Simulation Combined Approach to Police Patrol Services Staffing

Hanjing Zhang[1], Antuela Tako[1], Lisa M. Jackson[3], and Jiyin Liu[4]

1 School of Business and Economics, Loughborough University, Leicestershire, LE11 3TU, United Kingdom
h.zhang@lboro.ac.uk
2 School of Business and Economics, Loughborough University, Leicestershire, LE11 3TU, United Kingdom
a.takou@lboro.ac.uk
3 Aeronautical and Automotive Engineering, Loughborough University, Leicestershire, LE11 3TU, United Kingdom
l.m.jackson@lboro.ac.uk
4 School of Business and Economics, Loughborough University, Leicestershire, LE11 3TU, United Kingdom
j.y.liu@lboro.ac.uk

Abstract

Motivated by the squeeze on public service expenditure, staffing is an important issue for service systems, which are required to maintain or even improve their service levels in order to meet general public demand. This paper considers Police Patrol Service Systems (PPSSs) where staffing issues are extremely serious and important because they have an impact on service costs, quality and public-safety. Police patrol service systems are of particularly interest because the demand for service exhibits large time-varying characteristics. In this case, incidents with different urgent grades have different targets of patrol officers' immediate attendances. A new method is proposed which aims to determine appropriate staffing levels. This method starts at a refinement of the Square Root Staffing (SRS) algorithm which introduces the possibility of a delay in responding to a priority incident. Simulation of queueing systems will then be implemented to indicate modifications in shift schedules. The proposed method is proved to be effective on a test instance generated from real patrol activity records in a local police force.

1998 ACM Subject Classification G.1.10 Applications

Keywords and phrases Police patrol service system, Time dependent queue, Priority queue, Square root staffing, Simulation

Digital Object Identifier 10.4230/OASIcs.SCOR.2016.4

1 Introduction

Performance measures of the Police Patrol Service Systems (PPSSs) typically focus on response times, especially immediate response to incidents which occupy a large proportion of police patrol resources. According to the nature of reported incidents, call handlers in front desks divide incidents into four grades related to their urgency. They are emergency response incidents, priority response incidents, scheduled response incidents and incidents that could be resolved without deployment. Only the first two grades of incidents, emergency response incidents with threat to life and priority response incidents with necessary officer attendance, require immediate response. The common target for PPSSs in the UK is to attend to over 85% of emergency response incidents within 15 minutes and reaching over

5th Student Conference on Operational Research (SCOR'16).
Editors: Bradley Hardy, Abroon Qazi, and Stefan Ravizza; Article No. 4; pp. 4:1–4:11
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

80% of priority calls in 60 minutes [4, 16]. Unlike the well scheduled demand in industry manufacturing and transportation crews in airlines, police emergency services providers always face time-dependent service requests. Patrol officers must be ready to attend any assigned incident of which the time and place cannot be known in advance. Managers need to make reasonable predictions of required staff levels based on historical records, and allocate enough patrol officers to avoid poor performance with long periods of under-staffing. Long periods of over-staffing should also be avoided with the objective being to minimise the workforce costs.

Response times, as an important performance measure for PPSSs, consist of pre-travel delays and travel times. Accurate prediction of system performance requires information of both parts. Travel times usually depend on the distance between the scene of the crime and the patrol officers' current locations. The difference in speed, caused by many factors including types of roads, traffic situations and incident grades, where for emergency incidents blue lights and sirens can be used, will also influence the time spent on travelling [13, 14]. It is a bit more complicated to analyse pre-travel delays since lack of staff is the main contributing factor. Using queueing systems with time-dependent arrival rates can model pre-travel delays in patrol service systems adequately, but it is not easy to analyse because most queueing theory focuses on long-run steady-state behaviour of stationary queueing models, rather than queueing models switching between under-staffing and over-staffing. Discrete event simulation provides explicit numerical results to evaluate the system performance. However, as simulation models are always built case by case, it is quite time consuming to exhaustively simulate all possible system configurations. In this paper, we propose a scheduling method which combines both analytical results and simulation results for time-dependent priority queueing model to indicate appropriate shift schedules. Comparing with traditional staffing methods, the proposed method adjusts varying staff levels into staff levels of shift patterns to cope with the demand fluctuations.

Traditional methods for solving staffing problems usually assume staff levels on different periods are independent from each other. Staff level is determined to satisfy the service target in each period at lowest cost. Then an integer programming model is solved which fits an optimal shift schedule to these staff levels. As a simple approach, pointwise stationary approximation [8] was proposed for setting staff levels in call centres. This approximation provides a basic analytic support to determine the staff level with piecewise stationary arrival and short service times. Queuing processes in real life are more complex since service time may be relatively long. As a modification, Square Root Staffing (SRS) rule [11] is proposed to capture dependency effects of different periods and more accurate estimations for staffing requirements can be expected. This method has been proved to provide demand estimations that are more thoroughly estimated. However, in two-step staffing algorithms, required staff levels determined in the first step are fixed for generating shift schedules in the second step. Several solutions of required staff levels might exist that lead to shift schedules with different costs. Two-step staffing algorithms may lead to suboptimal solutions of shift schedules [8, 5].

Henderson and Mason [9] proposed a feed-back approach called iterative cutting plane method which updates required staff levels and shift schedules iteratively. It requires system service quality to be a non-decreased concave function of staffing. Continuing with their work, Atlason et al. [3] relax the concavity assumption to pseudo-concavity. The iterative cutting plane algorithm combines a queueing simulation and an integer programming model to generate required staff levels and shift schedules simultaneously. Queueing simulation tests the feasibility of current staff levels and refines staff level feasible set by adding more linear constraints. The integer programming solution provides simulation input in the next iteration.

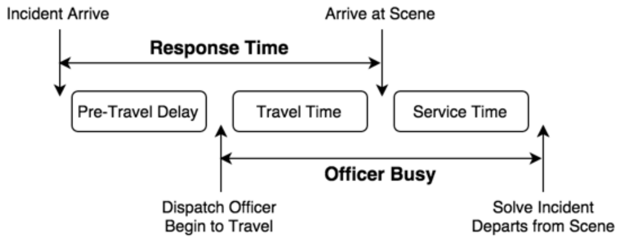

Figure 1 Response process in patrol police service systems.

This algorithm has been widely applied in many other service systems, such as healthcare [1] and delivery service systems [6]. Compared with two-step approaches, some suboptimal solutions for shift scheduling can be avoided in feed-back approaches, but there is an obvious disadvantage that feed-back approaches do not make full use of provided information at the starting point and large computation efforts are required to solve realistic problems.

Some other research has been conducted to construct shift schedules directly without using the concept of required staff level. Ingolfsson et al. [10] use Chapman Kolmogorov equations to evaluate performance of a given schedule. Gans et al. [7] apply integrated forecasting and stochastic programming to generate schedules for call centre staffing. An obvious advantage in applying a direct approach is that they skip the staffing requirement evaluations and focus on the final solution of shift schedules. It will largely reduce the feasible space and make it easier to explore the optimal solution.

In this paper, a new staffing algorithm is proposed to determine appropriate staffing level to fit shift schedules in PPSSs. This algorithm starts at a refined SRS algorithm. This refined square root indicates an appropriate staffing in each period and the corresponding shift schedule. Simulation will then be applied to improve the shift schedule towards the optimal solution.

The remainder of the paper is arranged as follows. Section 2 formally defines PPSSs. A mathematical model for the staffing problem is described in Section 3. A new staffing model based on SRS algorithm is proposed in Section 4 for solving patrol officer staffing problems. Section 5 introduces a simulation combined local search method to modify feasible shift schedules towards the optimal. An example case for PPSS problem is discussed in Section 6. The last section gives conclusions and identifies promising directions for further research.

2 Police patrol service system

Call handlers in contact management departments of police forces are often the first point of contact for the general public. In Leicestershire Police Force, call handlers deal with approximately 16,000 emergency calls and 750,000 non-emergency calls every year [15]. About half of these calls require immediate patrol officer attendance. Depending on reported information of an incident, call handlers will assess the situation and decide the urgent grade this incident should be. Among all the incidents requiring immediate patrol officer attendance, 70% are emergency incidents and the rest 30% are priority incidents. After the grading, free patrol officer will be informed for incident attendance. If no patrol officer is currently available, incidents will wait for service and pre-travel delays will be generated. The response process in PPSSs is illustrated in Figure 1.

When an incident is assigned to an available patrol officer, the patrol officer is labelled as busy until the assigned incident is solved. Response times are recorded as lags between the time incidents being recorded and the time patrol officers arrive at the scene. Incident arrival rate, incident service time and time spent on travel are predictable with the help of historical records. Identical to queueing systems, a patrol police response system consists of three main procedures: arrival, dispatch and service. The arrival of an incident is based on the time at which it is recorded. The service for each incident starts at the time that assigned patrol officers begin to travel to the incident and will last as long as it takes for the incident to be solved. Dispatch refers to the assignment of patrol officers to attend incidents.

PPSSs are similar to non-preemptive priority queues. Emergency incidents and priority incidents have separate logical waiting queues. When a patrol officer becomes free, the incident at the head of the emergency queue will be the next to be served unless all the waiting incidents are priority. The service of a priority incident will not be disrupted even if emergency incidents arrive. From previous experience in patrol response system operations, short delays in dispatching priority incidents are allowed to exist but emergency incidents always expect immediate responses because of travel times. The difference in target response times of emergency incidents and priority incidents is 45 minutes, this being the allowed short delay in responding to priority incidents. When a 24-hour working day is divided into thirty-two 45-minute periods, a modification can be made to tailor time-dependent staff levels into shift patterns by leaving some priority responses to the next period.

3 General staffing model

Integer programming, as a widely used model in workforce management provides an effective way to find a combination of shift staffing which minimizes workforce cost, subject to the target system performance. It is assumed that the whole planning horizon, usually one day, is divided into I non-overlapping periods. In each period $i \in I$, the service system is evaluated to make sure that it maintains a stable performance. Assume there are J predefined shifts. Matrix A records the relationships between shifts and periods.

$$A_{ij} = \begin{cases} 1, & \text{if period } i \text{ and the time of shift } j \text{ overlaps.} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Let c_j be the cost of working on shift j and x_j are the number of staff who work on shift j. The objective function is

$$\min \sum_j c_j x_j \ . \tag{2}$$

Besides x_j, another set of decision variables y_i are introduced to denote the lower bound of required staff in period i. A feasible combination of x_j should provide at least y_i working staff in period i, so

$$\sum_j A_{ij} x_j \geqslant y_i \ , \quad \begin{array}{l} \text{where } x_j \text{ and } y_i \text{ are integers} \\ \text{for all } i \in \{1, 2, \dots, I\} \text{ and } j \in \{1, 2, \dots, J\} \ . \end{array} \tag{3}$$

Minimum required staff levels should provide enough services to meet performance targets. In PPSSs, over 80% of priority incidents should be responded to within the targeted time. Emergency incidents require slightly faster services and at least 85% of them should receive successful responses. For ease of notation, set $\vec{y_i}$ as the vector of staff levels from period 1 to period i.

$$\vec{y_i} = (y_1, y_2, \dots, y_i), \text{ for all period } i \in \{1, 2, \dots, I\}. \tag{4}$$

Service quality in period i only depends on the value of $\vec{y}_i$. Any staffing beyond period i will not influence the system performance in period i. For a period $i \in \{1, \ldots, I\}$, an implicit functions $g_i^1(\vec{y}_i)$ is introduced to denote the actual successful response percentages for emergency incidents being reported in period i. Similar to $g_i^1(\vec{y}_i)$, function $g_i^2(\vec{y}_i)$ is for priority incidents. Constraints about service targets for emergency incidents and priority incidents can be formulated as follows,

$$g_i^1(\vec{y}_i) \geqslant 85\%; \quad g_i^2(\vec{y}_i) \geqslant 80\%, \quad \text{for all period } i \in \{1, 2, \ldots, I\} . \tag{5}$$

As stated before, incident arrival time, travel time and service time are random variables. Successful response percentages $g_i^1(\vec{y}_i)$ and $g_i^2(\vec{y}_i)$ are also random variables when the values of $\vec{y}_i$ are fixed so they can be approximated by their sample average $\hat{g}_i^1(\vec{y}_i)$ and $\hat{g}_i^2(\vec{y}_i)$. Assume simulations are performed K times with the same configuration of having $\vec{y}_i$ staff on shift from period 1 to period i and the successful response percentage of the kth iteration are denoted as $g_{ik}^1(\vec{y}_i)$ and $g_{ik}^2(\vec{y}_i)$. Constraint (5) can be replaced by

$$\hat{g}_i^1(\vec{y}_i) = \frac{1}{K}\sum_{k=1}^{K} g_{ik}^1(\vec{y}_i) \geqslant 85\%; \quad \hat{g}_i^2(\vec{y}_i) = \frac{1}{K}\sum_{k=1}^{K} g_{ik}^2(\vec{y}_i) \geqslant 85\%, \quad \forall i \in \{1, 2, \ldots, I\} . \tag{6}$$

Until now a general optimisation model for staffing problems is established to minimise shift costs (2) with respect to coverage constraints (3) and service target constraints (6).

4 Refined square root staffing model

In this section, the proposed approach which combines a refined SRS and simulation will be described. The general idea of this approach is to distribute required workload in each period and generate an initial shift schedule. Indicated by simulation results, the shift schedule can be further improved towards the optimal.

Consider a $s(t)$-server queueing system $M_t/G/s(t)$ with Poisson arrival process (M_t) and general service time (G). Assume incident arrival rate $\lambda(t)$ is a function of time t and incident service time μ is a random variable with a probability density function $f(\mu)$. According to the work of [11], the offered load at time t is $m(t)$ which means at time t there will be $m(t)$ servers busy on average.

$$m(t) \approx \int_0^t \Big\{\lambda(t-\mu) \cdot F^c(\mu)\Big\} d\mu, \text{ where } F^c(\mu) = \int_\mu^\infty f(v)dv . \tag{7}$$

An intuitive explanation for the value of $m(t)$ is as follows. Incidents that arrive at $(t-\mu)$ have the probability $F^c(\mu)$ of remaining in service at time t. Integrating incidents by their arrival time, the value of $m(t)$ is obtained. It is obvious that $m(t)$ provides a lower bound for long time steady state staffing in service systems without any customers who abandon their service requirements while waiting in queues. Allocating $m(t)$ servers at time t is not enough to achieve service targets. Additional servers need to be added in to improve system performances. As a general rule-of-thumb, SRS augments offered load $m(t)$ by an amount of staff that is proportional to the square root of the offered load.

$$s(t) \geqslant \left\lceil m(t) + \beta\sqrt{m(t)} \right\rceil , \tag{8}$$

where β is related to the delay probability $P\{\text{delay}\} = P\{N(0,1) \geqslant \beta\}$.

As stated in [17], SRS algorithm and its variations are remarkably effective for staffing in infinite server service systems and finite server service systems with abandon. In PPSSs, delayed incidents will keep on waiting even though they have been waiting for a long time. Previous delay will largely effect current system performance. Pre-travel delays in responding to priority incidents is acceptable. As defined in Section 2, a priority incident can be delayed in response by up to one period of 45 minutes. Theoretical analysis for time-dependent multi-grade queue system is too complex to perform, but when aligning target response time of priority incidents to emergency incidents, this multi-grade queue system is reduced to a single-grade queue and the SRS algorithm can be applied in analysing PPSSs. The main idea of the refined SRS is adjusting required staff level to fit shift schedules by setting intended delay in priority incident responses.

Set $t_0 = 0$ as the starting time of simulation, period $i \in \{1, 2, \ldots, I\}$ starts at t_{i-1} and finishes at t_i. Workload $M(i)$ is defined in (9) to represent the integral of offered load $m(t)$ in period i.

$$M(i) = \int_{t_{i-1}}^{t_i} m(v)dv, \quad \text{with } m(v) \text{ defined in (7)} . \tag{9}$$

As defined in Section 2, priority incidents in PPSS that account for about 70% of all immediate attendance incidents can wait up to one period of 45 minutes. Set w_i as the delayed workload from period i to period $i+1$ in responding to priority incidents. Assume there are z_i busy servers on average in period i, the provided workload as shown in the right hand side of (10) should be no less than the required workload as shown in the left hand side of (10).

$$M(i) + w_{i-1} - w_i \leqslant z_i \cdot (t_i - t_{i-1}), \quad \text{with } M(i) \text{ defined in (9)} . \tag{10}$$

Since only priority responses can be delayed, the workload related to priority incidents in period i sets an upper limit of w_i as shown in (11)

$$w_i \leqslant 70\% \cdot M(i) . \tag{11}$$

Constraint (10) and (11) align the service target for emergency incidents and priority incidents in response time, but emergency incidents still require slightly higher successful response percentage than priority incidents. Set the emergency incident target 85% as the common percentage for both grades of incidents, the successful response target for priority incidents 80% will be met but required staff levels may be slightly over estimated. Simulation modifications will help correct these over estimations. Based on the SRS algorithm presented in (8), for the number of average busy server z_i and the required staff level y_i, constraint (12) holds in period i

$$y_i \geqslant \lceil z_i + 1.03\sqrt{z_i} \rceil , \quad \text{where } P\{\text{delay}\} = 15\% = P\{N(0,1) \geqslant 1.03\} . \tag{12}$$

Please note that (12) is not a linear constraint because of square root part of z_i, but when z_i takes value from interval $[z_i^{\text{lower}}, z_i^{\text{upper}}]$, it can be approximated via the following linear constraint.

$$y_i \geqslant \left\lceil f(z_i^{\text{lower}}) + \frac{f(z_i^{\text{upper}}) - f(z_i^{\text{lower}})}{z_i^{\text{upper}} - z_i^{\text{lower}}} (z_i - z_i^{\text{lower}}) \right\rceil , \quad \text{where } f(z) = z + 1.03\sqrt{z} . \tag{13}$$

One way to achieve a reasonable lower bound z_i^{lower} is as follows. Assume in period i there is no work passed over from the previous period. Only emergency incidents that cannot be delayed in response will be attended in this period. It is the best lower bound achievable in terms of workload. Similarly, z_i^{upper} can be set as an upper bound when all workload that

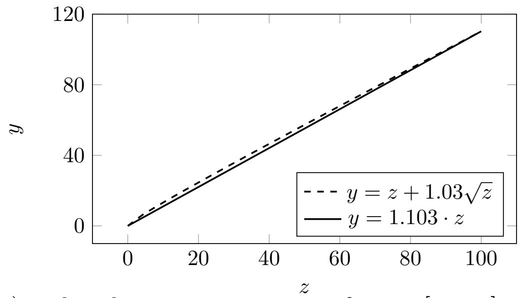

Figure 2 Function $f(z)$ and its linear approximation for $z \in [0, 100]$.

comes from priority responses initiated in the previous period is delayed to period i and there is no work carry over from period i to the next period.

Figure 2 plots function $f(z)$ in green and its linear approximation in blue for $z \in [0, 100]$. It indicates that constraint (13) provides a good linear approximation for constraint (12). The smaller the gap between z_i^{upper} and z_i^{lower}, the closer the approximation is to $f(z)$. Although y_i that satisfies (13) may still violate (12), a second-order cone programming [2] is simplified to a linear programming with these linear approximations. Instead of applying simulations to check the feasibility for service target constraint (6), based on the refined SRS algorithm, a combination of linear constraint (10), (11) and (13) are introduced to define feasible staff level y_i in period i. An integer programming model can be proposed to minimise the cost for shift schedules via objective function (2). Constraint (3) remains the same denoting the relationship between staff level y_i for period i and shift schedule x_j for shift j. This proposed linear programming model can be used to generate an initial shifts schedule which is close to the optimal shift schedule.

5 Simulation combined staffing modification

Simulation of queueing systems will then be implemented to indicate modifications in shift schedules. According to the response sequence in local police force, PPSSs are simulated by static priority queues in which emergency incidents can jump ahead of all priority incidents in the waiting queue but services for priority incidents will not be interrupted by new arrivals of emergency incidents. Simulator in this project is written in C++ using the standard procedure for discrete-event systems as suggested in [12], but it can be easily modelled by other simulation software as well.

In PPSSs, demand for service that comes from emergency incidents and priority incidents depends on an overall time-varying arrival rate $\lambda(t)$. Service process in queueing systems will start immediately when a patrol officer is available upon an incident arrival; otherwise the incident will join the queue. One should note that a service process in PPSSs actually starts at the time when an available patrol officer begins to travel to the assigned incident scene. That is to say the time one patrol officer spends on dealing with one incident also include the time patrol officers spend on travelling. Assume the time one patrol officer spends on one incident lasts long $\mu = \mu_1 + \mu_2$, where μ_1 denotes the time spend on travelling and μ_2 is the time to solve the incident at the scene. In a PPSS with time-varying staff level, the overall service rate $s(t)/\mu$ can be influenced by changing $s(t)$, the number of working patrol officers at time t. Figure 3 displays a simulation model for PPSSs.

Since patrol officers work in shift pattern, a shift schedule $\{x_1, x_2, \dots, x_J\}$ will decide the number of working patrol officers $s(t)$ at time t. The shift schedule obtained from refined

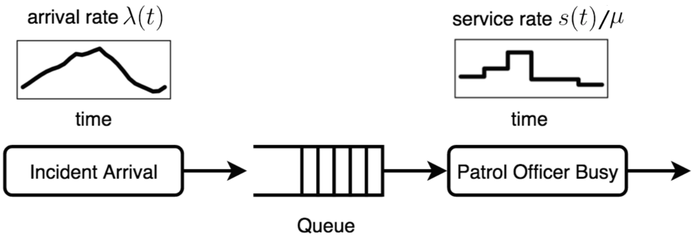

Figure 3 Police patrol service system simulation model.

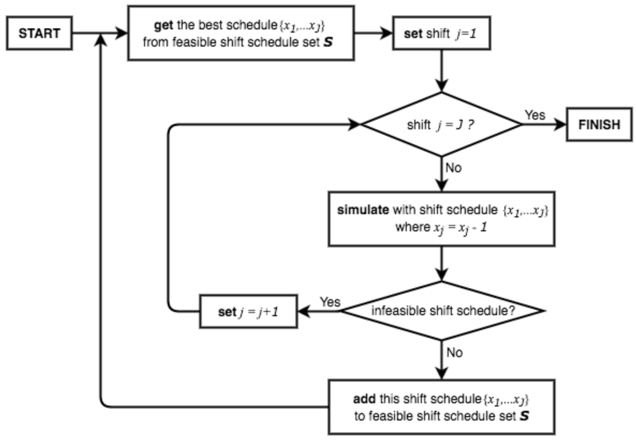

Figure 4 Simulation combined local search for staffing modification.

SRS model will be the first shift schedule operated in PPSS simulation model. For each period $i \in \{1, 2, \ldots, I\}$, both emergency incidents successful response percentage and priority incidents successful response percentage are compared with service targets. If the simulation system fails to meet service targets at period i, additional patrol officers will be arranged to work on related shifts; otherwise, less scheduled patrol officers may still be able to cope with the demand for service. The simulation combined staffing modification uses a local search algorithm as shown in Figure 4.

It starts at the initial shift schedule obtained from refined SRS model as stated in Section 4. Simulation is implemented to check whether the initial shift schedule can deliver qualified service. If service target is violated in period i, a new shift schedule will be generated to arrange one more patrol officer to work in period i. Local search modifications in staff level will continue until there is a feasible shift schedule that meets police patrol service targets at all time. Then local search are then applied to explore potential subtractions in the current feasible shift levels. Besides the easy implementation of local search, a linear objective function of shift levels is another main reason for applying local search. It is unlikely to get trapped in a local optimal shift schedule. Compared with other existing staffing algorithms, such as staffing via exhaustive simulation and iterative feed-back staffing, this local search integrated with refined SRS leads to a large computational saving.

6 Computational study

An example case for patrol staffing in Leicestershire police force is discussed in this section. According to the real patrol activity records in November 2010, input parameters for this simulation model are estimated as listed in Table 1 and Figure 5.

Table 1 Input parameters for police patrol service system simulation.

Urgency Grade	Arrival Rate *Poisson Process*	Service Time *Exponential Distribution*	Travel Time *Normal Distribution*
1. Emergency	$30\%\lambda(t)$/hour	1 hour *on average*	5 minutes *on average*
2. Priority	$70\%\lambda(t)$/hour		

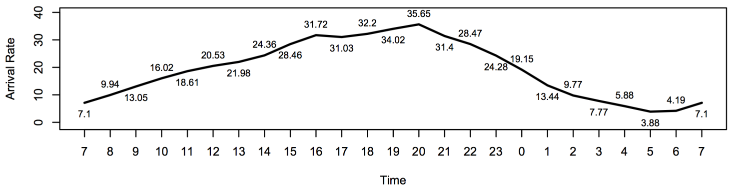

Figure 5 Incident arrival rate $\lambda(t)$ for police patrol service system simulation.

A working day in PPSSs starts at 7am and finishes at 7am the following day. Although PPSSs run continuously from one day to the next, the demand for service around 7am in the morning is very low and the system around that time is almost empty. For this reason, each working day can be viewed as a separated replication. Service levels for emergency incidents and priority incidents in all periods are estimated by performing multiple (1,000) independent replications of simulation. Every working day is divided into thirty-two 45-minute periods and service targets for emergency responses and priority responses will be estimated in all periods. Four different shifts can be performed by patrol officers. They are ‘Early’ shift from 07:00 to 15:00; ‘Day’ shift from 15:00 to 23:00; ‘Late’ shift from 19:00 to 03:00 and ‘Night’ shift from 23:00 to 07:00. The cost to perform an ‘Early’ shift, a ‘Day’ shift, a ‘Late’ shift or a ‘Night’ shift is set to £112, £112, £144 and £160, respectively. It considers related factors including hourly pay to patrol officer, vehicle fuel cost and shift type popularity.

In PPSSs, the time a patrol officer spent on an incident μ consists of the time for travelling μ_1 and the time spent on solving incidents μ_2. These two parts of service time come from two different distributions, normal distribution and exponential distribution. It is not easy to get an analytical form of the probability density function $f(\mu)$, but it can be estimated through random variables. Generate a number (10,000) of random variables for travel time μ_1 and generate the same number of random variables for service time μ_2. The distribution of μ can be estimated by random variable pairs of μ_1 and μ_2. Figure 6 plots the approximated function of $m(t)$ for a working day as defined in (7).

The initial shift schedule is generated by optimising the refined SRS model. Simulation combined with local search modifies the initial shift schedule towards the optimal. The initial and the optimal shift schedules for this example are summarised in Table 2.

In this example, the initial shift schedule generated by the refined SRS model is close to the optimal shift schedule, but it slightly over estimates the required staff level during rush hours around 19:00 to 23:00. The over estimation in staff level is mainly because PPSS does not strictly require immediate response without any pre-travel delay to achieve the service targets. With the help of the refined SRS model, the response target for all incidents is aligned so that more than 85% of incidents should be responded within 15 minutes. In this example case, travel times are assumed to follow a normal distribution with mean of 5 minutes, so the time spent on travelling for 85% incidents is less than 6 minutes. A pre-travel delay of up to 9 minutes is permitted. This over estimation can be diminished by tuning the value of safety parameter β in (8), but it is hard to be avoided analytically.

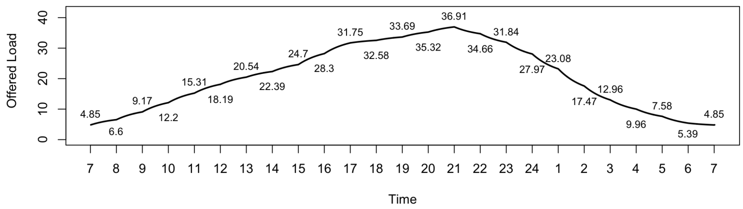

Figure 6 Offered load $m(t)$ for police patrol service system simulation.

Table 2 Initial shift schedule generated by refined square root staffing model.

	Early shift	**Day shift**	**Late shift**	**Night shift**
Initial Shift Schedule	19 officers	29 officers	11 officers	8 officers
Optimal Shift Schedule	19 officers	27 officers	9 officers	8 officers

7 Conclusion and future work

In this paper, a new algorithm is proposed to determine appropriate staffing levels in PPSSs. It starts at a refined SRS algorithm to indicate an initial shift schedule and then uses simulation combined with local search to modify the shift schedule towards the optimal. The proposed approach is believed to be more efficient comparing with other existing staffing algorithms. It is a hybrid method of two-step staffing and feed-back staffing. A good initial shift schedule generated will significantly reduce the computation time of simulation. It also skips the estimation of staff levels by modifying shift levels directly which may further reduce the computation time required to explore the optimal shift schedule.

The proposed staffing algorithm can also be revised to solve other staffing problems. The planning horizon can be extended from a single-day period to a multi-day period when some work initiated at the end of a day would be taken over by the staff scheduled to work at next day. The longer the planning horizon, the more the decision variables and the longer it will take to find the optimal solution via simulation only. Start the simulation with an informed initial shift schedule, as suggested in our approach, will largely reduce the computation time by skipping simulations of some unnecessary scenarios. The current model aligns the priority incident service target to the emergency incident service target and assumes that there is no pre-travel delay in responding to emergency incidents. However, the proposed staffing model is very sensitive to the time patrol officers spend on travelling, a potential improvement in the proposed staffing algorithm could be made by introducing reasonable pre-travel delay in emergency incidents as well.

References

1. Bernardetta Addis, Roberto Aringhieri, Giuliana Carello, Andrea Grosso, and Francesco Maffioli. Workforce management based on forecasted demand. In *Advanced Decision Making Methods Applied to Health Care*, pages 1–11. Springer, 2012.
2. Farid Alizadeh and Donald Goldfarb. Second-order cone programming. *Mathematical programming*, 95(1):3–51, 2003.
3. Júlíus Atlason, Marina A Epelman, and Shane G Henderson. Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science*, 54(2):295–309, 2008.
4. Suffolk Constabulary. Safety advice: police response policy. `http://www.suffolk.police.uk/safetyadvice/reportacrime/policeresponsepolicy.aspx`. Accessed: 2016-05-02.

5 Mieke Defraeye and Inneke Van Nieuwenhuyse. Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, 58:4–25, 2016.

6 Yixin Diao and Aliza Heching. Staffing optimization in complex service delivery systems. In *Network and Service Management (CNSM), 2011 7th International Conference on*, pages 1–9. IEEE, 2011.

7 Noah Gans, Haipeng Shen, Yong-Pin Zhou, K Korolev, A McCord, and H Ristock. Parametric stochastic programming models for call-center workforce scheduling. *Technical report*, 2009.

8 Linda V Green, Peter J Kolesar, and João Soares. Improving the sipp approach for staffing service systems that have cyclic demands. *Operations Research*, 49(4):549–564, 2001.

9 Shane G Henderson and Andrew J Mason. Rostering by iterating integer programming and simulation. In *Proceedings of the 30th conference on Winter simulation*, pages 677–684. IEEE Computer Society Press, 1998.

10 Armann Ingolfsson, Md Amanul Haque, and Alex Umnikov. Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research*, 139(3):585–597, 2002.

11 Otis B Jennings, Avishai Mandelbaum, William A Massey, and Ward Whitt. Server staffing to meet time-varying demand. *Management Science*, 42(10):1383–1394, 1996.

12 W David Kelton and Averill M Law. *Simulation modeling and analysis.* McGraw Hill Boston, 2000.

13 The Secretary of State for Transport. The road vehicles (construction and use) regulations. `http://www.legislation.gov.uk/uksi/1986/1078/made`. Accessed: 2016-05-02.

14 The Secretary of State for Transport. The road vehicles lighting regulations. `http://www.legislation.gov.uk/uksi/1989/1796/made`. Accessed: 2016-05-02.

15 Leicestershire Police. Call taker: responding to your calls 24 hours a day. `https://leics.police.uk/join-us/police-staff/call-taker`. Accessed: 2016-05-02.

16 West Midlands Police. Contact us: responding to your calls. `https://www.west-midlands.police.uk/contact-us/response-times/index.aspx`. Accessed: 2016-05-02.

17 Ward Whitt. What you should know about queueing models to set staffing requirements in service systems. *Naval Research Logistics (NRL)*, 54(5):476–484, 2007.