demand i will not be within the threshold of two or more facilities, thus making the facility viable in terms of a minimum threshold of customers. But, if all customers are within a threshold of a facility, then the market will be saturated with viable facilities (the goal of a franchisor). One might consider the first two objectives as being complementary, with potential to omit the second objective. But, Current and Storbeck (1994) argue that the second term can be used to tease out alternate optima when trading off the first and third objectives, conceivably an important issue.

7.7 Summary and Concluding Comments

Many models have been applied in facility location by retail chains, and they are often considered proprietary. Although books such as *Location, Location, Location: how to select the best site for your business* by Salvaneschi (1996) provide many details on the issues of site choice, the overall approach that is suggested could be classified as a guided seat-of-the pants process, once potential trade or market demand has been determined within 10 or 15 min of travel time form a potential location. Guides such as this book and others often neglect to state exactly how to address the multiple facility location problem, where sales at one facility might be cannibalized by another. Such guides also often lack good methods to estimate facility revenue, given a potential customer base within certain travel times to a site.

Large companies today may have a GIS department where volumes of data on customer purchases can be analyzed within a spatial context. The type of data that is collected by a retail chain includes a wealth of customer information, even knowing who comes to a facility and uses a coupon cut from a newspaper or printed from an online website. This data can and is being used in sophisticated ways to develop better patronage models that earlier works such as those in Salvaneschi (1996) did not anticipate. These companies are using models to select multiple facility locations simultaneously, as compared to the one at a time approach discussed by Salvaneschi (1996). Hodgson et al. (1996) apply the FCLP in such an extended manner, utilizing a road network and associated origin-destination travel in Edmonton, Canada. Spaulding and Cromley (2007) detail application of MAXCAP integrated into a GIS framework.

Estimating how many customers will be attracted to a new facility as compared to existing competitor facilities has been a problem that many researchers have attempted to tackle with varying degrees of success. Reilly (1929) was one of the first to delineate market areas using a gravity model. Market boundaries were assumed to be those points at which half of all customers travel to one competitor and half travel to the other competitor. The gravity model, and other models like a multinomial logit model, can take into account the differences in facility size as well as travel costs and other amenities. The sophistication of modeling customer demand has advanced considerably in the last 10 years in the era of big data. Customer demand models are now based upon first dividing the population into different lifestyle profiles, e.g., young married couples. Each type of facility attracts different proportions of people based upon their profiles, e.g., affluent-retired. Thus, the capture of demand can be modeled with a more refined approach (Benati 1999). But, one can still view this as a customer capture problem, and the model structures reviewed here can include more refined estimates of what is being captured. It is also possible to extend the capture process from complete capture and exact sharing of customers with portioning schemes a company experiences with their competitors.

References

Balakrishnan PV, Storbeck JE (1991) McTHRESH: modeling maximum coverage with threshold constraints. Environ Plan B: Plan Des 18(4):459–472

Benati S (1999) The maximum capture problem with heterogeneous customers. Comput Oper Res 26(14):1351–1367

Carnes T, Shmoys D (2008) Primal-dual schema for capacitated covering problems. In: International conference on integer programming and combinatorial optimization. Springer, Berlin, pp 288–302

Carreras M, Serra D (1999) On optimal location with threshold requirements. Socio-Econ Plan Sci 33(2):91–103

Christaller W (1933) Central places in Southern Germany. The pioneer work in theoretical economic geography (trans: Carlisle Baskinin 1966). Prentice-Hall, Englewood Cliffs

Chung CH (1986) Recent applications of the maximal covering location planning (MCLP) model. J Oper Res Soc 37:735–746

Chung CH, Schilling DA, Carbone R (1983) The capacitated maximal covering problem: a heuristic. In: Proceedings of 14th annual Pittsburgh conference on modeling and simulation, pp 1423–1428

Church RL (1980) Developing solid waste planning regions for the Tennessee Valley Authority. In: Proceedings of the 11th annual Pittsburgh conference on modelling and simulation, vol 11, pp 1611–1618

Church RL, Bell TL (1990) Unpacking central place geometry I: single level theoretical k systems. Geogr Anal 22(2):95–115

Colomé R, Lourenço HR, Serra D (2003) A new chance-constrained maximum capture location problem. Ann Oper Res 122(1):121–139

Current JR, Storbeck JE (1988) Capacitated covering models. Environ Plan B: Plan Des 15 (2):153–163

Current JR, Storbeck JE (1994) A multiobjective approach to design franchise outlet networks. J Oper Res Soc 45(1):71–81

d'Aspremont C, Gabszewicz JJ, Thisse JF (1979) On Hotelling's "Stability in competition". Econ: J Econ Soc 1145–1150

Dembski WA, Marks RJ (2009) Bernoulli's principle of insufficient reason and conservation of information in computer search. In: IEEE international conference on systems, man and cybernetics, 2009. SMC 2009, IEEE, pp 2647–2652

Drezner T, Drezner Z, Shiode S (2002) A threshold-satisfying competitive location model. J Reg Sci 42(2):287–299

Dupont P (1977/78) Laplace and the Indifference Principle in the *Essai philosophique des probabilités.* Rend Sem Mat Univ Politec Torino 36:125–137

Dwyer FR, Evans JR (1981) A branch and bound algorithm for the list selection problem in direct mail advertising. Manage Sci 27(6):658–667

Eiselt HA, Laporte G (1989) The maximum capture problem in a weighted network. J Reg Sci 29 (3):433–439

Geoffrion A, Bride RM (1978) Lagrangean relaxation applied to capacitated facility location problems. AIIE Trans 10(1):40–47

Gerrard RA (1996) The location of service facilities using models sensitive to response distance, facility workload, and demand allocation. PhD Dissertation, University of California, Santa Barbara

Gerrard RA, Church RL (1996) Closest assignment constraints and location models: properties and structure. Locat Sci 4(4):251–270

Gutiérrez-Jarpa G, Donoso M, Obreque C, Marianov V (2010) Minimum cost path location for maximum traffic capture. *Computers & Industrial Engineering* 58(2):332–341

Hansen P, Thisse JF (1981) Outcomes of voting and planning: Condorcet, Weber and Rawls locations. J Public Econ 16(1):1–15

Hodgson MJ (1990) A flow-capturing location-allocation model. Geogr Anal 22(3):270–279

Hodgson MJ, Rosing KE, Leontien A, Storrier G (1996) Applying the flow-capturing location-allocation model to an authentic network: Edmonton, Canada. Eur J Oper Res 90(3):427–443

Hong S, Kuby M (2016) A threshold covering flow-based location model to build a critical mass of alternative-fuel stations. J Transp Geogr 56:128–137

Hotelling H (1929) Stability in competition. Econ J 39(1929):41–57

Marković N, Ryzhov I, Schonfeld P (2015) Evasive flow capture: optimal location of weigh-in-motion systems, tollbooths, and security checkpoints. Networks 65(1):22–42

Pirkul H, Schilling D (1989) The capacitated maximal covering location problem with backup service. Ann Oper Res 18(1):141–154

Pirkul H, Schilling DA (1991) The maximal covering location problem with capacities on total workload. Manage Sci 37(2):233–248

Reilly WJ (1929) Methods for study of retail relationships. Research Monograph, 4, Bureau of Business Research, The University of Texas, Austin

ReVelle C (1986) The maximum capture or "sphere of influence" location problem: hotelling revisited on a network. J Reg Sci 26(2):343–358

ReVelle C, Serra D (1991) The maximum capture problem including relocation. INFOR 29 (2):130–138

ReVelle C, Murray AT, Serra D (2007) Location models for ceding market share and shrinking services. Omega 35(5):533–540

Salvaneschi L (1996) Location, location, location: how to select the best site for your business. Oasis Press/PSI Research, Grants Pass

Serra D, ReVelle C (1994) Market capture by two competitors: the preemptive location problem. J Reg Sci 34(4):549–561

Serra D, Marianov V, ReVelle C (1992) The maximum-capture hierarchical location problem. Eur J Oper Res 62(3):363–371

Serra D, Ratick S, ReVelle C (1996) The maximum capture problem with uncertainty. Environ Plan B: Plan Des 23(1):49–59

Serra D, ReVelle C, Rosing K (1999) Surviving in a competitive spatial market: the threshold capture model. J Reg Sci 39(4):637–650

Smogy C, Church RL (1985) Balancing access and service cover. Paper presented at the North American Meetings of the Regional Science Association, Philadelphia, PA
Spaulding BD, Cromley RG (2007) Integrating the maximum capture problem into a GIS framework. J Geogr Syst 9(3):267–288
Storbeck JE (1988) The spatial structuring of central places. Geogr Anal 20(2):93–110
Storbeck JE (1990) Classical central places as protected thresholds. Geogr Anal 22(1):4–21
Zeng W, Castillo I, Hodgson MJ (2010) A generalized model for locating facilities on a network with flow-based demand. Netw Spat Econ 10(4):579–611

Chapter 8
Continuous Space Coverage

8.1 Introduction

An important distinction in location analysis and modeling has long been discrete versus continuous approaches. In previous chapters, for the most part, the reviewed coverage problems have been discrete in the sense that the places at which a facility may be sited are known and finite in number, and the demand locations to be served are also known and finite. This has enabled discrete integer programming formulations of models to be developed, allowing for efficient and exact solution in many cases. In some circumstances, however, neither potential facility sites nor demand locations are necessarily known and finite. Thus, one aspect of a continuous space location model is that facilities may be sited anywhere. An example is depicted in Fig. 8.1 where all locations within the region are feasible. The implication is that there are an infinite number of potential facility locations to be considered, in contrast to an assumed finite set of potential locations in discrete approaches (see Chap. 2). Another aspect of a continuous space problem is that demand too is not limited to a finite set of locations. Rather, demand is assumed to be continuously distributed across geographic space, varying over a study area. One example of this is shown in Fig. 8.2, where the height of the surface reflects demand for service. As is evident in the figure, some level of demand can be observed everywhere and this varies across space. Another example is given in Fig. 8.3. The Census unit color reflects the amount of demand in each area. Within a unit the demand varies in some manner, but given limited knowledge, a small geographic area and relative homogeneity, it is often thought that demand is uniformly distributed in the unit. Thus, Fig. 8.3 reflects discontinuities in demand variability, but it remains varying across space. Irrespective of representation, the implication is that demand for service is everywhere, and in some cases can possibly be defined/described by a mathematical function. From a practical standpoint, the study area could be a demand region, collection of areal units, set of line segments, or any group of spatial objects. This clearly makes dealing with demand and its geographic variability particularly

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_8

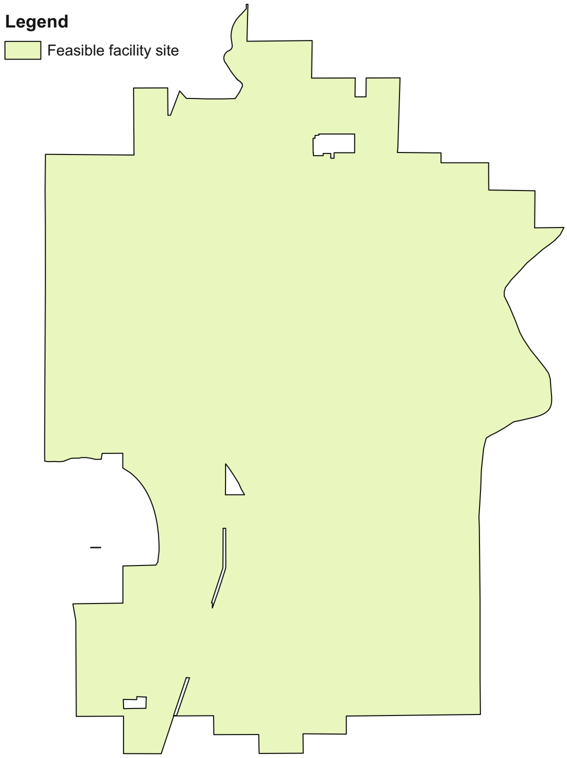


Fig. 8.1 Potential locations to site facilities

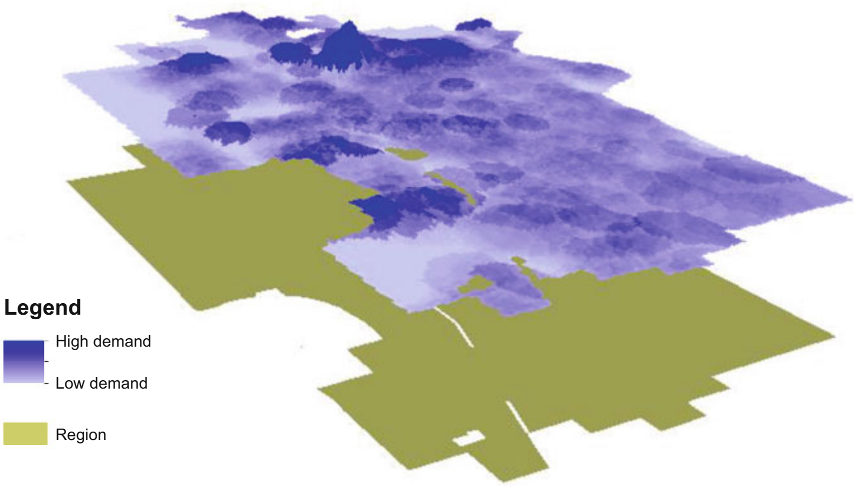


Fig. 8.2 Continuously distributed demand

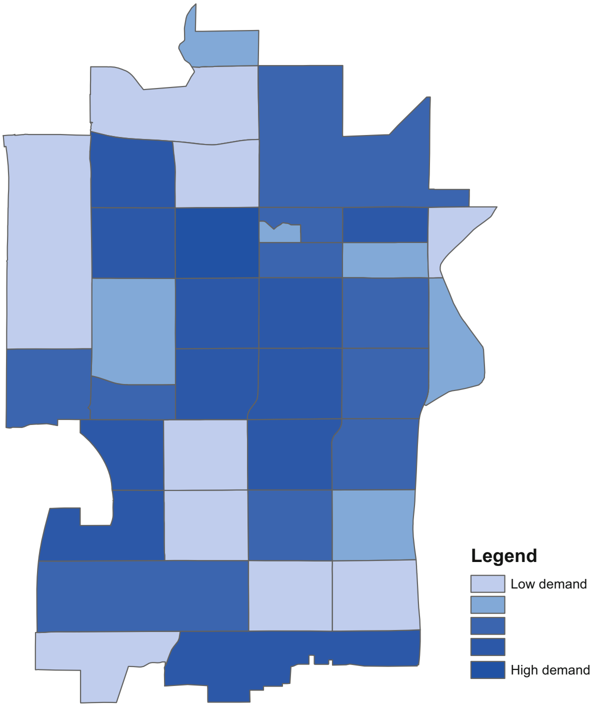


Fig. 8.3 Polygon unit based demand

challenging, especially when compared to a discrete representation view based on points. This contrasting view can be observed in Fig. 8.4 as centroids for each Census unit are identified, and demand is assumed to occur at these precise points in traditional modeling approaches.

There are a number of coverage problem situations where these two continuous space distinctions are highlighted. Consider the case of a forest region, where detection and response to fire is critical. Goodchild and Lee (1989) detail such a situation involving the placement or siting of fire watch towers. In theory, such towers could be sited anywhere, though some sites are no doubt more desirable/ appealing than others. Further, demand for service/detection, smoke or fire in this case, could occur anywhere across space. Of course, some locations may be more/ less prone to fire, depending on vegetation, fuels, closeness to roads, along high ridges, etc. Early detection and rapid response are essential for minimizing loss and harm, and this is only possible through the design of a good system of watch towers capable of monitoring/covering as much of the region possible given limited resources. Cellular telephone service is another continuous space coverage example (see Chap. 1). Cellular base stations often can be sited anywhere, and user calls (demand) are possible everywhere as well, provided there is coverage. Erdemir et al. (2008) sought to locate cellular base stations, explicitly considering that user calls can occur on any segment of road. Calls are therefore continuously distributed along

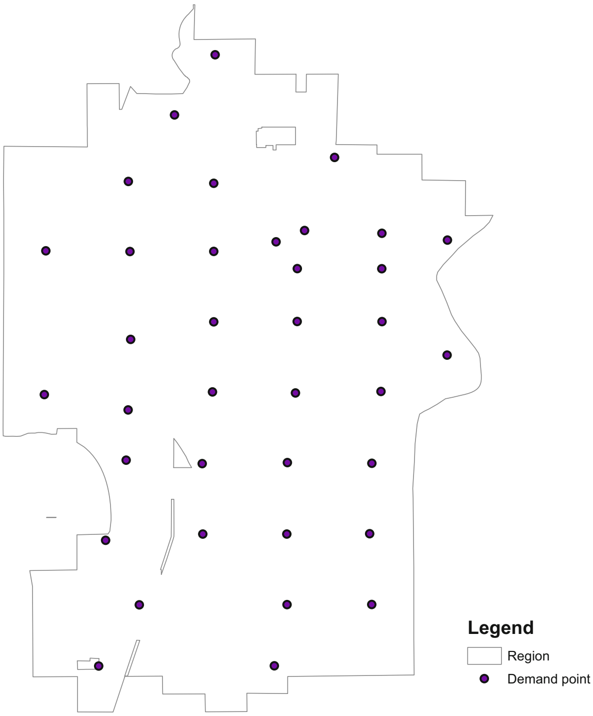


Fig. 8.4 Point based demand representation

streets in a road network. As is the case in fire detection, the goal in cellular system design is to provide the greatest coverage/service possible, but doing so at least cost. The city of Moore, Oklahoma, recently devastated by a vicious tornado, also reflects aspects of continuous coverage. Given broad scale damage and destruction, an important question in rebuilding is how many new sirens are needed for adequate community warning of severe weather or other threats, and where should they be placed. People could be anywhere in the community when danger arises and sirens can generally be sited almost anywhere. Murray and O'Kelly (2002) carry out warning siren siting along these lines, where demand for service exists everywhere in the region and sirens can be located anywhere in the region. Of course the continuous coverage context is not limited to fire watch towers, cellular base stations or emergency warning sirens. Similarly viewed situations can be found with Doppler weather equipment, wireless broadband, sensors and cameras, among others. What is consistent across all of these contexts is the two overarching issues expressed above for certain continuous coverage problems: facilities can be located anywhere; and demand for service may occur anywhere in space.

The literature associated with continuous space coverage is very broad. Classic coverage approaches have assumed demand to be at discrete point locations. Certainly this was the case for the most prominent modeling approaches, the LSCP (Toregas et al. 1971) and MCLP (Church and ReVelle 1974) detailed in Chap. 2. One of the early contributions to continuous space coverage was Kershner (1939) who wanted to determine the minimum number of circles with a given radius needed to cover a rectangle. This explicitly recognized that the entire region must be covered, suggesting demand distributed across the rectangle. ReVelle et al. (1976) note that the assumption of demand points in coverage modeling is problematic when demand is continuous in nature. They discuss extensions of the LSCP to address the case of continuous demand along a network. Other coverage extensions to address continuous demand along arcs in a network are detailed in Church and Meadows (1979), Hodgson (1990), Kuby and Lim (2005), Berman et al. (2007), Erdemir et al. (2008, 2010) and Capar et al. (2013). Aly and White (1978) and Benveniste (1982) explicitly highlight that the intent of coverage is to serve an area/region, not simply a set of representative points as often assumed. They state that errors are likely when point simplification of a region is relied upon (see also Daskin et al. 1989; Current and Schilling 1990). The reason is that points will get covered but not necessarily the space between points. This problematic feature has subsequently been demonstrated in Murray and O'Kelly (2002), Cromley et al. (2012) and Tong and Church (2012). Suzuki and Okabe (1995), Suzuki and Drezner (1996) and Wei et al. (2006) were interested in a problem related to coverage, the continuous *p*-center problem, assuming demand to originate in all places within a region. Murray (2005), Spaulding and Cromley (2007), Kim and Murray (2008), Tong and Murray (2009), Alexandris and Giannikos (2010), Murray et al. (2010), Cromley et al. (2012), Tong (2012), Yin and Mu (2012), Murray and Wei (2013) and Wei and Murray (2015, 2016) assume demand in a coverage model is continuously spread across a region and/or sub-areas. Finally, Murray et al. (2008a) and Matisziw and Murray (2009a, b) developed approaches that explicitly deal with continuous representation and service coverage of a region.

As noted previously, coverage model application is often based upon the assumption that potential facility sites are discrete points and their location is known in advance. A number of research efforts have recognized that a priori defined potential facility locations could be limiting and problematic in some planning contexts if there is no such restriction. Kershner (1939) allowed circles providing coverage to be centered anywhere. Worth noting is that the interest was only on how many circles were needed to cover an area (rectangle), seemingly unconcerned about exactly where these circles should be placed. Church and Meadows (1979) detailed a situation where facilities may be sited anywhere along a network of nodes and arcs. They specifically considered LSCP and MCLP variants where facility siting is on the network. Mehrez and Stulman (1982, 1984), Watson-Gandy (1982) and Church (1984) examined an extension of the MCLP where facilities may be sited anywhere in continuous space. This was done for the case where the coverage standard was measured according to the Euclidean metric and demand was represented as discrete points. Church (1984) and Drezner (1986) discuss a similar problem, with the distinction that the rectilinear distance metric is relied upon. Suzuki and Okabe (1995), Suzuki and Drezner (1996) and Wei et al. (2006) as noted above, examined a model capable of being used for coverage situations, the *p*-center problem. They explicitly assumed that facilities could be located anywhere in space. Murray and Tong (2007) developed an exact approach for the case when facilities can be sited anywhere in continuous space and demand is represented as polygons. Finally, it should be noted that there are interesting geometrical covering problems that have been proposed and solved (see Plastria 2002). This includes locating the smallest covering circle of a set of points (Elzinga and Hearn 1972), the smallest covering diamond using the rectilinear distance measure (Francis et al. 1992), the largest empty circle in a bounded region of discrete points (Toussaint 1983), and innovative graphical user interfaces designed to solve minimax location problems on the plane (Brady and Rosenthal 1980; Brady et al. 1983).

8.2 Problems

The general problem of interest in this chapter is to simultaneously site multiple facilities of the same basic type, each providing service having a decidedly spatial footprint. Examples of coverage provided by a spatial service facility include a district, delivery zone, viewshed, audible noise contour, cellular reception zone, emergency response area, etc. However, distinguishing the problem here from those emphasized in other chapters is that demand is continuously distributed, and present throughout a region, and facilities theoretically could be located almost anywhere in the region. The generic problem can therefore be stated as follows:

Site a system of facilities in order to provide coverage to demand across a region as effectively as possible.

There are a few implied aspects of the problem. One is siting facilities in the most efficient manner possible. Another is that each facility may essentially be sited anywhere in continuous space. As with the problems detailed previously, coverage may be complete in that all demand is served within the stipulated standard or it may be partial in the sense that not all demand is covered due to limited resources. There are others as well, but these are the two most prominent implied nuances.

For the problems addressed here, demand is defined to be continuously distributed and facilities can be sited anywhere, although there are many variations that could be and have been considered. The literature has generally focused on one or the other emphasis (locate anywhere or demand is continuously distributed), but there does exist recent work accounting for both albeit in limited ways.

8.2.1 Discrete Demand

In practice, the generic problem of interest in this chapter may be simplified in various ways, depending on the goals and intent of particular studies. A common simplification is to view demand as being discrete. One example is that demand occurs at a point, which has historically been assumed (see ReVelle et al. 1976; Church and Meadows 1979; Mehrez and Stulman 1982, 1984; Watson-Gandy 1982;

Church 1984). A more recent assumption is that demand occurs across a small geographic area that subdivides a region (Murray and Tong 2007; Murray et al. 2008a), such as Census blocks, block groups, tracts, etc. Each area (or polygon) has demand spread across its surface in some fashion.

Thus, some research assumes that demand is discrete, but maintains the continuous space property that facilities can effectively be located anywhere in the region. The major complication, of course, is that this means an infinite number of potential facility locations are possible. The work of ReVelle et al. (1976), Church and Meadows (1979), Mehrez and Stulman (1982, 1984), Watson-Gandy (1982), Church (1984) and Murray and Tong (2007) all represent approaches that assume some sort of discrete form of demand and recognize that it is possible to site facilities anywhere in space.

8.2.2 Discrete Potential Facility Locations

A different sort of simplification of the generic problem is to assume that potential facility locations are discrete and finite. This, in fact, has been a common assumption, as suggested in much of this book. Of course, this makes sense if after suitability analysis and other geographic screening there are only a few, finite number of sites where a facility may be placed. Demand for service may occur anywhere in geographic space, and may be in the form of a region, objects, areas, lines and points, but where facilities may be sited could be limited.

The work of Murray (2005), Spaulding and Cromley (2007), Kim and Murray (2008), Murray et al. (2008a, 2010), Tong and Murray (2009), Alexandris and Giannikos (2010), Cromley et al. (2012), Tong (2012), Yin and Mu (2012), Murray and Wei (2013) and Wei and Murray (2015, 2016) expressly consider this problem variant where demand is continuous but facility sites are discrete and finite.

8.2.3 Continuous Demand and Infinite Potential Facility Locations

While the majority of research on continuous space facility siting has employed various types simplifying assumptions to make problems more tractable and solvable, there have been a few select efforts that have maintained the spirit of the generic coverage problem. Work by Kershner (1939), Murray et al. (2008a) and Matisziw and Murray (2009a, 2009b) deal with problems where demand is continuously distributed and potential facility sites are unlimited, as reflected in the generic problem description. This does not mean, however, that no assumptions are made or needed. For example, there can be assumptions regarding the distribution of demand, specifically that it is uniformly distributed.

8.3 Formulations

There has been a considerable amount of attention devoted to modeling and solving continuous space coverage problems, as noted in the first two sections of this chapter. An important distinction that should be recognized by now is the two primary aspects of what continuous space can mean. First is that demand is continuously distributed in some manner across a region (Fig. 8.2). Demand for a facility can arise almost anywhere, suggesting an infinite number of locations. The second aspect is that potential facilities can effectively be sited anywhere in the region (Fig. 8.1). This means that there are an infinite number of locations that must be considered in the configuration of a facility service system. Thus, if one indicates that the planning context of interest involves continuous space coverage, both aspects of continuous space are implied. In practice, of course, there may well be circumstances that dictate that simplifying assumptions are appropriate, as suggested in the previous section.

With this in mind, there are two fundamental continuous space coverage problems related to the generic problem definition given in Sect. 8.1 that are discussed in the chapter. They are the continuous space generalizations of the LSCP and the MCLP detailed in Chap. 2. Of course, it is worth noting that other continuous space generalizations of coverage models are possible. In order to formalize these models, consider the following notation:

Φ = demand region;
i = index of demand sub-areas in region;
Δ = set of facility sites;
$f_i(\Delta)$ = demand coverage function for area i based on sited facilities Δ;
α_i = demand in area i;
j = index of sited facility;
(φ_j, λ_j) = location of facility j.

The decision variables are the coordinates of each facility, (φ_j, λ_j) . The set of located facilities, Δ, provides service coverage to the region. This demand coverage is measured using the function $f_i()$ for each demand area/object i, where the set of sited facilities Δ dictates what coverage can be provided based on their geographic locations. With this notation we can formulate the continuous space set coverage problem (CSSCP), an extension of the discrete LSCP.

$$\textit{CSSCP: Minimize} \quad |\Delta| \tag{8.1}$$

Subject to:

$$\iint f_i(\Delta)d\Phi = \alpha_i \quad \forall i \tag{8.2}$$

$$(\varphi_j, \lambda_j) \in \Delta \quad \forall j \tag{8.3}$$

The objective, (8.1), is to minimize the total number of facilities needed. Constraints (8.2) require that each sub-area be completely covered by the system of service facilities. Constraints (8.3) stipulate that located facilities must be in the demand region.

Major questions being addressed using the CSSCP are finding the most efficient number of facilities as well as their location needed to provide complete coverage of demand in the region. The membership of the set Δ is therefore unknown and to be determined. This means that it is not known a priori. Specifically, the size of the set of facilities, |Δ|, is to be determined by the model, similar to the objective of the LSCP. Coverage of each demand area/object is imposed in Constraints (8.2). The demand covered is derived through integration of the coverage of facilities over the entire area. This is required to equal the total demand in the area. That is, all demand must be suitably covered. The objective, potential facility locations and the area object function, $f_i()$, defining the distribution of demand combine to make the CSSCP highly non-linear in form, in contrast to the discrete structure of the LSCP. This, unfortunately, is the price that is paid for reducing assumptions about the nature of demand across space as well as recognizing that an infinite number of potential facility sites are possible.

Again, the CSSCP is a continuous space version of the LSCP, where demand is continuously distributed and facilities may be sited anywhere in the region. Objective (8.1) therefore reflects an intent to minimize total system cost by finding the smallest set of facilities that achieves complete coverage. As with the LSCP, the assumption is that the fixed and annual costs of a facility are basically the same, regardless of location. This means that cost minimization is achieved by simply minimizing the total number of needed facilities to provide desired service coverage to all regional demand. If costs varied significantly based on location, then the objective could be modified to account for this as follows:

$$\textit{Minimize} \;\; c(\Delta) \tag{8.4}$$

where $c()$ is a cost function associated with where facilities are sited, Δ. This objective would replace objective (8.1) in the CSSCP, giving a problem extension.

The second basic model is the continuous space version of the MCLP. Consider the following additional notation:

A_i = demand in area i;
p = number of facilities to be located.

As with the MCLP, there is an a priori specification of the number of facilities that is to be located, p. Further, given discrete objects, there is a total potential demand for each area, A_i. The formulation of the continuous space maximal coverage problem (CSMCP) follows.

$$\textit{CSMCP: Maximize} \quad \sum_i A_i \tag{8.5}$$

Subject to

$$\iint f_i(\Delta)d\Phi \geq A_i \quad \forall i \tag{8.6}$$

$$|\Delta| = p \tag{8.7}$$

$$(\varphi_j, \lambda_j) \in \Delta, \quad j = 1, \ldots, p \tag{8.8}$$

The objective, (8.5), seeks to maximize the total demand suitably served/covered. Constraints (8.6) track demand in each sub-area covered by the system of service facilities, Δ. Constraint (8.7) specifies that p facilities are to be sited. Constraints (8.8) stipulate that located facilities must be in the demand region.

As is the case when contrasting the MCLP from the LSCP, the distinction of the CSMCP is that investment potential may be limited. That is, complete service coverage of regional demand may be achieved only at an excessive cost, given the marginal returns of coverage on facility investment associated with the last 5–25% of coverage (see Church and ReVelle 1974; Grubesic et al. 2013). As a result, the total budget is limited in the CSMCP formulation. Also, the model must track or account for demand served by a facility configuration.

Noted above was the assumption that facility costs at each site are essentially the same in the CSSCP and CSMCP. If this is not the case, it is possible to account for this by extending the CSMCP as follows:

$$c(\Delta) = B \tag{8.9}$$

where B is the total budget for the siting facilities. This constraint would simply be substituted in the CSMCP for constraints (8.7).

A final comment regarding the above formulations of generic continuous space coverage models is that to apply such models in practice one would need to know the coverage function for each demand object, $f_i()$. This function, implicitly, would reflect knowledge of the distribution of demand. Yao and Murray (2013) discuss issues associated with spatially modeling a demand distribution and issues in their use in location modeling. This is not a trivial model consideration. Similar observations can be drawn for the incorporation and use of a cost function, $c()$, as well.

8.4 Simplification and Relaxation

The CSSCP and CSMCP represent two particular continuous space optimization problems. It has already been noted that extensions are possible to account for facility costs when such costs are not homogeneous across space. Similarly, the LSCP and MCLP extensions reviewed in this text and in the literature are also possible for the CSSCP and CSMCP. Such extensions will not be discussed here, rather the focus will now turn to how these models have been simplified and relaxed to support planning. The primary purpose of this section, then, is to highlight what has been done in terms of continuous space modeling. As will become evident, much of the work to date is based on discrete approaches. That is, discrete simplifications of the CSSCP and CSMCP have been pursued, often attempting to improve upon assumptions and/or spatial representation issues that arise in the application of traditional models when used to approximate a continuous space problem.

8.4.1 Discrete Demand and Discrete Potential Facility Locations

If it is assumed that demand is discrete and that potential facility locations are discrete, known in advance, then the CSSCP and CSMCP are potentially more manageable to solve. In particular, if demand is at point locations and potential facility locations are points as well, then the problems can be simplified to the classic LSCP and MCLP models. The model formulations for both of these problems have already been detailed in this text (see Chap. 2), so the details will not be repeated here. Nevertheless, this is an important observation as the continuous space problems and corresponding discrete space problems are conceptually the same, but differ due to assumptions regarding geographic representation and context. In essence, the LSCP is a special case of the CSSCP and the MCLP is a special case of the CSMCP. They are special cases because they impose assumptions about demand and potential facility sites being discrete rather than continuous.

The point, of course, is that the most popular approach for solving continuous space problems has been to assume that demand for service is represented as discrete objects, either points, lines or polygons, and that potential facility locations are also discrete and known in advance. With these assumptions, the LSCP or MCLP, as an example, could be applied as reflecting the intent of their continuous space counterpart, the CSSCP or CSMCP.

The implications of such simplification, e.g., solving a discrete problem when the actual problem is a continuous problem, is that errors may result. Empirical work by Murray (2005), Spaulding and Cromley (2007), Kim and Murray (2008), Murray et al. (2008a), Tong and Murray (2009), Alexandris and Giannikos (2010), Cromley et al. (2012) and Yin and Mu (2012), among others, highlights that coverage model results can vary depending on how demand is discretely represented and/or the form in which potential facility sites are defined. Theoretically, this relates to the well-known modifiable areal unit problem (MAUP) (see Murray 2005), but in fact is probably better characterized as error/uncertainty that results from using an approximate discrete model to represent what is really a continuous space problem. An issue of equal importance is the degree to which a discrete representation adequately represents a continuous problem when the demand function itself is subject to some level of error as compared to a continuous representation when subject to the same errors in the demand function.

8.4.2 Discrete Potential Facility Locations

To address the issue of error/uncertainty that results when applying a discrete model to solve what is actually a continuous space problem, further model refinements have been pursued. Assuming only discrete potential facility locations, a growing body of research has focused on devising coverage models that eliminate or minimize error and uncertainty in computed results. Noteworthy among these efforts are Murray (2005), Tong and Murray (2009), Murray et al. (2010), Alexandris and Giannikos (2010), Cromley et al. (2012) and Yin and Mu (2012). This work has recognized limitations that arise through demand discretization processes, given a set of discrete and known potential facility locations. In doing so, models have been proposed and developed that tend to be less sensitive to the shape and size of geometric objects/polygons structured to reflect demand, or ideal representations have been devised. Murray (2005) sought to address representation and error issues in the use of the LSCP. A summary and review is provided in Murray et al. (2010) that characterizes potential approaches for eliminating error and uncertainty as implicit and explicit coverage approaches.

8.4.2.1 Implicit Coverage

The implicit approach does not attempt to precisely account for which demand areas/objects are covered by which potential facilities in the model. Rather, the coverage that is implied by the siting of facilities at particular locations is structured, similar to what is done in the LSCP and MCLP, as an example. Consider the following additional notation:

k = index of coverage levels (1, 2, 3, . . ., K);
β_k = minimum acceptable coverage portion at level k;
Ω_{ik} = set of potential facilities j partially covering area i at least β_k;
γ_k = minimum number of facilities needed for complete coverage at level k;

$$X_j = \begin{cases} 1 & \text{if potential facility } j \text{ selected} \\ 0 & \text{otherwise} \end{cases}$$

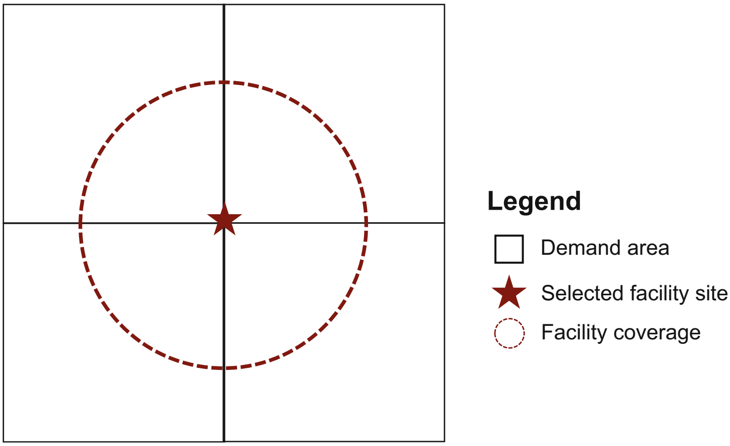


Fig. 8.5 Partial coverage of area objects

$$Y_{ik} = \begin{cases} 1 & \text{if area } i \text{ is covered at level } k \\ 0 & \text{otherwise} \end{cases}$$

The idea is that error occurs in models due to an inability to account for partial coverage of demand objects, yet multiple levels of partial coverage can equate to complete coverage of a demand object. Partial coverage of demand areas is illustrated in Fig. 8.5. All areas have some amount of service coverage, but none are fully or completely covered. The issue with discrete models like the LSCP and MCLP is that they treat coverage of demand in an all or nothing fashion. This is because these models were initially developed to deal with point based demand representations, and as a result there is no chance of partial coverage of a point. When demand is not a point, two or more facilities suitably covering portions of a demand object equate to sufficient service if the entire object is covered. Such a situation is shown in Fig. 8.6 where six facilities combine to provide service coverage to all portions of the demand areas. Structured models to better account for partial coverage therefore offer the potential to more accurately account for continuously distributed demand. An implicit version of the LSCP that incorporates partially served demand follows based on Murray (2005) (see also Murray et al. 2010).

$$\textit{LSCP-Implicit: Minimize} \quad \sum_j X_j \tag{8.10}$$

Subject to

$$\sum_{j \in \Omega_{ik}} X_j \geq \gamma_k Y_{ik} \quad \forall i, k \tag{8.11}$$

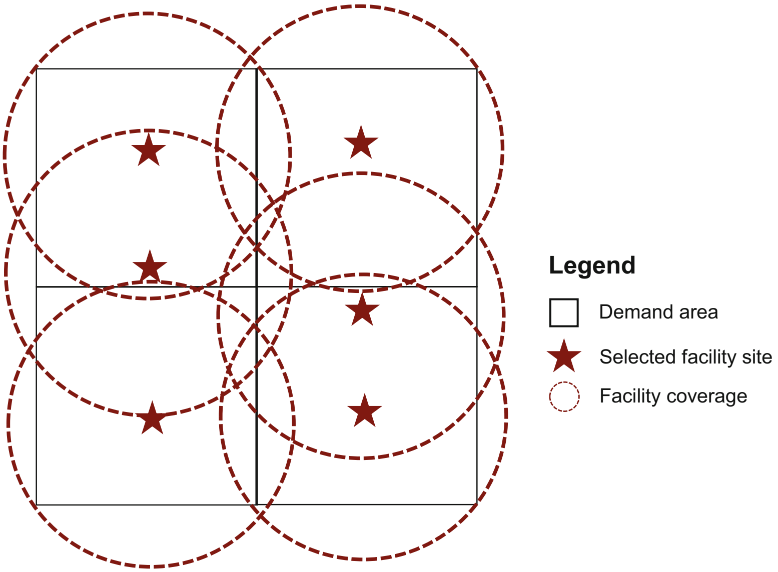


Fig. 8.6 Complete coverage of objects

$$\sum_k Y_{ik} = 1 \quad \forall i \tag{8.12}$$

$$X_j = \{0, 1\} \quad \forall j, \quad Y_{ik} = \{0, 1\} \quad \forall i, k \tag{8.13}$$

The objective (8.10) seeks to minimize the number of facilities needed. Constraints (8.11) link siting decisions to levels of coverage provided. Constraints (8.12) stipulate that one level of service is be provided to each demand area. Constraints (8.13) impose integer restrictions on decision variables.

This model differs from the LSCP in that multiple levels of partial coverage of an area translates into complete coverage under certain conditions. Constraints (8.11) account for this, where coverage at different levels is stipulated. For example, if $k = 4$, the interpretation would be that γ_4 number of facilities providing partial coverage would be necessary. One approach could be that $\gamma_4 = 4$, where at least four facilities providing partial service coverage, say 25% (β_4) or more of the demand area/object (but less than 100%), would be required in order for the demand object to be considered completely served. Constraints (8.12) then enforce that one particular level of coverage be achieved, either by a single facility ($k = 1$) or by a number of partial service facilities ($k > 1$).

Not surprisingly, a similar approach can be conceived of for the MCLP. Consider the following additional notation:

$$Z_i = \begin{cases} 1 & \text{if demand area } i \text{ covered} \\ 0 & \text{otherwise} \end{cases}$$

This decision variable enables tracking of what demand areas are suitably served. This is incorporated in a model as follows.

$$\textit{MCLP-Implicit: Maximize} \sum\nolimits_i g_i Z_i \tag{8.14}$$

Subject to

$$\sum\nolimits_{j \in \Omega_{ik}} X_j \geq \gamma_k Y_{ik} \quad \forall i, k \tag{8.15}$$

$$\sum\nolimits_j X_j = p \tag{8.16}$$

$$\sum\nolimits_k Y_{ik} = Z_i \quad \forall i \tag{8.17}$$

$$X_j = \{0, 1\} \quad \forall j, \;\; Y_{ik} = \{0, 1\} \quad \forall i, k, \;\; Z_i = \{0, 1\} \quad \forall i \tag{8.18}$$

The objective (8.14) is to maximize total demand suitably covered in the region. Constraints (8.15) link siting decisions to levels of coverage provided. Constraint (8.16) specifies the number of facilities to be sited. Constraints (8.17) stipulate that one level of service is provided to each demand area. Constraints (8.18) impose integer restrictions on decision variables.

The limitations, of course, are that discrete potential facility locations must be specified in advance. As noted previously, this has been found to introduce error/uncertainty if the actual problem allows for facilities to be sited anywhere in continuous space. Another issue is that computational difficulties have been observed in Murray (2005) as the number of partial coverage levels increases.

8.4.2.2 Explicit Coverage

As noted above, a somewhat different approach is to explicitly account for all possibilities of coverage. In particular, accounting for any and all partial facility service coverage of demand objects ensures that there are no omission errors due to spatial representation. Tong and Murray (2009) introduced this in the context of maximal coverage, formulating a model that accounted for all combinations of coverage provided to demand objects. For any combination of coverage up to a pre-specified level k, such models precisely track the portion of a demand object covered. Explicit coverage approaches are reviewed in Murray et al. (2010), including formulations for versions of the LSCP and MCLP.

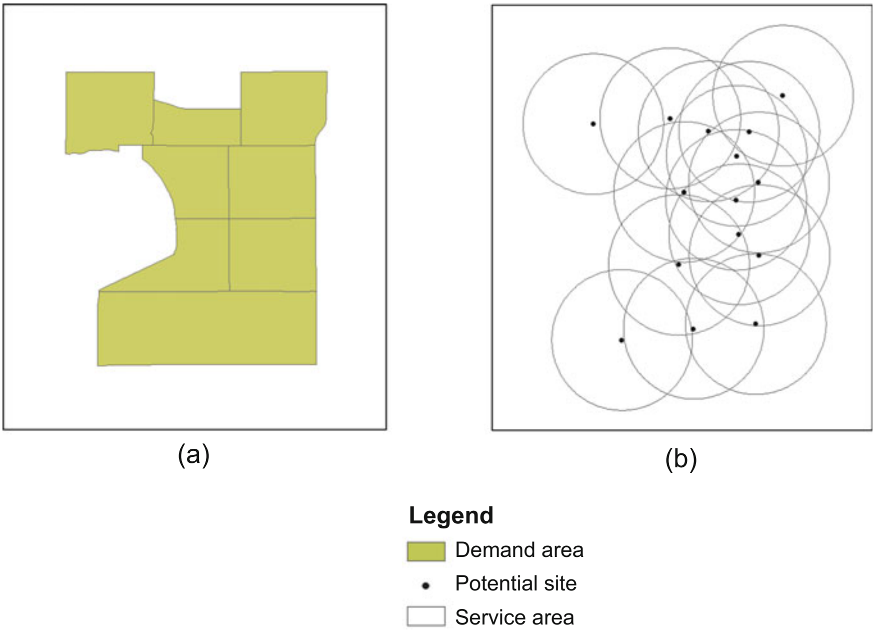


Fig. 8.7 Input layers for polygon overlay. (**a**) Demand areas. (**b**) Facility serve areas

Another explicit coverage approach has been introduced by Cromley et al. (2012) and Yin and Mu (2012) based upon data preprocessing using GIS overlay operations. Effectively, demand polygons/objects of the region are defined and delineated using the coverage areas for potential facility locations. An example is shown in Fig. 8.7, where two input layers must be combined to identify the unique polygons that result from this combination. Figure 8.8 shows two layers combined together. The overlay process is then applied and each uniquely defined area becomes an individual object, as shown in Fig. 8.9. With this preprocessing, discrete models like the LSCP and MCLP can then be directly applied. This eliminates issues arising from incomplete service coverage of an area/object because this can no longer happen. There is no partial coverage that results from the overlay process, and hence there is no opportunity for representation based error. Again, the notable limitation is that discrete potential facility locations are assumed. Computational issues are also considerable. Tong and Murray (2009) detail that the number of combinations grows quickly, leading to model instances that are difficult if not intractable to solve. Wei and Murray (2014) demonstrate that the number of demand units/objects that result from overlay, as an example, will generally be substantial, often exceeding computational capabilities, in identifying all of the individual unique units as well as the resulting optimization model that needs to be solved.

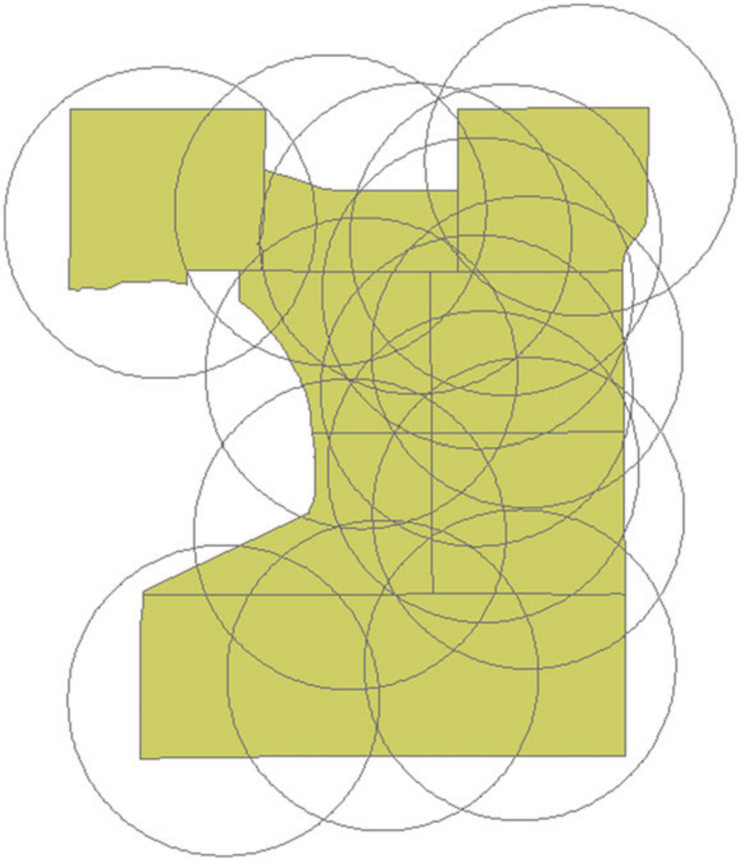


Fig. 8.8 Overlay of two layers: demand area and facility service coverage

8.4.3 Discrete Demand

When it is assumed that demand is discretely represented (i.e., points, lines, polygons or geometric objects), and they are small in comparison to the entire study region, this too has represented a particular continuous space coverage problem simplification that has enabled a discrete approach to be applied. With discrete demand objects, work by Church and Meadows (1979), Mehrez and Stulman (1982, 1984), Watson-Gandy (1982), Church (1984), Drezner (1986) and Murray and Tong (2007) derived locations in continuous space that are more preferable than others for facility siting under certain conditions. For example, Church and Meadows (1979) were interested in continuous space coverage along a network, Mehrez and Stulman (1982, 1984) and Church (1984) focused on demand points with facility coverage defined according to Euclidean distance, and Murray and Tong (2007) dealt with demand represented as any general polygonal object. These

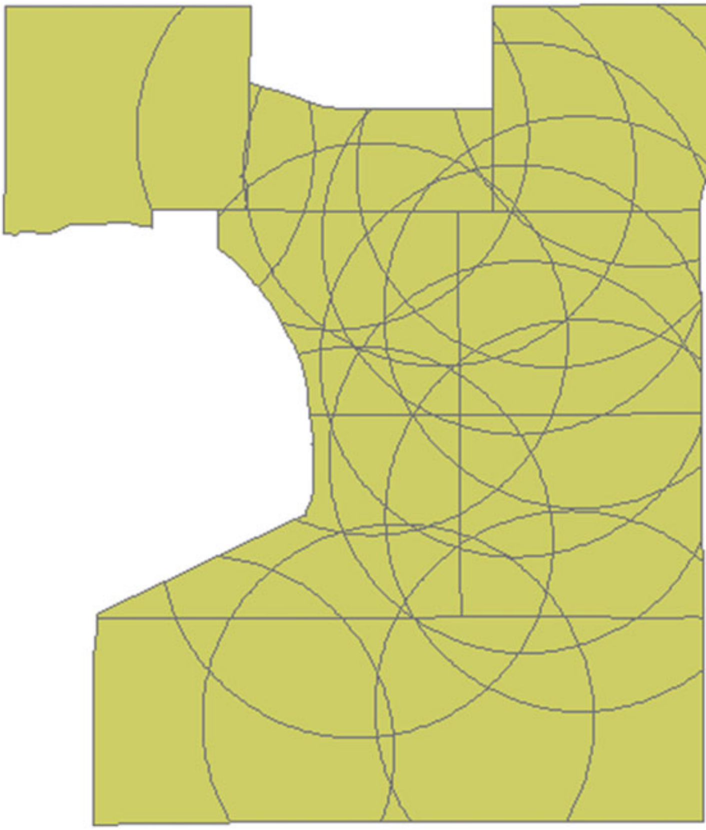


Fig. 8.9 New demand area polygons resulting from overlay process

approaches found that continuous space siting could be reduced from an infinite number of locations down to a finite set. Further, the finite, discrete sets of potential facility sites could be proven to contain the optimal continuous space siting solution. With a discrete set of potential facility sites along with discretely represented demand, it is possible to apply traditional discrete covering models, such as the LSCP and MCLP to solve such problems.

As noted above, the implication involved in simplifying continuous space as discrete demand objects can lead to erroneous results. Murray (2005), Spaulding and Cromley (2007), Kim and Murray (2008), Murray et al. (2008a), Tong and Murray (2009), Alexandris and Giannikos (2010), Cromley et al. (2012) and Yin and Mu (2012), among others, demonstrate the errors due to spatial representation of demand that occur in coverage models, if continuous space coverage is sought. Further, recent work by Wei and Murray (2014) suggest that the finite set is unique to the representation of demand, and therefore will not necessarily contain an optimal solution for the more general continuous space coverage of demand. Again, this means that errors are likely with such an approach.

8.5 Transformation and Solution

This chapter has emphasized that there is often interest in structuring and solving a continuous space coverage problem to address particular planning issues. Representing continuously distributed demand and accounting for an infinite number of potential facility sites makes such problems difficult to mathematically articulate, but also difficult to solve even when they can be sufficiently specified. The CSSCP and CSMCP are two examples of problems that are mathematically complex and difficult to solve exactly. For these reasons and others, simplifying assumptions have produced what can best be characterized as continuous space coverage problem variants, many with unique solution approaches. In this section approaches for transformation and solution are detailed. While noted that the combined assumptions of discrete demand and discrete potential facility locations represents one simplifying approach, the result is the traditional discrete space formulations, e.g., the LSCP and MCLP. A similar sort of conclusion follows for the assumption of discrete potential facility locations as well.

8.5.1 Finite Dominating Sets

The utilization of discrete demand in the form of points, lines, polygons or other geometric objects to represent continuously distributed demand has enabled a noteworthy approach for solving continuous covering problems to be devised. With discrete demand objects, it is possible to derive discrete potential facility locations that sufficiently represent continuous space. Such a discrete set of potential facility locations are known as a finite dominating set (FDS). This has been widely used (see Church and Meadows 1979; Mehrez and Stulman 1982, 1984; Watson-Gandy 1982; Church 1984; Drezner 1986; Murray and Tong 2007; Berman et al. 2007; Poetranto et al. 2009; Schobel et al. 2009; Tong and Murray 2009; Berman and Wang 2011; Tong 2012; Wei and Murray 2014) because it can be proven that an optimal solution to some coverage models using discrete demand will consist of members of the FDS. For demand represented as points, Church (1984) proved that the infinite number of potential facility sites could be reduced to a finite set of points for consideration in coverage maximization. Murray and Tong (2007) similarly proved that for any demand objects (e.g., point, line, and polygon) in service coverage optimization a finite set of potential facility sites could be identified to serve as a valid representation of continuous space facility sites.

While various names have been used to refer to a particular instance of the FDS, e.g. NIPS (Network Intersect Point Set), CIPS (Circle Intersecting Point Set), PIPS (Polygon Intersection Points Set), etc. depending on the authors, the focus here will be on the derivation of a FDS when any discrete spatial demand object is utilized.

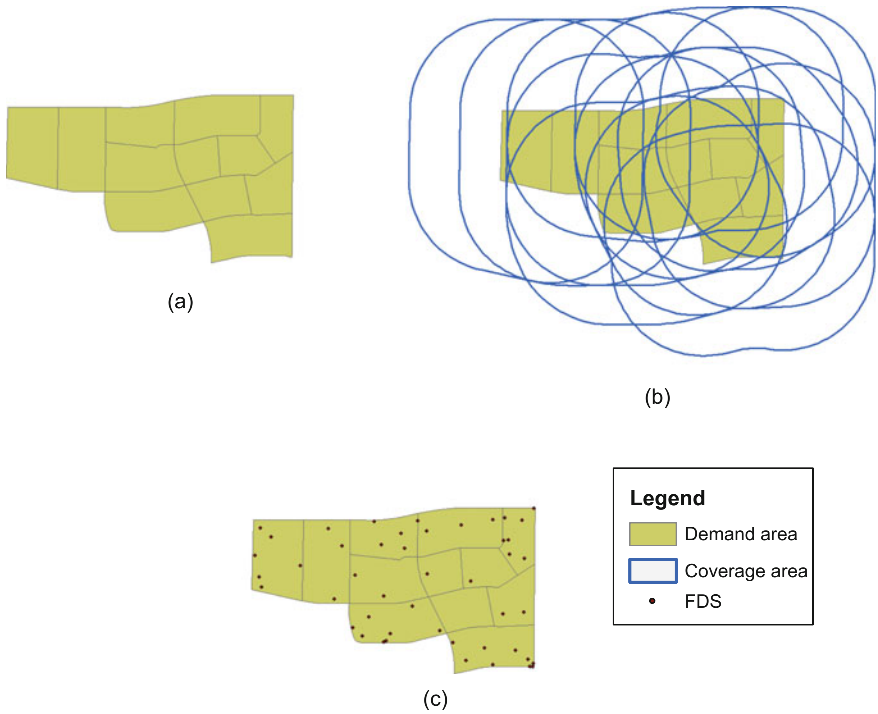


Fig. 8.10 Depiction of major steps of FDS process. (**a**) Demand areas. (**b**) Covering areas of demand. (**c**) FDS

Murray and Tong (2007) outline the PIPS approach, where the generalization of the FDS process is the following:

(a) Identify demand objects to be covered
(b) Derive covering areas around each demand object
(c) Find the intersection points of covering areas

An important thing to highlight is that the point set that results from step (c) is the FDS. While Step (a) is rather trivial, it is well-known that discrete representations of continuous space have the potential to impact any derived results. With that caveat in mind, step (b) involves finding the area for each object where siting a facility would provide complete coverage of the object. That is, siting a facility anywhere in this area would enable coverage of the object. This must be done for each demand object. With the collection of areas providing complete coverage to each individual demand object, step (c) involves finding all area intersection points. These points of intersection represent members of the FDS, as noted above. The reason for this is that siting a facility at an intersection point is as good or better than other points in an object covering area. It is possible to cover/serve multiple objects at intersection points, not just a single object. As a result, such points are superior to other points only providing coverage to a single object. For more details, consult Murray and Tong (2007).

An example of the FDS generation process is illustrated in Fig. 8.10. Given a set of demand objects (Fig. 8.10a), the resulting covering areas can be identified (Fig. 8.10b). In this example, the covering areas are circular. With the covering areas, the points of intersection can then be derived. This set is shown in Fig. 8.10c. This is the FDS for this set of defined demand objects. With both the demand objects and the FDS, a discrete coverage model may then be applied. As noted above, the LSCP and MCLP have been applied under these conditions as well as other coverage model extensions.

8.5.2 p-*Center*

When complete coverage of an entire region is required (i.e., CSSCP), it is possible to solve such a continuous space covering model as a p-center problem. Technically speaking, the p-center problem involves the location of p facilities such that the maximum distance any point within the region is from its closest facility is minimized. If the maximum distance associated with an optimal p-center solution satisfies the coverage requirements (and is a minimum value of p to achieve this), then that p-center solution is optimal to the respective CSSCP. Suzuki and Okabe (1995) and Suzuki and Drezner (1996) outline an approach, the Voronoi diagram heuristic, where facilities may be sited anywhere in order to serve the region. An assumption of Suzuki and Okabe (1995) and Suzuki and Drezner (1996) is that the region is a simple polygon. Wei et al. (2006) relax this assumption as well as address some practical application issues by extending the Voronoi diagram heuristic. The general heuristic is as follows:

1. Begin with $p = 1$
2. Generate randomly generated set of p points within the region and use this as the initial p center solution.
3. Construct a Voronoi diagram based on the current p center locations
4. Compute the 1-center for each Voronoi polygon, and use this set of 1-center solutions as the new p-center solution
5. If any center has moved, then go to step 3
6. If the solution does not completely cover the region based on the service standard, then increment p by one (i.e. $p = p + 1$) and go to step 2.
7. Stop

The heuristic begins by setting the number of facilities to be located at one in Step 1. Step 2 identifies the initial, randomly generated set of p facility locations within the region. Next, a partition of the region into p polygons is found (Step 3) based on the Voronoi diagram associated with the p sites. The Voronoi diagram represents an optimal allocation of areas in the region to their closest facility. Step 4 then finds the best facility site within each Voronoi polygon, which is the center of the smallest

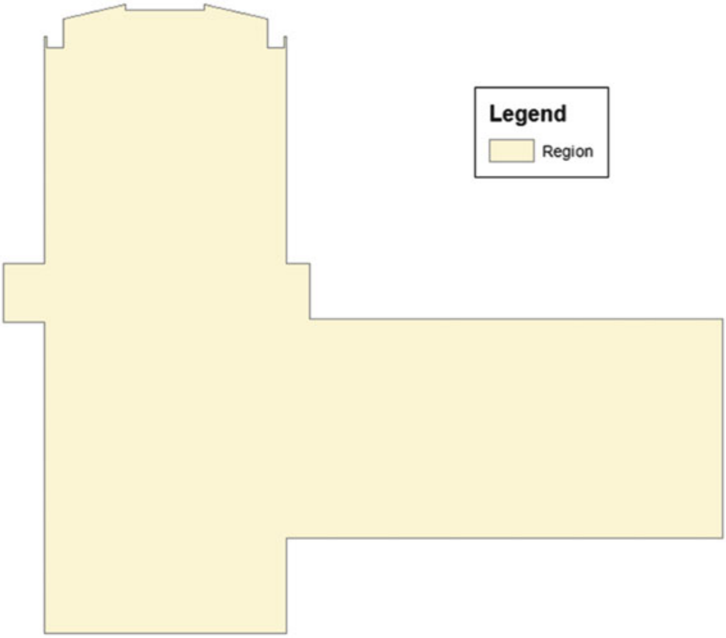


Fig. 8.11 Region to be served

disk that can completely enclose the polygon (a 1-center). This is then used as the new facility site for each associated Voronoi polygon. However, moving the facility site may alter optimal allocation, so this is assessed in Step 5. The process continues to cycle through Steps 2–5, until no facility site changes. Then, in Step 6 the level of coverage is computed, and if this solution does not completely cover the region, the number of facilities is increased and the heuristic starts over by generating a new random starting solution, and continues to iterate. Again, this approach may be used to solve a continuous space coverage problem where complete regional coverage is required. Unfortunately, there are no guarantees that this process will find an optimal set covering solution.

8.5.3 Skeleton

When complete regional coverage is not possible, then the CSMCP is of interest. A few recent approaches have been proposed to address maximal coverage situations. Murray et al. (2008b) and Matisziw and Murray (2009a, b) introduced the use of a skeleton, or medial axis, to indicate locations in continuous space where an optimal facility location can be found. Formally, the skeleton is the locus of all points equidistant to at least two nearest locations on the area boundary. Given the area shown in Fig. 8.11, the associated

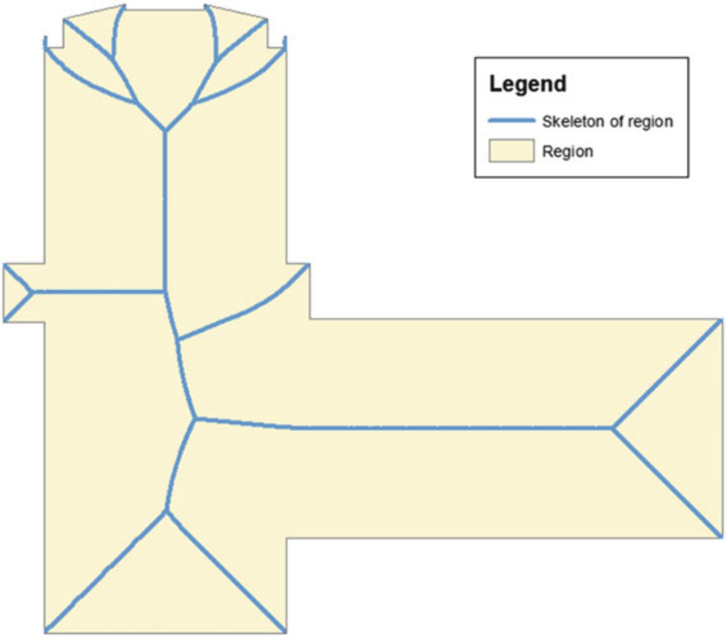


Fig. 8.12 Skeleton of region

skeleton is depicted in Fig. 8.12. Murray et al. (2008b) modified the above Voronoi diagram heuristic to search along skeletons to find better facility locations.

8.6 Summary and Concluding Comments

The emphasis of this chapter is on continuous space coverage modeling. Two important points emerge in continuous space coverage: facilities may be sited anywhere and demand is continuously distributed across the study area. The chapter reviewed two representative continuous space coverage problems, the CSSCP and CSMCP, giving both formulations and solutions techniques. It should be evident that continuous space aspects of the problems make them unique, and distinct from discrete formalizations like the LSCP and MCLP, respectively. Further, such problems are inherently difficult to formalize and challenging to solve. Interestingly a number of simplifying assumptions have been employed to make these continuous space problems more computationally (and conceptually) manageable. As a result, many problem variants can be found in the literature. This includes problem reformulation (i.e., implicit and explicit approaches), exploiting spatial properties (i.e., the FDS), and development of heuristic approaches like Voronoi diagram based heuristics.

References

Aly AA, White JA (1978) Probabilistic formulation of the emergency service location problem. J Oper Res Soc 29:1167–1179

Alexandris G, Giannikos I (2010) A new model for maximal coverage exploiting GIS capabilities. Eur J Oper Res 202:328–338

Benveniste R (1982) A note on the set covering problem. J Oper Res Soc 33:261–265

Berman O, Verter V, Kara BY (2007) Designing emergency response networks for hazardous materials transportation. Comput Oper Res 34:1374–1388

Berman O, Wang J (2011) The minmax regret gradual covering location problem on a network with incomplete information of demand weights. Eur J Oper Res 208(3):233–238

Brady SD, Rosenthal RE (1980) Interactive computer graphical solutions of constrained minimax location problems. AIIE Trans 12(3):241–248

Brady SD, Rosenthal RE, Young D (1983) Interactive graphical minimax location of multiple facilities with general constraints. AIIE Trans 15(3):242–254

Capar I, Kuby M, Leon VJ, Tsai YJ (2013) An arc cover–path-cover formulation and strategic analysis of alternative-fuel station locations. Eur J Oper Res 227:142–151

Church RL (1984) The planar maximal covering location problem. J Reg Sci 24:185–201

Church RL, Meadows ME (1979) Location modeling utilizing maximum service distance criteria. Geogr Anal 11:358–373

Church RL, Revelle C (1974) The maximal covering location problem. Pap Reg Sci 32:101–118

Cromley RG, Lin J, Merwin DA (2012) Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation. Int J Geogr Inf Sci 26 (3):495–517

Current J, Schilling D (1990) Analysis of errors due to demand data aggregation in the set covering and maximal covering location problems. Geogr Anal 22:116–126

Daskin MS, Haghani AE, Khanal M, Malandraki C (1989) Aggregation effects in maximal covering models. Ann Oper Res 18:115–140

Drezner Z (1986) The p-cover problem. Eur J Oper Res 26:312–313

Elzinga J, Hearn DW (1972) Geometrical solutions for some minimax location problems. Transp Sci 6(4):379–394

Erdemir ET, Batta R, Rogerson PA, Blatt A, Flanigan M (2010) Joint ground and air emergency medical services coverage models: a greedy heuristic solution approach. Eur J Oper Res 207:736–749

Erdemir ET, Batta R, Spielman S, Rogerson PA, Blatt A, Flanigan M (2008) Location coverage models with demand originating from nodes and paths: application to cellular network design. Eur J Oper Res 190:610–632

Francis RL, McGinnis LF, White JA (1992) Facility layout and location: an analytical approach. Pearson College Division, New York

Goodchild MF, Lee J (1989) Coverage problems and visibility regions on topographic surfaces. Ann Oper Res 18(1):175–186

Grubesic TH, Murray AT, Matisziw TC (2013) Putting a price on politics as usual: rural air transport in the United States. Transp Policy 30:117–124

Hodgson MJ (1990) A flow-capturing location-allocation model. Geogr Anal 22:270–279

Kershner R (1939) The number of circles covering a set. Am J Math 61:665–671

Kim K, Murray AT (2008) Enhancing spatial representation in primary and secondary coverage location modeling. J Reg Sci 48:745–768

Kuby M, Lim S (2005) The flow-refueling location problem for alternative-fuel vehicles. Socio Econ Plan Sci 39:125–145

Matisziw TC, Murray AT (2009a) Siting a facility in continuous space to maximize coverage of continuously distributed demand. Socio Econ Plan Sci 43:131–139

Matisziw TC, Murray AT (2009b) Area coverage maximization in service facility siting. J Geogr Syst 11:175–189

Mehrez A, Stulman A (1982) The maximal covering location problem with facility placement on the entire plane. J Reg Sci 22:361–365
Mehrez A, Stulman A (1984) An extended continuous maximal covering location problem with facility placement. Comput Oper Res 11:19–23
Murray AT (2005) Geography in coverage modeling: exploiting spatial structure to address complementary partial service of areas. Ann Assoc Am Geogr 95:761–772
Murray AT, Matisziw TC, Wei H, Tong D (2008b) A GeoComputational heuristic for coverage maximization in service facility siting. Trans GIS 12:757–773
Murray AT, O'Kelly ME (2002) Assessing representation error in point-based coverage modeling. J Geogr Syst 4:171–191
Murray AT, O'Kelly ME, Church RL (2008a) Regional service coverage modeling. Comput Oper Res 35:339–355
Murray AT, Tong D (2007) Coverage optimization in continuous space facility siting. Int J Geogr Inf Sci 21:757–776
Murray AT, Tong D, Kim K (2010) Enhancing classic coverage location models. Int Reg Sci Rev 33:115–133
Murray AT, Wei R (2013) A computational approach for eliminating error in the solution of the location set covering problem. Eur J Oper Res 224:52–64
Plastria F (2002) Continuous covering location problems. In: Drezner Z, Hamacher H (eds) Facility location: applications and theory. Springer, Berlin, pp 37–79
Poetranto DR, Hamacher HW, Horn S, Schöbel A (2009) Stop location design in public transportation networks: covering and accessibility objectives. TOP 17:335–346
ReVelle CS, Toregas C, Falkson L (1976) Applications of the location set covering problem. Geogr Anal 8:65–76
Schobel A, Hamacher HW, Liebers A, Wagner D (2009) The continuous stop location problem in public transportation networks. Asia Pac J Oper Res 26:13–30
Spaulding BD, Cromley RG (2007) Integrating the maximum capture problem into a GIS framework. J Geogr Syst 9(3):267–288
Suzuki A, Okabe A (1995) Using voronoi diagrams. In: Drezner Z (ed) Facility location: a survey of applications and methods. Springer, New York, pp 103–118
Suzuki A, Drezner Z (1996) On the p-center location problem in an area. Locat Sci 4:69–82
Tong D (2012) Regional coverage maximization: a new model to account implicitly for complementary coverage. Geogr Anal 44:1–14
Tong D, Church RL (2012) Aggregation in continuous space coverage modeling. Int J Geogr Inf Sci 26:795–816
Tong D, Murray AT (2009) Maximizing coverage of spatial demand for service. Pap Reg Sci 88:85–97
Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19:1363–1373
Toussaint GT (1983) Computing largest empty circles with location constraints. Int J Comp Inf Sci 12(5):347–358
Watson-Gandy CDT (1982) Heuristic procedures for the m-partial cover problem on a plane. Eur J Oper Res 11:149–157
Wei H, Murray AT, Xiao N (2006) Solving the continuous space p-center problem: Planning application issues. IMA J Manag Math 17:413–425
Wei R, Murray AT (2014) Evaluating polygon overlay to support spatial optimization coverage modeling. Geogr Anal 46(3):209–229
Wei R, Murray AT (2015) Continuous space maximal coverage: insights, advances and challenges. Comput Oper Res 62:325–336
Wei R, Murray AT (2016) A parallel algorithm for coverage optimization on multi-core architectures. Int J Geogr Inf Sci 30:432–450
Yao J, Murray AT (2013) Continuous surface representation and approximation: spatial analytical implications. Int J Geogr Inf Sci 27:883–897
Yin P, Mu L (2012) Modular capacitated maximal covering location problem for the optimal siting of emergency vehicles. Appl Geogr 34:247–254

Chapter 9
Disruption, Protection, and Resilience

9.1 Introduction

A number of location covering models have been developed to address problems of security and safety. Bell et al. (2011) apply the LSCP to identify aircraft alert sites to respond to security threats to the U.S. border and other key assets. Bao et al. (2015) dealt with the problem of locating watchtower locations for forest fire detection and monitoring. Zhang and Du (2012) describe an approach to locate a set of radars, and assign power levels to each of them so that all crucial points along a river are monitored, or covered, while minimizing the total power being used. Other interesting examples can be found in Agnetis et al. (2009), Bar-Noy et al. (2013), and Pan (2010). Addressed is the location of a set of facilities that can guard, protect or respond in an emergency, developed under the assumption that system components are always available and ready to act. In addition, it is almost always assumed that the network infrastructure will always be usable and that services such as emergency response can be made using the best routes on an unimpeded network. That is, the system always works and there are no disruptions to service.

Most of the time, one can assume that infrastructure will work as intended, but sometimes it will not because something has been disrupted. For example, Hurricane Matthew in 2016 caused significant damage and disruption to the road networks in North and South Carolina. Several sections of interstate highways were closed due to flooding. In Horry County, South Carolina alone, more than a dozen state highways and 160 county roads were closed during the storm due to flooding and downed trees. Such disruptions can last weeks, if not longer, when bridges have been compromised and dams have been breached. Another major disaster is the Tohoku earthquake and tsunami in 2011 that destroyed a number of towns in northern Japan, including the Fukushima Nuclear Power Plant, and decimated critical infrastructure. In Santa Barbara County, the January 2018 mudslides in Montecito resulted in 21 deaths and significant destruction, including the closure of U.S. 101 and frontage roads due to water, mud and debris. This is the primary route connecting Santa

© Springer International Publishing AG, part of Springer Nature 2018
R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_9

Barbara to southern cities, with the closure lasting for nearly 2 weeks. Whether it is by intention or by a natural event, disruption happens. This chapter addresses the potential for disruption and how it might be mitigated when planning and operating facilities. In the next section we present a model that identifies the worst case of disruption for an existing set of facilities. This is followed by a model that designs a covering system, recognizing that any one of the facilities in the system may be lost due to a natural or human-caused event. A more general model is then detailed that locates a set of facilities when multiple facilities may be lost to an attacker or natural disaster, with the objective of providing as much coverage as possible before and after a disruptive event. These models seek to make a system design as robust as possible. We then turn our attention to the possibility that other elements of supporting infrastructure may be lost or disrupted, compromising what is covered, not by the loss of a facility but by the loss of elements that make coverage possible across a network. Finally, we describe a model that chooses which facilities to fortify so that a system operates as effectively as possible in light of a disruption.

9.2 *r*-Interdiction

When one looks closely, there are a number of examples where facilities have been destroyed or damaged intentionally. In September 11, 2001, the World Trade Center towers were hit by airplanes piloted by terrorists. Among other impacts, this single event changed the operation systems of banks in the U.S. and the process of handling checks, because a sizable number of checks were destroyed in both planes as well as in the towers, making account reconciliation an almost impossible task. Today, all transactions between banks are handled electronically. In April 2013, snipers attacked the Metcalf Substation near San Jose, California and severely damaged 17 high voltage transformers. A quick response by the utility (PG&E) as well as the grid operator (California Independent Service Operator) in rerouting transmission into the San Francisco bay area averted a near collapse of the grid in Silicon Valley. Since that time, the U.S. government has asked that security at major substations be enhanced and even has begun the process stockpiling high capacity high voltage transformers for backup supply, as they can take a year or more to manufacture. A report to the U.S. Congress notes that high voltage transformers make up less than 3% of the transformers in U.S. power substations, but are involved in 60–70% of the nation's electricity delivery (Parfomak 2014). The Wall Street Journal (2014) notes that taking out just nine strategically selected substations across the U.S could bring darkness to most of the U.S.! Thus, even when a system is large, targeting only a small number of facilities may have significant impact (see also Murray and Grubesic 2007).

The idea of targeting a facility, which is commonly referred to as an asset, is an important subject of the military. When fighting a war, generals often select important targets, like bridges or factories, to destroy. The idea is that taking out the capacity to supply important needs of an army or impeding its movement by taking out a bridge can be decisive in winning a war. For example, in the U.S. Civil War, the battle of Glorieta Pass in 1862 (also called the "Gettysburg of the West"), several Union companies in a flanking action came across a Confederate supply train (80 supply wagons and 100 of horses and oxen). The Union company completely destroyed it as it was poorly guarded. This was considered to be a major element in the Confederates losing the battle and abandoning their attempt to take over the southwestern U.S. Thus, the interdiction of a supply wagon train was as decisive as fighting the actual battle. The natural question for an army or an attacker is: what should be interdicted to cause the greatest harm? There are two common ways in which this question can be asked, and depends on whether it is a transportation asset/link or a facility. Let's say that the target is a facility. Further assume that we have a set of *p* facilities that represent the system to be attacked and there are resources to attack *r* of these facilities. As an example, think of communication facilities that have a limit on how far each facility can communicate with demand (the maximal service standard *S*). Each demand will continue to receive broadcast communication as long as at least one facility exists after the attack as long as it is within the covering range of the demand. The problem at hand is to find which *r* of the facilities to attack and destroy in order to ensure the greatest amount of demand is without communication services. In a more general sense, we can define (Church et al. 2004):

Of the p different facility service locations, find the subset of r facilities, which when removed, maximizes the resulting drop in coverage.

This problem is called the *r*-facility Interdiction Covering (*r*IC) problem. To formulate this as a mathematical model, consider the following notation:

I = the set of demand points
Ω = the set of existing facilities
i = an index used to refer to a specific demand
j = an index used to refer to a specific facility location
a_i = the amount of demand at i
$N_i = \{j| \text{ facility } j \text{ covers demand } i\}$
$\Phi_i = |N_i|$, the number of existing facilities that cover demand i

$$s_j = \begin{cases} 1, & \text{if a facility located at } j \text{ is eliminated, i.e.interdicted} \\ 0, & \text{otherwise} \end{cases}$$

$$y_i^L = \begin{cases} 1, & \text{if demand at } i \text{ is no longer covered} \\ 0, & \text{otherwise} \end{cases}$$

We can now formulate this model in the following manner:

$$r\text{IC}: \qquad \text{Maximize} \sum_{i \in I} a_i y_i^L \tag{9.1}$$

Subject to:

$$\sum_{j \in \Omega} s_j = r \tag{9.2}$$

$$y_i^L \leq s_j \quad \forall i \in I \text{ and } j \in N_i \tag{9.3}$$

$$s_j \in \{0, 1\} \quad \forall j \in F \tag{9.4}$$

$$y_i^L \in \{0, 1\} \quad \forall i \in I \tag{9.5}$$

The objective of *r*IC, (9.1), is to maximize the amount of demand that will lose coverage after interdiction has occurred. Constraint (9.2) restricts the number of facilities that are interdicted or destroyed to equal *r*. Constraints (9.3) allow the decision variable y_i^L to equal 1 when all existing facilities that cover demand *i* are interdicted. If some facility at j' that covers demand *i* is not interdicted, then the variable $s_{j'}$ will be zero. Constraint (9.3) will force the variable y_i^L to be zero because: $y_i^L \leq s_{j'}$. Thus, only demands that have lost all of their existing covering facilities can be counted as having lost coverage. The integer restrictions on the decision variables are given in constraints (9.4) and (9.5). The restrictions on the y_i^L variables are listed for completeness, however, when solving this model they can be dropped as they will be zero-one in value as long as the s_j meet the binary integer restrictions. This problem can be easily solved with the use of any integer linear programming solver.

*r*IC is relatively compact, as the number of decision variable equals the number of demand points plus the number of existing facilities, and the number of constraints equals the number of times existing facilities cover demands ($\sum_{i \in I} \Phi_i$) plus the resource constraint on how many facilities can be interdicted. Constraints (9.3) can be condensed into a set of the following constraints:

$$\Phi_i y_i^L \leq \sum_{j \in N_i} s_j \quad \forall i \in I \tag{9.6}$$

which allows each y_i^L to equal one only when all facilities that cover demand *i* have been interdicted. Using constraints (9.6) instead of constraints (9.3) yields a model that has only one constraint for each demand in addition to the constraint on the number of facilities that are subject to interdiction.

One should easily recognize that when the number of facilities is quite small, that one can solve the above problem by enumeration. For example, if only 10 facilities exist and there are resources to interdict 3 of them, then there are only 120 different combinations of interdicting 3 facilities of the 10 facilities. But when the number of facilities is larger as well as the number of possible interdictions, then enumeration will probably be out of the question. For example, Church et al. (2004) note that identifying what to interdict when faced with 100 existing facilities and resources to interdict 10 of them, one would have to enumerate 28,848,458,598,960 possibilities. This is a task that could not be done except perhaps using a massively parallel machine for a sizable amount of time. This problem would be compounded further if one would want to develop a tradeoff curve between the lost coverage and the number of interdictions allowed.

To demonstrate the impact of interdiction on an existing system, Church et al. (2004) used the above model applied to the 55 node data set of Swain (1971). The

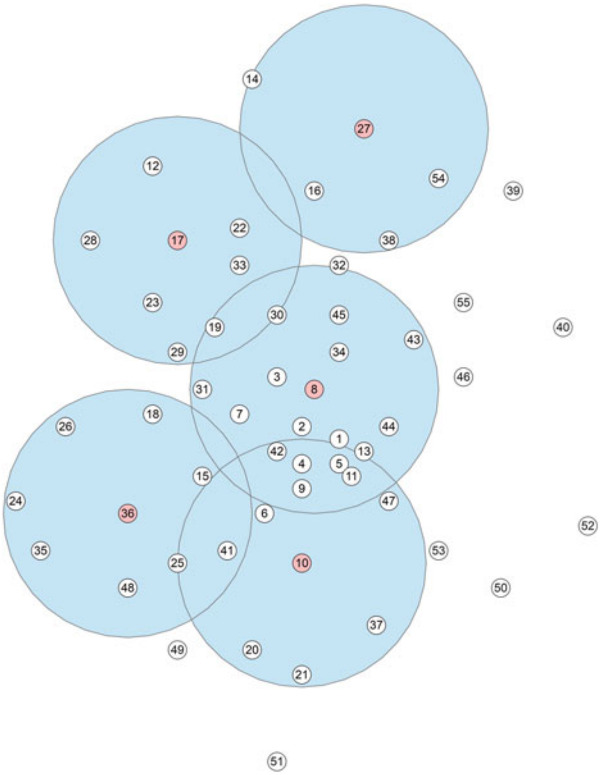


Fig. 9.1 An optimal solution to the MCLP problem when locating five facilities where the maximal service distance is 10. The solution involves sites 8, 10,17, 27, and 36. Coverage is 609 out of a possible 640

optimal five facility maximal covering solution entails facilities being placed at points 8, 10, 17, 27 and 36. This solution is depicted in Fig. 9.1. This configuration covers a total of 609 out of 640. Figure 9.2 depicts an interdiction solution where two facilities have been interdicted (facilities at points 8 and 10). Overall, coverage is reduced from 609 to 170, which means that remaining facilities cover only a fraction of what the original system covered.

9.3 Design Sensitive Facility Loss

In the previous section, we defined a model to maximize the greatest harm (as measured by loss of coverage) when striking or interdicting a fixed number of facilities. It can be applied to cases where a set of existing facilities covers all demand or only covers a subset of the demand. For instance, the example given in Fig. 9.1 depicts an interdicted MCLP solution, only partially covering demand distributed among 55 locations. A simple way of analyzing risk is to identify how well a system works after it loses any one of its components; that is, what is the worst case impact

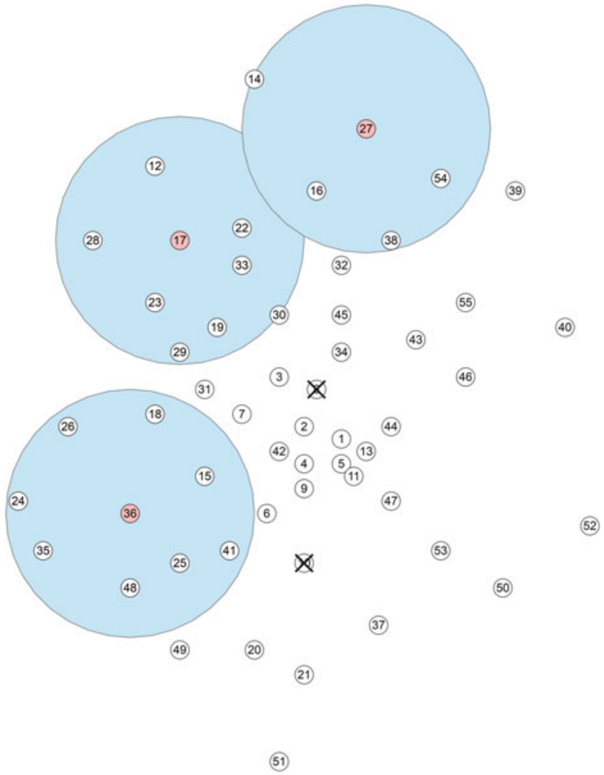


Fig. 9.2 An optimal rIC solution when interdicting the solution presented in Fig. 9.1 where $r = 2$. This solution involves the remaining sites of 17, 27, and 36 where coverage has been reduced to 170 out of 640

associated with the loss of one system component? This is an approach that has been used to assess possible impacts of systems such as rail transportation and electrical utility operations at the Risk Center of ETH in Zurich.[1] They call this method an $N - 1$ *analysis* of a system that contains N components. Within the context of a covering system, an $N - 1$ analysis would be to find the worst case loss in covering when one facility is destroyed. You can solve this problem as an $r = 1$ in rIC or simply enumerate all possibilities of losing one facility out of p. In some ways such a question might seem a bit short-sighted in that it focuses on the impact to a system when losing only one component instead of many. But this approach seems to be a good first cut approach in understanding what might happen if only one facility was lost. Obviously, when losing more than one system component or facility, the impacts will be even greater than just losing one. A second benefit of this approach is that one generates the range of impacts in the potential loss of a component, as well as identify

[1]Richard Church wishes to thank Prof. Heinimann of ETH for the gracious invitation to attend a day of meetings at the Risk Center of ETH in which a number of applications of the $N - 1$ analysis were presented.

the single most important component when the system is complete. Simply put, an $N - 1$ analysis is quite simple to perform and can be quite informative.

Assume that we have a system and the risk of losing one facility is somewhat significant and the risk of losing more than one facility is quite low or that if a facility is lost the chances are very high that it will be replaced before any other losses occur. Then, the overall design problem might be:

Locate a set of p facilities in order to maximize coverage, while ensuring that coverage losses are minimized when any one of the facilities is interdicted.

The configuration we seek is one in which total coverage is high when all facilities operate and coverage is also quite high when any one of the facilities is lost. We can call this an *p/p*-1 maximal covering design (*p/p*-1 MCD) problem involving the location of p facilities. We can count coverage losses in several different ways. Here, we will concentrate on minimizing the loss of coverage associated with a worst case event, or equivalently maximizing coverage after the loss of any one facility. To do this we need to introduce some new notation as well as modify a few terms used in the previous section. The most important property here is that each located facility is given an order, for example, $x_{jv} = 1$, represents the choice of site j as the v^{th} facility.

$J =$ the set of potential facility sites
$v = 1, 2, 3, \ldots, p$
$N_i = \{j| \text{ facility site } j \text{ covers demand } i\}$
$x_{jv} = \begin{cases} 1, & \text{if facility } v \text{ is located at site } j \\ 0, & \end{cases}$
$y_i = \begin{cases} 1, & \text{if demand } i \text{ is covered by the complete set of facilities} \\ 0, & \end{cases}$
$y_{iv} = \begin{cases} 1, & \text{if demand at } i \text{ is covered after facility } v \text{ is interdicted} \\ 0, & \text{otherwise} \end{cases}$
$w_1 =$ the importance weight associated with maximizing coverage provide by all p facilities
$w_2 =$ the importance weight associated with coverage after interdiction
$z^1 =$ the worst case amount of coverage remaining when any one of the facilities is struck

We can now formulate this model in the following manner:

$$p/p-1 \text{ MCD}: \qquad \text{Maximize } w_1 \sum_{i \in I} a_i y_i + w_2 z^1 \tag{9.7}$$

Subject to:

$$\sum_{j \in J} x_{jv} = 1 \quad \text{for} \quad v = 1, 2, \ldots, p \tag{9.8}$$

$$\sum_{v=1}^{p} x_{jv} \leq 1 \quad \forall j \in J \tag{9.9}$$

$$\sum_{v=1}^{p} \sum_{j \in N_i} x_{jv} \geq y_i \quad \forall i \in I \tag{9.10}$$

$$\sum_{v=1}^{p} \sum_{v \neq q} \sum_{j \in N_i} x_{jv} \geq y_{iv} \quad \text{for} \quad q = 1, 2, \ldots, p \quad \text{and} \quad \forall i \in I \tag{9.11}$$

$$\sum_{i \in I} a_i y_{iv} \geq z^1 \quad \text{for} \quad v = 1, 2, \ldots, p \tag{9.12}$$

$$x_{jv} \in \{0, 1\} \quad \forall j \in J \quad \text{and} \quad v = 1, 2, \ldots, p \tag{9.13}$$

$$y_{iv} \in \{0, 1\} \quad \forall i \in I \quad \text{and} \quad v = 1, 2, \ldots, p \tag{9.14}$$

$$y_i \in \{0, 1\} \quad \forall i \in I \tag{9.15}$$

The *p/p*-1 MCD uses the concept of *N/N*-1, that is, it involves finding a placement of p facilities, while maximizing coverage provided by these facilities as well as maximizing coverage after any one of these p facilities is lost. The first term of objective (9.7) involves maximizing coverage with the complete set of located facilities, while the second term involves maximizing the coverage provided when the most critical facility of the system is lost. These two objectives are weighted according to their importance. Thus, this model attempts to locate a set of facilities that cover a relatively high number of demands while also keeping coverage as high as possible in the event that any one facility is lost. Constraints (9.8) establish that there will be one facility location selected for the v^{th} facility. Constraints (9.9) ensure that each facility site will be selected at most once. Both constraints together specify the location of p-facilities. Constraints (9.10) define whether a given demand i is covered by the complete set of facilities, while constraints (9.11) define whether demand i is covered when the v^{th} facility of the p-facilities has been interdicted. This is an integer programming problem and modest sized problems can be solved by the use of an integer linear programming package. It should be noted that we have included the binary nature of the decision variables y_i and y_{iv} for completeness, but these restrictions can be relaxed by making them upper bounded variables (bounded by 1) and when solving to optimality all of these variables will be binary in value. It should also be noted that this type of model can be tightened somewhat by adding the following set of constraints (see O'Hanley and Church 2011):

$$\sum_{j=1}^{e} x_{jv} \geq \sum_{j=1}^{e} x_{jv+1} \quad \forall v = 1, 2, \ldots, p-1 \quad \text{and} \quad \forall e \in J \tag{9.16}$$

We assume here for convenience that each of the sites have been numbered from 1 to n. The purpose of this type of constraint is to eliminate a number of equivalent solutions that differ only by which site is assigned a specific position v. Essentially, the position assigned for a given site choice must be in the same order as the sites have been numbered. For example, the lowest ordered position of facilities in the solution, $v = 1$, will be assigned the facility with the lowest valued index of j that is selected for the configuration. The value of including this set of constraints increases as the value of p increases.

The above model handles the case when we are locating a system of facilities while dealing with the risk of losing one facility. It does not handle the case when there is a reasonable risk that more than one facility may be lost. This is addressed in the next section.

9.4 Planning for Greater Levels of Disruption

If the risk that more than one facility might be simultaneously interdicted is high, then one might be interested in designing the coverage system so that coverage is maximized by the original design, but also continues to be as high as possible when being subject to an r-facility interdiction attack. If we assume that interdiction might occur up to a level of three facilities, then our problem might be to maximize coverage while locating p-facilities, while also attempting to ensure that the resulting coverage level is as high as possible after our configuration is hit with the worst possible case of an attack of three-facilities. We know the if we design a system for being as resilient as possible when subject to the loss of three facilities, it should also provide a level of coverage that is higher than that associated with the loss of three facilities, when losing only one or two facilities. We can call this a robust design problem. To demonstrate that designs can be robust, consider the solution that was generated by Church et al. (2004) that is depicted in Fig. 9.3. This solution uses five facilities and covers 567 out of 640 when using a maximal service distance of 10. When one compares it to the optimal solution to this problem depicted in Fig. 9.1, one can see that coverage is not optimal but is still relatively high (567 vs. 609). Note that the two solutions (Fig. 9.1 and Fig. 9.3 have only one facility location in common (site 36). If the configuration in Fig. 9.3 is interdicted using the rIC model where r = 2, the resulting solution covers 475. This solution is depicted in Fig. 9.4. This rIC solution continues to cover 475 even after optimal interdiction which destroys 2 facilities. That is, there exist solutions which can maintain high levels of coverage even after optimal interdiction. Design models help to find such robust solutions.

This problem can be defined as a two person Stackelberg game, involving a leader and a follower. In this case the leader makes the decision as to where to place p-facilities. The follower then decides which facilities will be interdicted. For the follower, the objective is to cause the greatest disruption by solving an r-interdiction covering problem, that is, reduce coverage to the greatest extent possible upon interdicting r-facilities. Thus, the leader needs to decide where to place the p-facilities in order to provide a high level of coverage and do so in such a manner

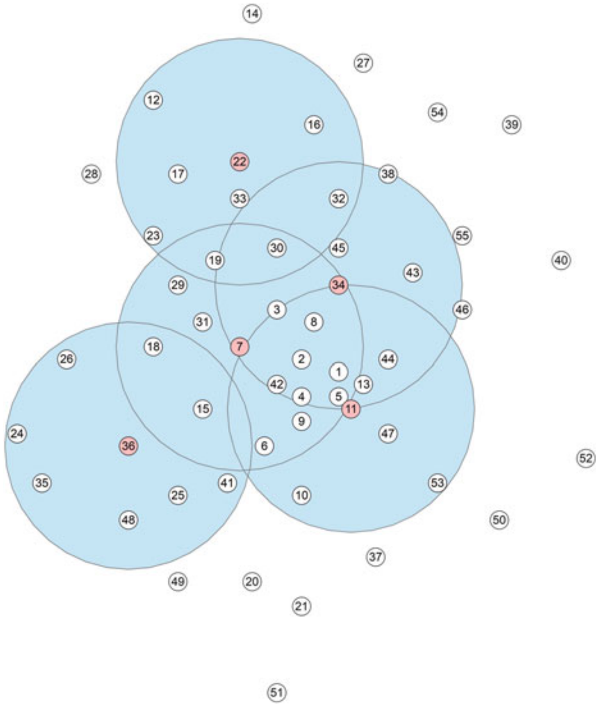


Fig. 9.3 A solution to the MCLP problem when locating 5 facilities where the maximal service distance is 10. This solution is not optimal, however coverage is still relatively high, where it covers 567 when an optimal solution covers 609 (see Fig. 9.1)

as to prevent as much as possible the damage to system coverage that can be inflicted by the follower.

Such a problem is best posed as a bi-level optimization problem. Bi-level optimization problems can be quite difficult to solve, as the complexity is compounded by the very nature of two antagonists in competition, one to do the greatest good in coverage, and the other to do the greatest harm in reducing coverage. It should be noted that the problem posed in the previous section can also be viewed as a two-person Stackelberg game, however, it was formulated as a single level optimization problem. For that problem, there are only p-ways that one facility can be eliminated. If the value of p is relatively small, possible loss levels associated with each of the p-cases can be tracked individually, without the overall problem becoming intractable. But the number of combinations of interdicting r facilities out of a configuration of p-facilities increases substantially as both r and p increase, which makes the type of formulation in the previous section unworkable. It also means that the model developed here is a generalization of the model given in the previous section. Using the terminology of the previous section we can call this a p/p-r Maximal Covering Design (p/p-r MCD) problem.

To formulate the bi-level optimization problem, we need to introduce a few new or redefined terms:

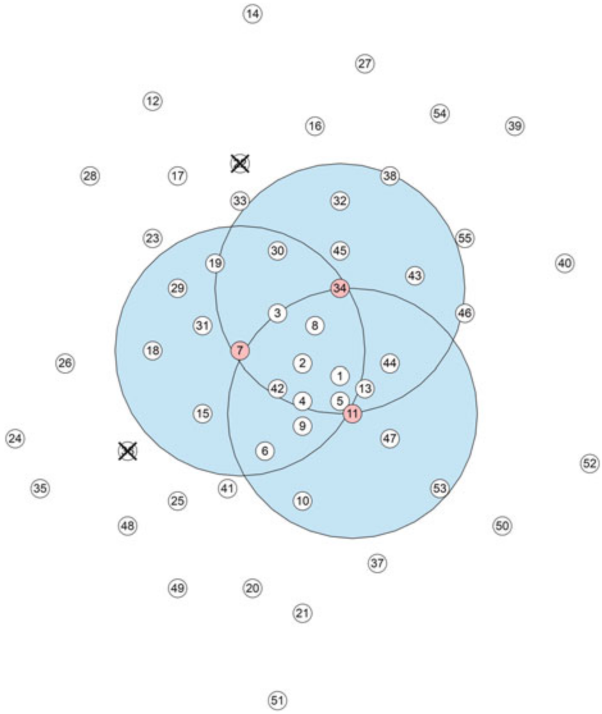


Fig. 9.4 An optimal rIC solution when interdicting the solution presented in Fig. 9.3 where $r = 2$. The solution involves the remaining sites of 7, 11, and 34 where coverage has been reduced to 475 out of 640. This interdicted solution outperforms the interdicted solution depicted in Fig. 9.2

z^r = the amount of coverage left after the worst case of losing r facilities occurs

$$y_{ir} = \begin{cases} 1, & \text{if demand at } i \text{ is still covered after } r \text{ facilities have been interdicted} \\ 0, & \text{otherwise} \end{cases}$$

We can now formulate the bi-level optimization model as follows (O'Hanley and Church 2011):

$$p/p - r\,\text{MCD}: \qquad \text{Maximize } w_1 \sum_{i \in I} a_i y_i + w_2 z^r \tag{9.17}$$

Subject to:

$$\sum_{j \in J} x_j = p \tag{9.18}$$

$$\sum_{j \in N_i} x_j \geq y_i \quad \forall i \in I \tag{9.19}$$

$$\text{Minimize } z^r = \sum_{i \in I} a_i y_i^r \tag{9.20}$$

Subject to:

$$\sum_{j \in J} s_j = r \tag{9.21}$$

$$y_i^r + s_j \geq x_j \quad \forall i \in I \quad \text{and} \quad j \in N_i \tag{9.22}$$

$$s_j \in \{0, 1\} \quad \forall j \in J \quad \text{and} \quad y_i^r \in \{0, 1\} \quad \forall i \in I \tag{9.23}$$

$$y_i \in \{0, 1\} \quad \forall i \in I \quad \text{and} \quad x_j \in \{0, 1\} \quad \forall j \in J \tag{9.24}$$

The *p/p* − *r* MCD as described above is a bi-level optimization problem. There are two objectives: one that represents the leader (9.17) and the other that represents the follower (9.20). The top level problem is comprised of the objective (9.17) and constraints (9.18) and (9.19) along with the leader's variables in conditions (9.24). Constraints (9.18) and (9.19) represent the constraints of a maximal covering location problem, where constraint (9.18) specifies that exactly *p*-facilities will be located and constraints (9.19) define what has been covered. If one or more facilities are located within the coverage range of demand *i*, then the sum on the left hand side of the inequality is 1 or greater. This allows the coverage variable y_i to be one in value. Otherwise, nothing covers demand *i* and the variable y_i is forced to be zero, indicating that demand *i* has not been covered. The first term of the objective (9.17) represents maximizing the level of coverage provided by the *p*-facilities that the leader locates.

The follower is represented by the objective (9.20), constraints (9.21) and (9.22), and the appropriate restrictions on the follower's variables given in conditions (9.23). The follower's problem is highlighted by the grey box. It is an optimization problem in and of itself and can be easily recognized as a *r*IC. The basic idea is that whatever the leader decides (based upon the values of the x_j variables), the follower is allowed to make decisions that impact the leader's decisions the most by solving a *r*IC. Constraints (9.22) differ from what is given in the *r*IC [constraints (9.3)], because in the *r*IC the facilities already exist. That model did not require location variables x_j. But, in this expanded design problem the *r*IC follower problem must respond to what is being located. Essentially, if a facility is located at a site *j* (i.e. $x_j = 1$) that covers some demand *i*, either that demand must be considered covered (i.e. $y_i^r = 1$) or that facility has been interdicted (i.e. $s_j = 1$). Since there is one such constraint for each facility site that covers a given demand *i*, demand *i* will not be covered when every site that has been selected that covers demand *i* has been interdicted or when no facility has been located to cover that demand. Note that the follower's objective is to minimize total coverage, so unless a y_i^r is forced to be at a value of 1, it will be valued at zero. The only way that such variables, y_i^r, will be one is when a facility has been located which covers that demand and when that facility has not been interdicted.

You can think of the *r*IC as representing an immediate response on the follower's part, which impacts the objective of the leader (9.17). The objective of the leader (9.17) maximizes a weighted sum of the coverage generated by the leader which is provided by all *p*-facilities as well as the sum of what remains covered once the follower has removed *r* of these *p*-facilities. The only way the leader can maximize the value of this objective is to find that set of *p*-facilities that provides a high level of coverage as well as provides as high as possible the level of coverage that remains after the follower's decision. The two terms of this objective are weighted in terms of their importance. If the risk is high in losing *r*-facilities then the importance weight for the second term of the objective might be relatively high as compared to the weight for the first term. Otherwise, the objective weight for the first term of the objective will probably be weighted at a higher value relative to the second term. It is also possible to solve for solutions which optimize the tradeoff between coverage of *p*-facilities and coverage after *r*-interdiction. Overall, this problem can be quite useful in designing robust maximal coverage systems.

The major problem is that most bi-level optimization problems are quite difficult to solve. Although structurally, they look like integer-linear programming problems. The real issue is that the leader (max) and follower (min) have competing objectives. When attacker problems do not contain integer variables, it is possible to represent the attacker as a dual optimization problem (max), allowing the two levels to be combined into single level, optimization problem. This cannot be done in most cases when the attacker problem contains integer variables, which in this instance it does. In general, specialized solution processes are necessary for this type of problem (O'Hanley and Church 2011) and most defender-attacker problems (Alderson et al. 2011), a topic which we do not delve into here. An expanded form of the problem has been proposed by Başdere et al. (2013) which involves the case where disruption is not complete.

9.5 Issues When Providing Complete Coverage

Before we move on to the next section where we discuss the possible loss of a network link and its impact on coverage, it is important to address the problem of how we might plan a system when we want to ensure coverage to all demand. Of course, without considering any issues of disruption, we can use the LSCP defined in Chap. 2 to minimize the number of needed facilities and position them in such a manner as to cover each and every demand. If we want to ensure that all demands are covered after the loss of one facility, then we could solve an LSCP where we require all demand to be covered at least twice. If each and every demand is covered twice, then losing any one facility in the system will reduce the level of coverage for those demands that were covered by the interdicted facility, but remaining coverage will still be at least once for all demands after interdiction.

We can also use forms of the models given in Sects. 9.3 and 9.4 to address issues of resilience associated with LSCP solutions. Let's suppose that we wish to design a system that covers all demand, and is relatively resilient in providing high levels of coverage after the possible loss of 1 facility. To address this type of problem, we first need to solve an LSCP problem. This yields the number of needed facilities in providing cover for all demand. Let's call this value $p_{\min}$. Thus, if our design should cover all demand before interdiction, we need to use at least $p_{\min}$ facilities. The model in Sect. 9.3 can be modified to design for complete coverage by removing all of the y_i variables from the problem. This essentially converts the problem objective to maximizing coverage after the loss of any facility. We also need to modify constraint (9.10) into the following:

$$\sum_{v=1}^{p} \sum_{j \in N_i} x_{jv} \geq 1 \quad \forall i \in I \tag{9.25}$$

This modified constraint will then ensure that the set of $p_{\min}$ located facilities covers all demand before any interdiction is considered. Thus, altogether these modifications will allow us to generate a solution to the LSCP which is as resilient as possible to the loss of one facility.

It is also possible to design systems that provide complete coverage and are as resilient as possible in providing coverage after the interdiction of r-facilities. This we can do using a modified form of the model given in Sect. 9.4. First we need to ensure that there are enough resources to ensure that all demands can be covered before interdiction. This requires setting $p \geq p_{\min}$ Second, the constraint (9.19) can be set to be greater than or equal to one, instead of greater than or equal to y_i. This will ensure that the configuration will cover all demand before the loss of r-facilities. Third, the first term of the leader's objective can be dropped as all demand will be covered by the initial design before interdiction. By making these changes, we can find a solution to the LSCP that provides as much coverage as possible after the loss of r facilities.

9.6 Coverage Loss When Supporting Infrastructure Is Lost

Military interdiction modeling has a long history in finding the most important links (road segments, bridges, etc.) that when removed cause the greatest impact to a moving army and its logistical support system. For example, the loss of a bridge may mean a detour is necessary and that may make that supply route longer. Corley and Sha (1982) were the first to address this problem which involves maximizing the length of the shortest path from a pre-specified origin to a pre-specified destination associated with interdicting a set number of arcs. Wollmer (1964) addressed the problem of interdicting a modest number of arcs in order to maximize the total cost of network flow between multiple sources and sinks. Both of these problems can be classified as Stackelberg games. The leader here is the attacker, who plans to take out a set of links. The follower will respond by routing their flow along the shortest existing path or across the network in order to minimize the overall cost of network flow between multiple sources and sinks. We can view this type of problem within a covering context when noting that coverage may be lost when a response across a system cannot be made within the maximal desired service time because some link has been destroyed. For example, it could happen that the shortest route from a demand i that is covered by an existing facility at j may be disrupted, because an arc on the shortest path between i and j is destroyed. Coverage of demand i may or may not be lost, due to this arc being interdicted. Demand i would still be covered if there was an alternate, disruption-free, path to facility j that can be traversed within the maximal service time, or an alternate, disruption-free, route exists to some other facility that can be traversed within the maximal service time. The basic idea is that after attack, routes to and from facilities and demand can be adjusted so that the greatest possible coverage levels remain. To determine what is covered based upon a selected set of interdicted arcs, consider the following notation, some of which is new and some of which is modified from what has been used previously in this chapter:

N = the set of nodes of the network, numbered from 1 to n
i,j = indices used to reference nodes of the network as either demand or facility locations, respectively
I = the set of nodes that are demands
Ω = the set of nodes that house an existing facility
d_{ij} = the shortest distance between demand i and an existing facility at j
S = the maximal service distance or time
$M_i = \{k \in N | d_{ik} \leq S\}$, the set of nodes that are within S distance before interdiction
$N_i = \{j \in F | d_{ij} \leq S\}$, the set of existing facility sites that are within S distance before interdiction
g,h,k = indices used to reference nodes of the network in general
A = the set of arcs (g, h) that define the network, where $g < h$
α_{gh} = the distance or travel time for arc (g, h) in the direction of g to h
α_{hg} = the distance or travel time for arc (g, h) in the direction of h to g

$$s(g,h) = \begin{cases} 1, & \text{if arc connecting } g \text{ and } h \text{ has been interdicted where } g < h \\ 0, & \text{otherwise} \end{cases}$$

$$t^i_{gh} = \begin{cases} 1, & \text{if arc } (g,h) \in A \text{ is on a potential coverage path for demand } i \\ 0, & \text{otherwise} \end{cases}$$

It is assumed that each undirected arc in the network is represented in both directions in terms of the path variables, t^i_{gh} and t^i_{hg}, one for each direction, but interdiction of that arc is represented by only one variable, $s(g,h)$. We can track whether a route exists to a given demand i from an existing facility $j \in N_i$ after interdiction with the following model structure:

$$\sum_{(i,g)\in A} t^i_{ig} + y^L_i = 1 \tag{9.26}$$

$$\sum_{(g,k)\in A} t^i_{gk} - \sum_{(k,h)\in A} t^i_{kh} = 0 \quad \forall k \in M_i \quad \text{where } k \neq i \quad \text{or} \quad k \notin N_i \tag{9.27}$$

$$\sum_{j\in N_i} \sum_{(h,j)\in A} t^i_{h,j} + y^L_i = 1 \tag{9.28}$$

$$\sum_{\substack{(g,h)\in A/\\ g,h\in M_i}} \alpha_{gh} t^i_{gh} \leq S \tag{9.29}$$

$$t^i_{gh} \leq 1 - s(g,h) \quad \forall (g,h) \in A \quad \text{where } g,h \in M_i \quad \text{when } g < h \tag{9.30}$$

$$t^i_{gh} \leq 1 - s(h,g) \quad \forall (h,g) \in A \quad \text{where } g,h \in M_i \quad \text{when } h < g$$

The structure of conditions (9.26) through (9.30) specify two types of conditions, the existence of a path from i to an existing facility at j within the maximal service distance or time S after interdiction, or that such a path does not exist. If a route does exist for i to a facility at $j \in N_i$, this route will be represented by the variables t^i_{gh} that equal one in value. Conditions (9.26)–(9.28) specify that the route must leave demand node i, use arcs that are within the distance S of node i before any interdiction, and reach a node $j \in N_i$. Thus, the search for a path from node i to a facility is constrained to be within a subset of arcs and nodes that were within the travel time or distance S before interdiction takes place. Note that constraints (9.27) specify that if the path enters node k where $k \notin N_i$ and $k \neq i$, then the path must leave that node. Conversely, if the path does not enter node k, it is prevented from leaving that node. These conditions are written for all possible intermediate nodes of the path (that are no further than S in length or time) for demand i, and are often called node balance or flow balance constraints. Constraint (9.29) specifies that this path must have a length less than or equal to the coverage standard. A key feature in this constraint set is that if an arc is interdicted, i.e. $s_{gh} = 1$, it is prevented from being used in constraint (9.30). What if a path of covering length does not exist that is shorter than S, then constraints (9.26) and (9.28) must be met by forcing y^L_i to be equal to one in value, indicating that coverage is lost due to interdiction.

Given that we can specify whether a given demand node is covered or not given a set of interdicted arcs, we now have the elements that we can structure the following problem:

Disrupt coverage of an existing system of facilities, F, as much as possible by interdicting r arcs.

We will call this the r-interdicted arcs covering problem, where the objective is to maximize the loss of coverage upon the interdiction of r-arcs (r-AIC). We can structure a model for this problem using the conditions formulated above as follows:

$$r - \text{AIC}: \qquad \text{Maximize } Z \tag{9.31}$$

Subject to:

$$\sum_{(g,h)\in A} s_{gh} = r \tag{9.32}$$

$$Z = \text{minimize } \sum_{i\in I} a_i y_i^L \tag{9.33}$$

Subject to: (9.26)–(9.30) for each $i \in I$

$$y_i^L \in \{0, 1\} \quad \forall i \in I \tag{9.34}$$

$$t_{gh}^i \in \{0, 1\} \quad \forall (g, h) \in A \quad \text{where } g, h \in M_i \tag{9.35}$$

$$s_{gh} \in \{0, 1\} \quad \forall (g, h) \in A \tag{9.36}$$

The r-AIC is a bi-level model. The top level represents the leader or attacker who attempts to disrupt the coverage being provided as much as possible by deciding which arcs to be interdicted. The limit of disruption is specified in constraint (9.32). Given the set of disrupted arcs, the follower or system operator attempts to reassign coverage routes so that as much of the demand is covered, by minimizing the amount of coverage that is lost. Altogether, this model is a very difficult problem to solve and possibly the only competitive approaches that will be developed to solve this problem will involve heuristic strategies.

9.7 Defensive Coverage

Coverage can be provided to a demand in a number of different ways. Sirens cover those areas which are within earshot. Communication towers provide coverage to those demand that can pick up a signal. Cameras cover what is within view and within a certain distance and resolution. For these types of covering systems, interdiction is meaningful in terms of facilities as these are the central elements that provide coverage. Other coverage systems are defined based upon being close enough so that travelling between a facility and a demand can be done within a maximal covering distance or time. In this section we focus on the latter definition of cover, that is, coverage access that is defined by a network, for instance access travel that occurs along a road network. In the previous section we modeled how to disrupt an existing covering system the most by interdicting a subset of the arcs. That problem is quite complex in and of itself without extending this to include additional elements of decision making. For example, consider the following problem:

Maximize the largest possible amount of coverage by locating a set of p-facilities, where coverage is defined by the worst case interdiction of r-arcs while using that set of p-facilities.

Like the problems described in Sects. 9.3 and 9.4, we can classify this as a maximal covering design (MCD) problem, while subject to an interdiction or r-arcs, or an $|A|/|A|$-r-MCD problem where |A| represents the number of arcs in the network. This problem involves a leader who makes decisions as to where to place the facilities, a follower who then decides which arcs to interdict, followed by the leader who determines the best level of coverage possible by routing each demand along the post interdicted network to their closest facilities. When demands can be served within the coverage standard by the shortest routes they are covered and when that is not possible, they are not covered. This general problem is a tri-level optimization problem, and fits nicely within a *defender-attacker-defender* model framework of Alderson et al. (2011). A very limited form of this problem, where $r = 1$ has been suggested by Berman et al. (2009). That is, their problem is to design a system of p-facilities that is as robust as possible in providing coverage while assuming that the worst case of losing 1 arc will occur. Such a problem would be very limited in application as compared to the general problem described above. Their limited problem was called the Defensive Maximal Covering Problem. We prefer to call it a facility location design problem, because there really is not any defensive measure being involved, like protecting a facility as compared to the problem of design in an attempt to keep operations as robust as possible. Overall the area of robust design, taking into account the possible impacts of lost road segments and or even facilities is just in the beginning phases. In the next section we turn our attention to defensive actions like the protection of an existing system.

9.8 Facility Fortification

There are a number of possible defensive actions that can be taken in order to protect a facility from harm, whether by some man-caused action or by some natural event. Facilities can be hardened to prevent incursion from a hostile force, built to withstand major earthquakes or firestorms, provided with backup generators, and even moved to higher ground. Such actions may make a facility more resilient in case of a disaster, as well as help to ensure that the entire system has not failed and still maintains some level of service.

Assume that we have a system that provides some type of service. It already exists and is based upon p-different facilities, represented by the set Ω. We will also assume that each of these facilities suffers from the same type of vulnerability, but that we can harden them so that they are fully protected from this potential harm. For instance, by changing the security system, we may make it impossible for an attack on the facility to be successful. This could apply, for example, to major substations where we can easily thwart an attack like the Metcalf substation in California.

Assume that further analysis suggests that attacks may involve strikes at r-different facilities and that we have resources to harden or fully protect only q of these facilities. Within the context of a coverage based system, we can define the following protection problem:

Maximize the largest amount of coverage possible after the interdiction of r-facilities by optimally fortifying q of the p-facilities.

The idea is that the defender has the resources to fully harden or protect q of the existing p-facilities. The decision is which of these facilities should be hardened so that the worst case loss of r of the p–q unprotected facilities results in the smallest loss of coverage. The defender or leader must decide which facilities to harden and the follower or attacker then decides which of the remaining unhardened facilities to strike so that as much coverage as possible is lost. Thus, the defender wishes to keep the attackers damage as low as possible by minimizing what losses will be produced by the attacker. We will call this problem the r-Interdiction Covering Problem with Fortification (rICF).

We can model this problem as a bi-level optimization problem reusing most of the notation defined above, except that we now need to introduce an additional decision variable:

$$f_j = \begin{cases} 1, & \text{if the existing facility at } j \in \Omega \text{ is fortified} \\ 0, & \text{otherwise} \end{cases}$$

The rICF model can then be formulated as:

$$r\text{ICF}: \qquad \text{Minimize } Z \tag{9.37}$$

Subject to:

$$\sum_{j \in \Omega} f_j = q \tag{9.38}$$

$$\text{Z} = \text{Maximize } \sum_{i \in I} a_i y_i^L \tag{9.39}$$

Subject to:

$$\sum_{j \in F} s_j = p \tag{9.40}$$

$$s_j + f_j \le 1 \quad \forall j \in \Omega \tag{9.41}$$

$$y_i^L \le s_j \quad \forall i \in I \quad \text{and} \quad j \in N_i \tag{9.42}$$

$$s_j \in \{0, 1\} \quad \forall j \in \Omega \tag{9.43}$$

$$y_i^L \in \{0, 1\} \quad \forall i \in I \tag{9.44}$$

$$f_j \in \{0, 1\} \quad \forall j \in \Omega \tag{9.45}$$

The objective, (9.37), seeks to minimize the amount of coverage that is lost due to the interdiction of an attacker, by altering what facilities would be subject to interdiction by allocating fortification resources. Constraint (9.38) specifies that the hardened facilities will be limited to q, which means that there will be p–q facilities left that are vulnerable to interdiction. The objective and constraints within the gray box represent the lower level problem of the attacker. The attacker's objective (9.34) is to maximize the coverage that is lost based upon what is interdicted. The attacker is limited to interdicting only r facilities in constraint (9.39). Constraints (9.41) prevent a fortified facility location from being interdicted and constraints (9.42) specify whether interdiction has resulted in the loss of coverage for each demand i. The remaining constraints specify the binary restrictions on the decision variables. Like many bi-level optimization problems, including the ones given in previous sections of this chapter, this problem can be quite hard to solve. The above formulation is similar to that of Dong et al. (2010) where they propose an implicit enumeration scheme to solve this problem.

Although we have not usually delved into the specifics of solution approaches in this book, one approach that has been proposed to solve a related problem is particularly applicable here and is itself based upon a covering problem. This related problem is the interdiction median problem with fortification (Church and Scaparra 2007). Scaparra and Church (2008) reasoned that in many cases that the number of interdicted facilities would be relatively small in number. If this was the case, it would be quite easy to enumerate all ways of interdicting a set of p-facilities. For example, if you have 25 facilities and interdiction could involve the destruction of 4 of the facilities, then there are only 12,650 different ways of interdicting these 25 facilities. We will call this the set of possible interdictions patterns (SPIP). Their solution approach was based upon the following observation:

Any pattern of interdicted facilities in the SPIP would be thwarted by the protection of any one of the facilities in that pattern, and this would mean that some other interdiction pattern on the list would generate a higher level of disruption.

That is, a pattern of interdiction can be prevented from occurring if at least one of the facilities in that pattern has been fortified or hardened. If we do harden one of the facilities in a pattern, then that pattern will no longer be as fruitful to an attacker as some other pattern.

Let us order the patterns in SPIP in order of decreasing harm or lost coverage, called the ordered list Θ. Thus, the top pattern on the list results in the highest loss of coverage and the pattern at the bottom of the list is associated with the least disruption or loss of coverage, due to the interdiction of r-facilities. If we do not protect one of the facilities that appears in the pattern at the top of the list, then our fortification plan will not thwart that interdiction pattern and the highest level of disruption will occur. Consider the following notation:

θ_k = the set of facilities associated with the k^{th} interdiction pattern of ordered list Θ, where $k = 1, 2, 3, \ldots, K$

$$\Psi_k = \begin{cases} 1, & \text{if all of the interdiction patterns from 1 to } k \text{ on list } \Theta \text{ are thwarted} \\ 0, & \text{otherwise} \end{cases}$$

By reusing most of the notation introduced in previous models of this chapter, we can formulate a special form of the *r*ICF problem as follows:

$$\text{Maximize} \quad \sum_{k=1}^{K} \Psi_k \tag{9.46}$$

Subject to:

$$\sum_{j \in \theta_k} f_j \geq \Psi_k \quad \forall k \tag{9.47}$$

$$\sum_{j \in \Omega} f_j = q \tag{9.48}$$

$$\Psi_k \geq \Psi_{k+1} \quad \forall k \leq K - 1 \tag{9.49}$$

$$f_j \in \{0, 1\} \quad \forall j \in \Omega \tag{9.50}$$

$$0 \leq \Psi_k \leq 1 \quad \forall k \tag{9.51}$$

The above model is an "ordered" MCLP. An interdiction pattern k is considered covered (or thwarted) when one of the facilities in configuration θ_k has been fortified. It is ordered in that pattern $k + 1$ is not considered thwarted, unless pattern k and others on the list above k have been thwarted as well. The objective (9.46) involves maximizing the number of interdiction patterns in the list Θ in the order of importance that can be thwarted. Essentially, this means that at some point as one moves down the list, the next most important interdiction pattern has not been thwarted and defines the amount of coverage lost by interdiction after the best use of fortification resources. Constraint (9.47) defines that pattern k can be considered thwarted, if some facility that is in pattern k has been fortified. Constraint (9.48) specifies that there will be only q fortified facilities, and constraints (9.49) specify that pattern $k + 1$ can be considered thwarted until pattern k has. It should be recognized that when fortifying facilities in the most disruptive patterns at the top of the list, any other patterns that utilize one of those fortified facilities, regardless of where they are on the ordered list, have been thwarted as well. But, the importance of fortification is that it does not matter if some less important patterns have been thwarted when some pattern higher on the list has not been thwarted. That is, the highest pattern on the list that has not been thwarted by fortification represents greatest impact of an attacker after fortification. That is why the order of the list is important and it is important to identify the highest position on the list that has not been fortified against. The objective then represents getting as far down the list in terms of thwarting each pattern. Scaparra and Church (2008) use this model to fortify a facility system that does not involve a covering system, but this is equally applicable here (see also Murray and Grubesic 2013). It is also important to note that there are special ways in which this model can be applied to reduce the size of the model being solved by strategically solving a sequence of problems. In addition, heuristic solution approaches have been proposed to either protect a system or interdict a system (Farahani et al. 2014; Mahmoodjanloo et al. 2016).

9.9 Applications

The models and approaches for making robust designs, fortifying facilities, and planning for possible disruption can be useful in preparing for both man-made and natural disasters. The interdiction models presented here can be helpful in estimating possible impacts and providing an understanding of what facilities or pieces of infrastructure are critical. The greatest impact that these models can have is where they are used to develop an understanding of just what are most critical for a system's operation. Often, such an understanding comes after experiencing a disaster, rather than being able to model and forecast such issues in advance. For instance, one area that is of increasing concern is the provision of emergency aid and shelters during major catastrophes. For example, Hurricane Mathew in 2016 flooded many roads and highways in the Carolinas and Hurricane Harvey in 2017 flooded many roads and highways in Texas. These events have shown that emergency resources were hard to deliver with such levels of transport system disruption. Even local resources and goods at retail centers were stranded, because access to and from these facilities were limited. The optimal location of emergency resource storage and distribution facilities is of immense value during such catastrophes. Fortifying facilities to withstand high winds, building them so that their operations are above flood level, and ensuring that their life support elements, like electricity, can be provided internally is of utmost importance. Critical elements in such systems can often be assumed to work, but their original design makes them vulnerable in the very instances that they are needed. For example, the disaster at the Fukushima Nuclear reactor was exacerbated by the fact that the emergency power generator was in a basement that flooded early in the tsunami disaster. This meant as the power plant systems of the nuclear reactors were lost and without operating backup generators, the control center was almost useless in limiting damage to the nuclear reactor. After tropical storm Allison hit Houston, Texas in 2011, the city experienced a blackout and medical centers were flooded, which forced many hospitals to evacuate all of their patients. Since that storm Houston hospitals have moved their electrical vaults and backup generators from basements to places that are considered to be above flood levels in addition to fortifying their properties against flood waters. This means that location, fortification and robust design are all elements that have an impact on a system's operation when it is needed the most. Although few interdiction and fortification models that have been vetted in the literature have been applied, since 9/11, there are notable examples where models have been applied to estimate the potential of worst case disasters in life-line support systems, like electricity and water. Maliszewski (2011) reviews some of these models with respect to their use in analyzing issues of homeland security. An interesting application of fortification can be found in Choi and Suzuki (2013) that involved a maximal covering approach to fortify a retail system. In addition, O'Hanley et al. (2007a, b) develop two models that use the concepts of expected covering and interdiction in designing a biological reserve system, that involves buying specific sites for a reserve system to hedge bets on losing critical species. Murray and Grubesic (2013) examined telecommunications infrastructure with respect to covering the most significant interdiction scenarios.

9.10 Summary and Concluding Comments

The research literature on vulnerability of facility based systems began with the work of Church et al. (2004). In that work two different models were introduced, the *r*-interdiction covering model and the *r*-interdiction median problem. Since that original work, modeling interdiction has expanded to a number of other systems, including hubs (Lei 2013), warehouses (Scaparra and Church 2012), telecommunications (Murray and Grubesic 2013), and electrical power systems (Salmeron et al. 2004), just to name a few. A recent review of this literature can be found in Murray and Grubesic (2007) and Scaparra and Church (2015), in addition to a more general review of disruption in supply systems by Snyder et al. (2006). Attention has also focused on problems of robust design and fortification.

References

Agnetis A, Grande E, Mirchandani PB, Pacifici A (2009) Covering a line segment with variable radius discs. Comput Oper Res 36:1423–1436

Alderson DL, Brown GG, Carlyle WM, Wood RK (2011) Solving defender-attacker-defender models for infrastructure defense. Department of Operations Research, Naval Postgraduate School, Monterey, CA

Bao S, Xiao N, Lai Z, Zhang H, Kim C (2015) Optimizing watchtower locations for forest fire monitoring using location models. Fire Saf J 71:100–109

Bar-Noy A, Rawitz D, Terlecky P (2013) Maximizing barrier coverage lifetime with mobile sensors. In: European symposium on algorithms. Springer, Berlin, pp 97–108

Başdere M, Aras N, Altınel İK, Afşar S (2013) A leader-follower game for the point coverage problem in wireless sensor networks. Eur J Ind Eng 7(5):635–656

Bell JE, Griffis SE, Cunningham WA, Eberlan JA (2011) Location optimization of strategic alert sites for homeland defense. Omega 39(2):151–158

Berman O, Drezner T, Drezner Z, Wesolowsky GO (2009) A defensive maximal covering problem on a network. Int Trans Oper Res 16(1):69–86

Choi Y, Suzuki T (2013) Protection strategies for critical retail facilities: applying interdiction median and maximal covering problems with fortification. J Oper Res Soc Jpn 56(1):38–55

Church RL, Scaparra MP (2007) Protecting critical assets: the *r*-interdiction median problem with fortification. Geogr Anal 39(2):129–146

Church RL, Scaparra MP, Middleton RS (2004) Identifying critical infrastructure: the median and covering facility interdiction problems. Ann Assoc Am Geogr 94(3):491–502

Corley HW, Sha DY (1982) Most vital links and nodes in weighted networks. Oper Res Lett 1 (4):157–160

Dong L, Xu-Chen L, Xiang-Tao Y, Fei W (2010) A model for allocating protection resources in military logistics distribution system based on maximal covering problem. In 2010 international conference on logistics systems and intelligent management, vol 1. IEEE, Piscataway, NJ, pp. 98–101

Farahani RZ, Hassani A, Mousavi SM, Baygi MB (2014) A hybrid artificial bee colony for disruption in a hierarchical maximal covering location problem. Comput Ind Eng 75:129–141

Lei TL (2013) Identifying critical facilities in hub-and-spoke networks: a hub interdiction median problem. Geogr Anal 45(2):105–122

Mahmoodjanloo M, Parvasi SP, Ramezanian R (2016) A tri-level covering fortification model for facility protection against disturbance in *r*-interdiction median problem. Comput Ind Eng 102:219–232

Maliszewski PJ (2011) Interdiction models and homeland security risks. J Homel Secur Emerg Manag 8(1): Article 4

Murray AT, Grubesic T (eds) (2007) Critical infrastructure: reliability and vulnerability. Springer, Berlin

Murray AT, Grubesic TH (2013) Fortifying large scale, geospatial networks: implications for supervisory control. In: Laing C, Badii A, Vickers P (eds) Securing critical infrastructures and critical control systems: approaches for threat protection. Information Science Reference, Hershey, pp 224–246

O'Hanley JR, Church RL (2011) Designing robust coverage networks to hedge against worst-case facility losses. Eur J Oper Res 209(1):23–36

O'Hanley JR, Church RL, Gilless JK (2007a) Locating and protecting critical reserve sites to minimize expected and worst-case losses. Biol Conserv 134(1):130–141

O'Hanley JR, Church RL, Gilless JK (2007b) The importance of in situ site loss in nature reserve selection: balancing notions of complementarity and robustness. Biol Conserv 135(2):170–180

Pan A (2010) The applications of maximal covering model in typhoon emergency shelter location problem. In 2010 I.E. international conference on industrial engineering and engineering management (IEEM). IEEE, Piscataway, NJ, pp. 1727–1731

Parfomak PW (2014) Physical security of the U.S. power grid: high voltage transformer substations, report 7–7500 (R4304), Congressional Research Service, Washington, DC

Salmeron J, Wood K, Baldick R (2004) Analysis of electric grid security under terrorist threat. IEEE Trans Power Syst 19(2):905–912

Scaparra MP, Church RL (2008) An exact solution approach for the interdiction median problem with fortification. Eur J Oper Res 189(1):76–92

Scaparra MP, Church R (2012) Protecting supply systems to mitigate potential disaster: a model to fortify capacitated facilities. Int Reg Sci Rev 35(2):188–210

Scaparra MP, Church RL (2015) Location problems under disaster events. In: Saldanha-da-Gama F, Stefan N, Gilbert L (eds) Location science. Springer International Publishing, Cham, pp 623–642

Snyder LV, Scaparra MP, Daskin MS, Church RL (2006) Planning for disruptions in supply chain networks. In: Johnson P, Norman B, Secomandi N (eds) Models, methods, and applications for innovative decision making. INFORMS, Baltimore, MD, pp 234–257

Swain R (1971) A decomposition algorithm for a class of facility location problems. PhD dissertation, Cornell University, Ithaca, NY

Wall Street Journal (2014) Assault on California power station raises alarm for potential for terrorism

Wollmer R (1964) Removing arcs from a network. Oper Res 12(6):934–940

Zhang Z, Du DZ (2012) Radar placement along banks of river. J Glob Optim 52(4):729–741

Chapter 10
Coverage of Network-Based Structures: Paths, Tours and Trees

10.1 Introduction

Generally speaking, location analysis and modeling is often characterized as a network based problem. Classic location books have reinforced this characterization, such as that by Handler and Mirchandani (1979) titled *Location on Networks* and that of Daskin (1995) titled *Network and Discrete Location*, as well as many other books and papers (the seminal work of Hakimi 1965, likely influenced this too). In fact, location modeling tools in ArcGIS, as an example, are part of the Network Analyst toolbox and strictly require an underlying network in order to structure and solve associated location (coverage) models. However, as reflected in much of this book, there is no inherent need or assumption that location decision making be restricted to a network, even if attribute data is derived from a corresponding network. This chapter deviates from much of the book in this respect, and assumes the underlying representation of an analysis region, including demand and potential facility locations, is based on a network. Most network location problems involve the positioning of one or more facilities at nodes or along arcs in order to optimize a level of service or access. Often, however, site selection is restricted to the nodes of the network. In the early 1980s Morgan and Slater (1980), Current (1981), and Slater (1982) broke new ground within location science when they independently suggested that facilities may be represented by some type of network structure. Examples include a path, a tree, a tour, or even the network itself. Slater (1982) suggested that facilities "... can be of an extended nature, rather than occupying a single point of a network." They suggested, for example, that facilities may be modeled as network paths representing railroad lines, pipelines, or transit routes. Slater (1982) proposed that four classes of locational problems should be considered within the context of network location: point-serves-point, point-serves-structure, structure-serves-point, and structure-serves-structure (see also Slater 1981, 1983). In this chapter we address the problem of designing a structure to serve points. One of the first examples in the literature is that of Morgan and Slater (1980) who defined

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_10

the "core" of a tree network as the shortest path connecting two endpoints of the tree, minimizing the sum of distances to all other nodes of the tree.

Current (1981) and Current et al. (1984) were the first to introduce the concept of locating a path in order to cover a set of demand nodes. This problem was called the shortest covering path problem. This concept was extended to a multi-objective path problem that maximizes path coverage while minimizing path length (Current et al. 1985). Since this early work, others have extended this concept to that of a covering salesman tour (Current and Schilling 1989), a maximal covering tour (Current and Schilling 1994) and a maximal covering tree (Hutson and ReVelle 1989, 1993). In this chapter we present the basic models of path, tour and tree covering along with a discussion of how such models might be used and applied. We also describe a more general problem of network investment/extension that is oriented towards increasing access coverage.

10.2 Shortest Covering Path

A considerable portion of John Current's research focused on the location of a structure within a network in order to provide service to a set of demand nodes (see Current 1981; Current et al. 1984). In much of this work, he defined service as being effective for a given node when the "structure" was located within a maximal service distance of that node. The quintessential problem of his research was the shortest covering path problem (SCPP):

Identify the minimum cost path through a network beginning at an origin and ending at a destination where each demand node is within some predetermined coverage standard, S, of the path.

This problem involves a prespecified origin, p, and destination, q. The objective is to find the shortest path starting at an origin and ending at a destination providing coverage to all network nodes within a maximal service standard, either based on distance or travel time. A path then is a route connecting an origin to a destination. If travel is restricted to a network of nodes and arcs, the path would consist of movement through the network, starting at the beginning node and moving along arcs through other nodes in the network until the destination node is reached. Unique is this case, however, is that the path provides coverage to other elements of the network not on the path. This clearly combines the concepts of the location set covering problem (LSCP) and a path problem. In essence the path must traverse the network in such a manner that it passes within the maximal service standard, S, of each node. The requirement that all nodes must be covered by the path is equivalent to the requirement that all demand (or nodes) must be covered by a set of located facilities in the LSCP. Whereas the LSCP seeks to cover all demand with the fewest facilities, the objective of the SCPP is to minimize the length of the path. In a network-based context, one can classify the LSCP as a point-serves-point problem, where facility sites and demand areas are nodes, and the SCPP can be classified as a structure-serves-point problem, where the path is the structure and points are nodes representing demand.

Assume that we have a network of nodes and arcs. In the original work of Current et al. (1984) the network was considered undirected. To refine this somewhat, we will assume here that the network consists of a mix of undirected arcs and directed arcs. Without loss of generality, if two nodes i and j are connected by an arc that is undirected, it can be replaced by two directed arcs, one for each direction. Our formulation is based upon the following notation, much of which is consistent with previous chapters:

N = set of network nodes
i,j = indices used to refer to specific nodes of the network
p = origin node of the path
q = destination node of the path
V = any subset of N
$A = \{ (i,j) |$ nodes i and j are connected by an arc that can be traversed$\}$
$T_k = \{ i \mid (i,k) \in A\}$, set of nodes where an arc can be traversed directly to node k
$F_k = \{ j \mid (k,j) \in A) \}$, the set of nodes that can be reached directly from node k
d_{ij} = proximity attribute for travel between nodes i and j, where $(i,j) \in A$
S = maximal service distance or travel time standard
$C_k = \{j| \quad d_{jk} \leq S$ where $(i,j) \in A\}$
$$x_{ij} = \begin{cases} 1, & \text{if path includes arc } (i,j) \\ 0, & \text{otherwise} \end{cases}$$

Using this notation, we can formulate the Shortest Covering Path Problem (SCPP) as follows:

$$\text{SCPP1}: \qquad Minimize \sum_{i \in N} \sum_{j \in F_i} d_{ij} x_{ij} \tag{10.1}$$

Subject to:

$$\sum_{j \in F_p} x_{pj} = 1 \tag{10.2}$$

$$\sum_{i \in T_q} x_{iq} = 1 \tag{10.3}$$

$$\sum_{i \in T_k} x_{ik} - \sum_{j \in F_k} x_{kj} = 0 \quad \forall k \in N \mid k \neq p, q \tag{10.4}$$

$$\sum_{j \in C_k} \sum_{i \in T_j} x_{ij} \geq 1 \quad \forall k \in N \mid k \neq p, q \tag{10.5}$$

$$\sum_{i \in V} \sum_{j \in F_i \cap V} x_{ij} \leq |V| - 1 \quad \forall V \subset N \mid 2 < |V| < n - 1 \tag{10.6}$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in A \tag{10.7}$$

The objective (10.1) of the SCPP minimizes the total length of all arcs selected for the path. The path starts at node p and terminates at node q. Constraint (10.2) specifies that the path uses an arc which leaves the origin node p. Constraint (10.3) specifies that an arc is selected entering node q, ensuring the path reaches the intended destination. Constraints (10.4) are associated with all intermediate nodes of the network. These constraints establish that an intermediate node cannot be an origin or a destination. That is, Constraints (10.4) require that if an arc is traversed to an intermediate node k, then an arc must be selected that leaves node k. The opposite is also true. If no arc is selected that enters node k, then no arc can be selected which leaves node k. Constraints (10.5) establish that the path must pass through at least one node that is within the maximum service standard of each intermediate node k. Given that the path starts at node p and ends at node q, means that those two nodes are automatically covered. Thus, these two nodes are not needed as the proximity requirements are already satisfied by their inclusion on the path. Constraints (10.6) deal with possible subtours, a sub-path that forms a loop (or cycle[1]) without connection to the origin-destination path. Subtours maintain constraints (10.2–10.5), but do represent an infeasibility. Constraints (10.6) prevent a subtour from being part of any solution. They were originally formulated by Dantzig et al. (1954) in an integer programming formulation of the travelling salesman problem. Finally, the path variables are restricted to be binary integer in constraints (10.7).

10.2.1 Subtour Issues

Solving the above formulation for any application instance beyond a very small network is generally out of the question as the number of constraints (10.6) increases exponentially with an increase in the number of nodes. Complete enumeration would result in nearly 2^n constraints, where $|N| = n$. This represents the various combinations of nodes that exist to uniquely connect a subtour, expressed as:

$$\binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n-2} + \binom{n}{n-1} \tag{10.8}$$

[1]Formally defined, a cycle is a path with the same beginning and ending node.

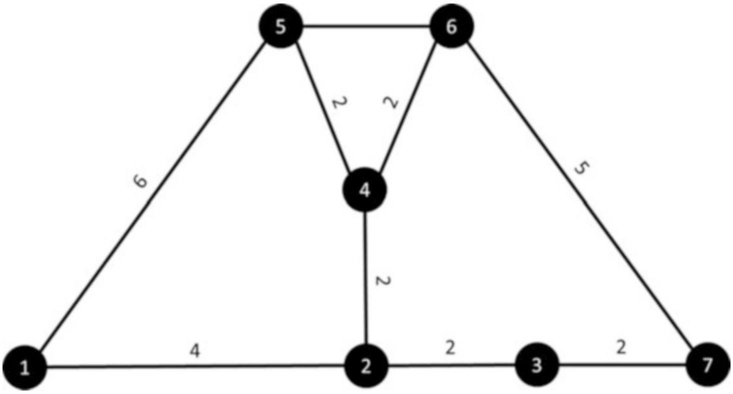


Fig. 10.1 Network of seven nodes and nine arcs (undirected)

This can be a very big number. For example, for a network with 15 nodes, $n = 15$, there would be 32,737 possible subtour constraints. A network of 50 nodes, $n = 50$, on the other hand would have over 1.1 quadrillion possible subtour constraints.

Current et al. (1994) formulated the SCPP problem with constraints (10.6) given above,[2] although the original work, Current et al. (1984), did not include them. The reason is that Current et al. (1984) implied in their discussion that such constraints would be used when necessary. They demonstrated that subtours do exist when solving the problem without constraints (10.6) and then used as few constraints (10.6) as possible to identify an optimal solution without any subtours.

To demonstrate the issue at hand, suppose that we apply this model without constraints (10.6) to an 8 node, 13 arc problem. This is depicted in Fig. 10.1. Assume that each arc can be traversed in either direction, with node 1 as the origin, node 8 as the destination, and a maximal service standard set at zero distance. This means that the shortest covering path needs to start at node 1, end at node 8, and visit each node. Solving the problem without constraints (10.6) yields the solution depicted in Fig. 10.2. Note that there is an unconnected loop or subtour, nodes 3 to 4 to 6 to 3. It is not connected to the path that travels from nodes 1 to 2 to 5 to 7 to 8. This is an optimal solution to the model without constraints (10.6), but it is not a feasible pathway. Thus, some or all of the constraints (10.6) are necessary to prevent infeasible subtours from occurring in a solution. This would mean up to 239 additional constraints to address subtour issues in this case.

There are three possible approaches to resolve/address the possibility of subtours occurring in a solution:

1. Include the entire set of constraints (10.6) in the model and solve

[2]Current et al. (1994) suggested that these constraints be written for all subsets V of N including N itself. But, a cycle comprised of the entire set of N nodes should never exist when solving the SCPP.

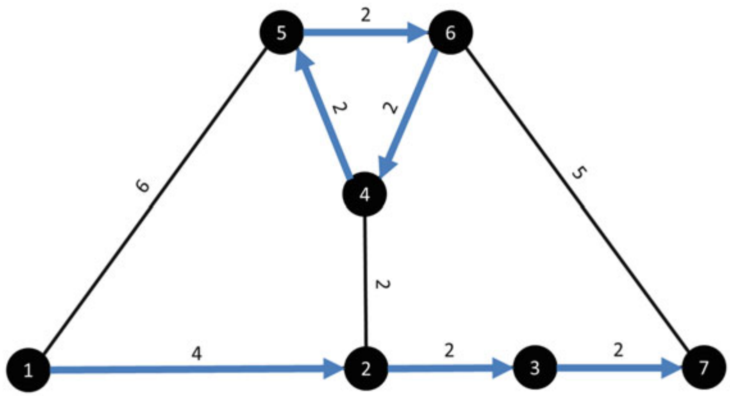


Fig. 10.2 A solution to SCPP1 when applied to the seven node, nine arc network [Constraints (10.6) not used]

2. Include only select constraints (10.6) in order to force a particular solution(s) to form feasible pathways without subtours
3. Reformulate the problem in a manner that constraints (10.6) are not necessary

The first approach is not practical as the number of constraints (10.6) grows exponentially as a function of the number of nodes, as noted above. Even for a very small network of 50 nodes the result is over 1.1 quadrillion constraints. Clearly, this would limit the size of any problem that could be solved. Current et al. (1984) realized this fact and suggested using the second approach. Basically, they first solve a problem without any constraints (10.6). For the eight node problem in Fig. 10.1, using the model without constraints (10.6) results in the solution containing a subtour, nodes 3 to 4 to 6 to 3 (Fig. 10.2). To prevent this subtour from occurring in a solution, the following constraint would be added:

$$x_{3,4} + x_{4,6} + x_{6,3} \leq 2 \tag{10.9}$$

This constraint prevents no more than two of the three arc selection variables $x_{3,\ 4}$, $x_{4,\ 6}$, and $x_{6,\ 3}$ from being one in value. This, in effect, would eliminate the occurrence of this specific loop in a solution. The basic idea is that after solving the model without constraints (10.6) either an optimal, feasible covering path is found or a solution is found comprised of a pathway along with one or more subtours. If any subtours appear in the solution, then one or more constraints (10.6) are added to prevent their occurrence and then the model is resolved. If this second solution contains subtours, then more constraints (10.6) are added, and the model is resolved. Essentially, one continues this process until the resulting solution is a connected path without any subtours. By solving this problem in an iterative manner, only those constraints (10.6) that are needed would be used.

One should note that constraint (10.9) is not an exact form of constraints (10.6), but rather a simplified form. Figure 10.3 depicts a hypothetical cycle of six nodes:

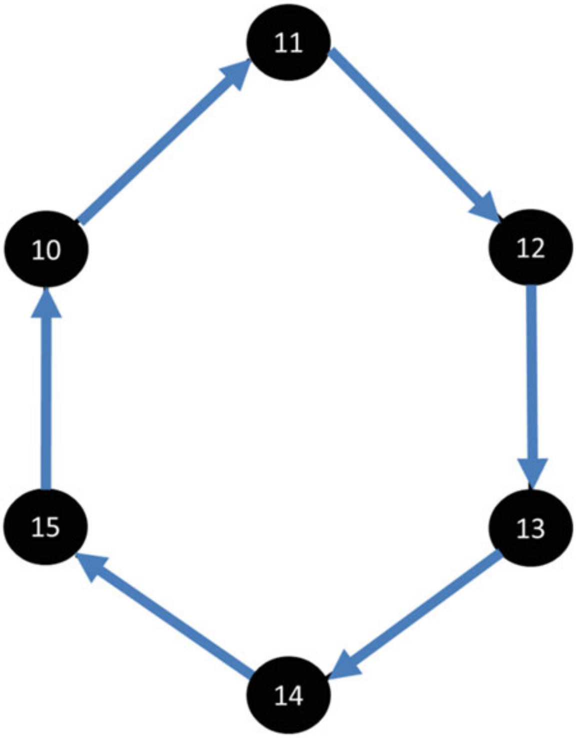


Fig. 10.3 Cycle consisting of six nodes

10 to 11 to 12 to 13 to 14 to 15 to 10. If this cycle was encountered in a solution, the Current et al. (1984) approach would involve adding the following constraint:

$$x_{10,11} + x_{11,12} + x_{12,13} + x_{13,14} + x_{14,15} + x_{15,10} \leq 5 \tag{10.10}$$

and then resolving the model. Of course, resolving this problem, without other stipulations, would probably result in a solution involving the same loop of nodes, but where the direction of the cycle would be reversed. Technically, this would require another constraint, namely:

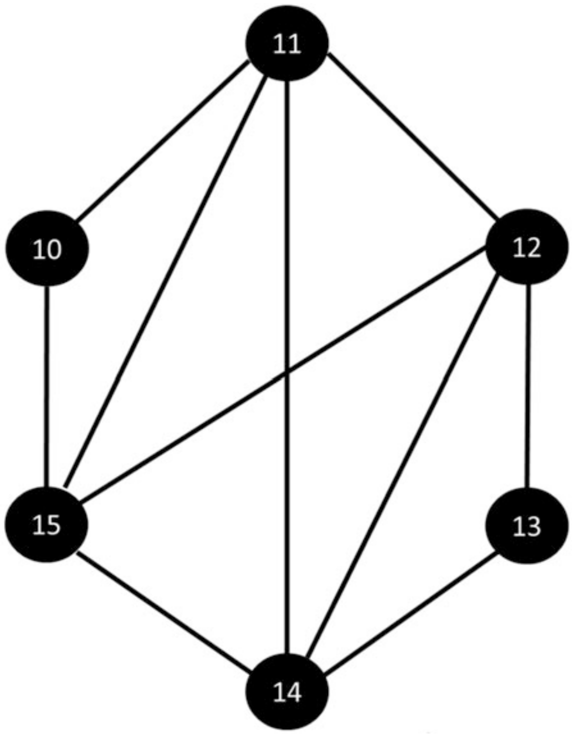
$$x_{10,15} + x_{15,14} + x_{14,13} + x_{13,12} + x_{12,11} + x_{11,10} \leq 5 \tag{10.11}$$

The proposed form of subtour breaking constraints (10.6) are somewhat more complex and would prevent both forward and reverse cycles as well as other possibilities for the set V, comprised in this case of nodes 10 through 15. Figure 10.4 depicts the cycle of nodes 10 through 15 along with all arcs that exist which connect any pair of nodes within this of six nodes. Given a set $V = \{10, 11, 12, 13, 14, 15, 16\}$, constraint (10.6) in this case would be:

$$\begin{aligned}&x_{10,11}+x_{10,15}+x_{11,10}+x_{11,12}+x_{11,14}+x_{11,15}+x_{12,11}+\\&x_{12,13}+x_{12,14}+x_{12,15}+x_{13,12}+x_{13,14}+x_{14,11}+x_{14,12}+\\&x_{14,13}+x_{14,15}+x_{15,10}+x_{15,11}+x_{15,12}+x_{15,14}\quad\leq\;5\end{aligned}\tag{10.12}$$

Essentially, this constraint allows at most 5 arcs to be used in a solution among all of the arcs which interconnect the six nodes in the set V. Thus, this constraint will cover both conditions (10.10) and (10.11) as well as other combinations, including sets of possible subtours spanning the six node set. For example, Fig. 10.5 depicts two subtours, 10 to 11 to 15 to 10 and 12 to 13 to 14 to 12, which would be prevented from simultaneously occurring as well. Any subtour or combination of subtours which spans the set V will be prevented in this case. This includes, for example, the subtour 10 to 11 to 14 to 13 to 12 to 15 to 10, or the three subtours 10 to 11 to 10, 12 to 13 to 12, and 14 to 15 to 14. Consequently, it makes sense to use the general form of constraints (10.6) for subtour elimination rather than the simplified form used in Current et al. (1984).

Fig. 10.4 Modified network of arcs



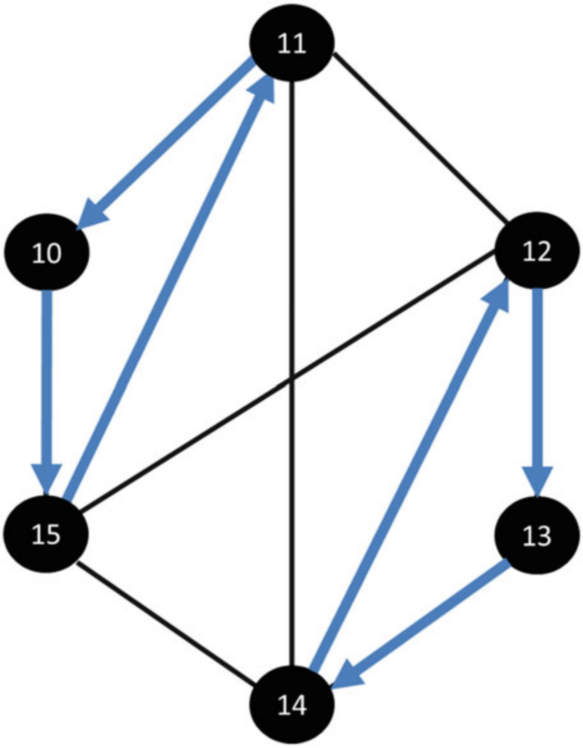


Fig. 10.5 Two subtours among the set of nodes that would be prevented by imposing constraint (10.12)

10.2.2 Alternatives for Eliminating Subtours

There are alternate forms in which the shortest path covering problem can be cast in order to eliminate the need for constraints (10.6). One of the ways is to use an approach based upon a construct of Vajda (1961) originally developed for the travelling salesman problem. Consider the following decision variable representing the order in which arcs are used along a route:

$$x_{ijt} = \begin{cases} 1, & \text{if the } t^{th} \text{ arc in the path travels from } i \text{ to } j \\ 0, & \text{otherwise} \end{cases}$$

Using this decision variable, it is possible to formulate a variant of the SCPP, inspired in part by the related work of Curtin and Biba (2011). The formulation follows:

$$\text{SCPP2}: \qquad \textit{Minimize} \sum_{i} \sum_{j \in F_i} \sum_{t=1}^{n-1} d_{ij} x_{ijt} \tag{10.13}$$

Subject to:

$$\sum_{j \in F_p} x_{pj1} = 1 \tag{10.14}$$

$$\sum_{i \in T_q} \sum_{t=1}^{n-1} x_{iqt} = 1 \tag{10.15}$$

$$\sum_{i \in T_k} \sum_{t=1}^{n-1} x_{ikt} \leq 1 \quad \forall k \tag{10.16}$$

$$\sum_{j \in F_k} \sum_{t=1}^{n-1} x_{kjt} \leq 1 \quad \forall k \tag{10.17}$$

$$\sum_{i} \sum_{j \in F_i} x_{ijt} \leq 1 \quad \forall t < n \tag{10.18}$$

$$\sum_{i \in T_k} x_{ikt} - \sum_{j \in F_k} x_{kjt+1} = 0 \quad \forall k \neq p, q \& t < n \tag{10.19}$$

$$\sum_{j \in C_k} \sum_{i \in T_j} \sum_{t=1}^{n-1} x_{ijt} \geq 1 \quad \text{for all} \quad k \in N \quad \text{where} \quad k \neq p, q \tag{10.20}$$

$$x_{ijt} \in \{0, 1\} \quad \forall (i, j) \in A, t < n \tag{10.21}$$

The objective (10.13) of SCPP2 minimizes path distance based upon the arcs selected for the path, equivalent to the SCPP1. Each arc (i,j) is represented by a set of variables, x_{ijt}. If a given arc is selected, $x_{ijt} = 1$, then the path travels from node i to node j along arc (i,j) as the t^{th} arc in the path. Accordingly, the order of the arcs encountered along the path is tracked by the index t. As the path must begin at the origin p, then the first arc chosen in the path must leave node p. This property is maintained in constraint (10.14). The path must eventually arrive at the destination node q. Since we do not know in advance how many arcs are to be used in the path, we do not know the order nor the number needed. Consequently, constraint (10.15) ensures that the path reaches the destination node q using some order t, but that is not fixed in value. Constraints (10.16) maintain that each node is entered at most once over all possible levels of path order. Constraints (10.17) maintain that a path can leave a node at most once. Constraints (10.18) maintains that at most one arc is chosen for a given order. That is, no two arcs can be assigned the same order in the path. Constraints (10.19) are key in this model variant. It establishes that if a specific intermediate node k is entered on the t step or order, there must be an arc selected to leave that node using an arc selected with order $t + 1$. This constraint ensures that the path is a continuous and that the order of arcs selected along the path increases in increments of one. There is no requirement that all sequence orders be used. Rather, when the path reaches the destination at a specific sequence order, $t = q$, the decision variable values associated with $t + 1$ or higher will be zero. Constraints (10.20) ensure that all nodes other than the origin and destination must be covered or served by the path. Finally, constraints (10.21) present the binary integer restrictions on the decision variables.

Technically speaking, constraints (10.17) are redundant considering constraints (10.16). This is due to the fact that if each node can be entered at most once along the path based upon constraints (10.16), then constraints (10.19) will prevent a node from being left more than once. Collectively, the constraints prevent any subtours from occurring in a solution. Thus, subtour elimination constraints are not needed in SCPP2. Further, the path is also not permitted to cross itself either, a condition that is imposed in SCPP1 as well.

The size of SCPP2, in terms of constraints is on the order of $n^2 + n$, which is considerably smaller than that of SCPP1. Unfortunately, the SCPP2 model contains $n - 1$ times more integer variables than SCPP1. Thus, the cost of reducing the number of constraints in SCPP1 is significant. There exist several alternate ways in which SCPP can be formulated based upon approaches relied upon in structuring the traveling salesman problem. They include the use of a flow-based structure to ensure no subtours (Gavish and Graves 1978), a two-commodity flow model (Finke et al. 1983), and variables established to track node sequencing (Miller et al. 1960). Virtually, all models suffer from dimensionality issues and/or lack of a strong linear programming relaxation. Orman and Williams (2004) have tested a number of formulations for the travelling salesman problem and conclude that the best approach appears to be the original approach of Dantzig et al. (1954), where subtour breaking constraints are included as needed. They suggest solving a problem sequentially, adding constraints when one or more subtours arise, and continuing until no subtours exist. The resulting solution is then feasible, and an optimal traveling salesman problem solution. Their results lend credence to the approach adopted by Current (1981) and Current et al. (1984), suggesting that it is likely the best method to formulate and solve the classic form of the SCPP.

10.3 Salesman and Tour Coverage

In a previous section we provided two relatively different formulations for the SCPP. Both formulations are based in part on past travelling salesman problem research. Current and Schilling (1989) proposed a counterpart to the travelling salesman problem called the Covering Salesman Problem (CSP). They describe the CSP as follows:

Identify the minimum cost tour of a subset of n cities such that every city not on the tour is within some predetermined covering distance standard, S, of a city that is on the tour.

Solutions to the classic travelling salesman problem must visit each city on the tour. The CSP therefore represents a relaxation of this assumption and instead requires that the tour come sufficiently close to each city. A tour then is a special type of path that must visit each node in the network. Application of the CSP could involve finding the shortest route for a mobile health clinic to travel within a rural region of villages, where the route stops at villages that are sufficiently close to all remaining villages.

Formulating this problem is quite straight forward given the previous discussion of the SCPP1. In particular, constraints (10.2) and (10.3) are eliminated and constraints (10.4) and (10.5) are modified to hold for all nodes $k \in N$ as there is no pre-specified origin or destination. Solving this problem will yield the shortest cycle covering all nodes. It is important to note that the SCPP2 cannot be easily modified in order to solve for a covering salesman tour because it is not known a priori how many arcs will be needed for the covering salesman tour.

The CSP does not require that a specific node be a starting (and ending) location, and Current and Schilling (1989) anticipated that in many cases this might be desirable. They also suggested there might be other nodes/cities required to be part of the tour. To accomplish this, all that is required is to add a constraint (10.2) for each such node. While the SCPP2 cannot be easily modified to solved for a covering tour, if a starting city is known (e.g., node p), then a covering route can be found by modifying constraints (10.15) to hold for node $q = p$ and by enforcing constraints (10.19) and (10.20) for all nodes except p.

Current and Schilling (1989) also suggested that in some cases the cost of a route may be more than the sum of the distances of the arcs that are traversed, as it may well be a function of the number of stops on the route. Such a possibility can be easily addressed as a second objective. This second objective would equal the number of arcs chosen for the path, multiplied by the cost of stopping. Current and Schilling (1989) indicated that one could then generate a tradeoff between the number of stops and the length of the route. Note that solving a covering salesman problem where the route consists of exactly T stops can be also found using a modified form of SCPP2, where constraints (10.14) and (10.15) are dropped, constraints (10.18) are converted into equalities and invoked for all orders from 1 to T, constraints (10.19) are imposed for all orders 1 to $T - 1$, constraints (10.20) are enforced for all nodes, and the following constraint is added:

$$\sum_{i \in T_k} x_{ikT} - \sum_{j \in F_k} x_{kj1} = 0 \quad \forall k \tag{10.22}$$

This constraint forces the arc used in order T to enter the node left in order 1, thereby making a loop. The modified model is structured so that it is a path that starts at some endogenously determined starting node, uses exactly T arcs, and returns to that starting node using the path order T arc. When solving this model for a range of values T one can generate a tradeoff between salesman route length and the number of nodes/stops used in the route.

ReVelle and Laporte (1993) proposed the traveling circus problem, which is related to the covering salesman problem. This was later called the covering tour problem (CTP) in Gendreau et al. (1997), and this naming convention is adopted here. A formal description is as follows:

Identify the minimum cost tour that visits a prespecified set of nodes and covers a prespecified set of nodes.

Formal specification of CTP is facilitated by the use of the following three sets of nodes (or vertices in a graph):

V = the set of nodes that can be visited along the route or tour
Ω = set of nodes that must be visited along the route/tour
Φ = set of nodes that must be covered by the route/tour

The CTP seeks to find the minimum length tour that visits all nodes Ω and covers all nodes Φ. It should be noted that such tours may not exist for a given graph. If sets V and Φ include all nodes, N, and the set Ω is empty, then the CTP is equivalent to the CSP. If Ω is not empty, then the CTP is equivalent to a special case of the CSP where certain nodes must be visited. Gendreau et al. (1997) provide an alternative approach for formulating the CTP, differing from the two models given above, and propose a branch and cut algorithm as well as a heuristic for solution. It should be noted that there are several related problems that have appeared in the literature. They include the tour cover problem on a graph involving node costs and not arc costs (Arkin et al. 1993), the geometric covering salesman problem where the vertex set of possible tour points is not a discrete set of points but planar regions (Arkin and Hassin 1994), the prize collecting travelling salesman problem (Fischetti and Toth 1988), the selective travelling salesman problem (Laporte and Martello 1990), and the generalized covering salesman problem (Golden et al. 2012).

10.4 Maximal Coverage Variants

Given coverage path equivalents for the LSCP, extension to account for partial coverage of demand areas (nodes) should not be surprising. Current et al. (1985) proposed an extended path problem that allowed for some nodes not to be covered. Whereas the shortest covering path must cover all nodes, this extension allowed for some nodes not to be covered. This will be referred to here as the Maximal Covering–Shortest Path Problem (MCSPP). It is described as follows:

Given a path origin and destination, find the path that starts at the origin and terminates at the destination that minimizes path distance and maximizes path coverage.

This is a two objective problem, formulated by extending the SCPP. To demonstrate how this is done, consider the following additional notation:

a_k = demand at node k

$y_k = \begin{cases} 1, & \text{if demand at node } k \text{ is covered within standard } S \\ 0, & \text{otherwise} \end{cases}$

Z_1 = objective function value corresponding to total coverage of the path

Z_2 = objective function value indicating the total distance of the path

Given this notation, the MCSPP is formulated as follows:

$$\text{MCSPP}: \quad \textit{Maximize}\ Z_1 = \sum_{\substack{k \in N \\ k \neq p,q}} a_k y_k \tag{10.23}$$

$$\textit{Minimize}\ Z_2 = \sum_{i \in N} \sum_{j \in F_i} d_{ij} x_{ij} \tag{10.24}$$

Subject to:

$$\sum_{j \in C_k} \sum_{i \in T_j} x_{ij} \geq y_k \quad \forall k \neq p,q \tag{10.25}$$

$$y_k \in \{0,1\} \quad \forall k \neq p,q \tag{10.26}$$

and constraints (10.2, 10.3, 10.4, 10.6, and 10.7).

This model is essentially the same as the SCPP1 except that the possibility has been added that a given node and its demand may not be covered by the path. This is done with the use of the coverage variable y_k introduced above. Essentially the path has to be within S of a given node k in order for the coverage variable y_k to be equal to one. This is established in constraint (10.25). Current et al. (1985) proposed solving this model without constraints (10.6) in the same manner as the shortest covering path problem. When doing so, however, it is to be expected that subtours will occur, just as was the case in the SCPP solution process. This means that constraints would be necessary like those of constraints (10.6) to prevent subtours. So, as new subtours are encountered, additional subtour elimination constraints are added and the problem resolved. Eventually, a solution will be obtained that is optimal and feasible (i.e. containing no subtours). This approach is exactly the same as that described for the SCPP previously. An optimal MCSPP solution is given in Fig. 10.6 involving the City of Richardson, TX. This solution was generated with emphasis on coverage as compared to path distance. Thus, the path is less direct between the origin and the destination nodes and tends to meander as a means of increasing path coverage.

A complicating feature of the MCSPP deserving some attention is the fact that is has two objective functions, (10.23) and (10.24). The implication is that there may be many different solutions that could be considered optimal with respect to some

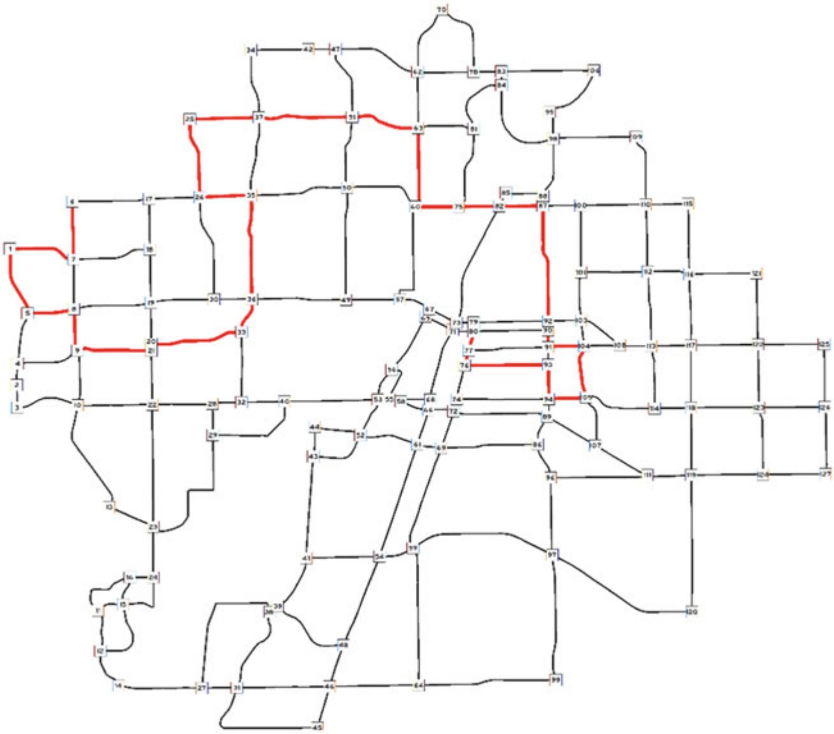


Fig. 10.6 A solution to MCSPP applied to the City of Richardson, TX. This path emphasizes coverage over that of path distance, where the origin node is on the outside boundary of Richardson and the destination node represents the transit center near the center of the city

combination of preference ratings of one objective over the other. Such solutions are known as Pareto optima. Further, it may be possible to summarize these alternative optima using a graphic that is often termed objective space. This is a plot of all solutions in terms of their values of objective (10.23) and objective (10.24), with one of the objectives represented by the x-axis and other on the y-axis. Altogether these solutions represent a tradeoff that exists between the two objectives. The challenge of course is identifying these tradeoff solutions. Cohon (1978) discusses methods for solving multiobjective optimization problems, noting that two options are the weighting method and the constraint method. Rather than get into the details of solution methods at this point, we will leave the discussion at merely highlighting that solving models that involve two or more objectives is itself a challenging task, and the subject of much ongoing research. Details for generating tradeoffs for path problems can be found in Medrano and Church (2014, 2015).

While the MCSPP is challenging to solve on many different fronts, one issue affecting this has to do with the number of associated decision variables. Under certain conditions, reformulation with fewer decision variables may be possible. For example, suppose that the covering standard is zero, meaning that the path needs to travel to each node that it covers. The design of a light rail train route connecting villages is a good example of when such an extension might be necessary, where light rail does not effectively serve a village unless it stops there. Current et al. (1985) called such a problem extension the maximum population shortest path problem and created a special formulation for this case. They reasoned that a pathway would enter a node at most once on its way to the destination. If we know that a pathway enters a node at most once, then that node is reached and "covered" when an arc is chosen entering that node.

Using the notation described above for the SCPP and the MCSPP, a variant is possible. For consistency purposes, this is simply referred to as the MCSPP2:

$$\text{MCSPP2}: \qquad \textit{Maximize } Z_1 = \sum_{k \neq p,q} \sum_{i \in T_k} a_k x_{ik} \tag{10.27}$$

$$\textit{Minimize } Z_2 = \sum_{i \in N} \sum_{j \in F_i} d_{ij} x_{ij} \tag{10.28}$$

Subject to:

$$\sum_{i \in T_k} x_{ik} \leq 1 \quad \forall k \neq p,q \tag{10.29}$$

and constraints (10.2), (10.3), (10.4), (10.6), and (10.7)

Consistent with MCSPP, the MCSPP2 has two objectives: maximize coverage, (10.27), and minimize path length, (10.28). Constraints (10.29) ensure that the path enters each node (not the origin or destination) at most once. This allows us to eliminate the need for the coverage variable y_k, and reconstruct the objective (10.27) solely in terms of the path variables, x_{ij}. Thus, this is a more compact model than when the maximal service distance is greater than 0 in value. One may wonder why this construct cannot be used for the MCSPP as well. The reason why is that the maximal covering path may travel close to a node more than once, and thus cover that node several times. If constraints (10.29) are modified for non-zero coverage distances, if would take the following form:

$$\sum_{j \in C_k} \sum_{i \in T_j} x_{ij} \leq 1 \quad \forall k \neq p,q \tag{10.30}$$

This, in essence, would prevent a path from coming close to a given node k more than once. If any arc is less than the maximal service standard, then condition (10.30) would actually prevent that arc from ever being used in a solution. That is, using that arc would cover the node by its entry, and be covered by the fact the nearby neighboring node was entered as well in a previous stage or order.

Another variant to be detailed is associated with a tour instead of a path. Current et al. (1994) referred to this as the maximal covering tour problem (MCTP). This problem can be described as follows:

Given a network of n nodes, where each node has an associated amount of demand and each arc has a traversal cost, find the route which visits exactly p-nodes while minimizing route traversal cost and maximizing route coverage.

They chose to restrict the tour to visit exactly p-nodes, as this was a special case of what has been termed the median tour problem. The median tour problem involves finding a route which visits exactly p-nodes while minimizing the route cost as well as the sum of the shortest distances of those nodes not selected for the route in travelling to their closest node on the route.

The MCTP can be formulated by modifying the MCSPP. As the MCSPP is bi-objective, involving the minimization of path distance and the maximization of coverage, these are the same objectives of the MCTP. Formulation requires the removal of the origin and destination constraints, (10.2) and (10.3), in the MSCPP and enforce the path continuation constraints (10.4) for all nodes. In addition, it is necessary to add the following constraint:

$$\sum_{i} \sum_{j \in F_i} x_{ij} = p \tag{10.31}$$

This constraint ensures that exactly p nodes will be selected for the tour. Thus, the entire model can be formulated as bi-objectives, (10.23) and (10.24) subject to constraints (10.4), for all nodes, (10.6), (10.7), (10.25), (10.26), and (10.31).

10.4.1 Trees

Thus far, network structures have been reviewed consisting of paths and tours. Recall that a path is a route from an origin to a destination, traveling along arcs in the network through nodes. A tour is a particular type of path where the origin is also the destination, and all nodes in the network must be visited. Differing somewhat from a path or tour is a network structure known as a tree. Formally defined, a tree is a sub-network that is connected and contains no cycles (or loops). The problem of locating structures that provide coverage was extended to tree graphs in Hutson and ReVelle (1989). They defined two types of coverage, direct and indirect. In direct coverage, only nodes connected to the tree are considered served. That is, the service distance standard is assumed to be zero. In the indirect coverage case it is assumed that nodes relatively close to the tree structure, but not attached to it, are considered served. For example, if the tree structure represents a hydrogen fuel distribution network, then the nodes of the tree are refilling stations. Accordingly, hydrogen vehicles from nearby nodes of the tree can access the refueling system as can those nodes that are part of the tree structure. Indirect coverage can therefore be defined for maximal covering standards greater than 0. The distinction here is somewhat akin to the difference between the MCSPP2 (i.e., direct coverage) and the MCSPP (i.e., combined direct and indirect coverage). Hutson and ReVelle (1989) also suggested that indirect coverage could also include those nodes not on the tree but are within a specified distance of an arc on the tree. Worth noting as well is that indirect coverage provided by arcs may be applicable to problems where system access is based upon all tree structure elements and not just the nodes of the tree (see Curtin and Biba 2011).

Covering tree location problems are bi-objective, unless complete coverage must be provided. In the case of finding a complete direct covering tree on a graph, it makes sense to find the least cost tree that reaches all nodes. Since the tree must cover (reach) all demand, and assuming that the cost of the tree is the sum of the distances of the arcs selected for the tree, then this problem reduces to one of finding the minimal spanning tree of a graph. This can be done by very efficient algorithms (Kruskal 1956). However, if a discrete set of points distributed across a planar region is to be covered by a direct covering tree, then the problem is that of finding the optimal Steiner tree on the plane. Thus, the graph version of this problem is quite simple, and the planar version of the problem is quite complex, computationally.

Hutson and ReVelle (1993) suggested that minimal cost spanning trees could serve as a "backbone" network for service delivery problems. For example, tree structures have been applied for the provision of communication, water, and limited service road networks (Church and Current 1993). Given an efficient spanning tree, Hutson and ReVelle (1989, 1993) proposed two problem variants:

1. minimal cost covering subtree on a spanning tree backbone—*Identify those arcs which form a connected subtree on a tree graph, which minimizes cost and covers all nodes within a specified maximal service distance*
2. maximal covering subtree on a spanning tree backbone—*Identify those arcs which form a connected subtree on a tree graph, which simultaneously minimizes cost and maximizes demand coverage*

The more general form of the subtree design/selection problem is the latter case, and will be formulated as the maximal direct covering subtree problem (MDCSTP). We can rely on notation already used, although simplification is possible through reference to each arc on the tree being numbered $j = 1, 2, 3, \ldots, m$. Consider then the following notation:

$$x_j = \begin{cases} 1, & \text{if arc } j \text{ has been selected for the maximal covering tree} \\ 0, & \text{otherwise} \end{cases}$$

$I_{h,i} = \{j | \text{ arc } j \text{ is on the path between nodes } h \text{ and } i\}$

$P_i = \{j | \text{ arc } j \text{ can cover node } i\}$

The MDCSP was originally formulated by Hutson and ReVelle (1989) as follows:

$$\text{MDSCP}: \qquad Minimize\ Z_1 = \sum_j c_j x_j \tag{10.32}$$

$$Maximize \;\; Z_2 = \sum_i d_i y_i \tag{10.33}$$

Subject to:

$$y_i \leq \sum_{j \in P_i} x_j \tag{10.34}$$

$$y_h + y_i \leq x_j + 1 \quad \forall i, h, j \in I_{h,i} \tag{10.35}$$

$$x_j = \{0, 1\} \quad \forall j \tag{10.36}$$

$$y_i = \{0, 1\} \quad \forall i \tag{10.37}$$

Objective (10.32) minimizes the costs of the arcs selected for the covering subtree. Objective (10.33) maximizes the demand covered by the selected subtree. Constraints (10.34) establish whether a given demand has been covered. Constraints (10.35) ensure that the arcs that have been selected produce a subtree that is connected. Constraints (10.36) and (10.37) impose binary integer restrictions on decision variables. The key constraint in this model is (10.35). The basic idea that the path between any two nodes on a spanning tree is unique, otherwise there would be a cycle within the tree graph and the graph would not be a tree. So, if two nodes, h and i, are covered by the subtree, then for the subtree to be connected, all arcs on the path between nodes h and i must be selected for the subtree, otherwise the subtree will not be connected. Essentially, if both nodes are covered, $y_i = 1$ and $y_h = 1$, this condition forces any $x_j = 1$ where arc j is on the path between nodes h and i.

This model is deceptively simple, but the number of constraints (10.35) that are necessary to ensure that the subtree is connected can be quite large. Hutson and ReVelle (1989) solved this problem using a weighting approach, where a positive weight is used for the coverage objective and a negative weight is used for the tree cost objective. By varying the weights, and solving this model for each set of weights, they were able to generate a tradeoff between subtree coverage and subtree cost. Because the number of constraints (10.35) can be very large, Hutson and ReVelle (1989) formulated a model that could optimally add to an existing subtree, requiring a limited number of constraints. They suggested that this smaller model could be solved $n - 1$ times, where the existing subtree was one of the $n - 1$ arcs. By solving for each of the arcs as an existing tree on which to build an optimal subtree, the globally best covering subtree can then be identified as the best of these $n - 1$ solutions.

Church and Current (1993) developed a different formulation for the maximal direct and indirect covering subtree problems. Rather than using a variable to represent the selection of an arc for the subtree, they employed variables that represented a clipping/trimming off process of portions/branches of the initial tree graph in order to reveal the optimal subtree. Clipping a branch of a subtree was based upon the direction in which a branch was clipped. Their model involved only $2n - 2$ binary variables and $n - 1$ constraints, which eliminate the need for a large number of constraints (10.35). They also developed a polynomial bounded algorithm (on the order of n^2) that could solve both the indirect and direct covering tree problems, eliminating the need for solving an integer programming problem.

10.5 Arc Improvement

Murawski and Church (2009) extended the concept of coverage to a network of specific node locations as well as arcs when dealing with a problem of health clinic access in Ghana. Their concern was directed to the fact that existing health clinics were often inaccessible in the wet season as roads were of poor quality and were impassible, an issue raised in Oppong (1996). Whereas Oppong (1996) had proposed to relocate health clinics to increase clinic access by taking into account road quality, Murawski and Church (2009) proposed improving roads in order to increase access to existing clinics. Their problem of infrastructure investment can be defined as:

Maximize the total demand who can be provided with service coverage by determining which network arcs are to be improved, subject to a limited road investment budget.

This problem is called the maximal covering network improvement problem (MC-NIP). We can formulate this model with some redefined notation:

i,j,k = indices of network nodes
c_{ij} = cost of upgrading arc $(i,j) \in A$
$U = \{ (i,j) | (i,j) \in A \text{ and } c_{ij} > 0\}$, the set of arcs that are not upgraded
B = budget for arc improvement
F = {set of existing facility sites}
D = {set of nodes without upgraded access to a facility}
$F_i = \{ j \mid d_{ij} \leq S \text{ where } j \in F \}$
a_i = population of node i
Ω_{ij} = { set of arcs on possible path between nodes i and j}

$$H_{ij} = \begin{cases} 1, \text{if arc } (i,j) \text{ is upgraded} \\ 0, \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1, \text{ if node } i \text{ and facility } j \text{ are connected by an upgraded arc} \\ 0, \text{otherwise} \end{cases}$$

$$P_{i'j'}^{ij} = \begin{cases} 1, \text{if upgraded route between } i \text{ and } j \text{ uses arc } (i',j') \\ 0, \text{otherwise} \end{cases}$$

The above variables are defined for only those cases which apply. For example, if a village is already provided all-season access to a health clinic services within the maximum coverage distance, then it is unnecessary to have a coverage variable for that village. The same is true for road links that are already considered all-season passable. Murawski and Church (2009) note that the choice of direction (clinic to village coverage via paved road segments) is somewhat arbitrary and that the model could be formulated using path directions from villages to clinics.

The model formulation is as follows:

$$\text{MC-NIP}: \qquad Maximize \sum_{i \in D} \sum_{j \in F_i} a_i y_{ij} \tag{10.38}$$

Subject to:

$$\sum_{(i',i) \in \Omega_{ij}} P^{ij}_{i'i} = y_{ij} \quad \forall i \in D, j \in F_i \tag{10.39}$$

$$\sum_{\substack{(i',k) \in \Omega_{ij} \\ (k,i') \in \Omega_{ij}}} P^{ij}_{ak} - \sum_{\substack{(k,j') \in \Omega_{ij} \\ (j',k) \in \Omega_{ij}}} P^{ij}_{kb} = 0 \quad \forall i \in D, j \in F_i\ k \neq i \& k \neq j \tag{10.40}$$

$$P^{ij}_{i'j'} + P^{ij}_{j'i'} \leq H_{i'j'} \quad (i',j') \in \Omega_{ij} \mid i \in D, j \in F_i, i \neq i', i \neq j', j \neq i', j \neq j' \tag{10.41}$$

$$\sum_{(i',j') \in \Omega_{ij}} \left(P^{ij}_{i'j'} + P^{ij}_{j'i'}\right) d_{i'j'} \leq S y_{ij} \quad \forall i \in D, j \in F_i \tag{10.42}$$

$$\sum_{j \in F_i} y_{ij} \leq 1 \quad \forall i \in D \tag{10.43}$$

$$\sum_{(i,j) \in U} c_{ij} H_{ij} \leq B \tag{10.44}$$

$$H_{ij} \in \{0,1\} \quad \forall (i,j) \in U \tag{10.45}$$

$$0 \leq P^{ij}_{i'j'} \leq 1 \& 0 \leq P^{ij}_{j'i'} \leq 1 \quad \forall i \in D, j \in F_i, (i',j') \in \Omega_{ij} \tag{10.46}$$

$$y_{ij} \in \{0,1\} \quad \forall i \in D, j \in F_i \tag{10.47}$$

Objective (10.38) seeks to maximize the number of people that are provided upgraded service coverage by virtue of arc investment decisions. Access coverage is provided to a node by the existence of a service facility and an upgraded arc path from that facility to a specific node, as long as it is with the service coverage standard. The facility nodes already have “all-season” access and others might already have such access due to existing upgraded arcs. The remaining nodes can be classified as to whether they can be provided upgraded coverage access through arc investments, D, or whether they are too far from a facility to ever be considered accessible independent of the arc condition. The model addresses only those nodes in D. Constraints (10.43) ensure that a demand node is considered to be covered no more than once, even though it may be possible to travel two different upgraded access routes from a facility to that demand node. Constraint (10.44) limits the investment in all upgrading to no more than an allowable budget. The bulk of the remaining constraints track which nodes are provided coverage by virtue of which arcs are selected for upgrading. To achieve this it is necessary to define if a path exists between a facility and a node that is no longer than S and is comprised entirely of upgraded arcs. Path variables, $P_{i'j'}^{ij}$, are used to define a feasible path between node i and facility j. Such variables represent traveling along an arc connecting i' and j'. When $P_{i'j'}^{ij} = 1$, then all-season accessible travel from nodes i' and j' from facility j to demand node i is possible. If an accessible path connects node i to facility j, then constraint (10.39) allows y_{ij} to equal 1. If this is not the case, y_{ij} will remain at zero. Constraints (10.40) maintain that paths are connected through the typical node-balance constraints used in previously defined models. At every feasible intermediate node between node i and facility j, when a path enters an intermediate node k, the path must exit that same node k. Altogether, a feasible path between node i and facility j starts at node j, and must reach and stop at demand node i. Whenever a feasible path does not exist between i and j, then all path variables $P_{i'j'}^{ij}$ associated with i and j are zero in value. Path variables are not restricted in value when traversing any existing all season road links. But, path variables are restricted to be zero, when their associated arcs need upgrading and they have not been upgraded This is established in constraint (10.41). That is, a path from facility j to node i can use a specific arc (i', j') only if it is already upgraded, or has been upgraded. The last significant constraint (10.42) restricts the distance of the path selected between demand i and facility j to be less than or equal to the standard S, otherwise the path is not feasible. Finally, constraints (10.45), (10.46), and (10.47) indicate restrictions on decision variables.

Collectively, the MC-NIP allows no more than one feasible path comprised of upgraded arcs between any facility and node $i \in D$. Any such path must be no longer than the service coverage standard, S. This model was applied to the same region that was modeled by Oppong (1996), and enabled substantial increases in coverage to be provided by modest investments in network improvements.

10.6 Applications

The covering path problems are particularly appealing in that they represent design goals found in transit routing, light rail routing and rail station location, emergency/disaster relief supply routes where the nodes along the route are distribution points for emergency supplies, and mobile clinics for vaccination and health screening. The area of transit design has been addressed by a number of researchers starting with the conceptual models first proposed by Current et al. (1984). Boffey and Narula (1998) and Wu and Murray (2005) have suggested multi-path maximal covering shortest route models that optimize several routes simultaneously in order to cover a region. The maximal covering shortest path structure has also been expanded in a number of ways for extending transit routes, maximizing coverage of origin-destination travel, stop location, and arc coverage (see Murray and Wu 2003; Church et al. 2005; Matisziw et al. 2006; Curtin and Biba 2011).

The covering tour problems also capture the essence of design problems as planning security patrol routes and rural postal delivery routes using clustered/sets of mail boxes. For instance, suppose an industrial plant and grounds were divided into small sectors, and the objective was to determine a security route which was able to view each of the areas as stops along the route. In some communities, mail boxes for neighborhoods are being clustered and located at a specific point. One can think of the shortest covering tour problem as one that captures the essence of the route moving from cluster box locations and returning to the starting office, while ensuring that each customer is no further than a certain distance from their closest location of clustered mailboxes. The same type of problem involves drop box locations for banks, where drop boxes are placed so that customers can make large cash deposits after bank hours. There are also a number of potential applications for the generalized covering tour problem where a return frequency differs among nodes or areas.

10.7 Summary and Concluding Comments

Slater (1982) classified location problems based upon points serving points, points serving structures, structures serving points, and structures serving structures. This chapter highlighted the fact that the shortest covering path problem, as an example, can be conceived to be a structure (i.e., path in this case) serving a set of points (i.e., nodes in this case). The emergency warning siren location problem described in Chaps. 1 and 8 involves structures (i.e., siren coverage areas defined as a polygon) serving a structure (a region defined as a polygon). In this chapter we have discussed problems on a network that involve a structure (path, tour, tree and arc upgrading) that serve discrete points of demand represented as nodes of a network. Other problems can be derived and defined within this framework, many of which have not yet been explored in any formal manner. That is, they remain open research problems in terms of how best to structure and solve in support of planning and management decision making.

The models that have been formulated in this chapter for optimizing maximal covering paths and maximal covering tours have been based upon an assumption that optimal paths or routes never backtrack or cross over themselves to form cycles/loops. This assumption has been made explicit in Curtin and Biba (2011) as they ensure that a covering path never crosses over itself by constraining the path so that it enters a node no more than once. The use of sub-tour elimination constraints in the covering path models of Current et al. (1984, 1985) also imposes that loops will not occur for feasible solutions. However, Niblett (2016) and Niblett and Church (2016) have questioned such an assumption, highlighting that the shortest covering path problem of Current et al. (1984) may result in an infeasible solution, when in fact a feasible solution exists, or even result in finding a sub-optimal solution. That is, there are instances where the optimal solution in fact requires a cycle to achieve maximum coverage. Niblett and Church (2016) have presented a new, revised formulation for the shortest covering path problem capable of identifying the true optimal solution, whether it involves a cycle or not. They have also shown how to reformulate the node-arc path covering model of Curtin and Biba (2011) so that it can find optimal solutions when optimal solutions may involve an associated cycle.

Our emphasis in this chapter has been to describe the fundamental problems that have been developed to date associated with the location of a network structure, like a path, tour or tree that is designed to cover some of all of the nodes of the network. We have also presented integer-linear programming formulations for these fundamental models. It should be noted heuristics and optimal algorithms have been developed for some of these problems, and remains an active research area.

References

Arkin EM, Hassin R (1994) Approximation algorithms for the geometric covering salesman problem. Discret Appl Math 55:197–218

Arkin EM, Haldorsson NM, Hassin R (1993) Approximating tree and tour covers of a graph. Inf Process Lett 47:275–282

Boffey B, Narula SC (1998) Models for multi-path covering-routing problems. Ann Oper Res 82:331–342

Church R, Current J (1993) Maximal covering tree problems. Nav Res Logist 40:129–142

Church R, Noronha, V, Lei T, Corrigan W, Burbidge S, Marston J (2005) Spatial and temporal utility modeling to increase transit ridership: final report to Caltrans. VITAL, University of California, Santa Barbara, CA

Cohon JL (1978) Multiobjective programming and planning. Academic, New York

Current JR (1981) Multiobjective design of transportation networks. PhD dissertation, The Johns Hopkins University, Baltimore, MD

Current JR, Schilling DA (1989) The covering salesman problem. Transp Sci 23:208–213

Current JR, Schilling DA (1994) The median tour and maximal covering tour problems: formulations and heuristics. Eur J Oper Res 73:114–126

Current J, ReVelle C, Cohon J (1984) The shortest covering path problem: an application of locational constraints to network design. J Reg Sci 24:161–184

Current JR, ReVelle CS, Cohon JL (1985) The maximum covering/shortest path problem: a multiobjective network design and routing formulation. Eur J Oper Res 21:189–199

Current J, Pirkul H, Rolland E (1994) Efficient algorithms for solving the shortest covering path problem. Transp Sci 28(4):317–327

Curtin KM, Biba S (2011) The transit route arc-node service maximization problem. Eur J Oper Res 208:46–56

Dantzig G, Fulkerson R, Johnson S (1954) Solution of a large-scale traveling-salesman problem. J Oper Res Soc Am 2:393–410

Daskin MS (1995) Network and discrete location: models, algorithms, and applications. Wiley, New York

Finke GA, Claus A, Gunn E (1983) A two commodity network flow approach to the travelling salesman problem. In Proceedings of the 14th South Eastern conference on combinatorics, graph theory and computing, Atlantic University, FL

Fischetti M, Toth P (1988) An additive approach for the optimal solution the prize-collecting salesman problem. In: Golden BL, Assad AA (eds) Vehicle routing: methods and studies. North Holland, Amsterdam

Gavish B, Graves SC (1978) The travelling salesman problem and related problems. Working Paper OR-078-78. Operations Research Center, MIT Press, Cambridge, MA

Gendreau M, Laporte G, Semet F (1997) The covering tour problem. Oper Res 45:568–576

Golden B, Naji-Azimi Z, Raghavan S, Salari M, Toth P (2012) The generalized covering salesman problem. INFORMS J Comput 24:534–553
Hakimi SL (1965) Optimum distribution of switching centers in a communication network and some related graph theoretic problems. Oper Res 13(3):462–475
Handler GY, Mirchandani PB (1979) Location on networks: theory and algorithms. MIT Press, Cambridge, MA
Hutson VA, ReVelle C (1989) Maximal direct covering tree problems. Transp Sci 23:288–289
Hutson VA, ReVelle C (1993) Indirect covering tree problems on spanning tree networks. Eur J Oper Res 65:20–32
Kruskal JB Jr (1956) On the shortest spanning subtree of a graph and the travelling salesman problem. Proc Am Math Soc 7:48–50
Laporte G, Martello S (1990) The selective travelling salesman problem. Discret Appl Math 26:193–207
Matisziw TC, Murray AT, Kim C (2006) Strategic route extension in transit networks. Eur J Oper Res 171:661–673
Medrano FA, Church RL (2014) Corridor location for infrastructure development: a fast bi-objective shortest path method for approximating the Pareto frontier. Int Reg Sci Rev 37 (2):129–148
Medrano FA, Church RL (2015) A parallel computing framework for finding the supported solutions to a biobjective network optimization problem. J Multicrit Decis Anal 22 (5–6):244–259
Miller CE, Tucker AW, Zemlin RA (1960) Integer programming formulation of travelling salesman problems. J ACM 3:326–329
Morgan CA, Slater PJ (1980) A linear algorithm for a core of a tree. J Algorithm 1:247–258
Murawski L, Church RL (2009) Improving accessibility to rural health services: the maximal covering network improvement problem. Socio Econ Plan Sci 43:102–110
Murray AT, Wu X (2003) Accessibility tradeoffs in public transit planning. J Geogr Syst 5:93–107
Niblett TJ (2016) On the development of a new class of covering-path models. PhD dissertation, UCSB, Santa Barbara, CA
Niblett TJ, Church RL (2016) The shortest covering path problem a new perspective and model. Int Reg Sci Rev 39:131–151
Oppong JR (1996) Accommodating the rainy season in third world location-allocation applications. Socio Econ Plan Sci 30:121–137
Orman AJ, Williams HP (2004) A survey of different integer programming formulations of the travelling salesman problem. Working Paper No. LSEOR 04.67. London School of Economics and Political Science, London
ReVelle CS, Laporte G (1993) New directions in plant location. Stud Locational Anal 5:32–58
Slater PJ (1981) On locating a facility to service areas within a network. Oper Res 29:523–531
Slater PJ (1982) Locating central paths in a graph. Transp Sci 16:1–18
Slater PJ (1983) Some definitions of central structures. Lect Notes Math 1073:169–178
Vajda S (1961) Mathematical programming. Addison-Wesley, London
Wu C, Murray AT (2005) Optimizing public transit quality and system access: the multiple-route, maximal covering/shortest-path problem. Environ Plann B Plann Des 32:163–178

Chapter 11
Grand Challenges

11.1 Introduction

The main objective in writing this text was to describe many of the developments in location science that have been proposed, modeled, and applied using a covering framework. Covering models are based upon the use of some standard of service, ranging from a maximal response time standard in locating ambulances to a minimal decibel standard used in placing warning sirens. Many covering problems can broadly be classified into two types: (1) cover each and every demand using the smallest number of facilities, the Location Set Covering Problem (LSCP); or (2) maximize the demand that is covered while locating a fixed number of facilities, the Maximal Covering Location Problem (MCLP). These two problems and related models emerged in the early 1970s and have formed the basis for a considerable portion of this book. Applications have involved the location of surveillance cameras, cell phone towers, fire stations, health clinics, ambulances, and sirens, just to name a few. They have even included the selection of biological reserve sites, designing teeth color shades, and laying out fabric cutting patterns. Altogether, the location science literature reflects a rich history and long term development of covering models, involving the fields of geography, engineering, business, computer science, health planning, environmental science, conservation biology, regional science and economics, and urban planning.

Many of the problems and associated models described in this book demonstrate how covering models have been extended to address a wide range of real world issues and concerns. Many of the detailed models represent sophisticated extensions of the LSCP and the MCLP, addressing expected coverage, multiple level coverage, probabilistic coverage, and capacity constrained coverage, among others. The chapters of this book are organized conceptually around different issues, such as the underlying problem domain (e.g. discrete or continuous), the operational characteristics of facility service (e.g., congested systems), and varying value of coverage benefits (e.g., discrete versus gradual). Each chapter has provided a sizable list of

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_11

references from which this work has been based, but of course space limitations mean our referenced work is not exhaustive. Finally, this book has focused principally on the models that have been structured and not much on how they have been solved. Often such models have been tackled through the use of commercial mixed integer linear programming software. However, many exact and heuristic approaches that have been developed beyond the use of commercial optimization software. As our emphasis has been on describing different types of models that have been developed, not on the solution approaches per se, our thinking was that those interested in a specific model could then consult the literature to see the approaches employed rather than codifying all of the nuances developed for specific problems.

Even though this literature is quite substantial, there are a number of challenges for future research. We outline seven major areas in this short concluding chapter to focus attention on what we consider to be some of the grand challenges in covering model research per se, and location science in general.

11.2 Big Data

Over the past 40 years, location science has evolved from something that was very data poor (or data non-existent) to what many now call a data rich environment. For example, in retail, data from credit card purchases, address of residence, and other customer information can be gathered to develop a profile of a customer and those who live in a service area (Wedel and Kannan 2016). They can merge that with data of the property parcel having that address, and even information purchased from third party vendors to accurately classify a customer and identify the types of purchases they are likely to make. They can even estimate how much disposable income individuals have. In the past one was lucky if they could get the residences of their customers or commercial property square footage in an area. The past model of retail attraction and possible "capture" was that of Huff, a simple gravity model. Now this model can be far more sophisticated. But, even this picture is far more complicated than what it appears. Data rich is something of a misnomer as large volumes of data cannot always be translated into a sophisticated and statistically significant model. For example, how much of the credit card information is really meaningful when it is used to translate into the demand for a "bricks and mortar" bookstore? Also, how much and what types of data will be held by a company and considered proprietary? Even the techniques of the past like "reading a grocery store" discussed in Salvaneschi's book on *Location, Location, Location* are still useful techniques. Even the simple posting of sales volumes, like those for an individual Home Depot store can be useful. Salvaneschi (1996) noted that the use of Reilly's law has merit, but it is "like calculating the velocity and direction of a soccer ball in a vacuum versus the velocity and direction during a hurricane." He concluded that "great inaccuracies will occur when other factors of equal importance are ignored." Today, we have the possibility to ensure that our models are accurate from the perspective of supporting them with relevant and timely data.

People use the term Big Data to characterize a data rich environment, implying that big means that traditional data processing techniques and data storage systems cannot handle the volume, velocity, and/or variety of sources and types. There are challenges in just analyzing and visualizing such data. Even the issue of whether it is useful can be hard to address. For example, cell phone data can be used to analyze customer traffic patterns in a store, including which aisles they travel down and how much dwell time is taken in front of specific product shelves. Unfortunately, this does not directly translate into a customer model of demand. Even though the Internet of Things is touted as being the new frontier of data, most of it will be deemed useless for a given task. In fact, we have become so used to the idea that we will have lots of data, but the important and meaningful information remains hidden in the data and will require sophisticated methods to identify just what it is and how it can be used. Just to drive home this point, Marr (2016a) (see also Marr, 2016b), in a contributed piece in Forbes magazine, noted that:

> The data volumes are exploding, more data has been created in the past two years than in the entire previous history of the human race . . . Data is growing faster than ever before and by the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet.

From clicks on Google's search engine, eBay, and Amazon to cell phone location information to instantaneous face recognition for images/video, the world will be awash in data and some of this will be useful in location modeling, especially retail location and capturing customers. The Internet of Things that will connect us will provide data that will make it possible to better understand the needs within public services as well as how to operate businesses in a more efficient manner. For example, call data from EMS and fire dispatch centers used to be secure and privately held, not accessible for anyone except those within their organizations. Now, such data can be easily acquired on the Internet, where personal identifying data has been scrubbed. This will undoubtedly lead to improved models of emergency response and facility placement while meeting standards of coverage as more insight is possible.

11.2.1 Algorithms for Larger, More Nuanced Problems

Big data and data-rich environments mean that location problems may be defined using demand represented not by 200 or 2000 objects, but 200,000 or even 2,000,000 objects. The same may hold true for potential facility sites for some problems, although one may argue that problems often have a significantly larger number of demand points than potential facility sites. In a recent study by Church and Li (2016) they represented the major urban area of Los Angeles County by a mesh of approximately 6500 points. They used this set of points to find a theoretical minimum number of needed fire stations satisfying standards based coverage requirements. Solving this problem as a MCLP using general purpose mixed-integer

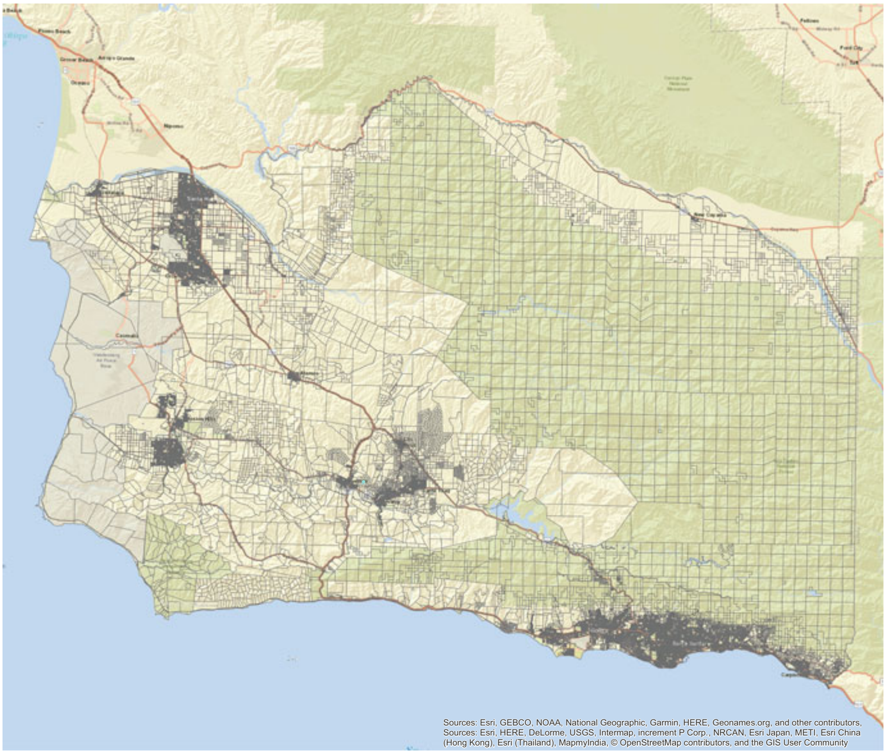


Fig. 11.1 Land parcels in Santa Barbara County

linear programming software, as an example, literally taxed the capabilities of the software, requiring days/weeks to solve. This problem was not what one would classify as large, but it is clearly large within the context of current general purpose software. For illustrative purposes, Fig. 11.1 shows the 130,208 parcels in Santa Barbara County. Again, not exactly big data relative to what is common in many planning and analysis environments, but large in the context of coverage modeling. Extending the range of problem sizes that one can solve to optimality will be a challenge. It is possible that utilizing existing techniques in a high performance computing environment (e.g., a large number of parallel processors) may help to extend the range of problem sizes that one can solve optimally. Further, it is necessary to further develop new and more capable solution strategies, both optimal and heuristic, for current and future "large" problems. Altogether, there are a wide variety of existing covering models that should be the target for newer, faster solution methods capable of extending the effective size of problems that can be solved within a reasonable amount of computational time.

There also exists the need to extend the capabilities of solving larger problems in specialized problem domains, like continuous surfaces, planes, and 3-D space. Although the problem of camera placement, as an example, is well understood for surveillance, no system today is designed to automatically solve this problem in a large area in order to minimize installation and maintenance costs. This problem and others have yet to be addressed with capable solution algorithms. The era of big data makes this challenge even more relevant.

11.2.2 Balancing the Use of Actual and Aggregated Data

Aly and White (1978), Benveniste (1982), Daskin et al. (1989) and Current and Schilling (1990) were the first to raise issues regarding representational error in coverage modeling. Daskin et al. (1989) and Current and Schilling (1990) addressed potential distortion/representational issues in aggregating demand points as a precursor step to solving a covering model. The basic idea is that through data aggregation, one can create a smaller, but solvable problem. The work of Murray and O'Kelly (2002) demonstrated representational issues in using discrete points to approximate a region when covering continuous space (a different type of aggregation). In either case, representational errors occur. The former involves aggregation of data points, and the latter involves using points to represent areas. One way to address such issues is to suggest that there should be no aggregation of spatial data nor any approximation of an area as doing so involves representational error. But, if no methods of aggregation are allowed or no methods that simplify spatial representation are employed, the resulting complexity of many problems will mean that only small data sets or regions can be evaluated using models. Clearly, there is a realistic balance between problem representation and possible representational error in solving location covering problems. This issue becomes more acute as challenges in the era of big data emerge. What is needed is a better understanding of aggregation techniques, resulting levels of error, and possible standards for what is appropriate for problem representation and data aggregation. This also includes the development of new and potentially better ways of aggregating spatial data or approximating a region before the application of a model is structured and solved. Having good methods for spatial representation and data aggregation will help extend the effective capabilities to apply covering models to larger problems.

11.3 Developing Better Models and Service Assumptions

Box (1976) is famous for stating that "... all models are wrong ..." while at the same time stating that "... some are useful." This assessment applies as well to models developed in location science. When it comes to location modeling in general, and location covering modeling specifically, there is a dichotomy in approaches depending on whether the model and intended application is in the private sector or the public sector. ReVelle et al. (1970) stated that "... private sector and public sector location problems are alike in that they share the objective of maximizing some measure of utility to the owners while at the same time satisfying constraints on demands and other conditions." While in the private sector such problems mostly center around the objective of maximizing efficiency and profits, the public sector involves issues similar to the private sector plus "... the additional dilemma that goals, objectives and constraints are no longer easily quantifiable, nor are they even in commensurate units or easily defined." Later, Liebman (1976) when echoing the previous assessment stated that "... public systems problems are frequently ill defined and have fuzzy constraints and vague multiple objectives ..." and concluded by calling them "wicked." In any public setting, one might wonder why any model would be useful. Brill (1979) added some clarity to this issue when discussing the development of models in the public sector. He concluded that there were two substantive lingering issues: (1) failure to consider equity, and (2) the empirical shortcomings in estimating benefits and costs. Brill (1979) argued that public sector problems are inherently multi-objective, but that there would be difficulties "... in achieving a complete representation of a planning problem using a mathematical formulation." His answer to that was that analysts should seek to develop models to generate solutions on the tradeoff curve or surface between the stated objectives as well as identifying significantly different alternatives. It was further argued that there could be solutions which were nearly Pareto-optimal, but different in decisions that could be on the tradeoff surface if the set of objectives was actually complete in its representation.

Although many of the models detailed in this book were developed for the private sector, an equally large number of them were developed for public sector use. A quick review of public sector oriented covering location models suggests a potential lack of an explicit equity objective, although a few were developed to provide a sense of equity (see Chap. 6). Few, if any, of the public sector side are defined as multi-objective, when indeed they are. In addition to this there has been an almost complete reliance on using a surrogate measure to represent that "value" of coverage, e.g., the number of people covered or served within a desired maximal service distance. The use of a surrogate measure of service or social benefit means that the empirical shortcomings raised by Brill (1979) in estimating benefits and costs still holds true some 40 years later. The fact that there has been little if any attempts to include or develop better estimates of benefit for such models is a major weakness. Chapter 6 does introduce the fact that coverage benefits should be modeled across different distance ranges or standards, and demonstrates how such properties can be structured, but there has not been a push to apply or further develop these models even though they represent an important step to better capture public benefits.

The development of solution approaches for new and existing location covering contexts must also evolve in the sense that faster solution times and optimal solutions are nice, but little attention has been devoted to the importance of alternative solutions that can support planning and decision making. For example, identification of high performing but significantly different spatial alternatives is seen as essential in many public sector contexts. We may well find that it is common for a problem to have two solutions that achieve similar levels of coverage while at the same time involving facility sites that are quite different spatially. The idea is that we need to extend our focus to modeling approaches where identification of such occurrences is possible, enabling better support for decision making. Such a concern is not new, with Brill et al. (1982) carrying out related work for land use planning. Scaparra et al. (2014) is a good example of recent work attempting to address the need for supporting decision making through the identification and exploration of close to optimal solutions that spatially vary.

We started this section by referring to the statement of Box (1976) that all models are wrong. We know that there are limitations to what can be modeled, yet there is nothing inherently wrong with a simple model. Many location covering models are quite simple in structure but powerful. However, there do exist weaknesses that should be addressed, as noted above. Altogether, there are significant challenges for researchers and practitioners to better structure, solve, and apply location covering models.

11.4 Problem Transformation

There can be benefits to reformulation or transformation of a problem. This may involve the use of different mathematical constraints to impose a condition, as was the case with the ACLP in Chap. 5, or the structuring of coverage using allocation variables, as was the case for the CLSCP and CMCLP in Chap. 7. Church (1974) was the first to demonstrate how to transform a covering problem into a p-median problem. Since the worst case p-median problems are provably hard to solve, belonging to the class of NP-hard, the simple polynomial transformation process suggested by Church (1974) provides a simple proof of NP for the MCLP. Since the initial work in transforming the MCLP into a p-median problem, Hillsman (1984) showed how to transform an LSCP into an equivalent p-median problem. Church and Weaver (1986) demonstrated that many multi-level covering models, including the MEXCLP (Chap. 4), are equivalent to the vector-assignment p-median problem. Since those initial efforts, the connections between various covering models and forms of the p-median problem and simple plant location problem have been made. In some circumstances using a p-median structure incorporating allocation variables in Balinski constraints provides a tighter model than the original covering model. This can help to speed solution times when using general purpose solvers for certain problems. More generally, such transformations make a covering problem harder to solve as this requires the addition of a sizable number of allocation variables along with an increased number of constraints as well. But, what these transforms provide is a gateway to all of the heuristics and algorithms that have been developed to solve p-median problems and their variants, including the hierarchical median, vector assignment median, capacitated median, and stochastic median problems. Optimal approaches include decomposition, Lagrangian relaxation, and dual ascent. Heuristic approaches for the p-median problem run the gamut, and presumably will not need to be modified to solve many types of covering problems. This includes such well-known approaches as Teitz and Bart (interchange), variable neighborhood search, Tabu search, genetic algorithms, simulated annealing, and GRASP with path relinking as examples.

One caveat when suggesting the use of *p*-median techniques in solving covering problems through the process of transformation is that they are relatively untested in terms of their robustness in finding optimal and near optimal solutions. Are there properties which make a covering problem harder to solve when the distance matrix of a *p*-median is contorted? Do such transformations produce barrier like properties for heuristic search methods due to routines relying on closest assignments, a mainstay for the *p*-median problem, where a large, albeit contrived, number of ties in transformed distances are likely? Will such heuristics need to be further modified to handle transformed distance matrices in the same robust way that they handle non-transformed distance matrices? Some preliminary analysis on the part of the authors suggests that the questions remain unanswered to date.

11.5 Flexible and Accessible Location Application Software Systems

There are many applications of covering models in the literature. Some of these approaches were proof of concept models, some were one of a kind applications designed to support a specific purpose in decision making, and others involved the use of specialized software like GIS (geographic information systems). By and large, many of these approaches are difficult to duplicate or replicate without the skills and expertise to develop and apply. This does not mean that covering models themselves are difficult to understand or potentially apply, but that there are impediments that make such analyses difficult to accomplish in current computing and software environments. For example, an application may represent the use of several types of software. This means that someone might need to be proficient with many tools to accomplish the task, or that a team of individuals with the requisite expertise would need to be assembled to complete the task. The main reason for this is that location models, solution software, and mapping/visualization systems are not well integrated or are not flexible enough to address the task at hand. This unfortunately means that application often requires greater effort than what is desirable, at least in the context of general accessibility in the public sector.

To explain this issue in more detail, suppose that data is sourced from several vendors and imported into a GIS for data manipulation and management. A GIS can also be used to display results, by mapping configurations of facilities and depicting who is covered and who is not. Some GIS also contain software modules that can be used to solve location-allocation problems as well as pure location problems like the LSCP and MCLP. An example is ArcGIS, produced by Esri. In particular, the Network Analyst extension contains location-allocation functions which can be used to solve a *p*-median problem, an attendance maximizing problem and the MCLP, among others options. If you want to solve an MCLP using this function, as an example, one should understand how this package solves the problem. Implemented in ArcGIS is a location-allocation approach designed to solve the *p*-median problem using a GRASP-inspired heuristic. To solve any of their problem variants, the MCLP in particular, the software first transforms the distance matrix so that it is an equivalent *p*-median problem. Unfortunately, a heuristic like GRASP offers no guarantee as to whether it has identified an optimal solution, just that the solution is the best that it was able to find. While not common knowledge, the Esri heuristic is executed 128 times, and reports to the user the best found in the 128 applications of the approach. The advantages of such a heuristic are: (1) good solutions to the *p*-median problem appear to be found; and, (2) it is designed to solve large sized problems. The main disadvantage is that the heuristic has never been sufficiently tested as to its effectiveness in solving the range of possible problems it allows to be structured, including the MCLP and LSCP, masquerading as a *p*-median problem. Another issue is that the location-allocation function cannot be used to solve most of extended forms of the MCLP and LSCP detailed in this book, only the two classic forms of location covering. Finally, not all forms of the classic LSCP and MCLP can be solved using this software. For example, the location-allocation feature in ArcGIS is designed to solve problems with an underlying transportation network, not a continuous plane, making the CSSCP (continuous space set covering problem) and CSMCP (continuous space maximal covering problem) described in Chap. 8 virtually impossible to address.

An alternative is to solve covering problems outside of GIS, perhaps employing a general-purpose mixed integer linear programming software package, like XPRESS, IBM-CPLEX, or Gurobi. But, this means that data must first be exported from the GIS, then processed and/or converted into a form that is usable by the solver packages. This often requires interfacing with some type of model builder software (e.g., OPL, IVE, AMPL, Lingo, GAMS, MatLab, etc.). Once a model has been created using input data, it can then be solved. If a solution file is generated, it can subsequently be parsed to obtain associated information, including selected facility locations, demand covered, etc. This then enables import to GIS for further analysis, display, summary, etc. This is or can be a tedious process, and requires expertise in using model building software, data manipulation, and interfacing with GIS. This represents a loosely coupled approach as all of the data links, etc. are done as tasks by the analyst. The advantage of this approach is that one can use the solution method of choice, either a general purpose software package or a specially designed heuristic or algorithm.

The important point is that application of a location covering model may not be an easy task to accomplish using existing software without extensive experience. Although there are some location packages that are designed for specific purposes and are relatively easy to use, most are designed for a specific type of data and solve a specific model. What is needed is greater functionality within GIS, perhaps even integrating a general purpose model builder (e.g., OPL, IVE, AMPL, Lingo, etc.) and solver (e.g., Cplex, XPRESS, Gurobi, etc.) into the software. Even development of stand-alone software packages are important to consider, offering potential to make it easy to import spatial, temporal and attribute data in order to structure and solve covering models, specifically, and location-allocation models in general. The barriers to covering model application need to be lowered. This is an important issue for future research and development to address. Great strides have been made in the area of spatial analytics, particularly spatial statistics, to make open source code and products generally available to users. Examples include PySAL and GeoDA, with use and application increasing substantially as a result.

11.6 GIS Developers

In the previous section we described some of the issues in applying location covering models. The major impediments in developing applications include the need for expertise in modeling/optimization, data creation, GIS, data visualization, etc. For some rudimentary problems like LSCP and MCLP, it is possible to set up, solve, and display results of a specific application, all within select commercial GIS packages. But this does not extend to other location covering models, or even to a more advanced problem domain, such as Euclidean and continuous space coverage problem variants. There is little flexibility in what can be solved, limited techniques for problem solution, and issues with the simple visual depiction and summary of solutions. For example, although one can fix out certain facility sites (or fix in certain sites) before solving, this functionality is not simple to use. Map-based displays leave little flexibility to test parameters when solving a problem. For example, if one wants to view $p = 5$, 6, and 7 solutions, each of these problems need to be set up individually, run, and displayed. There are no easy ways to click on a facility location and move it (or close it), as an example, then resolve. Such functionality is at the heart of management plan development.

The reasons for this are clear. GIS software is designed to support data creation, storage, management, manipulation and display. Analysis capabilities, while vastly improved over the last 30 years, remains limited. Because of the difficulty in carrying out analysis, users have been forced to develop functions/add ins that carry out tasks, or have settled on simple map-based displays for a user to make a decision without supporting analytics. Perhaps there is a perception that the market does not exist to support further interface development, but new products like Microsoft's Power BI appear to address some of these issues. Overall, the potential for GIS developers to support location analytics remains significant, but has far to go if software is to make a meaningful impact on analysis and decision making processes.

11.7 Artificial Intelligence and Machine Learning

Samuel (1959) described machine learning as giving "… computers the ability to learn without being explicitly programmed." This is, in itself, an exciting possibility. But how is learning actually structured within a computer? Machine learning relies heavily on statistical modeling. It is hard to simply classify all of the types of machine learning in a short paragraph or two, however, one major distinction can be made between supervised machine learning and unsupervised machine learning. Whereas the former involves giving a machine a "training set" of data from which to develop rules or a model that can be used to predict some outcome or choice, the latter involves giving the machine all of the data and the machine proceeds to find patterns in that data. Often, there is some structure as to how a machine "learns." A simple type of machine learning situation involving a location problem might be to give a machine a training set of good solutions and bad solutions, each categorized as to being good or bad, and then attempting to learn the rules that make something good or bad. An advanced problem might be that a machine not only learns which is good or bad, but then tries to identify new candidate solutions which would be classified as being good or even excellent. The fact is that machine learning may add a new perspective to how we view and solve problems. For example, it may be possible to take the decisions made by Home Depot in their selection of store locations and predict which locations they might select next or in the future.

11.8 Summary and Concluding Comments

Overall, there are a number of significant research and development challenges in location science, and in particular coverage-based modeling. These challenges include the development of better and more capable algorithms and heuristics, addressing representation issues, embracing the techniques of machine learning as well as producing better and more useful software in order to promote the transfer of research to the application and decision making processes. There are gaps in our knowledge base, especially as to whether solving transformed problems are competitive and equally robust to that of direct solution procedures. We also face a challenge to make public sector models more appealing for general use by exploring equity issues as well better estimates and metrics of service benefit.

References

Aly AA, White JA (1978) Probabilistic formulation of the emergency service location problem. J Oper Res Soc 29(2):1167–1179
Benveniste R (1982) A note on the set covering problem. J Oper Res Soc 33(3):261–265
Box GE (1976) Science and statistics. J Am Stat Assoc 71(356):791–799
Brill ED Jr (1979) The use of optimization models in public-sector planning. Manag Sci 25(5):413–422
Brill ED Jr, Chang SY, Hopkins LD (1982) Modeling to generate alternatives: the HSJ approach and an illustration using a problem in land use planning. Manag Sci 28(3):221–235
Church RL (1974) Synthesis of a class of public facilities location models. PhD Dissertation, The Johns Hopkins University, Baltimore, MD
Church RL, Li W (2016) Estimating spatial efficiency using cyber search, GIS, and spatial optimization: a case study of fire service deployment in Los Angeles County. Int J Geogr Inf Sci 30(3):535–553
Church RL, Weaver JR (1986) Theoretical links between median and coverage location problems. Ann Oper Res 6(1):1–19
Current JR, Schilling DA (1990) Analysis of errors due to demand data aggregation in the set covering and maximal covering location problems. Geogr Anal 22(2):116–126
Daskin MS, Haghani AE, Khanal M, Malandraki C (1989) Aggregation effects in maximum covering models. Ann Oper Res 18(1):113–139
Hillsman EL (1984) The p-median structure as a unified linear model for location—allocation analysis. Environ Plan A 16(3):305–318
Liebman JC (1976) Some simple-minded observations on the role of optimization in public systems decision-making. Interfaces 6(4):102–108
Marr B (2016a) Big Data: mind-boggling facts everyone must read. Forbes (posted online Sept 2015)
Marr B (2016b) Big data in practice. Wiley, Chichester
Murray AT, O'Kelly ME (2002) Assessing representation error in point-based coverage modeling. J Geogr Syst 4(2):171–191
ReVelle C, Marks D, Liebman JC (1970) An analysis of private and public sector location models. Manag Sci 16(11):692–707
Salvaneschi L (1996) Location, location, location: how to select the best site for your business. Oasis Press/PSI Research, Grants Pass, OR
Samuel A (1959) Some studies in machine learning using the game of checkers. IBM J Res Dev 3(3):210–229
Scaparra MP, Church RL, Medrano FA (2014) Corridor location: the multi-gateway shortest path model. J Geogr Syst 16(3):287–309
Wedel M, Kannan PK (2016) Marketing analytics for data-rich environments. J Mark 80:97–121

Index

© Springer International Publishing AG, part of Springer Nature 2018

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6