PERGAMON

Computers & Operations Research 29 (2002) 667–684

computers & operations research

www.elsevier.com/locate/dsw

A simulated annealing approach to police district design

Steven J. D'Amico[a], Shoou-Jiun Wang[b], Rajan Batta[c], Christopher M. Rump[c,*]

[a]*Blasland, Bouck & Lee, Inc. 6723 Towpath Road, Syracuse, NY 13214, USA*
[b]*First USA Bank, DE1-1027, 201 North Walnut Street, Wilmington, DE 19801, USA*
[c]*Department of Industrial Engineering, University at Buffalo, State University of New York, 342 Lawrence D. Bell Hall, Box 602050 Buffalo, NY 14260, USA*

Abstract

This paper considers the problem of redistricting or redrawing police command boundaries. We model this problem as a constrained graph-partitioning problem involving the partitioning of a police jurisdiction into command districts subject to constraints of contiguity, compactness, convexity and size. Since the districting affects urban emergency services, there also exist quality-of-service constraints, which limit the response time (queue time plus travel time) to calls for service. Confronted with the combinatorial challenge of the districting problem, we propose a simulated annealing algorithm to search for a "good" partitioning of the police jurisdiction. At each iteration of the algorithm, we employ a variant of the well-known PCAM model to optimally assign the patrol cars and assess the "goodness" of a particular district design with respect to some prescribed performance measures. This approach differs from the well-known Hypercube queuing model, which simply evaluates the performance of a user-specified district design and allocation. A computational case study using data from the Buffalo, New York, Police Department reveals the merits of this approach.

Scope and purpose

Two of the primary concerns of urban police departments are the effective use of patrol cars and the workload balance between officers in different geographical districts. In recent years, a well known, public domain software package based on the Patrol Car Allocation Model (PCAM) has been developed. PCAM was designed to help police management specify the number of patrol cars that should be on duty at various times of the day on each day of the week in each district. For long-term planning, police management also faces the thorny problem of designing these districts. To address this problem, we employ a simulated annealing search method to determine the geographic boundaries between the police districts. PCAM is used to evaluate the "goodness" of each district design encountered in this search, where "goodness" involves minimizing the disparity between the maximum workload and the minimum workload of the police officers. Working with the Buffalo, New York, Police Department, we were able to significantly reduce officer workload disparity while maintaining current levels of response time.

* Corresponding author. Tel.: +1-716-645-2357; fax: +1-716-645-3302.
E-mail address: crump@eng.buffalo.edu (C.M. Rump).

0305-0548/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0305-0548(01)00056-9

Keywords: Districting; Graph partitioning; Simulated annealing; Urban emergency services; Patrol car allocation; Hypercube queuing model

1. Introduction

Operations Research plays an important role in the optimal deployment of scarce urban resources. Emergency services such as police, fire and ambulance services are highly labor intensive, with personnel costs accounting for more than 90% of operating costs [1]. In addition, for urban police departments, typically over half of all these personnel resources are used for patrol car operations [2]. Therefore, a major goal of police management is to find an equitable and efficient distribution of these patrol resources across their city or jurisdiction.

To make this distribution more manageable, police departments face the mammoth task of partitioning their jurisdiction into command districts or precincts. Each command district usually has a headquarters and commanding officer to oversee its police operations. Districts are further subdivided into patrol sectors or beats, with at least one patrol car assigned to each beat. On an even smaller scale, sectors are composed of reporting districts, the smallest geographical area for which police statistics are kept and reported to interested parties. These atomic elements, which for many cities coincide to census block groups, are the geographical building blocks from which sectors and districts are defined [3]. Viewing each of these so-called R-districts as a node connected to adjacent R-districts via arcs, we can build a graph to represent the police jurisdiction. The districting problem then takes the form of a graph-partitioning problem, which is known to be NP-complete [4].

Despite its complexity, the districting problem must be solved in a variety of contexts, such as in deciding sales territories, school districts, political districts and, in our case, police command districts. Hence, there exists much literature on districting problems, particularly in the political realm. Williams [5] provides a review of this vast literature on political redistricting problems. Escaping review, George et al. [6] describes an iterative location/allocation method applied within a geographic information system (GIS) to electoral districting in New Zealand. More recently, Mehrotra et al. [7], building on the earlier work of Garfinkel and Nemhauser [8], use a column generation approach to select political districts. Similar methods have also been used in designing school districts [9] and sales territories [10].

This article, an outgrowth of preliminary work done by one author for a Master's thesis [11], develops a new approach to police districting. Our methodology, like some political studies (e.g., Nagel [12]), attempts to improve an existing district design by using local search techniques that involve swapping population units—in our case R-districts—from one district to a neighboring district. However, we employ the more recently developed search technique of simulated annealing (cf., e.g., Reeves [13]). Simulated annealing, like other metaheuristics such as tabu search and genetic algorithms, has been successful in finding reasonably good solutions for many combinatorial optimization problems. In fact, upon completion of this work, we found a recent paper by Bozkaya, Erkut, and Laporte [14] that applies tabu search to the political districting problem.

Districting problems all share similar concerns for geographical compactness, contiguity, and some measure of equal "size", whether size is measured in terms of area, population or customers. A combination of these measures determines the overall effectiveness of a particular districting design. For example, Kaiser [15] measures compactness with a moment of inertia of the area of a district. Horn [16] minimizes the total length of district perimeters to find the most compact partitioning. Mehrotra et al. [7] employs a branch-and-price methodology to evaluate the "non-compactness" cost of a proposed political district.

What distinguishes the police districting problems from these other applications is an additional consideration for patrol officer workloads and response times to calls for service (CFS). These considerations require incorporating queueing measures; officer workloads constitute the utilization of servers, and response times constitute customer waiting times. These average queueing measures are highly non-linear functions of call rates and service times and therefore yield a difficult non-linear optimization problem. We search for a geographic partitioning that addresses these queueing-related concerns, with common districting constraints on geographical contiguity, compactness, convexity, and size.

The remainder of this paper is organized as follows. The next section explains how we exploit well-known patrol car allocation models to address the queueing concerns. Section 3 discusses the implementation of our simulated annealing approach. In particular, we define feasibility rules for neighboring solutions that ensure some amount of geographical contiguity, compactness, convexity, and uniform size. In Section 4, we present the results of a case study of this methodology on data collected from the Buffalo, New York, Police Department. The final section offers conclusions and suggestions for future work.

2. Patrol car allocation

A traditional approach for allocating patrol cars involves splitting car resources in direct proportion to the total calls for service, giving equal weight to all priorities. A pitfall of this so-called hazard method is that it ignores queueing effects [1]. Fewer officers may be required to serve an area with a large call volume, simply because the service time (including travel) for those calls is smaller than in other low-call-volume areas.

Recognizing the need to account for queueing effects, we turned to well-known queueing models that have been developed in the past 25 years for deployment of police resources. Larson's [17] seminal Hypercube Queueing Model has been widely used to design police patrol beats and sectors. Its early computer implementation [18] has been migrated to the PC environment in what is called the Desktop Hypercube (DH). The DH software developed by Sacks has enabled the Orlando Police Department to balance officer workloads and response times across patrol beats while maintaining neighborhood boundaries [19]. The Hypercube model has been implemented in several other cities as well. Dizengoff [20] used a prototype version of the DH in an analysis of the Chapel Hill, North Carolina Police Department.

Treating each patrol car as a distinguishable entity, the Hypercube model is able to generate, for each car, performance statistics such as the travel time to incidents, workload, and proportion of dispatches outside its assigned beat. Two weaknesses of the Hypercube model are its inability to handle call priorities or to use time-dependent rather than steady-state input data.

A competing model is the PCAM model developed by Chaiken and Dormont [21,22]. The PCAM model is capable of prescribing the number of patrol cars to allocate to each command district for any particular hour of the week. The model handles priority calls-for-service (CFS) data and time-dependent input. It is also flexible enough to be implemented by police departments that have a wide variety of command area configurations, shift schedule constraints and dispatching policies.

The public domain PC software implementation of PCAM [23] is a modified version of the original [24]. The PCAM software does not require user expertise in a particular computer programming language for its implementation. PCAM users are able to specify a performance objective, performance constraints, or a combination of performance constraints and a performance objective. Performance objectives include minimizing average response time to a call, average queue time for a call or the fraction of calls delayed in queue. Constraints allow for imposing upper bounds on such performance measures as workload percentage, fraction of calls delayed in queue, average travel time, as well as response time or queuing delay for lower priority calls.

There are practical drawbacks to using either PCAM or DH for district design. The drawback to a practical use of the DH is that it is a descriptive tool rather than a prescriptive (optimizing) tool. The DH requires that the user specify the sector or beat for each patrol car, in essence, the partition. The DH then evaluates the performance of that design. It is up to the user to then decide whether or not the performance is good enough. Likewise, although it is an optimizing tool, PCAM only prescribes an optimal allocation of cars to a defined region, not what the region should look like. Thus, although these tools provide a quick performance evaluation of a particular partition of the city, the combinatorial challenge of finding an optimal partition still exists. It is with this in mind that we turned to a simulated annealing search routine.

3. Simulated annealing approach

Our police-districting problem—like all districting problems—is a combinatorial challenge. Due to concerns for queueing performance the objectives and constraints on our problem are highly non-linear. Thus, to search for a "good" solution of this problem, we propose a simulated annealing heuristic approach. At each iteration of the simulated annealing algorithm, PCAM is applied to find an optimal allocation of patrol cars to a given partition.

The simulated annealing algorithm starts with an initial solution, for instance the current district partitioning. At each iteration, a prospective solution is generated from the current partition by reassigning an atom (R-district) located on the border between two adjacent districts. If the prospective solution's objective value is better than that of the best solution, then the prospective solution is saved as the best solution and becomes the current solution. If the prospective solution is not better than the best solution, then it may become the current solution regardless with an acceptance probability determined by the so-called temperature parameter (cf., e.g., [13]). In a simulated annealing algorithm, the temperature gradually decreases as the algorithm progresses, thereby decreasing the acceptance probability. Thus, the effect is to reject inferior solutions at greater frequency as the algorithm progresses.

The probability of inferior solution acceptance, p, at each iteration of the simulated annealing algorithm is given by

$$p = \exp\left(-\frac{v(s) - v(s_0)}{t}\right),$$

where $v(s)$ = objective value of the prospective solution, s; $v(s_0)$ = objective value of the current solution, s_0; t = temperature.

Eventually, it is desired that the algorithm stop with an optimal or near-optimal solution. The overall effect of the simulated annealing algorithm is to begin searching the solution space in a highly random manner, often replacing the current solution with inferior solutions. This random solution-replacement process provides the algorithm freedom to explore possible solutions and prevents confinement to a local minimum.

However, as the algorithm progresses, the chance that an inferior solution displaces the current one gradually decreases. If the algorithm was provided enough chances to explore all possible solutions, then this decrease is analogous to keeping increasingly superior solutions. As the probability of accepting an inferior solution decreases further, the algorithm tends to avoid drastic moves among possible solutions, and begins searching locally for a superior solution in the "neighborhood" of the current solution.

The neighborhood of a solution is defined as follows. Given a particular current solution and its associated value, the simulated annealing algorithm requires a systematic method for choosing another solution in the next iteration. The process of choosing the next solution is managed by defining a *neighboring* solution. In this application, a neighbor of a current solution is a partition that is identical to the current solution except that one atom (R-district) is reassigned from a district to an adjacent district. The neighbor definition will result in one atom, on the border between two adjacent districts, being swapped from one district to the other district at each iteration. With a neighbor defined, the algorithm has the means for moving from one solution to the next. However, to keep the algorithm from favoring particular neighbors and possibly precluding the best solution from being visited, all the neighbors of a particular current solution must be given an equal chance of being visited. This requires that we randomly select which neighbor to visit next. The set of all neighbors adjacent to the current solution is referred to as its *neighborhood*.

At each iteration, the current solution is assumed to be feasible. To insure that the next solution chosen is also feasible (i.e., whether or not it is eligible to replace the current solution), it must be clear what conditions constitute feasibility. In this application, feasibility depends on both a quality-of-service constraint as well as several desirable physical attributes for each geographical district. In particular we describe the constraints as follows:

1. *Response Time*—The average response time in each district during all weekly time blocks in the week should not exceed a specified upper bound.
2. *Size*—The ratio of the areas of the largest and smallest districts should not exceed a specified upper bound.
3. *Contiguity*—Each district should remain connected.
4. *Compactness*—The ratio of the longest Euclidean path and the square root of the area should not exceed a specified upper bound.

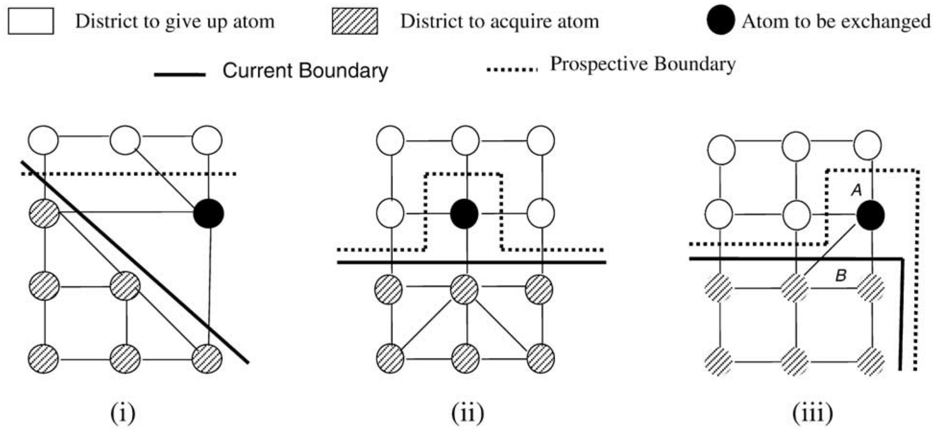

Fig. 1. Examples of convexity violations.

5. *Convexity*—The atom added should not create a protrusion out of the new district, and the atom removed should not create an indentation in the old district.

The response time constraint is needed to ensure that the solution suggested at the end of the procedure is implementable—we do want to obtain a better workload distribution at the expense of very poor quality of service. The size constraint avoids the possibility of producing, for example, one large district and several small ones. The contiguity constraint is necessary to keep patrol officers from crossing district boundaries. The compactness constraint avoids gerrymandering, ensuring that a district will have a relatively round rather than long and slender shape. Aside from practicality and aesthetics, a technical reason to require compactness is that the PCAM algorithm uses the square root law to estimate travel time. This estimation technique is accurate only for compact areas [25].

Of the five constraints, the convexity constraint is the only one that is somewhat unclear. We have to clarify the terms "protrusion" or "indentation". The following three transfer situations (each illustrated separately in Fig. 1) are considered infeasible situations that violate convexity:

(i) an exchange which causes an atom to be too far away from the closest atom in its district;
(ii) an exchange which causes an atom to be adjacent to only one other atom in its district;
(iii) if A is the exchanged atom and B is an atom adjacent to A in the receiving district, then B must be adjacent to at least four atoms (including A) in its district.

To avoid severe indentations, we also developed the following procedure to check convexity. First, we determined a direction vector's head by averaging the coordinates of the atoms in the giving district that are adjacent to the exchanged atom. Similarly, to obtain the coordinates of the direction vector's tail we averaged the coordinates the atoms in the receiving district that are adjacent to the exchanged atom. With the direction vector and the exchanged atom, we formed a new coordinate system as shown in Fig. 2. We considered the exchange infeasible if both Quadrants 2 and 3 had atoms from the giving command district.

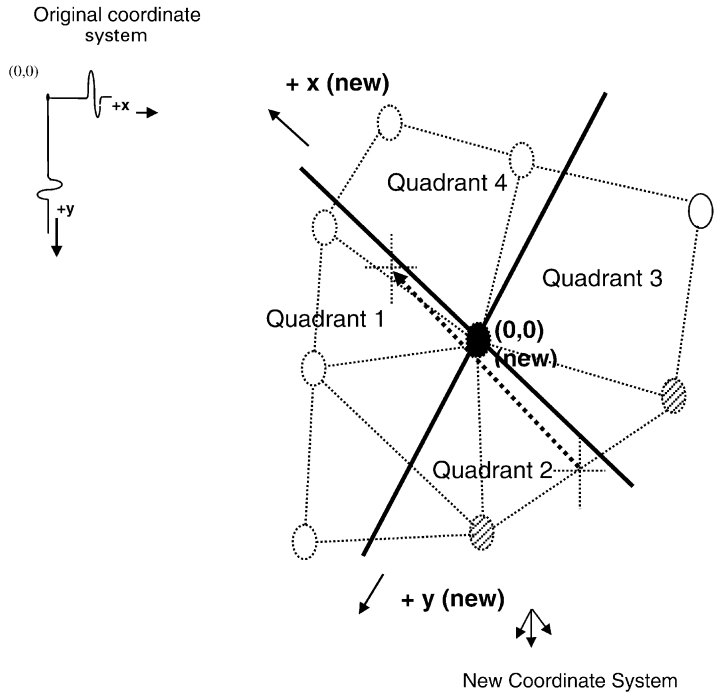

Fig. 2. Coordinate transformation to check convexity.

With these feasibility constraints in place, the simulated annealing algorithm can be performed. Dowsland [26] offers a generic statement of the algorithm, in the form for our application. The performance of the simulated annealing algorithm, however, depends heavily on several algorithmic parameters. These parameters include the temperature, the rate at which the temperature decreases (and hence p decreases), and the termination conditions such as a pre-specified total number of iterations. These parameters must be adjusted in order to improve the efficiency of the algorithm. To aid in this task, experimentation on a trial-and-error basis is applied in most applications. White [25], however, describes a more refined technique involving computing the standard deviation of the changes in the objective value.

4. Buffalo police case study

Buffalo is the second-largest city in the State of New York with a population of about 330,000. The City of Buffalo is located on the shores of Lake Erie and comprises approximately 40 square miles. The Buffalo Police Department (BPD) consists of 533 police officers and deploys between 33 and 54 police cars at any point of time. In 1997, there were about 350,000 CFS.

Prior to 1996, the BPD had eight geographical command districts. Through a planning process that involved police chiefs and legislatures in the city government, BPD reorganized into five command districts named A–E. Each of the districts is divided into four sectors or patrol beats numbered 1–4, which are the designated areas in which particular patrol cars are assigned. The

BPD has further divided the City of Buffalo into 409 so-called R-districts, for data tracking purposes. (Each dot in the adjacency graph depicted in Fig. 3 represents the geographic center of an R-district.) These R-districts are relatively small (on the order of a census block group) and form the atoms upon which the sectors and districts are defined [3].

Historically, the BPD has determined command district boundaries from experience. Typically, boundaries lie along major streets and are drawn so that the CFS volumes are approximately the same across the various geographical commands. Since this method ignores differences in service times and non-patrol duties, it fails to balance the workload of geographical commands.

The problem we faced is that of partitioning the City of Buffalo into five contiguous districts, each comprised of R-districts, and deciding how many patrol cars should be allocated to each time block of the five districts.

4.1. Objectives

The chief concern of the BPD administrators was the current disparity in officer workloads, defined to be the percentage of on-duty time spent responding to calls (as opposed to time spent on patrol or non-emergency calls). Fig. 4 shows the reason for their concern. This figure displays the distribution of workload for the BPD's nearly 6,500 weekly carhours, which translates into about 38 patrol cars on an average shift. Particularly disturbing is that over 25% of the time during a week, officers are under-utilized with workloads below 26% (the left tail of the figure). Likewise, nearly 18% of the time, officers are overworked with workloads above 42% (the right tail of the figure).

These results show great variability among officer workloads. From a standpoint of officer morale, it is certainly undesirable for some patrol officers to be responding to emergencies only 20% of the time, whilst their fellow officers, perhaps in a different district or assigned to a different shift, being utilized for emergencies in excess of 50% of their time.

Therefore, the objective of this study, as directed by BPD management, was to minimize the disparity between the maximum and minimum workloads of patrol officers, their internal customers. A secondary concern was to constrain the maximum average response time to calls from their external customers. Unfortunately, PCAM is not able to minimize workload disparity directly. However, since it is possible to constrain workload we chose to impose a maximum workload constraint while minimizing average response time. By successively lowering the workload bound as far as possible, we were, in essence, able to minimize the maximum workload. We deemed this a suitable surrogate to minimizing workload disparity.

4.2. Data manipulation and model validation

According to the current allocation of the BPD, the five districts have the same time block structure. There are four time blocks: 07:00–16:00, 16:00–21:00, 21:00–02:00, and 02:00–07:00. During a time block, the number of patrol cars scheduled for duty does not change. PCAM allows the user to define any number of days to represent a cycle. To account for variations in the call rate and allocations for a given time of day during the week (e.g., weekend days typically yield higher call rates), seven days was chosen as a cycle. For convenience,

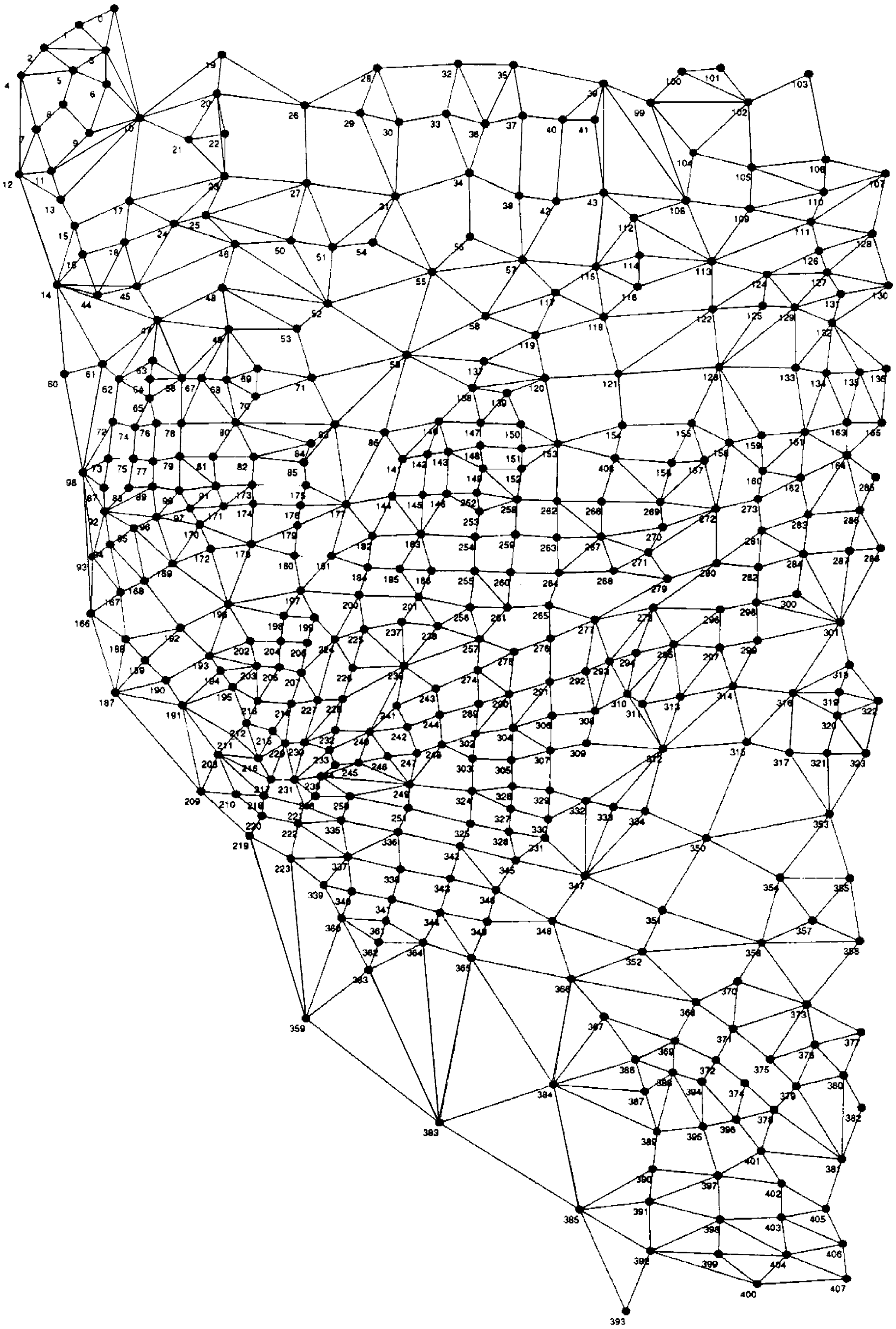

Fig. 3. Adjacency graph induced by R-districts.

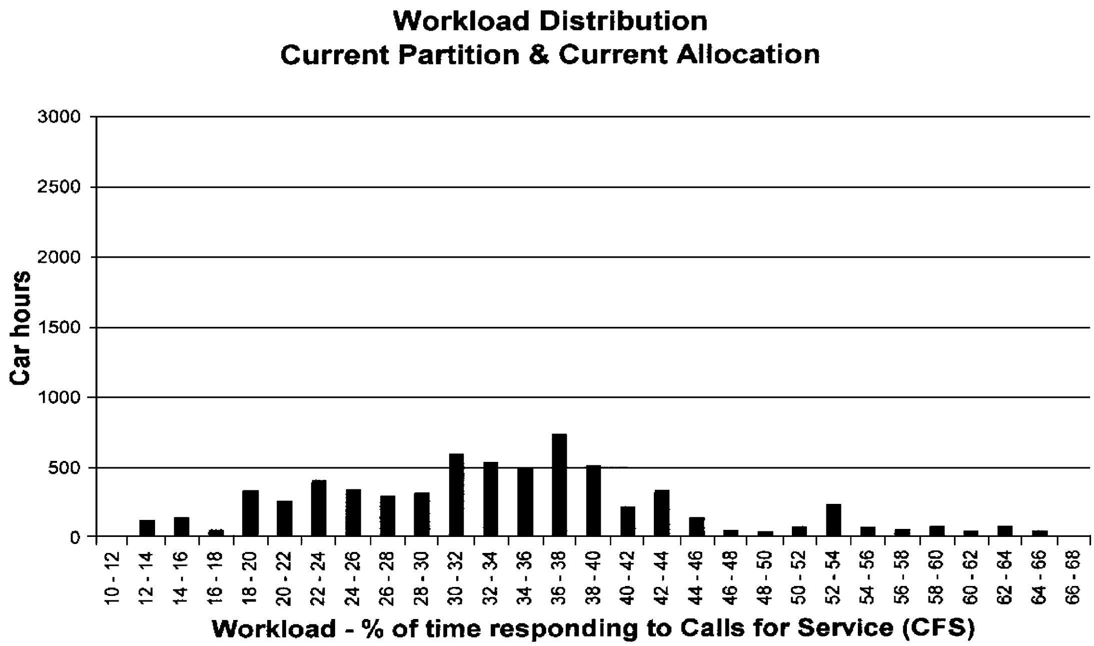

Fig. 4. Workload distribution: current partition with current allocation.

Sunday was chosen to be the first day of the weekly cycle. Therefore, the PCAM cycle began Sunday morning at 07:00 and ended the following Sunday morning at 07:00.

Data for the model was made available to us by the BPD. The analysis is based upon data for the 6-month period from February 1, 1997 through July 31, 1997. We first needed to verify that PCAM was a good predictor of the workload performance measure. Applying PCAM to the data set provided by the BPD and comparing its workload predictions for the 140 time blocks (4 blocks per day $*$ 7 days per week $*$ 5 districts) to the BPD actual workload statistics revealed only a 0.5% deviation, on average. Compared with the current workload disparity of $64.2 - 12.5 = 51.7\%$, this is a very small error. We therefore concluded that PCAM's workload calculations were very accurate. (The implementation of PCAM is not trivial. This accuracy was a result of prior tailoring and fine tuning of PCAM at the BPD prior to this district design study.)

4.3. Simulated annealing parameters

In keeping with the current BPD district design, we defined the following feasibility parameters. The average response time in any given time block was constrained to be no more than 29 min. Also, no district was allowed to be more than 2.5 times the size of any other district. For compactness, the square root of a district's area to its longest Euclidean path was kept below 1.5. For convexity, each atom was required to be within 0.3 miles of its nearest neighboring atom within its own district. Any exchanges that violated any one of these restrictions were not considered feasible "moves" in the simulated annealing algorithm.

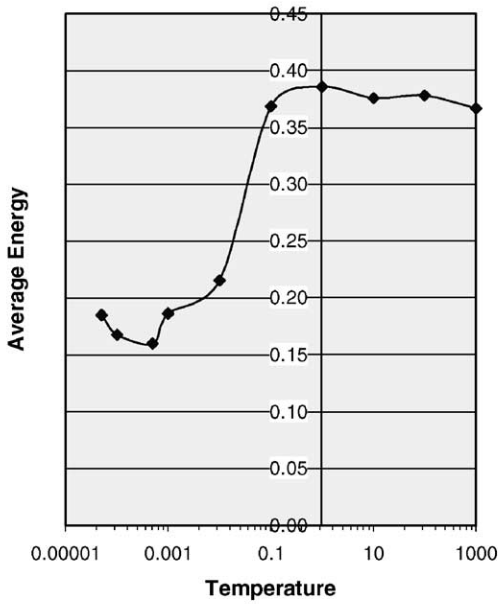

Fig. 5. Annealing curve.

A critical issue in running a simulated annealing algorithm is to determine the initial and the ending temperatures. In our experiments, we first kept the temperature constant and ran the procedure for a large number of iterations. The average energy (objective value) for each constant temperature was then calculated. From the annealing curve depicted in Fig. 5 we can see that when keeping the temperature higher than 0.1 over a large number of iterations, the average energy remains about 37%. Therefore, we considered temperatures over 0.1 extremely high. Under such high temperatures the procedure moves to feasible neighbors randomly even though the neighbors have high energy. Also, 37% can be interpreted to be the average energy of all the possible solutions. The energy for temperatures 0.05, 0.02, 0.017, and 0.016 is 36%, 36%, 33%, and 26%, respectively. Based on this observation, we conclude that in our case temperatures between 0.017 and 0.1 are reasonable choices for the initial temperature.

Starting from such an initial temperature, we then employed a temperature reduction function $\alpha(t) = at$, $0 < a < 1$, over time t. As the temperature dropped, the average energy decreased. This phenomenon is expected since now we are forcing the method gradually to seek better solutions. Notice that as the temperature drops to 0.0001, the average energy values start to increase again. This is due to the rarity of available states at the low tail of the energy distribution and the propensity of the algorithm to get stuck at a local minimum at low temperature. We concluded that 0.0001 is a suitable ending temperature.

4.4. Experimental results

Using the initial temperatures between 0.01 and 0.12, and ending temperature 0.0001, several trials consisting of $n = 400$ K iterations were executed with the current geographical commands

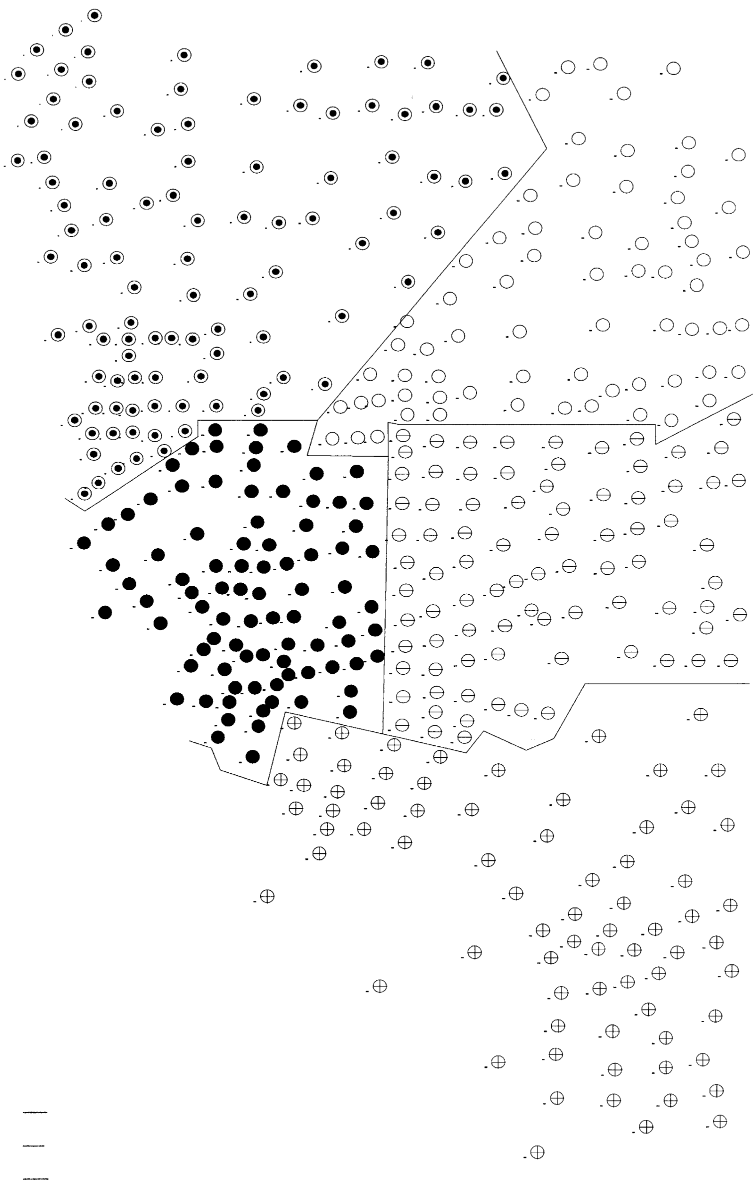

Fig. 6. Current geographical patrol district configuration.

as the initial partition. Upon execution, workload disparities between 0.1413 and 0.1504 were achieved. The best solution with the minimum achieved disparity of 0.1413 occurred with an initial temperature of 0.1. It took a 450 MHz personal computer about 2 h to implement each simulated annealing procedure.

Figs. 6 and 7 depict the current and this best partition, respectively. The R-districts are shaded by district, with lines partitioning the city according to the current partition. Notice that the best

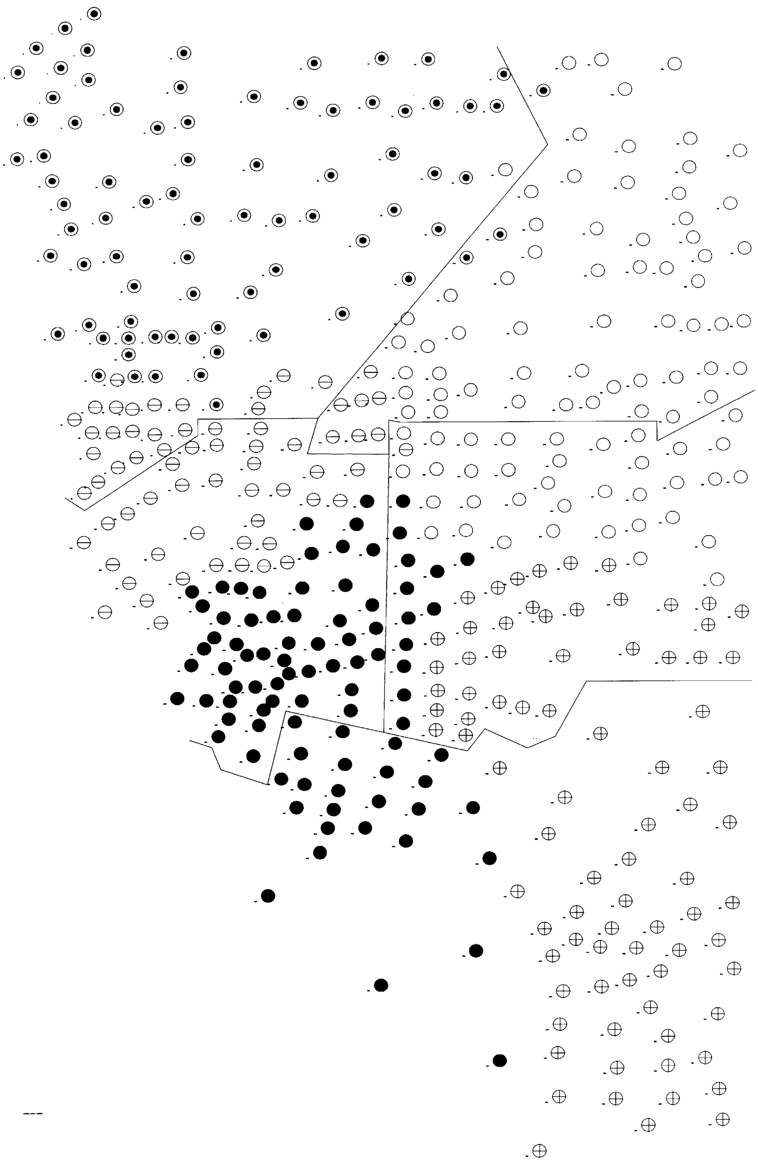

Fig. 7. Best geographical patrol district configuration.

partition has evolved from the current partition primarily by a westward evolution (towards the left of Fig. 7) of the R-districts currently residing in District C of Fig. 6. As a result, many of the darkly shaded R-districts currently in District B have pushed south into current District A.

Notice also that the current partition appears more geographically compact, with a less significant disparity in district size. Despite the current similarity in size, most time blocks in current Districts B and C have a higher workload than time blocks in the other districts. This is due in part to the fact that current District B (West/Downtown Buffalo) has the highest CFS rates, followed by current District C (East Buffalo). In contrast, currently both Districts D and E (North Buffalo) have smaller CFS rates. In these districts, some time blocks have very low workloads even though only one patrol car is assigned to them. The partition suggested by our algorithm allows Districts A (South Buffalo), D, and E to share the workload of Districts B and C. Even though the resulting configuration does not look as geographically compact as the current BPD command configuration, it does improve the workload disparity.

These improvements in workload disparity can also be seen in terms of the workload distribution, when compared to the current situation depicted in Fig. 4. When the patrol cars are optimally allocated to the BPD's current districts (Fig. 8(a)), the high tail of the workload distribution from Fig. 4 is eliminated, and the low tail is reduced significantly. However, some time blocks still have a workload as low as 12%. These low workloads result from the fact that one of the districts has a relatively small CFS rate, yet must be patrolled by a minimum number of cars in order to meet the imposed response time constraint.

By repartitioning the city, our algorithm does improve the workload disparity by raising the lowest workload from around 12% to 26% (see Fig. 8(b)), while lowering the highest workload from under 44% to under 42%. The maximum workload disparity thus improved from about 30% to about 14%. At first glance, it is perhaps discouraging that the maximum workload did not drop more than it did. However, our improved workload disparity apparently must come at the expense of aggregating some R-districts having high CFS rates with other districts having low CFS rates.

4.5. Model robustness

To avoid initial bias induced by always starting the algorithm at the current partition, some random partitions were used to initialize our algorithm. The final results differed by less than 1%.

Alternative possibilities for moving from one solution to the next were considered as well. In addition to the swaps of one R-district atoms at each iteration, we also performed swaps of size k atoms, for $k = 2, 5, 10, 15$. Here, feasibility is checked in two stages. In the first stage, k individual atoms were swapped one at a time, with a check for geographic feasibility (as discussed in Section 3) before each swap. In the second and more time-consuming stage, PCAM was used to check for response time feasibility for the new configuration resulting from these k swaps. The motivation of multiple swaps is to save on the total number of iterations each containing a time-consuming PCAM call. Trials at $k = 2$ and 5 yielded solutions close to (and often identical to) the solutions found for $k = 1$. The swaps at $k = 10$ and 15 performed gradually worse.

Also, there was some concern that if the feasibility constraints described in Section 3.3 were too strict, the algorithm would not be very "active" in moving around, possibly trapping itself at a local minimum near the initial solution. To alleviate this concern, we ran the algorithm while ignoring all of the geographic feasibility constraints except for contiguity. The empirical lower bound we obtained was approximately 13%, which does not improve our current objective value by more than 1%. Therefore, we conclude that the strict feasibility constraints and initial solutions are not critical to the objective value in this situation and the current objective is relatively close to the empirical lower bound.

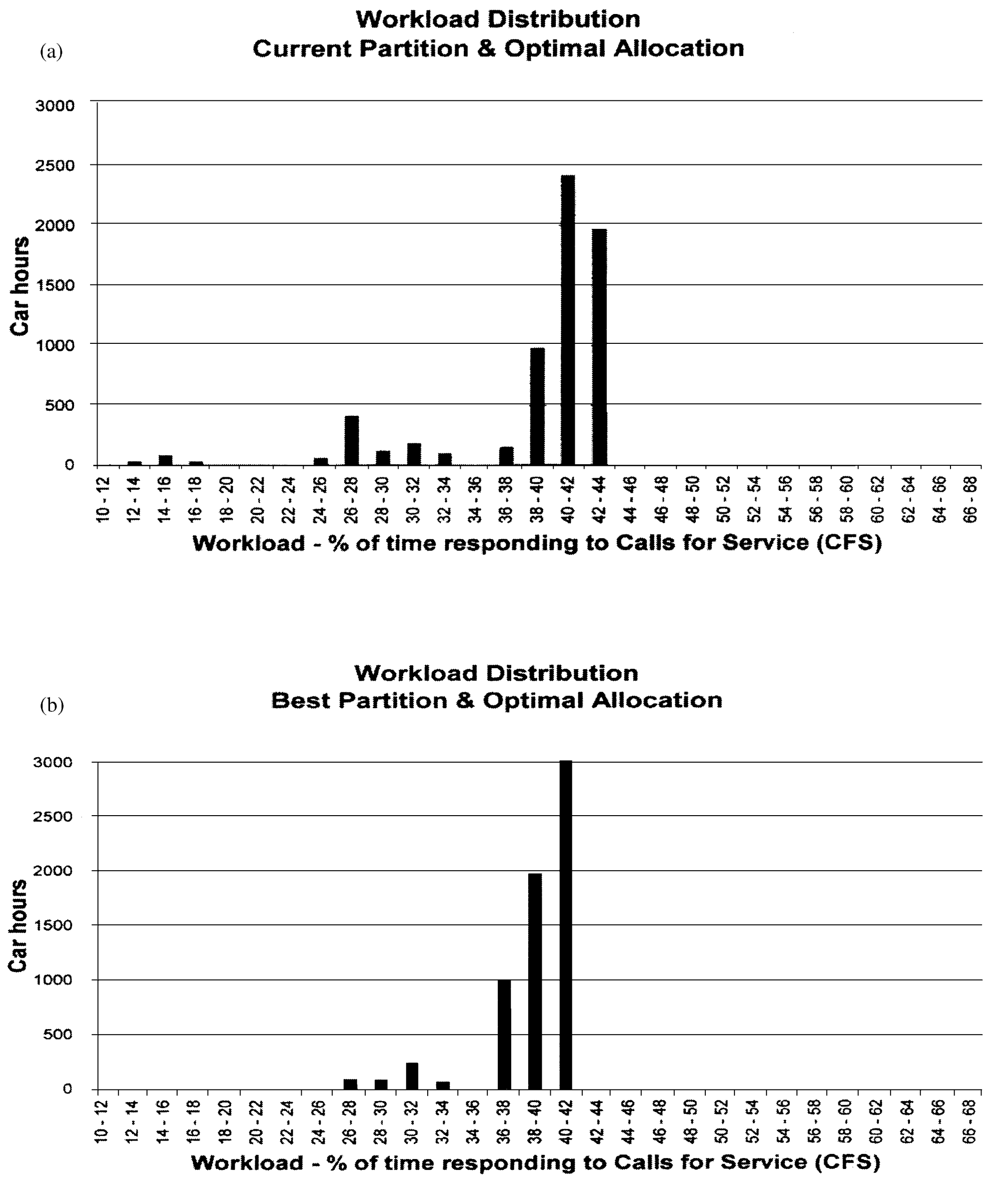

Fig. 8. Workload distribution: current vs. best partitions.

5. Conclusions and suggestions for further work

We proposed a simulated annealing search methodology for police command districting. At each iteration of the simulated annealing heuristic, we utilized a variant of the well-known PCAM to allocate patrol resources to the current district partition and thereby evaluate the expected performance of that design. A case study on data collected from the BPD established the merits of this approach.

Using PCAM to allocate cars optimally to the current BPD districts goes a long way towards minimizing disparity in officer workloads—time spent answering calls for service—across districts with work shifts. Our experience suggests that the current maximum workload disparity of over 50% (some officer workloads are higher than 60%, others however near 10%) can be drastically improved by reallocating cars more effectively. As shown in Fig. 8(a), a proper allocation greatly reduces the variance in workload.

This is not to say our search for improved district boundaries is without merit. Under optimal car allocations, we were able to find an improved district design that lowered the disparity among officer workloads from 30% to only 14%. Also, the proportion of small workloads under 36% was greatly reduced. Hence, officer workloads were more uniformly balanced (primarily between 36% and 42%) across all districts and work shifts. At the same time, our response time feasibility constraints ensured no increase in the maximum response time of 29 min under current BPD operations. Management at BPD is currently reviewing this redistricting proposal for their long-range planning.

A side benefit of the proposed districts is that they help alleviate an annoying dispatching quirk present in the current district configuration. Currently, any calls for service (primarily auto accidents) on the Buffalo Skyway heading south along Lake Erie on the western edge of District A are dispatched to officers from District B. This is due to the fact that the entrance ramp to the Skyway is located in District B just a few blocks from BPD Headquarters in downtown Buffalo. The proposed configuration alleviates this problem by covering the Skyway with one district.

A suggestion for possible improvement of the algorithm is to start with a partition that minimizes the average boundary of each geographical command. This is to make the command shape "round" and is expected to decrease response time by reducing travel time. Further improvements could be achieved by introducing some geographical information system (GIS) software to help estimate the statistics needed in applying PCAM. For example, PCAM estimates the travel time to the incident by square root laws. GIS software could provide more accurate information about the streets and district areas, thereby offering better estimates of travel time. The authors are presently engaged in this activity.

Acknowledgements

This research was partially supported by a grant from the Buffalo Police Department (BPD) under then Commissioner R. Gil Kerlikowske and then Chief of Staff Kevin Comerford. In addition, we received much help and guidance—particularly with the data collection—from Jim Kaufmann in the BPD MIS Department. We thank them all for their valuable support.

References

[1] Swersey AJ. The deployment of police, fire, and emergency medical units. In: Pollock SM, Rothkopf MH, Barnett A, editors. Handbooks in operations research and management science: operations research and the public sector, Vol. 6. Amsterdam: North-Holland Press, 1994. p. 151–200.

[2] Green L. A multiple dispatch queueing model of police patrol operations. Management Science 1984;30: 653–64.
[3] Sarac A, Batta R, Bhadury J, Rump CM. Reconfiguring police reporting districts in the city of Buffalo. OR Insight 1999;12(3):16–24.
[4] Papadimitriou CH, Steiglitz K. Combinatorial optimization: algorithms and complexity. Englewood Cliffs, NJ: Prentice-Hall, 1982.
[5] Williams Jr JC. Political redistricting: a review. Papers in Regional Science 1995;74(1):13–40.
[6] George JA, Lamar BW, Wallace CA. Determining New Zealand electoral districts using a network-based model. Operational Research Society of New Zealand, Proceedings 1993;29:276–83.
[7] Mehrotra A, Johnson EL, Nemhauser GL. An optimization based heuristic for political districting. Management Science 1998;44(8):1100–14.
[8] Garfinkel RS, Nemhauser GL. Optimal political districting by implicit enumeration techniques. Management Science 1970;16(8):495–508.
[9] Ferland JA, Guenette G. Decision support system for a school districting problem. Operations Research 1990;38(1):15–21.
[10] Fleischmann B, Paraschis JN. Solving a large scale districting problem: a case report. Computers and Operations Research 1988;15:521–33.
[11] D'Amico SJ. Allocation of police patrol cars in the city of Buffalo. M.S. thesis, Department of Industrial Engineering, State University of New York at Buffalo, Buffalo, NY, 1997.
[12] Nagel S. Simplified bipartisan computer redistricting. Stanford Law Review 1965;17:863–99.
[13] Reeves CR. Modern heuristic techniques for combinatorial problems. New York: Halsted Press, 1993.
[14] Bozkaya B, Erkut E, Laporte G. A tabu search heuristic for political districting. Research Report No. CRT-99-12, Centre de recherché sur les transports (CRT), 1999.
[15] Kaiser H. An objective method for establishing legislative districts. Midwest Journal of Political Science 1966;X:200–13.
[16] Horn MET. Solution techniques for large regional partitioning problems. Geographical Analysis 1995;27: 230–48.
[17] Larson RC. A hypercube queuing model for facility location and redistricting in urban emergency services. Computers and Operations Research 1974;1:67–95.
[18] Larson RC. Hypercube queuing model: program description. New York City Rand Institute, New York, 1975.
[19] Sacks SR, Grief S. Orlando magic. OR/MS Today 1994;21:30–2.
[20] Dizengoff D. Efficient patrol allocation through improved sector design at the Chapel Hill Police Department. M.S. thesis, Department of Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC, 1991.
[21] Chaiken JM, Dormont P. A patrol car allocation model: background. Management Science 1978a;24: 1280–90.
[22] Chaiken JM, Dormont P. A patrol car allocation model: capabilities and algorithms. Management Science 1978b;24:1291–300.
[23] Chaiken JM, Walker WE. Patrol car allocation model: user's manual. Rand Corporation Publication R-3087/2-NIJ, 1985.
[24] Chaiken J, Walker WE, Dormont P. A patrol car allocation model. Version 2.1, Rand Corporation (Modified for the Hartford Police Department by S. Sacks, The University of Connecticut, Storrs, CT), 1995.
[25] White SR. Concepts of scale in simulated annealing. Proceedings IEEE International Conference on Computer Design 1984;11:646–51.
[26] Dowsland KA. Simulated annealing. In: Collin RR, editor. Modern heuristic techniques for combinatorial problems. New York: Wiley, 1993.

Steven J. D'Amico is a consultant at Blasland, Bouck & Lee, Inc. in Syracuse, New York. He received a B.S. in Physics from the State University of New York at Binghamton before attending the State University of New York at Buffalo where he earned an M.S. in Industrial Engineering with a concentration in Operations Research.

Shoou-Jiun Wang is a credit/risk analyst at First USA Bank. He received a B.S. in Mathematics from National Cheng-Kung University, Tainan, Taiwan before attending the State University of New York at Buffalo where he

earned an M.S. in Statistics and a Ph.D. in Industrial Engineering with a concentration in Operations Research. His research interests are in applied operations research.

Rajan Batta is Professor and Chair of the Department of Industrial Engineering, University at Buffalo, State University of New York. He received his B.Tech. in Mechanical Engineering from the Indian Institute of Technology, New Delhi, India, and his Ph.D. in Operations Research from the Massachusetts Institute of Technology, Cambridge, Massachusetts. His research interests lie mainly in the area of Applied Operations Research where he has published approximately 50 journal papers. He has also been active professionally and serves as a member of the editorial advisory board for *Computers & Operations Research*. His research has been funded by several agencies and corporations including the National Science Foundation, the National Institute of Justice, Lockheed Martin Federal Systems, the Center for Transportation Injury Research, and Federal Express.

Christopher M. Rump is an Assistant Professor in the Department of Industrial Engineering at the University at Buffalo, State University of New York. He received a B.S. in Applied Mathematical Sciences from Texas A&M University in 1989 and a Ph.D. in Operations Research from The University of North Carolina at Chapel Hill in 1995. His research interests include queueing decision models, game theory and applied probability modeling, with application to computer/communication systems, manufacturing and public sector service systems.