A Workforce Allocation and Scheduling Model for Sequential Service Systems with Disjoint Service Areas and Resource Sharing

Randy Grivel[+], Ronald G. Askin[+], Jorge A. Sefair[*]

[+] School of Computing and Augmented Intelligence, Arizona State University, {rgrivel@asu.edu, ron.askin@asu.edu}

[*] Department of Industrial and Systems Engineering, University of Florida, jorge.sefair@ise.ufl.edu

This paper is scheduled to appear in a special issue of the Annals of Operations Research in honor of Professor John Mulvey of Princeton University for his many and lasting contributions to the field of operations research and especially to optimization methods and modeling applications to business and government settings.

A Workforce Allocation and Scheduling Model for Sequential Service Systems with Disjoint Service Areas and Resource Sharing

Abstract: Service and manufacturing systems typically involve processing entities (hereafter referred to as customers) through one or more human-operated service stations, where the arrival process of customers exhibits stochastic and nonstationary but estimable behavior. Achieving operational efficiency requires configuring service capacity to accommodate arrivals in a way that minimizes both service and customer waiting time costs. In this paper, we investigate the optimal configuration of service channels for such a system, along with the prescription of server work schedules and allocation to service stations. We propose a tractable input-output model to approximate queue dynamics, designed to be integrated into an optimization framework for configuration and workforce planning without introducing additional decision variables. The problem is motivated by the operation of Security Screening Checkpoints (SSCPs), which perform passenger and baggage screening at multiple locations in major U.S. airports.

1 Introduction

Optimization models have played a critical role in workforce management across a variety of service systems. The foundational work by Dantzig (1954) introduced a generalized set cover model for workforce allocation. Later, Bartholdi (1981) argued that cyclic workforce scheduling problems are efficiently solvable when resources are continuously available across consecutive periods, but become NP-complete when resource availability is intermittent. Since Dantzig's pioneering work, workforce allocation models have expanded to include additional features, enhancing their fidelity to real-world systems while also increasing the complexity of the problems being modeled (Van den Bergh et al., 2013).

This paper presents a deterministic workforce allocation and scheduling model where jobs must visit a sequence of servers across a disjoint set of service areas that share a resource pool. Base work schedules are fixed for each worker, but break times and overtime decisions are determined by the model in accordance with labor rules to minimize both overtime costs and customer wait times. When moving resources between service areas, the model accounts for non-negligible transfer times.

Although our models are broadly applicable to many multi-location service systems with sequential servers, we center our analysis on the operation of Security Screening Checkpoints (SSCPs). Airport security operations present a complex logistical challenge, requiring the integration of advanced screening technologies with proactive queue and staffing management. These systems must balance the dual objectives of delivering a smooth passenger experience while maintaining the highest levels of security to prevent unauthorized items from being brought onboard. Today, pre-boarding screening of passengers and baggage is a standard procedure at nearly every airport worldwide, requiring multiple stages of interaction with security personnel. In 2023, the U.S. Transportation Security Administration (TSA) screened over 858 million passengers, 1.9 billion carry-on items, and 484 million checked bags (TSA, 2023). This was accomplished despite workforce attrition that occurred during the pandemic. With a mandate to "protect the nation's transportation systems to ensure freedom of movement for people and commerce," as defined in the Aviation and Transportation Security Act of 2001, the TSA faced significant technology and workforce challenges to maintain timely yet thorough security screening.

Relative to previous research, this paper provides the following extensions and contributions:

- Overtime and break schedules are determined for each worker, constrained by work rules and worker attributes such as gender and certifications.
- Worker movement between service locations is optimized, accounting for idle transit times.
- Productivity and resource requirements resulting from service configuration changes are incorporated into decision-making.
- Staffing needs are determined endogenously, based on the operational configuration selected for each service location and time. If desired, hiring options implying calling-in off-duty workers can be included in the model by use of the additional resources variables included in the formulation.
- Stochastic queuing behavior is approximated with a computationally tractable approach and integrated into the optimization scheme.
- Quality of service constraints on maximum time in the system are enforced in the model.
- A high-fidelity optimization model is shown to prescribe optimal worker schedules and assignments while balancing customer service performance and overtime costs.

2 Literature Review

In a review of staffing and scheduling methods for service systems with non-stationary demand, Defraeye and Van Nieuwenhuyse (2016) identified three common approaches to determining optimal staffing decisions: (1) determining staffing requirements beforehand and then optimizing staffing and scheduling decisions, (2) using an iterative process where system performance is evaluated with fixed staffing and then updated to improve performance, and (3) embedding both scheduling decisions and staffing requirements into a single model. The review noted that most papers employed the first method, following a similar formulation to Dantzig (1954). For example, Eitzen et al. (2004) determined staffing decisions for a multi-skilled workforce for known workforce requirements. Cordeau et al. (2010) presented a method to schedule telecommunication technicians based on task difficulty and skills. A review of physician scheduling by Erhard et al. (2018) found that most papers ignore how the order of completed jobs impacted future staffing needs. An exception is Brunner et al. (2010), who addressed physician staffing by creating schedules for known staffing requirements, incorporating features like flexible shift start and end times and overtime.

Papers using the second method typically iterate between a variation of a set cover model and a stochastic evaluation method, such as a simulation, until a termination criterion is met (Defraeye and Van Nieuwenhuyse, 2016). An example of this approach was presented by Kolesar et al. (1975), who used a linear program alongside a simulation to determine police patrol car routing. This work has since been expanded, and most relevant to this paper, Feldman et al. (2008) presents methods for staffing multiserver queuing systems.

The third method generally results in more realistic and intuitive models but also makes the problems more difficult to solve (Defraeye and Van Nieuwenhuyse, 2016). Some approaches use optimization techniques that embed a simulation-based algorithm into the solution process, as demonstrated by Avramidis et al. (2010). Their work addresses issues such as employee skill levels, breaks, and shift allocation, while also showing that in many cases, an iterative method can yield locally optimal solutions.

The staffing problems in the reviewed literature have incorporated factors such as worker skills, shift start and end times, and the scheduling of breaks and overtime. We also account for worker attributes and their routing across various locations. In a similar context, Begur et al. (1997) incorporated nurses' skill levels and assigned them to patients' homes based on specific care needs. Al-Yakoob and Sherali (2007) presented a mixed-integer program to schedule employees across geographically dispersed locations over a multi-day horizon. Mulvey (1979) proposed a workforce allocation model that assigns faculty members to academic quarters and courses while incorporating their individual preferences. A survey by Castillo-Salazar et al. (2014) found that while some studies consider employee skills in efforts to minimize travel time, workforce routing research often neglects job backlogs and queue dynamics.

Related to workforce allocation problems is the tour scheduling problem (TSP), in which long-term schedules are created by assigning workers to shifts over an extended planning horizon. TSPs can incorporate break assignments (Bechtold and Jacobs, 1990) and employee skills (Emmons and Burns, 1991; Hung, 1994; Narasimhan, 1997; Rong, 2010). However, the long-term nature of TSPs can reduce the fidelity of day-to-day operational goals. In our framework, long-term shift schedules are predetermined, while the focus is on optimizing daily decisions regarding breaks, overtime, and location assignments.

There is a diverse body of literature on staffing queuing systems with time-varying demand. Jennings et al. (1996) present a method for determining staffing requirements in systems with fluctuating job demand. Green et al. (2007) conducted an analysis of staffing methods for call centers, using approximation methods, heuristic rules, and iterative approaches. He et al. (2016) developed an algorithm for staffing decisions in service systems with non-Poisson, non-stationary arrivals. Heemskerk et al. (2022) employed an infinite-server approach combined with a square-root staffing rule for many-server systems. In general, these studies do not account for recourse decisions when available staffing or system constraints prevent the recommended configurations from being implemented. While queuing works such as Wallace and Whitt (2005) and Izady and Worthington (2012) consider job routing within the queuing system, they focus on routing jobs to servers rather than routing workers to service areas. Queue performance metrics are often used to determine staffing demand, but related works typically do not address worker scheduling or routing decisions, nor do they explicitly consider scenarios where the arrival rate temporarily exceeds the processing rate due to limited resources. Our problem requires modeling queues where arrival rates can temporarily exceed processing capacity without customer loss, an issue rarely addressed in analytical queuing models.

The model presented in this paper addresses the workforce routing and allocation problem for a series of sequential servers with disjoint service areas. This problem is motivated by the U.S. airport security screening process. Airport check-in counters face a similar challenge, as demand is realized some time before a flight's departure, requiring staffing decisions to meet this demand. Stolletz (2010) introduced a binary programming model that generates employee schedules for airport check-in counters, balancing demand with employee preferences. Building on prior work, Stolletz and Zamorano (2014) and Zamorano et al. (2017) developed models for check-in operations that account for worker qualifications and task assignments, where each task depends on the completion of the preceding one.

In the field of airport security screening, Martonosi (2011) developed a method for determining policies to open and close service stations in order to manage queues throughout the day. Hanumantha et al. (2020) proposed a model to optimize staffing and configuration decisions for the TSA security screening process. Building on this work, Jiao et al. (2024) introduced a time-expanded network to coordinate labor decisions across multiple service areas and used a data-driven piece-wise linear clearing function to estimate the number of passengers processed. The formulation presented in this paper extends the work of Jiao et al. (2024) by incorporating and analyzing the effects of breaks, overtime, and labor classification decisions.

Very few papers in the literature use deterministic models to solve service system problems where jobs visit disjoint service areas with a shared pool of resources. Two closely related works are those by Wu et al. (2023) and Jiao et al. (2024). Wu et al. (2023) presented a staffing and scheduling model for a system with disjoint service areas, where jobs visit only a single service node. Jiao et al. (2024) studied a similar system but focused on a two-stage service process. However, neither of these papers addresses the scheduling of breaks, overtime, or the required qualifications (i.e., skills) of the available staff.

The process studied in this paper consists of jobs that must visit a sequence of servers before exiting. The stochastic nature of arrival and service times is captured by a deterministic function that predicts the number of jobs processed per period. Resources required to staff the servers can move between service locations within a known transit time. A fixed set of workers is available for each shift, with predetermined start and stop times. Work break requirements are determined by the model based on available resources and system-wide needs. The model also includes options for allowable overtime and prescribes their use. Worker-specific certifications are required for certain workstations. Additionally, workstations may have limited buffer capacities, require varying levels of human resources, follow ramp-up and ramp-down rules, and operate under multiple configurations that affect service rates and maximum service times. To our knowledge there is no model that integrates all of these features into a single framework.

The remainder of this work is organized as follows. Section 3 outlines the key features of the service system under study and presents the corresponding mathematical programming model. In Section 4, we apply the model to a case study on security screening operations using data from Miami International Airport. Section 5.1 details a series of experiments that assess various model features across different problem sizes.

3 System Assumptions and Mathematical Model

3.1 Service System

The service system consists of disjoint areas where "jobs", in the form of an entity requiring service, enter the system. After entering, a job receives service at a series of sequential service nodes before exiting. We refer to each service received by a job at any node as a "service step." Each area is considered job-disjoint, meaning that once a job begins service in a given area, it must complete all required services within that area before exiting. Figure 1 illustrates a system with three service areas, two of which have three service nodes, while the remaining area has two service nodes.

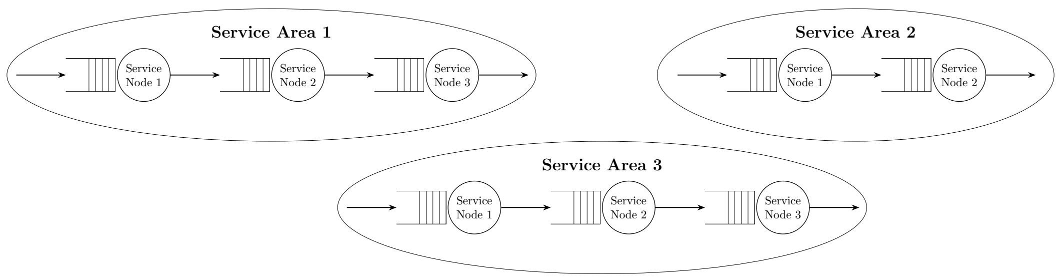


Figure 1: A system with three service areas.

For simplicity, each service step is represented by a single service node, which can operate under multiple configurations. Each configuration has a unique processing rate and resource requirements. Queues may form at any service node, with finite queue buffers potentially present between sequential nodes. This design allows for the inclusion of various service node processing regimes within the same framework. A service area can contain any number of service nodes, and each node can have different operating configurations. Configurations can represent different levels of parallel servers or operating policies.

Our goal is to determine the optimal service configurations at each node over time, along with the break and overtime schedules for each worker to support these configurations. The objective is to minimize the total customer wait time and overtime costs, given fixed base schedules. The proposed optimization model computes wait time as a function of the selected configuration, using an embedded clearing function (Jiao et al., 2024) . Additionally, in the motivating environment, maximum "job" (i.e., customer) wait times are mandated when feasible.

3.2 Resource dynamics

Resources must be allocated to different system areas to meet the resource requirements of each service node for a given configuration. A configuration refers to the number of service channels open and their operating mode, which determines the resource needs at each service area during a given period. The time horizon is divided into equal-length periods, with the arrival rates, service rates, and configuration remaining constant within each period.

We assume that each resource is only available for a specific portion of the planning horizon, such as a worker on a shift. The time a resource is available is referred to as its shift, and the number of resources available for each shift is known in advance. Some resources require breaks during the day, and for the purposes of this paper, we assume that multiple options for break schedules, which comply with legal or operational regulations, are known beforehand. Additionally, some resources can work beyond their regular shifts, referred to as overtime, which incurs additional costs to the system. Similar to break schedules, all options for overtime schedules are assumed to be known ahead of time. The model then selects the optimal overtime and break plans for each shift.

Resource sharing is incorporated into the model by assuming that a resource located at a given service area and node requires an integer number of periods to travel to another service area and node. Depending on managerial preferences and the real-world dynamics of the system, it is possible to pool resources, such as allowing travel within a service area to occur within the same period, or restricting movement between areas that are deemed too far apart. The spatiotemporal resolution of the model can be adjusted for finer granularity (e.g., finer time discretization or service nodes representing specific workstations). However, this increased detail may lead to greater computational complexity in the model.

Figure 2 illustrates a time-expanded network representation of resource movement options for two shifts with known break times. This network extends the model presented in Jiao et al. (2024) by incorporating additional operational aspects, such as breaks, overtime, and worker classifications. The network is organized into horizontal layers representing service areas and vertical layers representing time periods. In this example, the system includes three service areas (horizontal layers for each shift) and eight periods (vertical layers across shifts), with predefined travel times between areas. The source and sink nodes denote the start and end of the shift, respectively. During break periods, resources are routed to break nodes, preventing their use for a specified number of periods. The remaining nodes represent a service area at a given time period. Horizontal arcs indicate that the resource remains in the same location from one period to the next (as indicated by the node label), while arcs connecting nodes across different horizontal layers represent movement between service areas. These arcs account for travel time, which may extend beyond a single period. The resource exits the network via an arc connecting to the sink node at the end of the shift.

The time-expanded network allows for the inclusion of overtime hours, which may be used before and after a regular shift. Additional nodes are appended to the periods immediately preceding or following an existing shift. Figure 2 provides an example of optional overtime schedules that begin after the shift. The first option consists of two additional periods, while the second allows for three additional periods. The nodes labeled as "Sink/Source" represent the end of the regular shift and the beginning of the the overtime period, if used. The sink nodes represent the end of the overtime period. Each service area node requires a specific number of workers based on the operational configuration selected for the period.

Additional features in the time-expanded network capture various operational aspects. Each shift may include multiple break options, which requires duplicating the horizontal layers for each shift and break

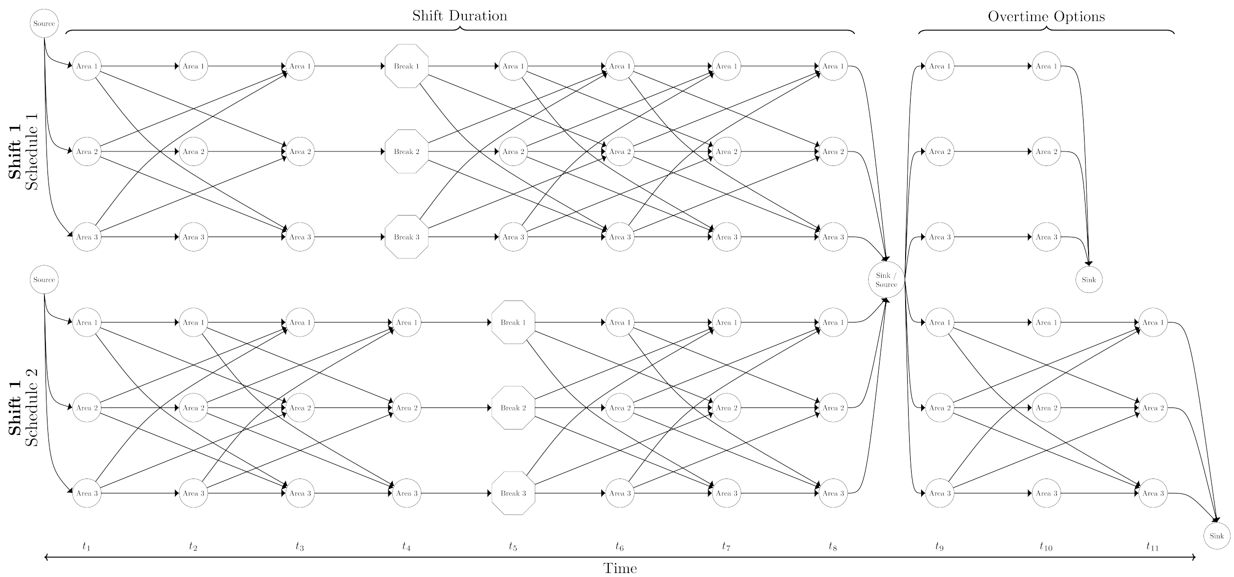


Figure 2: Break and overtime schedules.

option. These layers can be further combined with overtime choices. From a network-flow perspective, the flow through the network represents workers assigned to the same shift. This flow, corresponding to the number of workers on a shift, enters the network at the source node and exits at the sink node for that shift (e.g., "Sink/Source" node in Figure 2). If needed, a portion of the regular shift flow can extend to overtime nodes. Since workers are categorized by type (i.e., skills) and shift, the network flow dynamics are formulated as a multi-commodity flow problem.

3.3 Mathematical Model

This section presents the proposed mathematical programming model. Tables 1, 2, and 3 define the notation for sets, parameters, and decision variables, respectively. The model is formulated as a mixed-integer program. Not all constraints and variables are required in every instance; they can be omitted depending on the desired model options, decisions, and specific information sought.

The set I represents all service areas, and for each $i \in I$, the set J_i contains all corresponding service nodes. The set T contains all periods within the planning horizon. Additionally, it is essential to track the available configurations for each service node. The collection of possible configurations for service area i and service node j is represented by the set $\mathcal{K}_{ij}$.

Set	Description
I	Service areas defined as $\{1, 2, \ldots, \|I\|\}$
J_i	Service nodes for area $i \in I$ defined as $\{1, 2, \ldots, \|J_i\|\}$
$\mathcal{K}_{ij}$	Configurations available for service area $i \in I$ and node $j \in J_i$
T	Planning periods
Ω	Shifts (Consisting of both regular and overtime shifts)
D	Resource distinction (e.g., features)
Γ	Resource classifications defined as a collection of one of more distinctions
Γ_d	Set of classifications that include distinction $d \in D$, $\Gamma_d \subseteq \Gamma$
O_ω	Schedule options (i.e., breaks and overtime options) available within shift $\omega \in \Omega$
$N_{\omega o \gamma}$	Nodes in the time expanded network available for resources on shift $\omega \in \Omega$, schedule $o \in O_\omega$, and classification $\gamma \in \Gamma$
$N^1_{\omega o \gamma}$	Nodes in $N_{\omega o \gamma}$ that are neither source nor sink nodes
$N^2_{\omega o \gamma}$	Source nodes in $N_{\omega o \gamma}$
$N^3_{\omega o \gamma}$	Sink nodes in $N_{\omega o \gamma}$
$E_{\omega o \gamma}$	Arcs in the time expanded network available for resources on shift $\omega \in \Omega$, schedule $o \in O_\omega$, and classification $\gamma \in \Gamma$

Table 1: Sets

Parameter	Description
r_{ijk}	Nominal processing rate of service node $j \in J_i$ of area $i \in I$ under configuration $k \in \mathcal{K}_{ij}$
$\tilde{r}_{ijk_1k_2}$	Processing rate adjustment when service node $j \in J_i$ in area $i \in I$ changes configuration from k_1 to k_2, $k_1,\ k_2 \in \mathcal{K}_{ij}$
λ_{it}	Job arrivals to the first service node in area $i \in I$ at time $t \in T$
$\beta^\ell_1, \beta^\ell_2, \beta^\ell_3$	Estimated coefficients for linear function ℓ to predict the number of jobs processed in a period
τ	Maximum number of periods a job is allowed to wait in any queue
Π_{ijk}	Maximum queue size at the end of a period when node $j \in J_i$ in area $i \in I$ operates under configuration $k \in \mathcal{K}_{ij}$
M	Minimum number of periods a configuration must remain selected
$h_{\omega\gamma}$	Number of resources of classification $\gamma \in \Gamma$ available on shift $\omega \in \Omega$
$H_{\omega\gamma}$	Maximum number of additional resources of classification $\gamma \in \Gamma$ for shift $\omega \in \Omega$
L_{ijkd}	Demand of resources with distinction $d \in D$ at node $j \in J_i$ of area $i \in I$ to operate configuration $k \in \mathcal{K}_{ij}$
c_{ij}	Cost of holding a customer in queue at node $j \in J_i$ of area $i \in I$ for one period
$\tilde{c}_{\omega o \gamma}$	Cost of hiring resources of classification $\gamma \in \Gamma$ to work schedule $o \in O_\omega$ of shift $\omega \in \Omega$
$\hat{c}_{ijk}$	Operational cost for using configuration $k \in \mathcal{K}_{ij}$ in node $j \in J_i$ of area $i \in I$ for one period

Table 2: Parameters

Decision Variable	Indices	Description
$x_{ijkt} \in \mathbb{B}$	$\forall i \in I,\ j \in J_i,\ k \in \mathcal{K}_{ij},\ t \in T$	Equal to 1 if configuration k is chosen to operate node j in area i at time t, 0 otherwise
$\mu_{ijt} \in \mathbb{R}^+$	$\forall i \in I,\ j \in J_i,\ t \in T$	Effective processing rate, i.e., customers processed per unit of time (e.g., per 15-minutes), of node j in area i at time t
$y_{ijk_1k_2t} \in \mathbb{R}^+$	$\forall i \in I,\ j \in J_i,\ k_1, k_2 \in \mathcal{K}_{ij},\ t \in \{2, 3, \ldots, \|T\|\}$	Equal to 1 if configuration k_1 is selected at time $t-1$ and configuration k_2 is selected at time t for node j in area i, 0 otherwise
$Q_{ijt} \in \mathbb{R}^+$	$\forall i \in I,\ j \in J_i,\ t \in T$	The estimated queue at the end of period t at node j of area i, with $Q_{ij0} = 0$ by assumption
$p_{ijt} \in \mathbb{R}^+$	$\forall i \in I,\ j \in J_i,\ t \in T$	Estimated number of jobs processed during period t at node j of area i
$\tilde{Q}_{ijtt'} \in \mathbb{R}^+$	$\forall i \in I,\ j \in J_i,\ t' \in T,\ t \in \{\max\{t' - \tau, 1\}, \ldots, t' - 1\}\}$	The number of customers in queue at time t at area i node j that are scheduled to be processed at time t'
$\delta^r_{ijkt} \in \mathbb{B}$	$\forall i \in I,\ j \in J_i,\ k \in \mathcal{K}_{ij},\ t \in T,\ r \in \{u, v\}$	Binary variables used to enforce that configurations can only change after a predetermined number of periods
$\eta_{\omega o \gamma} \in \mathbb{Z}^+$	$\forall \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma$	Number of resources of classification γ in shift ω assigned to schedule o
$\tilde{\eta}_{\omega o \gamma} \in \mathbb{Z}^+$	$\forall \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma$	Number of additional resources of classification γ allocated to schedule o of shift ω
$f_{nv,\omega,o,\gamma} \in \mathbb{Z}^+$	$\forall \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma,\ nv \in E_{\omega o \gamma}$	Flow of resources of classification γ on schedule o of shift ω along edge (n, v)

Table 3: Decision Variables

The proposed model is presented in (1)-(25), where the variable-type constraints are omitted for simplicity.

$$\min \sum_{i\in I}\sum_{j\in J_i}\sum_{t\in T} c_{ij}Q_{ijt} + \sum_{\omega\in\Omega}\sum_{o\in O_\omega}\sum_{\gamma\in\Gamma}\tilde{c}_{\omega o\gamma}\tilde{\eta}_{\omega o\gamma} + \sum_{i\in I}\sum_{j\in J_i}\sum_{k\in\mathcal{K}_{ij}}\sum_{t\in T}\hat{c}_{ijk}x_{ijkt} \tag{1}$$

s.t.

$$\sum_{k\in\mathcal{K}_{ij}} x_{ijkt} = 1, \quad \forall i\in I,\ j\in J_i,\ t\in T, \tag{2}$$

$$\mu_{ij1} = \sum_{k\in\mathcal{K}_{ij}} r_{ijk}x_{ijk1} - \sum_{k_1\in\mathcal{K}_{ij}}\sum_{\substack{\{k_2\in\mathcal{K}_{ij}:\\ r_{ijk_1}>r_{ijk_2}\}}} \tilde{r}_{ijk_1k_2}y_{ijk_1k_2 2}, \quad \forall i\in I,\ j\in J_i \tag{3}$$

$$\begin{aligned}\mu_{ijt} = &\sum_{k\in\mathcal{K}_{ij}} r_{ijk}x_{ijkt} - \sum_{k_1\in\mathcal{K}_{ij}}\sum_{\substack{\{k_2\in\mathcal{K}_{ij}:\\ r_{ijk_1}<r_{ijk_2}\}}} \tilde{r}_{ijk_1k_2}y_{ijk_1k_2t}\\ &-\sum_{k_1\in\mathcal{K}_{ij}}\sum_{\substack{\{k_2\in\mathcal{K}_{ij}:\\ r_{ijk_1}>r_{ijk_2}\}}} \tilde{r}_{ijk_1k_2}y_{i,j,k_1,k_2,t+1}, \quad \forall i\in I,\ j\in J_i,\ t\in\{2,3,\ldots,|T|-1\}\end{aligned} \tag{4}$$

$$\mu_{ij|T|} = \sum_{k\in\mathcal{K}_{ij}} r_{ijk}x_{ijk|T|} - \sum_{k_1\in\mathcal{K}_{ij}}\sum_{\substack{\{k_2\in\mathcal{K}_{ij}:\\ r_{ijk_1}<r_{ijk_2}\}}} \tilde{r}_{ijk_1k_2}y_{ijk_1k_2|T|} \quad \forall i\in I,\ j\in J_i \tag{5}$$

$$x_{i,j,k_1,t-1} + x_{ijk_2t} - 1 \le y_{ijk_1k_2t}, \quad \forall i\in I,\ j\in J_i,\ k_1,\ k_2\in\mathcal{K}_{ij},\ t\in\{2,3,\ldots,|T|\} \tag{6}$$

$$x_{i,j,k_1,t-1} \ge y_{ijk_1k_2t}, \quad \forall i\in I,\ j\in J_i,\ k_1,\ k_2\in\mathcal{K}_{ij},\ t\in\{2,3,\ldots,|T|\} \tag{7}$$

$$x_{ijk_2t} \ge y_{ijk_1k_2t}, \quad \forall i\in I,\ j\in J_i,\ k_1,\ k_2\in\mathcal{K}_{ij},\ t\in\{2,3,\ldots,|T|\} \tag{8}$$

$$\beta_1^\ell Q_{i,1,t-1} + \beta_2^\ell\mu_{i1t} + \beta_3^\ell\lambda_{1t} \ge p_{i1t}, \quad \forall i\in I,\ t\in T,\ \ell=1,2,3 \tag{9}$$

$$\beta_1^\ell Q_{i,j,t-1} + \beta_2^\ell\mu_{ijt} + \beta_3^\ell p_{i-1,j,t} \ge p_{ijt}, \quad \forall i\in I,\ j\in\{2,3,\ldots,|J_i|\},\ t\in T,\ \ell=1,2,3 \tag{10}$$

$$Q_{i,1,t-1} + \lambda_{it} - p_{i1t} = Q_{i1t}, \quad \forall i\in I,\ t\in T \tag{11}$$

$$Q_{i,j,t-1} + p_{i,j-1,t} - p_{ijt} = Q_{ijt}, \quad \forall i\in I,\ j\in\{2,3,\ldots,|J_i|\},\ t\in T \tag{12}$$

$$Q_{ijt} = \sum_{t'=t+1}^{\min\{|T|,t+\tau\}} \tilde{Q}_{ijtt'}, \quad \forall i\in I,\ j\in J_i,\ t\in\{1,\ldots,|T|-1\} \tag{13}$$

$$\tilde{Q}_{ijtt'} \le p_{ijt'}, \quad \forall i\in I,\ j\in J_i,\ t'\in T,\ t\in\max\{t'-\tau,1\},\ldots,t'-1 \tag{14}$$

$$Q_{ijt} \le \sum_{k\in\mathcal{K}_{ij}} \Pi_{ijk}x_{ijkt}, \quad \forall i\in I,\ j\in J_i,\ t\in T \tag{15}$$

$$x_{ijkt} - x_{i,j,k,t+1} + \delta^{\mathrm{u}}_{i,j,k,t+1} - \delta^{\mathrm{v}}_{i,j,k,t+1} = 0, \quad \forall i\in I,\ j\in J_i,\ k\in\mathcal{K}_{ij},\ t\in\{1,2,\ldots,|T|-1\} \tag{16}$$

$$\sum_{k\in\mathcal{K}_{ij}} \delta^r_{ijkt} \le 1 - \sum_{k\in\mathcal{K}_{ij}} \delta^s_{i,j,k,t+m}, \quad \forall i \in I,\ j \in J_i,\ t \in \{1,2,\dots,|T|-M\}, \qquad m \in \{1,2,\dots,M\},\ r,s \in \{\text{u, v}\} \tag{17}$$

$$\delta^r_{ijkt} = 0, \quad \forall i \in I,\ j \in J_i,\ k \in \mathcal{K}_{ij},\ t \in \{|T|-M+1,\dots,|T|\},\ r \in \{\text{u,v}\} \tag{18}$$

$$\delta^r_{ijkt} = 0, \quad \forall i \in I,\ j \in J_i,\ k \in \mathcal{K}_{ij},\ t \in \{2,\dots,M\},\ r \in \{\text{u, v}\} \tag{19}$$

$$\sum_{o\in O_\omega} \eta_{\omega o\gamma} \le h_{\omega\gamma}, \quad \forall \omega \in \Omega,\ \gamma \in \Gamma \tag{20}$$

$$\sum_{o\in O_\omega} \tilde{\eta}_{\omega o\gamma} \le H_{\omega\gamma}, \quad \forall \omega \in \Omega,\ \gamma \in \Gamma \tag{21}$$

$$\sum_{\{u\in N_{\omega o\gamma}:(u,n)\in E_{\omega o\gamma}\}} f_{un,\omega,o,\gamma} - \sum_{\{v\in N_{\omega o\gamma}:(n,v)\in E_{\omega o\gamma}\}} f_{nv,\omega,o,\gamma} = 0, \quad \forall n \in N^1_{\omega o\gamma},\ \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma \tag{22}$$

$$\sum_{\{v\in N_{\omega o\gamma}:(n,v)\in E_{\omega o\gamma}\}} f_{nv,\omega,o,\gamma} = \eta_{\omega o\gamma} + \tilde{\eta}_{\omega o\gamma}, \quad \forall n \in N^2_{\omega o\gamma},\ \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma \tag{23}$$

$$\sum_{\{u\in N_{\omega o\gamma}:(u,n)\in E_{\omega o\gamma}\}} f_{un,\omega,o,\gamma} = \eta_{\omega o\gamma} + \tilde{\eta}_{\omega o\gamma}, \quad \forall n \in N^3_{\omega o\gamma},\ \omega \in \Omega,\ o \in O_\omega,\ \gamma \in \Gamma \tag{24}$$

$$\sum_{\gamma\in\Gamma_d}\sum_{\omega\in\Omega}\sum_{o\in O_\omega}\sum_{\{(u,v)\in E_{\omega o\gamma}:g_{\omega o}(v)=(i,j,t)\}} f_{uv,\omega,o,\gamma} \ge \sum_{k\in\mathcal{K}_{ij}} L_{ijkd}x_{ijkt}, \quad \forall i \in I,\ j \in J_i,\ t \in T,\ d \in D \tag{25}$$

The three terms of the objective function (1) respectively minimize the total cost associated with customer wait time, optional worker call-ins/overtime (if allowed), and the operational cost of a configuration. Constraints (2) ensure that, for each period, exactly one configuration is selected for any service node within a given service area. The decision variable μ_{ijt} captures the effective processing rate for each configuration, as a function of the nominal processing rate r_{ijk} for each area $i \in I$, node $j \in J_i$, and configuration $k \in \mathcal{K}_{ij}$. To model productivity loss when changing configurations, we introduce the penalty $\tilde{r}_{ijk_1k_2}$, defined for each pair of configurations $k_1, k_2 \in \mathcal{K}_{ij}$ for service area $i \in I$ and service node $j \in J_i$. This penalty accounts for the costs associated with shutting down or starting up capacity. For example, when capacity is reduced, infrastructure may need to close before the end of the period to allow workers to safely shut down equipment. Similarly, when capacity is increased, it takes time for the new infrastructure to reach full operational capacity, as workers need time to turn on equipment. The effective processing rate, including the penalties for opening and closing areas, is captured through constraints (3) – (8). In these constraints, if the selected configuration remains unchanged between periods t and $t+1$, the effective processing rate μ is given by the nominal rate r associated with that configuration. However, if there is a change in configuration, at least one y-variable is equal to one, and the processing rate is adjusted according to the corresponding $\tilde{r}$ values.

The parameter λ_{it} represents the arrival rate of jobs to service area $i \in I$ at time $t \in T$. Jobs enter a service area only at the first service node, while for subsequent nodes, arrivals are equal to the number of jobs processed from the preceding node. We use p_{ijt} to denote the number of jobs processed at node $j \in J_i$ of service area $i \in I$ at time $t \in T$. The decision variable Q_{ijt} denotes the number of jobs in the queue at service area $i \in I$ and node $j \in J_i$ at the end of period $t \in T$. We assume that the system is empty at the beginning of the planning horizon (i.e., $Q_{ij0} = 0$). Constraints (9)–(10) incorporate a set of hyperplanes trained offline using data from multiple simulations. These hyperplanes provide a deterministic approximation of the maximum number of jobs that can be processed, based on factors such as stochastic processing rates under a given configuration, stochastic new job arrival rates, and initial queue sizes. In this work, we employ three hyperplanes for each service area and node, following the approach in Jiao et al. (2024). Constraints (11)-(12) ensure that the queue dynamics at any service area and node are consistent with the initial queue, arrivals, and the number of jobs processed.

The decision variable $\tilde{Q}_{ijtt'}$ represents the number of jobs in the queue at the end of time period t that are scheduled to be processed in time period t' ($t' > t$) in service area i and service node j. The parameter τ denotes the maximum number of time periods a job can wait before being scheduled for service. Constraints (13)-(14) ensure that all jobs are scheduled within τ periods of their arrival in the queue. These $\tilde{Q}$-variables and constraints are necessary when there are strict requirements on the maximum waiting time before a job is processed. However, they can be eliminated—along with the associated constraints—when this feature is not a part of the operational performance metrics.

The parameter Π_{ijk} denotes the maximum number of jobs that can be waiting at node j in service area i when configuration k is selected. This is enforced through constraints (15), which limit the maximum queue size. Additionally, if a service node changes its operating configuration, it is possible to enforce a minimum duration during which the new configuration must remain active after a change. The parameter M defines the minimum number of periods a configuration must remain selected following a change. By using binary variables δ^r_{ijkt} along with constraints (16)-(19), the model prevents frequent configuration changes. If a new configuration is adopted at time $t+1$ (i.e., $x_{ijkt} = 0$ and $x_{i,j,k,t+1} = 1$), then the corresponding δ^u-variable is set to one. Conversely, if a configuration is no longer used (i.e., $x_{ijkt} = 1$ and $x_{i,j,k,t+1} = 0$), the corresponding δ^v-variable is set to one. Constraints (17) ensure that if a configuration change occurs at time t, no additional changes can take place for the next M time periods (or until the end of the planning horizon, if fewer than M periods remain). Constraints (18) and (19) prevent any configuration changes from occurring during the first and last M periods of the planning horizon, respectively.

In our modeling approach, a time-expanded network tracks the flow of labor, ensuring that breaks are allocated, overtime is permitted, and worker classification demands are satisfied. The set Ω represents all shifts, while the set O_ω contains all schedules for shift ω, including different break schedules that can be assigned to a resource for the day.

Each worker may have multiple distinctions, where each distinction represents a different attribute. Table 4a provides examples of distinctions that a worker might have in a system representing the airport security screening process. All possible distinctions are stored in the set D. The resource classifications are defined as collections of elements from D, represented by the set Γ. Thus, if a worker belongs to classification $\gamma \in \Gamma$, they possess all the distinctions included in γ. Table 4b presents examples of worker classifications that could be used in an airport security screening system.

Worker Distinction
Male
Female
Checkpoint Certification
Baggage Certification
Dual Certification

(a)

Worker Classification
{Male, Checkpoint Certification}
{Male, Baggage Certification}
{Male, Dual Certification}
{Female, Checkpoint Certification}
{Female, Baggage Certification}
{Female, Dual Certification}

(b)

Table 4: An airport security screening example of the distinction set D and the classification set Γ

The time-expanded network contains a node for each combination of service nodes and periods within the time horizon. It can be viewed as a collection of individual networks, one for each combination of $\omega \in \Omega$, $o \in O_\omega$, and $\gamma \in \Gamma$. Each individual network consists of a subset of nodes and arcs relevant to a specific triplet (ω, o, γ), denoted by $N_{\omega o \gamma}$ and $E_{\omega o \gamma}$, respectively. The set of nodes is divided into three disjoint subsets: $N^1_{\omega o \gamma}$, $N^2_{\omega o \gamma}$, and $N^3_{\omega o \gamma}$, which represent the regular nodes (non-sink and non-source), source nodes, and sink nodes, respectively. The regular nodes $n \in N^1_{\omega o \gamma}$ can be mapped to a unique triplet $g(n) \mapsto (i, j, t)$, where i is a service area, j is a service node, and t is a time period. In the case of breaks, which are also regular nodes, the mapping represents a dummy area where labor cannot be utilized. Integer flow variables $f_{uv,\omega,o,\gamma}$ model the movement of labor, where arc $(u, v) \in E_{\omega o \gamma}$ represents the possible movement of workers between locations across time periods, in accordance with operational rules. For example, an arc $(u, v) \in E_{\omega o \gamma}$ with $g(u) \mapsto (i_1, j_1, t)$ and $g(v) \mapsto (i_2, j_2, t+1)$ indicates that a worker can leave service node j_1 in service area i_1 at time t and arrive at service node j_2 in service area i_2 at time $t+1$, provided they have certification γ to work at the destination and both time periods fall within their assigned schedule o and shift ω.

For any given shift $\omega \in \Omega$ and resource classification $\gamma \in \Gamma$, there may be resources that have already been allocated and are thus considered a fixed cost over the planning horizon. In some systems, workers are scheduled full-time for shifts several weeks in advance and are regarded as pre-allocated resources. The number of pre-allocated workers is represented by the parameter $h_{\omega\gamma}$. If these resources are insufficient to meet the system's demand, the model can assign existing resources to work overtime at an additional cost. To accommodate this, one or more overtime shifts can be created and linked to the beginning, the end, or both, of each shift that permits overtime schedules. These overtime shifts are added to the set of shifts. To model the potential flow of workers between a regular shift and overtime, the sequence of nodes representing a schedule (within a shift) is connected to the appropriate "Sink/Source" node based on the overtime start and end time. Figure 2 illustrates two overtime schedules added at the end of a regular shift with two break schedules, but the concept is analogous for overtime added before the beginning of a shift. The parameter

$H_{\omega\gamma}$ defines the maximum number of additional resources of classification γ that can be allocated to shift ω, and the integer decision variable $\tilde{\eta}_{\omega o \gamma}$ captures the number of allocated additional resources.

Constraints (20) ensure that the resources allocated to a shift do not exceed the pre-allocated resources. The limit on the resources available to hire, in excess of the pre-allocated resources, is enforced by constraints (21). We assume that overtime shifts will not have any pre-allocated workers and labor will be "hired" in an amount not to exceed the labor available for the corresponding regular shift. Constraints (22)-(24) follow a traditional multi-commodity flow-balance approach to track the flow of labor through the network. The pre-allocated and additional workers act as the supply at the source nodes and the demand at the sink nodes, ensuring that the number of workers entering the network equals the number of workers exiting at the end of the shift or after overtime is completed. The regular nodes serve as transshipment nodes, indicating that the same number of workers entering a location at a given time either remains in that location for one more period, moves to a different location, finishes the shift, or begins overtime. This construction can be seen as a multi-commodity flow structure, where a commodity is a triplet (ω, o, γ).

Service systems often impose operating constraints based on the classifications or certifications of resources. For example, in an airport security screening process, each screening lane at a checkpoint requires at least one male and one female worker to conduct passenger screening (e.g., pat-downs). Moreover, some certifications are required to operate certain equipment. To capture these real-world constraints, the set $\Gamma_d \subseteq \Gamma$ is defined as the collection of all resource classifications in Γ that include distinction $d \in D$. The parameter L_{ijkd} denotes the required number of resources with distinction d to operate configuration $k \in \mathcal{K}_{ij}$ at service node $j \in J_i$ in service area $i \in I$. More generally, worker skills are incorporated by first defining the set of all skills, D, and then creating subsets of Γ representing combinations of skills that workers may possess, such that $\Gamma = \bigcup_{d \in D} \Gamma_d$. Each set Γ_d is a collection of worker classifications containing skill $d \in D$. For any service area $i \in I$, service node $j \in J_i$, and configuration $k \in \mathcal{K}_{ij}$, the required amount of labor with skill type $d \in D$ is denoted by L_{ijkd}, which must be fulfilled by workers with classification Γ_d.

Incorporating worker skills into the model allows us to capture the spatiotemporal dynamics of resources in systems where certain worker classes can operate specific configurations that other classes cannot. Constraints (25) ensure that the spatiotemporal demands for each worker classification are met according to the selected configurations for operating service nodes in each period.

4 Case Study - Security Screening Operations at an Airport

4.1 Description and input data

The Transportation Security Administration (TSA) operates in U.S. airports to ensure the safety of federal airspace through various routine screening procedures. Every passenger undergoes a two-stage screening process. First, passengers proceed to a security checkpoint where a Transportation Security Officer (TSO) verifies their documents at a Travel Document Checking (TDC) station. Once their documents are cleared, passengers move on to a Screening Lane (SL), where their person, personal items, and carry-on luggage are checked for potential security threats. The proposed model was validated using a case study based on masked data from one of the busiest days at Miami International Airport (MIA). Each security checkpoint is treated as a service area with two sequential service nodes: the first node represents the TDC stations, and the second node represents the SLs. An additional service area with a single node was included to capture the checked baggage screening process. We use a 24-hour planning horizon, divided into 15-minute periods.

Our case study includes four security checkpoints, each operated under a specific configuration that defines the number of TDCs and SLs active during any given period. The possible configurations for TDCs and SLs are constrained by the space limitations at each checkpoint. In the real system, a third step in the screening process is used for secondary screening; however, in this model, that behavior is assumed to be captured within the processing rate of the SLs (second service node).

The staffing requirements depend on the selected configuration. Each open TDC requires one TSO. For SLs, the number of TSOs needed is based on the number of open lanes in the configuration. If the number of open SLs is even, six TSOs are assigned per pair of SLs. If the number is odd, six TSOs are assigned to each pair, with an additional four TSOs allocated to the remaining SL. This arrangement is due to SLs being configured in pairs that share an Automated Imaging Technology (AIT) scanner and a walk-through metal detector (WTMD). Additionally, TSA regulations specify the required number of male and female TSOs at each checkpoint based on the number of open SLs, which is a resource distinction in set D. This ensures proper staffing for pat-down procedures when passengers are flagged for secondary screening. Another distinction accounted for in our model is TSO certification, which differentiates between those certified to work in the checkpoint area, the baggage area, or both (dual certification).

Passenger processing rates in each service area depend on the chosen configuration. Each screening lane is assumed to process 180 passengers per hour, while each TDC processes 250 passengers per hour. When a lane opens, its processing capacity is reduced by 50% during the first 15 minutes to account for the ramp-up (setup) effect. Similarly, a 50% reduction is applied in the period before a lane closes to account for the ramp-down effect. These adjustments capture the operational protocols involved in opening or closing a screening lane, including equipment activation and report completion. Neglecting these effects may lead to overestimated processing rates, potential understaffing of SLs, and overly optimistic predictions of wait times and queue lengths. Table 5 presents the impact of these ramp-up and ramp-down effects on processing rates.

Processing Rates						
Period	1	2	3	4	5	6
Lanes Open	1	1	2	2	2	1
Rate / 15 min without ramp effects	45	45	90	90	90	45
Rate / 15 min with ramp effects	45	45	67.5	90	67.5	45

Table 5: Comparison of configuration changes with and without a ramp affect.

To minimize operational disruptions, the model requires each selected configuration to remain in place for at least one hour before any changes can occur. Frequent reconfigurations are operationally undesirable for both TSOs and passengers, as they require additional tasks such as adjusting the waiting area layout and redirecting passengers to the appropriate service areas.

A fraction of passengers arriving at the airport will have checked luggage. Based on feedback from TSA analysts, the luggage screening process is less sensitive to delays than checkpoint screening, allowing us to disregard queues in the baggage area. The demand for baggage-area-certified TSOs can be estimated *a priori*. This demand is determined by an external algorithm and fixed in advance, as described in Appendix B. Constraints (25) are also applied in the baggage area, with required labor set according to the method outlined in Appendix B.

Since TSO rotations within a checkpoint occur regularly without significant productivity loss, we assume that idle time for officers moving between service nodes (TDCs and SLs) within the same checkpoint is negligible and can be accommodated within the same 15-minute period. Travel between distinct checkpoints incurs a delay. Traveling to a neighboring checkpoint requires one 15-minute period during which the TSO is unavailable, while travel to more distant checkpoints requires two 15-minute periods. In the case study, the baggage processing area is treated as a neighboring checkpoint to all other checkpoints in the analysis.

In practice, a semiannual process determines the allocation of TSOs to shifts; therefore, the associated workforce cost is treated as exogenous to the model. Masked data for TSO shifts at MIA is used in this case study. For each shift, we consider multiple break schedules. Shifts lasting at least 8.5 hours include two 15-minute breaks and one 30-minute break. Shifts between 6 and 8.5 hours include two 15-minute breaks, while shifts longer than 4 hours but less than 6 hours have one 15-minute break. Shifts exactly 4 hours long do not include breaks. Table 6 summarizes the shift lengths, the number of TSOs assigned to each shift, and the possible break schedules options. The amount of labor allowed for overtime schedules is capped by the pre-determined labor assigned to the regular shifts. Each TSO can work 1, 2, or 4 hours of overtime, either before or after the regular shift. Overtime is compensated at a standard rate of \$60 per hour, based on the average TSO overtime pay. We assume that adding overtime hours does not increase the number of breaks; however, this assumption can be adjusted through input parameters.

Shift Length (h)	Number of Shifts	Break Schedule Options Per Shift	Number of TSOs
10.5	2	4	18
8.5	15	4	221
7.5	2	2	2
6.5	2	2	2
5.5	2	2	2
5	6	1	14
4.5	1	1	1
4	14	1	30

Table 6: The characteristics of shifts and breaks for the security screening case study.

Figure 3 presents the predicted baseline arrivals for each checkpoint, derived from the flight schedule, operational parameters, and historical data (Hanumantha et al., 2020).

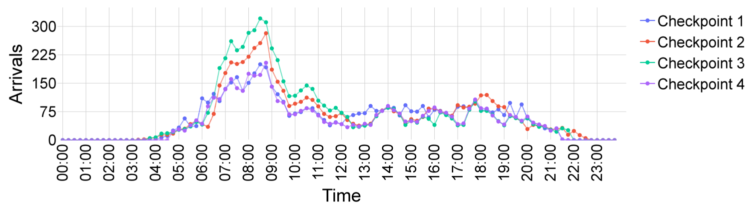


Figure 3: Arrivals for the airport security screening case study.

There is no strict requirement limiting the number of periods a passenger can wait at the airport before beginning processing. As a result, constraints (13)-(14), along with the $\tilde{Q}$-variables, are excluded from the model. While precise wait time rates are difficult to determine, we follow the U.S. Department of Transportation's estimates (USDOT, 2016) and adjust the recommendation for inflation, penalizing the queueing variables (Q_{ijt}) at a rate of \$15 for each passenger still in the queue at the end of a period. The operational costs, represented by the $\hat{c}$-parameters, correspond to estimated equipment usage costs per lane and TDC per period, excluding personnel expenses.

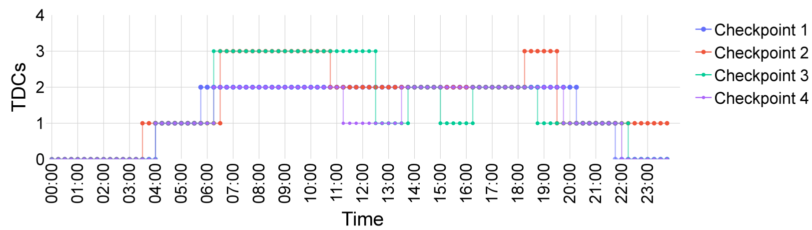
4.2 Results

The results presented were obtained using the Gurobi solver on a conventional desktop running Windows 10, equipped with an AMD Ryzen 7 3700X 3.59 GHz processor and 32 GB of RAM. Solutions were computed within a 1% optimality gap.

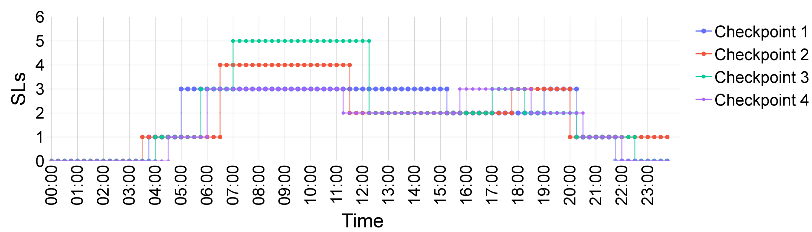
Movement of TSOs through the airport strains the system in two ways. First, the idle time during movement, as captured by the model, results in a loss of available labor. Second, an operational plan with excessive movement throughout the day increases the difficulty of adhering to the plan during implementation, which is undesirable for supervisors who typically prefer minimizing staff movement. To illustrate the effect of TSO movement, the model was first run allowing movements at any time of day. After two hours, the solver found no feasible solutions and achieved a lower bound for the objective function value of \$274,348. With all other input data remaining the same, movement was restricted to only after the completion of a break. This version of the model terminated at an optimality gap of 0.65% with an objective value of \$276,249 in just 40 seconds. The original model, allowing movement at any time, found a feasible solution with less than a 1% optimality gap in 468 seconds after being warm-started with the solution from the movement-after-breaks-only model.

The remaining discussion in this section pertains to the solution where movement was only allowed after breaks. By reducing the number of movement options, the solver was able to find a solution that was both implementable and near-optimal. This does not suggest that every model can disregard significant movement considerations to improve solution time; rather, practitioners should carefully evaluate whether all possible movement options are truly necessary. The impact of movement decisions (i.e., no movement allowed, movement only after breaks, or movement anytime) will be explored further in Section 5.1.

Configuration decisions are arguably the most important decisions to provide to managers for their daily planning. Figure 4 summarizes the configuration decisions for TDCs and SLs throughout the day in the case study. These results demonstrate the dynamic nature of configuration decisions, which generally align with the passenger arrival profile shown in Figure 3.



(a) TDC configurations



(b) SL configurations

Figure 4: Configurations for the security checkpoints in the case study.

The model also outputs the estimated queue lengths. The maximum queue for the day is 804 passengers, with an average queue of 101 passengers during periods when the queue is non-zero. The maximum queue occurs at checkpoint three, where there are three TDC stations available and a maximum of five operational SLs. With all TDCs and SLs in use, each service node has a maximum processing rate of 750 and 900 passengers per hour, respectively. Based on these rates, the queue is expected to be cleared in no less than $60(804/900) = 53.6$ minutes. In the TSA screening process, a 50-minute wait would typically be considered unacceptable, triggering concern for the end user. However, it is important to note that this case study is based on data from the busiest and atypical day at the airport. As a mitigation measure, the TSA could deploy canine units or additional TSOs to the checkpoints to expedite the screening process and reduce queues. These additional actions could be incorporated into the model as alternative configuration options or deployed when wait times exceed acceptable thresholds.

The model prescribes overtime for only 18 TSOs, with one TSO working 2 hours of overtime and the rest working 1 hour. In practice, determining when TSOs should take breaks is often a pain point in managing the security screening system. These decisions are typically left to floor managers to make in real-time. Breaks are usually scheduled when queues are manageable, which may not coincide with reasonable meal times and may be scheduled too close to the beginning or end of a shift. This uncertainty in break times leads to dissatisfaction among TSOs, a growing concern due to high turnover in recent years.

By objectively determining break schedules in accordance with organizational rules, the model alleviates some of the challenges the TSA faces in scheduling breaks. The user only needs to input break schedules that ensure fair and equitable rules for TSOs, and the model will allocate workers to the schedules that place the least strain on the system. Figure 5 displays four break options for a shift from 4:15 a.m. to 12:45 p.m, as well as the passenger arrivals (λ) and number of TSOs assigned to each schedule (Column "TSOs On Schedule"). For this particular shift, the model aims to allocate TSOs in a way that avoids scheduling breaks during periods of high checkpoint traffic. However, given the fixed break options, TSOs from other shifts also provide coverage during these peak periods. Figure 6 illustrates a sequence of seven consecutive periods within a time-expanded network of a single shift. Initially, TSOs are distributed across three checkpoints before taking a break, after which they regroup at Checkpoint 2. Note that not all TSOs are immediately available post-break, as they must travel to Checkpoint 2 from their previous locations.

In assigning break times, managers must consider the certifications of the remaining TSOs, future arrival streams, TSOs on other shifts, and staffing needs or available TSOs in other areas. Coordinating these decisions manually is complex, if not impossible. Although the model aggregates workers of the same classification in the time-expanded network, the operational plan (i.e., model solution) can be post-processed to determine the schedules of individual TSOs throughout the day. Additionally, examining system performance over several days can help TSA analysts identify when and how the system might fail due to significant deficiencies in worker distinctions, potentially prompting cross-training or hiring requests.

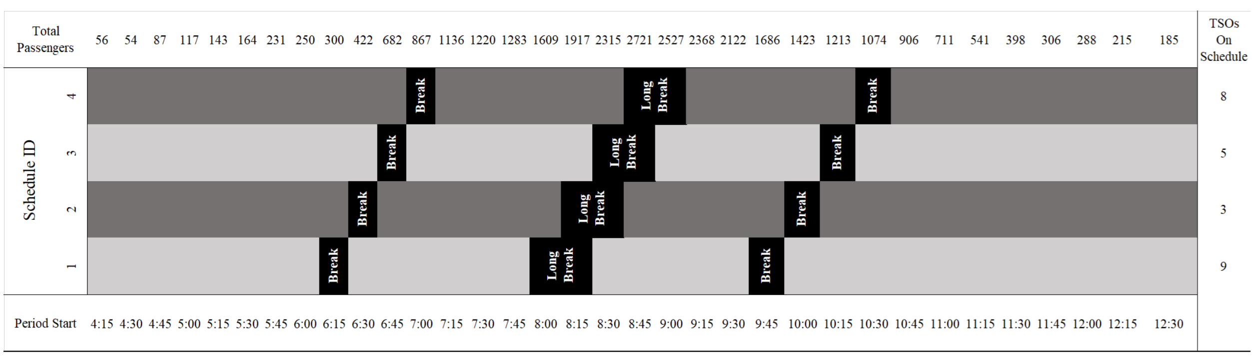


Figure 5: Break options for an example shift.

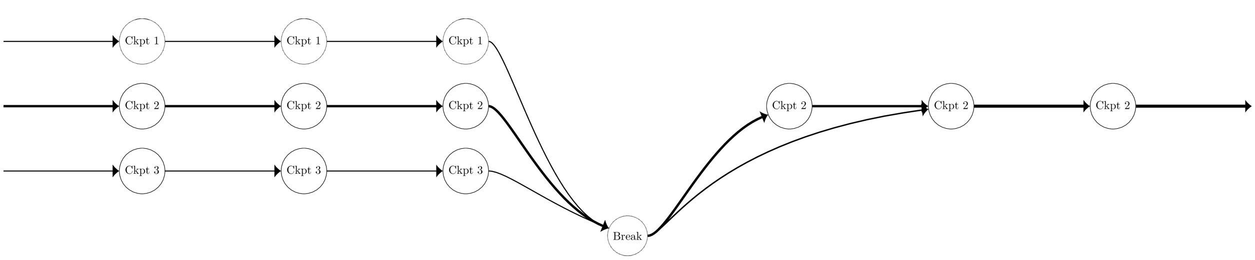
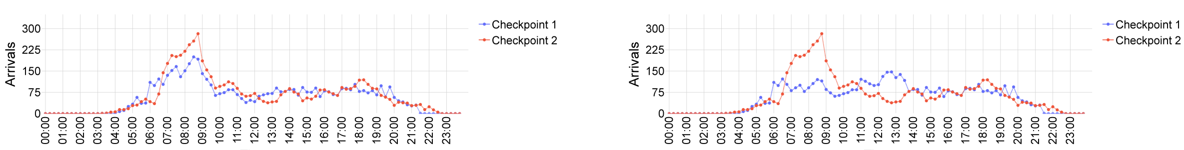


Figure 6: Portion of a time expanded network for a single shift.

5 Computational Experiments

This part of the study is rooted in our experience working with real-world decision-makers and the practical questions they often raise. Their concerns centered on two main areas: the scalability of the model and the opportunity cost of adding features relative to the computational resources required. The computational analysis offers quantifiable evidence of the impact of increased staffing levels, which is essential for justifying such operational decisions and requesting budgetary authority. We designed a suite of computational experiments to better understand how different model features impact solution time, cost, and information completeness. In practice, some decisions such as break times could be removed from the model to reduce computation time at the expense of requiring supervisors to make judgments. Fast solution times are especially critical for what-if analyses, where the model must be run repeatedly to evaluate various scenarios (e.g., passenger arrival patterns, officer call-offs, K9 deployments). Slow models hinder this process. In the event of unexpected disruptions—such as flight delays or checkpoint incidents—the model may need to be re-run with updated inputs. Rapid computation is crucial to enable timely decisions. Our analysis helps identify which model features (factors) can be feasibly incorporated for realistic problem instances. One factor was varied over three levels, while five other factors were varied over two levels each. Each unique combination of factor levels is referred to as *design point*.

Worker movement through the airport was tested at three levels of flexibility. The most flexible level allowed movement between checkpoints at any time, the intermediate level allowed movement only after a break, and the least flexible required workers to remain in the area where they started their shift. In addition to potentially impacting solution time, this factor is also important in practice, particularly for TSA managers, who prefer to move TSOs only when operationally necessary. Two arrival patterns were used in the experiments. The first was based on masked data from the four checkpoints case study at MIA, where peak-time arrivals overlap (see Figure 3). This overlap may not occur at all airports or at all times. The second arrival pattern introduced a two-hour shift in peak times for two of the checkpoints, resulting in a non-overlapping pattern. This adjustment allows us to induce movement in worker deployment by creating dynamic labor demands across both space and time. Figure 7 depicts the arrivals used for this experiment.



(a) Synchronized baseline arrivals for 2 checkpoints

(b) Shifted arrivals for 2 checkpoints

Figure 7: Arrival patterns for the 2 checkpoint experiment

Since two different arrival patterns were used, a set-covering model was employed to determine an initial allocation of workers to shifts for each pattern. The *set covering* model used for determining staffing levels is described in Appendix A. In practice, staffing decisions are driven by historical peak days and budget constraints, but a method to estimate a lower bound on worker requirements was needed for the experiment. The resulting staffing level represents an *optimistic lower bound* on actual needs, as it does not account for worker distinctions or labor availability losses due to breaks. In practice, the actual workforce is substantially larger than this minimum to mitigate the risk of understaffing. The number of workers per shift is varied across two levels. The first level involves increasing the staffing suggested by the set covering model (i.e., the lower bound) by 20% to account for additional labor needs due to breaks, movement, or overtime. The second level applies the same approach but increases staffing by 40% for each shift.

Table 7 presents the shift lengths, the number of break options for each shift length, and the distribution of shifts used in the experiments. The shift distribution is based on masked data from the TSA. The computational experiments include two options for break schedules. In the first option, the model is allowed to assign break schedules to workers within a shift out of a predetermined set that satisfies operational rules (e.g., union agreements). In the second option, break schedules are pre-assigned and treated as input data. In this case, breaks are assigned randomly using a uniform probability distribution.

For each shift, workers are assumed to be available for 1 to 6 hours of overtime in one-hour increments, either before or after their scheduled shift. We modeled two scenarios to assess the impact of overtime policies: the first permits all overtime options, while the second prohibits overtime entirely.

Shift Length (h)	Possible Break Schedules	Percentage of All Shifts	15min Breaks	30min Breaks
10.5	6	5.00%	2	1
8.5	5	75.00%	2	1
7.5	4	0.50%	2	0
6.5	3	1.00%	2	0
5.5	3	1.00%	1	0
5	2	4.50%	1	0
4.5	2	13.00%	0	0

Table 7: Shift description and distribution.

Each worker's classification can be tracked in the model. A TSO may be trained to work exclusively in the baggage screening area, the checkpoint area, or in both. Additionally, there are gender requirements for passenger screening at each checkpoint, which depend on the number of open screening lanes. To determine the impact of these features, we solve the model with both certification and gender considerations included. We also solve a case where neither gender nor training considerations are factored into the model.

A full factorial experiment with five runs at each design point was conducted to understand how different factors in the input data impact solution quality and time. There were a total of $3 \cdot 2^5 = 96$ different design points. Each of the five runs at any given design point varied by the noise added to the arrival patterns. For each time period and service area, the expected arrivals λ_{it} were used to generate a random instance of arrivals. This was achieved by sampling a uniform random variable over the interval $[0.5\lambda_{it}, 1.5\lambda_{it}]$, and using the resulting value as the deterministic arrival for that run. This procedure was repeated five times for each of the arrival patterns, and the same five instances were used to test other factors.

Tables 8 and 9 summarize all the sets and parameters used in our experiments. Processing capacities across checkpoints in the generated instances were higher compared to the MIA case study. This adjustment was made because the airport experienced long queues during the busiest parts of the day, causing nearly every TDC and SL to operate at full capacity in most periods. Additionally, the number of available TDCs was increased to provide the model with more flexibility and expand the range of decisions.

Set	Description
I	$\{1,2,3\}$, where area 3 is a baggage area
J_i	$\{1,2\}$ for $i \in \{1,2\}$ and $\{1\}$ for $i=3$
$\mathcal{K}_{ij}$	$\{1,2,3\}$ for $i=1,\ j \in J_i$, $\{1,2,3,4\}$ for $i=2,\ j \in J_i$ and $\{1\}$ for $i=3,\ j=1$
T	$\{1,2,\ldots,96\}$
D	{ male, female, SSCP qualified, baggage qualified, both qualifications }
Γ	{male, SSCP qualified}, {male, baggage qualified}, {male, both qualifications}, {female, SSCP qualified}, {female, baggage qualified}, {female, both qualifications}
Ω	Generated by set cover model in Appendix A augmented with shifts representing overtime.
O_ω	Break or Overtime schedules for shift $\omega \in \Omega$
$N_{\omega o \gamma}$	Nodes of the time expanded network of resource traveling for shift $\omega \in \Omega$, schedule $o \in O_\omega$, and classification $\gamma \in \Gamma$

Table 8: Sets used in the computational experiment for 2 checkpoints.

Parameter	Description
r_{ijk}	225 pax/hr for each TDC and 200 pax/hr for each SL
$\tilde{r}_{ijk_1k_2}$	50% of the processing rate per TDC or SL for a single period
λ_{it}	Arrivals to the TDC derived from masked data
β	$\beta_1^1 = 0.99318$, $\beta_2^1 = 0.00291$, $\beta_3^1 = 0.99019$, $\beta_1^2 = 0.50791$, $\beta_2^2 = 0.45905$, $\beta_3^2 = 0.50493$, $\beta_1^3 = 0.00299$, $\beta_2^3 = 0.99932$, and $\beta_3^3 = 0$
Π_{ijk}	Unlimited for TDCs and 25 passengers per SL
M	1 hour (4 periods)
$h_{\omega\gamma}$	Determined by the model in Appendix A. No predetermined workers on an overtime shift.
$H_{\omega\gamma}$	$h_{\tilde{\omega}\gamma}$ for overtime shifts where $\tilde{\omega}$ is the regular shift corresponding to overtime shift ω. No additional hires allowed for regular shifts.
L_{ijkd}	1 TSO per TDC; 6 TSOs for every two SLs, 4 TSOs for a single SL. Each SL pair open requires one male and one female TSO. TSOs needed at baggage determined in advance (see Appendix B. TSOs must be certified to work at a checkpoint or the baggage area.
c_{ijt}	Cost of queueing at $60 per hour
$\tilde{c}_{\omega\gamma}$	Cost of overtime at $60 per hour per TSO
$\hat{c}_{ijk}$	Operational cost at $20 per TDC and SL open per hour

Table 9: Parameters used in the computational experiment for 2 checkpoints.

Sections 5.1 and 5.2 present the performance of the model under various feature combinations for two and four checkpoints. These configurations reflect realistic operational scenarios for the TSA. For instance, Terminal 4 at Phoenix Sky Harbor Airport (PHX) operates four checkpoints, where reassigning officers between terminals is generally impractical due to long transit times. Most regional airports across the United States operate with only one or two checkpoints.

5.1 Results for Two Checkpoints

Table 10 reports the average time required to find a solution within a 1% optimality gap across all design points, while Table 11 presents the average ratio of the solution value compared to that of the most unrestricted design point. The unrestricted design points for each combination of staffing and arrivals are used as benchmarks for comparison. As expected, models with higher operational resolution—allowing movement at any time, breaks, and overtime—are more challenging to solve. This represents an important trade-off, as Table 11 demonstrates the cost benefits of increased flexibility, such as allowing overtime or letting the model determine break schedules. The version of the model with the highest operational resolution includes 4,032 binary variables, 43,852 integer variables, and 20,152 continuous variables, along with a total of 34,416 constraints. In contrast, the model with the lowest operational resolution contains the same number of binary and continuous variables but only 840 integer variables and a total of 22,928 constraints.

Two Checkpoint Average Termination Time (Seconds)			Lower Bound + 20% Staffing				Lower Bound + 40% Staffing			
			All Distinctions		No Distinctions		All Distinctions		No Distinctions	
			Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks
Movement Anytime	Overtime	Break Decisions	86	89	49	35	73	54	37	23
		Fixed Breaks	60	52	30	24	44	33	21	18
	No Overtime	Break Decisions	3,596	893	2,504	124	421	58	30	20
		Fixed Breaks	1,125	450	325	78	124	57	26	11
Movement After Breaks	Overtime	Break Decisions	20	14	17	9	17	10	13	10
		Fixed Breaks	20	15	14	12	17	11	10	9
	No Overtime	Break Decisions	1,217	134	2,625	42	399	10	15	10
		Fixed Breaks	1,753	204	432	46	144	43	13	5
No Movement	Overtime	Break Decisions	21	11	15	8	16	9	11	7
		Fixed Breaks	17	15	16	12	15	10	11	8
	No Overtime	Break Decisions	665	106	1,483	29	1,220	22	19	9
		Fixed Breaks	444	199	531	45	333	43	15	5

Table 10: Average time to find a solution within 1% gap with two checkpoints.

Using overtime incurs an additional cost, but this can be beneficial if it leads to reduced wait times. This is evident in most design points in our experiment, where a relatively small overtime expense helps lower queueing costs. The average queue and overtime costs for all design points are shown in Table 12. In some cases, the savings are significant, such as in row 1, column 1, where $7,344 spent on overtime results in a savings of $357,493 in queueing costs. In practice, the solutions for design points without overtime and extreme queueing costs would be considered impractical. A somewhat counterintuitive result arises from the analysis of the impact of overtime decisions on solution time. With overtime allowed, a 1%-optimal solution was found, on average, in 24 seconds, whereas the time increased to 460 seconds on average when overtime

Two Checkpoint Ratio of Achieved Objective Value to Unrestricted Objective Value			Lower Bound + 20% Staffing				Lower Bound + 40% Staffing			
			All Distinctions		No Distinctions		All Distinctions		No Distinctions	
			Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks
Movement Anytime	Overtime	Break Decisions	1.044	1.016	1.000*	1.000*	1.037	1.006	1.000*	1.000*
		Fixed Breaks	1.091	1.033	1.023	1.010	1.073	1.015	1.020	1.005
	No Overtime	Break Decisions	6.215	1.772	3.893	1.459	1.500	1.026	0.999	1.001
		Fixed Breaks	8.862	2.490	5.437	1.797	2.180	1.107	1.394	1.006
Movement After Breaks	Overtime	Break Decisions	1.048	1.021	0.999	1.001	1.038	1.008	1.000	1.001
		Fixed Breaks	1.091	1.037	1.022	1.010	1.074	1.016	1.020	1.004
	No Overtime	Break Decisions	7.261	1.781	4.356	1.462	1.673	1.027	0.998	1.001
		Fixed Breaks	10.124	2.513	5.835	1.805	2.277	1.257	1.396	1.005
No Movement	Overtime	Break Decisions	1.050	1.023	1.002	1.002	1.045	1.010	1.004	1.000
		Fixed Breaks	1.092	1.039	1.024	1.010	1.079	1.018	1.019	1.005
	No Overtime	Break Decisions	9.293	1.794	5.947	1.475	3.114	1.033	1.022	1.004
		Fixed Breaks	11.442	2.566	7.227	1.806	3.944	1.278	1.394	1.005

* indicates a design point used as the "unrestricted" design point for comparisons.

Table 11: Average ratio of the achieved objective value to the best objective value for the least restrictive design point (same arrivals and staffing level) with two checkpoints.

was prohibited. This suggests that overtime decisions may help the model find quicker incumbents, which speeds up the overall solution process. This effect is particularly pronounced when labor is at a low level, making it more challenging to find integer solutions that accommodate arrivals.

Using fixed breaks speeds up the model by reducing the complexity of scheduling decisions, as there are fewer decisions to make. On average, the time to solve instances with break decisions is 340 seconds, while instances with fixed breaks solve in 145 seconds. The trade-off for using fixed breaks, however, is a decrease in solution quality, reflected in an increase in the average objective value by $28,103.22.

As expected, an increase in the number of available workers leads to a decrease in the objective function value. When only looking at design points that use staffing levels 40% above the lower bound, a benefit of $11,137.13 is realized on average in the objective function value when break decisions are not randomly chosen. The corresponding average computational time is only 62 seconds. This means that there is the potential for realizable cost savings by allowing the model to prescribe break schedules without an excessive increase in computing needs. We also observe that increasing the workforce mitigates the impact of the lack of overtime when it is not allowed. In systems where overtime is generally undesirable but necessary to maintain service standards, staffing levels must be adjusted to align with operational demands.

Two Checkpoint Average Queueing Cost (Average Overtime Cost)			Lower Bound + 20% Staffing				Lower Bound + 40% Staffing			
			All Distinctions		No Distinctions		All Distinctions		No Distinctions	
			Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks
Movement Anytime	Overtime	Break Decisions	$60,958 ($7,344)	$75,993 ($2,772)	$60,377 ($4,824)	$75,674 ($1,764)	$60,908 ($2,124)	$75,728 ($720)	$60,344 ($312)	$75,690 ($324)
		Fixed Breaks	$60,996 ($10,404)	$76,145 ($3,948)	$60,714 ($6,108)	$75,990 ($2,232)	$60,990 ($4,308)	$75,851 ($1,344)	$60,519 ($1,368)	$75,976 ($408)
	No Overtime	Break Decisions	$418,451 (-)	$139,647 (-)	$262,178 (-)	$114,034 (-)	$93,343 (-)	$78,101 (-)	$60,605 (-)	$75,996 (-)
		Fixed Breaks	$598,277 (-)	$195,902 (-)	$364,141 (-)	$140,962 (-)	$136,584 (-)	$84,430 (-)	$86,094 (-)	$76,437 (-)
Movement After Breaks	Overtime	Break Decisions	$60,969 ($7,608)	$76,020 ($3,144)	$60,507 ($4,680)	$75,614 ($1,884)	$60,934 ($2,172)	$75,926 ($756)	$60,402 ($276)	$75,958 ($192)
		Fixed Breaks	$60,878 ($10,512)	$76,014 ($4,440)	$60,637 ($6,108)	$76,023 ($2,244)	$60,911 ($4,440)	$75,951 ($1,356)	$60,464 ($1,404)	$75,901 ($456)
	No Overtime	Break Decisions	$489,143 (-)	$140,389 (-)	$293,165 (-)	$114,253 (-)	$104,223 (-)	$78,065 (-)	$60,527 (-)	$76,021 (-)
		Fixed Breaks	$683,366 (-)	$197,578 (-)	$392,289 (-)	$141,513 (-)	$142,906 (-)	$96,037 (-)	$86,216 (-)	$76,345 (-)
No Movement	Overtime	Break Decisions	$61,000 ($7,668)	$76,001 ($3,312)	$60,532 ($4,884)	$75,631 ($1,968)	$60,962 ($2,616)	$75,898 ($864)	$60,390 ($492)	$75,786 ($276)
		Fixed Breaks	$61,018 ($10,500)	$76,020 ($4,536)	$60,706 ($6,180)	$75,992 ($2,268)	$60,950 ($4,668)	$75,814 ($1,560)	$60,402 ($1,440)	$75,845 ($540)
	No Overtime	Break Decisions	$625,938 (-)	$141,516 (-)	$400,587 (-)	$115,478 (-)	$196,198 (-)	$78,567 (-)	$62,273 (-)	$76,288 (-)
		Fixed Breaks	$772,547 (-)	$201,799 (-)	$487,977 (-)	$141,563 (-)	$249,515 (-)	$97,775 (-)	$86,149 (-)	$76,373 (-)

Table 12: Average queue cost and overtime cost for all runs of the two checkpoint model.

Intuitively, the more flexible the movement of TSOs through the airport, the better the solution quality achieved by the model. Allowing movement only after breaks was sufficient for achieving most of the movement savings while reducing computational effort. Table 13 compares the objective function values under different TSO movement rules. When additional workers are available, the savings are less significant, as sufficient labor can be allocated to each service area without needing to move TSOs. This is expected, as once the number of workers reaches a certain level, the limiting factor for processing passengers becomes the system's capacity, not the availability of workers.

The time to find a solution is expected to decrease on average when movement is restricted. However, this is not always the case, as shown by the largest average solution time occurring when movement is completely restricted with staffing 40% above the lower bound. The differences in solution times with a 1% gap are presented in Table 14. The model with fewer hires consistently requires more computation time than the model with more hires, regardless of movement restrictions. This indicates that movement rules could potentially be anticipated based on input data, allowing for predefined critical periods to be established in

Movement Allowed	All Instances	LB+20% Staffing	LB+40% Staffing
Any Time	$130,683	$179,726	$81,640
After Breaks	$138,440	$193,411	$83,469
None	$158,432	$220,679	$96,186

Table 13: Average objective values under different TSO movement rules.

Movement Allowed	All Instances	LB+20% Staffing	LB+40% Staffing
Any Time	330.22	594.94	65.5
After Breaks	228.44	410.90	45.97
None	167.82	226.08	109.55

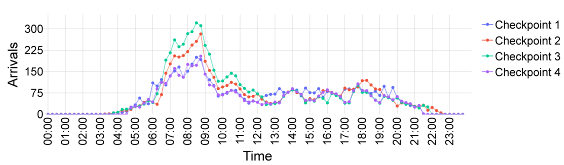
Table 14: Average solution time in seconds under different movement rules.

advance.

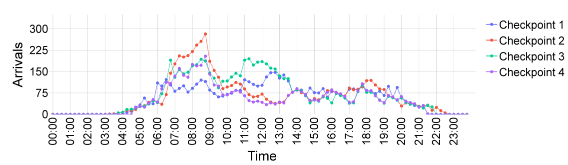
With the shifted peaks in arrivals, the average solution time is significantly reduced to just 65 seconds. In contrast, the baseline arrivals, with overlapping peaks, have an average solution time of 419 seconds. Including worker distinctions in the model increases the average solution time to 300 seconds, compared to 185 seconds when distinctions are not considered. Additionally, there is an average increase of $46,362 in the objective value when comparing the model results with and without this feature.

5.2 Results for Four Checkpoints

We also conducted experiments using instances with four checkpoints. Figure 8 depicts the arrival patterns for these checkpoints, based on data from MIA. We use the same noise-adding strategy to the arrivals and maintained the same procedure for generating break schedules as in the two-checkpoint analysis. All other model features are unchanged.



(a) Baseline arrivals



(b) Shifted arrivals

Figure 8: Arrivals patterns for the 4 checkpoint experiment

Despite the model having more decision variables, we observe the same solution pattern: instances reach a 1% optimality gap faster when overtime is allowed. In contrast, most instances without overtime fail to reach a solution within a 1% gap after two hours. Additionally, in the majority of design points, even a marginal overtime expenditure leads to significant reductions in queueing costs. The largest model used in the four-checkpoint experiments includes 8,640 binary variables, 205,536 integer variables, and 91,984 continuous variables, with a total of 90,124 constraints. The smallest model has the same number of binary and continuous variables but only 2,020 integer variables and 50,700 constraints. Table 15 shows the average solution time to reach a 1% optimal solution and the corresponding optimality gap for all design points. It should be noted that, given the model's size, the inclusion of movement decisions (whether after breaks or at any time) tends to slow down the model when additional labor is available.

Four Checkpoint Average Termination Time (Average Gap)			Lower Bound + 20% Staffing				Lower Bound + 40% Staffing			
			All Distinctions		No Distinctions		All Distinctions		No Distinctions	
			Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks
Movement Anytime	Overtime	Break Decisions	7,200 s (8.42%)	7,071 s (1.77%)	4,916 s (2.27%)	1,993 s (0.94%)	3,287 s (0.85%)	2,932 s (0.91%)	879 s (0.81%)	304 s (0.88%)
		Fixed Breaks	7,200 s (6.87%)	6,540 s (1.23%)	2,715 s (0.65%)	1,646 s (0.98%)	2,306 s (0.89%)	1,310 s (0.89%)	317 s (0.89%)	314 s (0.87%)
	No Overtime	Break Decisions	7,200 s (43.50%)	7,200 s (-)	7,200 s (30.92%)	7,200 s (26.74%)	7,200 s (35.13%)	7,200 s (15.73%)	3,822 s (0.91%)	1,501 s (0.86%)
		Fixed Breaks	31 s (ˆ)	7,200 s (44.36%)	7,200 s (40.11%)	7,200 s (36.38%)	7,200 s (39.77%)	7,200 s (19.06%)	7,200 s (20.73%)	1,768 s (0.83%)
Movement After Breaks	Overtime	Break Decisions	1,091 s (0.89%)	516 s (0.99%)	419 s (0.88%)	142 s (0.96%)	177 s (0.89%)	110 s (0.96%)	87 s (0.93%)	43 s (0.60%)
		Fixed Breaks	1,781 s (0.69%)	479 s (0.95%)	235 s (0.77%)	187 s (0.96%)	101 s (0.94%)	118 s (0.97%)	45 s (0.98%)	53 s (0.85%)
	No Overtime	Break Decisions	7,200 s (35.94%)	7,200 s (27.92%)	7,200 s (30.08%)	7,200 s (21.48%)	7,200 s (27.02%)	7,201 s (5.77%)	2,050 s (0.88%)	312 s (0.89%)
		Fixed Breaks	0 s (ˆ)	7,200 s (35.31%)	7,200 s (39.15%)	7,200 s (30.54%)	7,200 s (33.18%)	7,200 s (4.67%)	6,225 s (8.08%)	311 s (0.88%)
No Movement	Overtime	Break Decisions	457 s (0.87%)	451 s (0.98%)	252 s (0.90%)	149 s (0.99%)	102 s (0.94%)	90 s (0.92%)	63 s (0.74%)	37 s (0.85%)
		Fixed Breaks	1,714 s (0.93%)	704 s (0.97%)	241 s (0.77%)	190 s (0.95%)	56 s (0.91%)	98 s (0.97%)	44 s (0.78%)	36 s (0.86%)
	No Overtime	Break Decisions	7,200 s (35.60%)	7,200 s (37.62%)	7,200 s (35.12%)	7,200 s (28.48%)	7,200 s (24.75%)	6,162 s (5.49%)	4,347 s (1.07%)	400 s (0.90%)
		Fixed Breaks	0 s (ˆ)	7,200 s (53.33%)	7,200 s (44.88%)	7,200 s (33.45%)	7,200 s (34.79%)	7,200 s (8.45%)	6,612 s (8.86%)	721 s (0.88%)

- indicates a design point where no feasible solution was found after two hours.
ˆ indicates a design point where the model was infeasible due to the inability to satisfy labor constraints.

Table 15: Average time to find a solution within 1% gap and average gap in the four checkpoint experiment.

To compare how various factors impact objective quality, we followed the same analysis as in the two-checkpoint case, calculating the ratio of the objective value achieved for each design point relative to that achieved at the least restrictive design points. Four different design points were used for the most relaxed model, based on arrival patterns and staffing levels. Table 16 shows the achieved ratios and the number of instances not solved to within 1% after two hours (in parentheses). The four design points used for constructing the ratios are also noted in the table. The combination of allowing movement only after breaks, along with overtime and endogenous break decisions, provided the best performance across all model variants.

Four Checkpoint Ratio of Achieved Objective Value to Unrestricted Objective Value (Instances not Solved to 1%)			Lower Bound + 20% Staffing				Lower Bound + 40% Staffing			
			All Distinctions		No Distinctions		All Distinctions		No Distinctions	
			Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks	Synced Arrival Peaks	Shifted Arrival Peaks
Movement Anytime	Overtime	Break Decisions	1.184 (0)	1.062 (0)	1.002* (1**)	1.000* (0)	1.025 (0)	1.028 (0)	1.000* (0)	1.000* (0)
		Fixed Breaks	1.299 (0)	1.106 (0)	1.040 (0)	1.015 (0)	1.053 (0)	1.037 (0)	1.022 (0)	1.003 (0)
	No Overtime	Break Decisions	12.415 (3)	- -	6.195 (4)	2.026 (0)	2.244 (1)	1.408 (0)	1.031 (0)	1.036 (0)
		Fixed Breaks	ˆ ˆ	5.918 (1)	7.744 (3)	2.579 (0)	2.898 (2)	1.586 (0)	1.580 (0)	1.085 (0)
Movement After Breaks	Overtime	Break Decisions	1.160 (0)	1.067 (0)	1.003 (0)	1.003 (0)	1.031 (0)	1.032 (0)	1.002 (0)	0.999 (0)
		Fixed Breaks	1.293 (0)	1.112 (0)	1.042 (0)	1.017 (0)	1.057 (0)	1.039 (0)	1.023 (0)	1.004 (0)
	No Overtime	Break Decisions	12.295 (0)	4.740 (0)	6.260 (0)	2.080 (0)	2.288 (0)	1.596 (0)	1.043 (0)	1.039 (0)
		Fixed Breaks	ˆ ˆ	5.704 (0)	7.975 (0)	2.640 (0)	3.027 (0)	1.752 (0)	1.547 (0)	1.093 (0)
No Movement	Overtime	Break Decisions	1.197 (0)	1.079 (0)	1.006 (0)	1.010 (0)	1.032 (0)	1.038 (0)	1.003 (0)	1.002 (0)
		Fixed Breaks	1.384 (0)	1.125 (0)	1.042 (0)	1.020 (0)	1.059 (0)	1.044 (0)	1.020 (0)	1.005 (0)
	No Overtime	Break Decisions	12.547 (0)	6.109 (0)	7.109 (0)	2.495 (0)	2.370 (0)	1.798 (0)	1.081 (0)	1.053 (0)
		Fixed Breaks	ˆ ˆ	8.600 (0)	9.125 (0)	2.908 (0)	3.263 (0)	2.022 (0)	1.561 (0)	1.113 (0)

* indicates a design point used as the "unrestricted" design point for comparisons.
** indicates a design point where one instance did not solve to a 1% gap in two hours and additional time was allowed to calculate the ratios.
\- indicates a design point where no feasible solution was found after two hours.
ˆ indicates a design point where the model was infeasible due to the inability to satisfy labor constraints.

Table 16: Average ratio of the achieved objective value to the best objective value for the least restrictive design point (same arrivals and staffing level) and number of instances not solved to 1% with four checkpoints.

As model size increases, some design points struggle to find high-quality solutions within a two-hour time limit. Implementing fixed break schedules can reduce solution time, sometimes with minimal impact on solution quality. This effect is evident in Tables 15 and 16 for some design points that allow overtime. Across both the two and four checkpoint experiments, the findings indicate that overtime is essential for achieving both timely and high-quality solutions.

The solution quality when allowing worker movement at any time, compared to movement only after breaks, shows only a marginal difference. However, for instances that reach a 1% optimality gap, the solution time is significantly shorter when movement is restricted to after breaks or not allowed at all. Consistent with the two-checkpoint case, these results suggest that the timing of movement decisions should be carefully considered, as strategically allowing movement can lead to quicker identification of high-quality solutions.

6 Conclusions and Future Work

This paper presents a model for optimizing a service system with disjoint service areas and shared resources, inspired by the operation of security checkpoints in U.S. airports. The case study illustrates the potential impact that optimization tools can have on the operation of government agencies.

In the modeled service system, each service area contains several disjoint service nodes that jobs must visit. Resources are classified by shift and certifications, with the ability to have scheduled time off during a shift or optional overtime. The model incorporates various domain-specific constraints and generates worker schedules, including breaks and work locations. An embedded queue throughput model accounts for finite buffers and provides estimates of throughput by period, as well as customer wait times.

A case study was presented to demonstrate the applicability of the model using realistic data from an actual airport security screening process. Two sets of experiments were conducted, each with modifications to the input data from the case study. The two-checkpoint scenario, showed that the model could be solved in a reasonable amount of time across all design points. The four-checkpoint scenario proved more difficult to solve under all design points. In both experiments, computational time was found to be reasonable for practical applications but sensitive to the number of features included in the model.

The model presented here can be generalized to any system with sequential service steps. With minor modifications to the job routing, the model can also accommodate rework processes and job splitting, where jobs are not routed in a strictly linear fashion.

As observed in both the case study and the multi-checkpoint experiments, the model can sometimes be solved faster when movement is restricted. Allowing worker moves only following breaks provides most of the value of movement flexibility without excessive computational requirement. Another key finding is that solution time depends on the slack in resource availability. Under tight labor conditions, modest overtime expenditures can substantially reduce wait times. This highlights the importance of incorporating overtime decisions, which adds realism by providing flexible workforce in case of external disruptions (e.g., delayed incoming flights). The model enables exploration of the trade-off between additional workers and the use of overtime to maintain a given customer service level while minimizing costs. While overtime incurs additional costs, these are generally offset by reductions in queue-related costs. Although ignoring workforce distinctions can improve solution times, it frequently leads to infeasible or impractical staffing plans. From an operational standpoint, the most balanced model–offering both realism and tractable computation–includes overtime, movement after breaks, workforce distinctions, and break decisions. This setup performs robustly across various arrival patterns and staffing levels. Although it is possible to modify the proposed model to generate candidate break schedules as an output, this task is left for future work.

Future work could explore more aspects of the stochastic nature of arrivals at checkpoints. In the context of the TSA study, this approach enables more realistic modeling of passenger behavior across different traveler categories (e.g., leisure vs. business, families vs. solo travelers) by incorporating uncertainty in their arrival times at the checkpoint before departure. Algorithmic extensions might also be developed to handle larger instances with more screening locations. Reformulating the model and network construction to allow for the implicit consideration of breaks could potentially yield high-quality solutions more quickly. While the selection of basic schedules (start times, lengths, and quantities) falls outside the scope of this model, as those decisions are made in a semiannual bidding process for the case study, the model can still be used in the bid proposal process to define the appropriate set of schedules and their quantities based on forecasted arrivals.

Acknowledgements

The authors would like to express their appreciation to the TSA analysts, particularly Nicholas Kobilansky and Matthew Howley, for their assistance in understanding the complexity and operational parameters of airport security screening checkpoints. Additionally, the authors would like to thank Jian Jiao, a PhD student at the University of Florida, for his contributions to the development of the baggage staffing model. The authors gratefully acknowledge the constructive and insightful feedback provided by the Main Guest Editor, Stavros Zenios, and the two anonymous reviewers.

Compliance with Ethical Standards

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-08-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. Authors declare that they have no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Al-Yakoob, S. M. and Sherali, H. D. (2007). Mixed-integer programming models for an employee scheduling problem with multiple shifts and work locations. *Annals of Operations Research*, 155(1):119–142.

Avramidis, A. N., Chan, W., Gendreau, M., L'Ecuyer, P., and Pisacane, O. (2010). Optimizing daily agent scheduling in a multiskill call center. *European Journal of Operational Research*, 200(3):822–832.

Bartholdi, J. J. (1981). A guaranteed-accuracy round-off algorithm for cyclic scheduling and set covering. *Operations Research*, 29(3):501–510.

Bechtold, S. E. and Jacobs, L. W. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. *Management Science*, 36(11):1339–1351.

Begur, S. V., Miller, D. M., and Weaver, J. R. (1997). An integrated spatial dss for scheduling and routing home-health-care nurses. *Interfaces*, 27(4):35–48.

Brunner, J. O., Bard, J. F., and Kolisch, R. (2010). Midterm scheduling of physicians with flexible shifts using branch and price. *IIE Transactions*, 43(2):84–109.

Castillo-Salazar, J. A., Landa-Silva, D., and Qu, R. (2014). Workforce scheduling and routing problems: literature survey and computational study. *Annals of Operations Research*, 239(1):39–67.

Cordeau, J.-F., Laporte, G., Pasin, F., and Ropke, S. (2010). Scheduling technicians and tasks in a telecommunications company. *Journal of Scheduling*, 13(4):393–409.

Dantzig, G. B. (1954). Letter to the editor - a comment on edie's "traffic delays at toll booths". *Journal of the Operations Research Society of America*, 2(3):339–341.

Defraeye, M. and Van Nieuwenhuyse, I. (2016). Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, 58:4–25.

Eitzen, G., Panton, D., and Mills, G. (2004). Multi-skilled workforce optimisation. *Annals of Operations Research*, 127(1–4):359–372.

Emmons, H. and Burns, R. N. (1991). Off-day scheduling with hierarchical worker categories. *Operations Research*, 39(3):484–495.

Erhard, M., Schoenfelder, J., Fügener, A., and Brunner, J. O. (2018). State of the art in physician scheduling. *European Journal of Operational Research*, 265(1):1–18.

Feldman, Z., Mandelbaum, A., Massey, W. A., and Whitt, W. (2008). Staffing of time-varying queues to achieve time-stable performance. *Management Science*, 54(2):324–338.

Green, L. V., Kolesar, P. J., and Whitt, W. (2007). Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management*, 16(1):13–39.

Hanumantha, G. J., Arici, B. T., Sefair, J. A., and Askin, R. (2020). Demand prediction and dynamic workforce allocation to improve airport screening operations. *IISE transactions*, 52(12):1324–1342.

He, B., Liu, Y., and Whitt, W. (2016). Staffing a service system with non-poisson non-stationary arrivals. *Probability in the Engineering and Informational Sciences*, 30(4):593–621.

Heemskerk, M., Mandjes, M., and Mathijsen, B. (2022). Staffing for many-server systems facing non-standard arrival processes. *European Journal of Operational Research*, 296(3):900–913.

Hung, R. (1994). Single-shift off-day scheduling of a hierarchical workforce with variable demands. *European Journal of Operational Research*, 78(1):49–57.

Izady, N. and Worthington, D. (2012). Setting staffing requirements for time dependent queueing networks: The case of accident and emergency departments. *European Journal of Operational Research*, 219(3):531–540.

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

Jiao, J., Arici, B. T., Sefair, J. A., and Askin, R. G. (2024). Workforce and service optimization in a multistage service system with dynamic arrivals and operational constraints. Technical Report. University of Florida.

Kolesar, P. J., Rider, K. L., Crabill, T. B., and Walker, W. E. (1975). A queuing-linear programming approach to scheduling police patrol cars. *Operations Research*, 23(6):1045–1062.

Martonosi, S. E. (2011). Dynamic server allocation at parallel queues. *IIE Transactions*, 43(12):863–877.

Mulvey, J. M. (1979). Strategies in modeling: A personnel scheduling example. *Interfaces*, 9(3):66–77.

Narasimhan, R. (1997). An algorithm for single shift scheduling of hierarchical workforce. *European Journal of Operational Research*, 96(1):113–121.

Rong, A. (2010). Monthly tour scheduling models with mixed skills considering weekend off requirements. *Computers & Industrial Engineering*, 59(2):334–343.

Stolletz, R. (2010). Operational workforce planning for check-in counters at airports. *Transportation Research Part E: Logistics and Transportation Review*, 46(3):414–425.

Stolletz, R. and Zamorano, E. (2014). A rolling planning horizon heuristic for scheduling agents with different qualifications. *Transportation Research Part E: Logistics and Transportation Review*, 68:39–52.

TSA (2023). TSA 2023 year in review. 2023.

USDOT (2016). Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis. Technical report, U.S. Department of Transportation. `https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-valuation-travel-time-economic`. Last Accessed March 2025.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., and De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367–385.

Wallace, R. B. and Whitt, W. (2005). A staffing algorithm for call centers with skill-based routing. *Manufacturing & Service Operations Management*, 7(4):276–294.

Wu, Z., Liu, R., and Pan, E. (2023). Server routing-scheduling problem in distributed queueing system with time-varying demand and queue length control. *Transportation Science*, 57(5):1209–1230.

Zamorano, E., Becker, A., and Stolletz, R. (2017). Task assignment with start time-dependent processing times for personnel at check-in counters. *Journal of Scheduling*, 21(1):93–109.

A Modified Set Covering Model for Minimal Staffing Needs

A modified set-covering model was used to determine staffing needs as input to the workforce allocation model. While this model is not considered optimal for the system, it serves as a heuristic for estimating staffing needs when no other method is available. The modified set-covering model presented here was used to establish a lower bound on staffing needs given that the real staffing levels cannot be disclosed.

The sets, subsets, decision variables, and parameters for this set-covering model are entirely independent of any references made elsewhere in this paper. Below are the definitions of the sets, subsets, decision variables, and parameters used in the modified set-covering model.

Set/Subset	Description	Data Used
S	All shifts	One shift of every length starting on the hour
K	Classifications available for workers	All classifications shown in Table 4b
T	Planning periods	96 different 15 minute intervals over 24 hours
L	All shift lengths	10.5, 8.5, 7.5, 6.5, 5.5, 5, 4.5 hour-long shifts
S_ℓ	Subset of shifts of length $\ell \in L$	

Table 17: Sets and Subsets

Parameter	Description	Data Used
c_s	Cost of hiring a worker for shift $s \in S$	Goal to minimze number of workers per shift length. Shifts starting before 6am or after 6pm cost 1.1 workers. All other shifts cost 1 worker.
π_k	Cost multiplier for certification $k \in K$	A 5% penalty for workers qualified to work in either baggage or a checkpoint. No penalty otherwise.
a_{st}	Indicates if shift $s \in S$ covers time $t \in T$	Set to 1 if the shift covers the period, 0 otherwise.
d_{kt}	Worker demand of certification $k \in K$ at $t \in T$	Estimated beforehand as the minimum of either the number of workers required to operate enough screening lanes to clear all passengers in 15 minutes or the maximum number of screening lanes available.
p_ℓ	Percentage of a shift type required to be length $\ell \in L$	See Table 7.
ϵ	Amount of allowable slack on the shift distribution percentages	Slack of 0.01%

Table 18: Parameters

Decision Variable	Description
$x_s k$	Integer number of workers hired with certification $k \in K$ on shift $s \in S$

Table 19: Decision Variables

The modified set cover model is given in (26)-(29). The objective function (26) minimizes the total cost of labor used. Constraints (27) ensure that enough workers with certification $k \in K$ are hired at time $t \in T$

to cover the respective worker demand. Constraint (28) forces the model to select shifts according to a predetermined distribution of shift lengths within some tolerance ϵ.

$$\min \sum_{s \in S} \sum_{k \in K} \pi_k c_s x_{sk} \tag{26}$$

$$\sum_{s \in S} a_{st} x_{sk} \geq d_{kt}, \quad \forall k \in K, t \in T \tag{27}$$

$$\sum_{s \in S_\ell} \sum_{k \in K} x_{sk} \geq (p_\ell - \epsilon) \sum_{s \in S} \sum_{k \in K} x_{sk}, \quad \forall \ell \in L \tag{28}$$

$$x_{sk} \in \mathbb{Z}^+ \tag{29}$$

To determine the number of TSOs required for a checkpoint (the d-parameters), we use a heuristic that aims to clear all passengers in the queue and all new arrivals within 15 minutes, assuming all passengers are ready at the beginning of the period. For each period, a configuration is selected that meets this goal. If there is insufficient capacity to process everyone within 15 minutes, the configuration with the largest capacity is chosen, and any unprocessed passengers are carried over to the next period, where the process is repeated. The required labor for the checkpoint is determined by the labor needed to operate the configuration prescribed by the heuristic. Appendix B describes the method used for determining the required labor for the baggage area. Both of these values are then used as input for the set-cover model.

This method does not account for break scheduling and assumes that either overtime assignments in the main model or the addition of extra labor can compensate for the loss of labor caused by breaks. Additionally, the method assumes that travel between checkpoints is negligible. These factors significantly affect real-world system behavior, which is why a more complex model is required for both staffing and workforce allocation decisions.

B Calculating Required Labor for the Baggage Area

In addition to checkpoint screening, baggage screening is a key part of TSA operations. For the case study, TSA requires the allocation of Transportation Security Officers (TSOs) to baggage processing units (BPUs) throughout the day. These TSOs must hold a special certification for baggage screening, distinguishing them from those assigned to checkpoint screening.

Passengers' checked baggage is routed to the BPUs after drop-off at airline counters or curb stations. The TSA baggage screening begins when the passengers' checked baggage arrives at the processing unit at a bag processing center. All checked bags undergo a security inspection, and some may require secondary inspection. The processing rate for secondary screening is distinct, and the processing capacity of these additional screening units is considered unlimited.

Modeling the operation of the baggage processing center requires a set of assumptions. We assume that all baggage from various checkpoints within the same terminal is directed to one or more baggage processing centers within that terminal. Each center operates under a First-In, First-Out (FIFO) policy. When the volume of baggage arrivals exceeds the facility's maximum processing capacity, the excess baggage is queued and processed in the next time interval. Operational hours are also a critical consideration; baggage arriving before the start of daily operations is processed during the first working interval. The modeling approach allows for an unlimited number of screening units, ensuring that baggage requiring secondary screening is handled within the same period. The planning horizon for the baggage screening area aligns with the periods used in the workforce planning model.

Table 20 provides a detailed description of the parameters and variables in the model. With these, we formulate a mathematical model to calculate the number of TSOs needed. The volume of baggage arriving in each 15-minute time interval is calculated based on the passenger arrival data. The expected number of bags at baggage processing center j at time t is given by

$$N_{jt}^{b} = q_j \sum_{i \in I} \sum_{t' \in T(t)} \Delta_{i,t'-t} \beta_{it'} p_{it'}, \tag{30}$$

where $T(t) = \{t-L, \cdots, t-1, t, t+1, \cdots, t+U\}$. Set $T(t)$ represents the possible periods during which the baggage can arrive at the baggage processing units, ranging from L periods before to U periods after the passenger arrives at the checkpoint (period t). The expected number of bags waiting (unprocessed) at the end of time t is given

$$N_{jt}^{w} = \left(N_{jt}^{b} + N_{j,t-1}^{w} - P\mu_j^{b} \right)^{+}, \tag{31}$$

where $(\cdot)^{+} = \min\{\cdot, 0\}$. The expected number of rechecked bags (secondary inspection) at time t is given by

$$N_{jt}^{r} = k \min\left\{ N_{jt}^{b} + N_{j,t-1}^{w}, P\mu_j^{b} \right\}. \tag{32}$$

Equations (30)-(32) help determine the required number and type of baggage processing units, as well as the number of TSOs needed. The number of TSOs required to operate the equipment and perform the screening procedures at baggage processing center j at time t is given by

$$d_{jt} = h \min\left\{\left\lceil \frac{N^b_{jt} + N^w_{j,t-1}}{\mu^b_j} \right\rceil, P\right\} + h \left\lceil \frac{N^r_{jt}}{\mu^r_j} \right\rceil, \tag{33}$$

where the first term captures the TSOs required for primary inspection, and the second term accounts for those needed for secondary inspection.

A critical requirement of the system is that all baggage must be processed within no more than two periods (30 minutes) after arrival at the baggage facility. This ensures that bags can be loaded on time before flight departures. We enforce this requirement by ensuring that the maximum number of bags left in the queue never exceeds the processing capacity of the facility at any time. To meet the two-period processing requirement, it must hold that $N^w_{jt} \leq P\mu^b, \forall t \in T$, meaning that all bags waiting will be processed within one period. If this condition is not met, some bags already waiting at time t will not be processed in the next period, resulting in a processing delay beyond the stipulated two-period time frame.

Symbol	Definition
p_{it}	Expected passenger arrivals at checkpoint i at time t
β_{it}	Expected number of bags per passenger arriving at checkpoint i at time t
d_{jt}	Number of TSOs needed to process bags at bag processing center j at time t
q_j	Percentage of bags processed at bag processing center j
L	Minimum number of periods between the arrival of the bag of screening and the passenger of the checkpoint
U	Maximum number of periods between the arrival of the bag of screening and the passenger of the checkpoint
Δ_{il}	Percentage of bags arriving within l periods (before or after) with respect to the passenger arrival at checkpoint i
t	Planning period
t_0	Start time of baggage screening
P	Number of available bag processing units
k	Expected percentage of recheck bags
h	Number of TSOs needed to operate a bag processing unit
μ^b_j	Processing rate (bags) per bag processing unit per period at primary screening of bag processing center j
μ^r_j	Processing rate (bags) per bag processing units per hour secondary inspection at bag processing center j
N^b_{jt}	Expected number of bags at bag processing center j at time t
N^r_{jt}	Expected number of recheck bags at bag processing center j at time t
N^w_{jt}	Expected number of bags waiting at bag processing center j at the end of time t

Table 20: Parameters

We illustrate our calculations using passenger arrival data from a busy day at Miami International Airport. This data, spanning checkpoints North 1 to North 4, forms the basis for subsequent calculations. Additionally, parameter data relevant to baggage processing is extracted from information available for the

North Inline East Bag Zone.

Figure 9 presents various examples of baggage earliness arrival curves, passenger arrival predictions, and the corresponding calculations for bag arrivals and the number of TSOs needed. We depict four distinct baggage earliness arrival curves for analysis. These curves capture the percentage of bags arriving at the baggage processing unit as early as L periods (60 minutes) before the passenger's arrival at the checkpoint and up to U periods (30 minutes) after. Bag arrivals later than the passenger's arrival at the checkpoint account for late drop-offs and the corresponding travel time to the screening facility. The bimodal curve reflects a situation where passengers drop their bags either very early or late relative to their arrival at the checkpoint. The left-skewed curve indicates a tendency for passengers to drop their bags earlier, while the right-skewed curve captures a propensity for later arrivals. The skewed bimodal curve suggests that most bags are being dropped 30 minutes before the passenger's arrival at the checkpoint, while a subset is dropped later. These scenarios illustrate the relationship between the timing of bag arrivals and the staffing levels required to accommodate fluctuating demands imposed by different baggage drop-off behaviors.

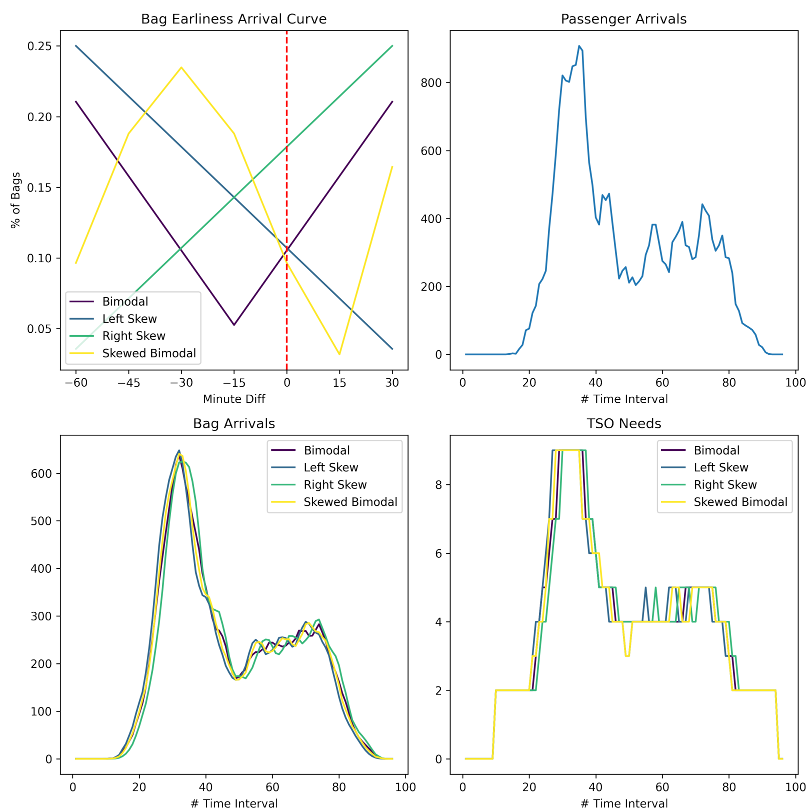
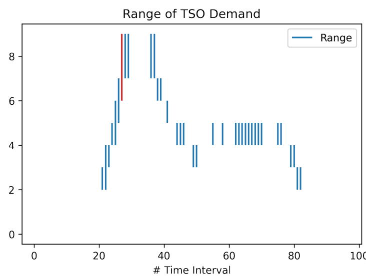
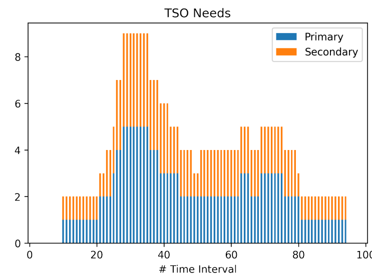


Figure 9: Analysis of Predictions & Staffing Needs by Bag Earliness Arrival Curve

Figure 10a presents the maximum difference in the number of required TSO personnel across different baggage earliness arrival curves over various time intervals, typically not exceeding 2 TSOs. In most time periods, the difference in the number of TSOs required for screening is less than or equal to 1. The maximum difference of 3 occurs just before the peak. This figure illustrates the sensitivity of TSO staffing calculations to different passenger behaviors related to baggage drop-off. Figure 10b presents the forecasted number of TSOs required for primary and secondary screening areas over various time intervals using the skewed bimodal earliness arrival curve.



(a) TSO Demand Range



(b) Stacked TSO Needs (Skewed Bimodal)

Figure 10: TSO demand range and needs.