On the Capacity of Highway Checkpoints: Models for Unconventional Configurations

Weihua Gu *a, Michael J. Cassidy a, Yuwei Li b

a 416 McLaughlin Hall, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, United States

b Department of Transportation Planning, Shanghai Maritime University, 1550 Haigang Boulevard, Shanghai 201306, China

Abstract

Widening a highway toll plaza, border crossing or other checkpoint by adding travel lanes and vehicle-processing stations in parallel fashion can be expensive, and sometimes even infeasible. Rather than expanding laterally, checkpoint capacities can be increased by branching or staggering the vehicle-processing stations, or by placing them in tandem. Analytical models are formulated to estimate the vehicle-processing capacities achievable via these layouts. The models indicate that tandem designs tend to produce the highest capacities among the three alternatives. Other insights are unveiled as well.

Keywords: checkpoint capacity, toll plaza, border crossing, tandem queueing system

1 Introduction

Border crossings in a highway network are often sources of severe congestion. For example, the vehicle delays at crossings between the U.S. and Canada often average between 30~60 minutes. The economic losses that result are estimated to be in the billions of dollars per year (Ontario Chamber of Commerce, 2005). The negative social and environmental impacts only add to these costs. Hence, the benefits achieved by increasing the vehicle-processing capacities of border crossings can be significant. The same can be said of toll plazas and other types of highway checkpoints.

Checkpoint capacities can be improved by adding service booths (i.e., vehicle-processing stations), together with the attendant travel lanes, in parallel fashion. However, the extra lateral space required for this kind of expansion can be expensive, and sometimes impossible to acquire.

In light of this, the present paper examines three alternative layouts for service booths: tandem, staggered and branch. These are each illustrated in Figs. 1a–c. For the tandem layout (Fig. 1a), a vehicle need only stop to be served at one of two paired booths. Visual inspection of all three figures reveals how the three alternatives allow for the addition of service booths by claiming additional space in the longitudinal direction (only), along the road itself. This space, rather than lateral space, tends to be more readily available at checkpoint locations. Therefore, the alternatives are often relatively inexpensive to construct.

* Corresponding author. Tel.: +1 (510) 931-6646; fax: +1 (510) 643-8919.
E-mail address: weihuagu@berkeley.edu.
Given this advantage, the alternatives have rightfully enjoyed some degree of attention in the literature. Yet, the attention has been limited. For example, mention is made of staggered- and of branch-booth designs in Rubenstein (1983) and Schaufler (1997), respectively. Yet to our knowledge, the literature does not include models for estimating the capacities of these designs.

A capacity model has previously been formulated for tandem-booth layouts (Hall and Daganzo, 1983). It indicates that adding a second booth in tandem along each travel lane can increase vehicle-processing rates at toll plazas, and that by inserting vehicle-storage buffers between each set of tandem booths, capacities can increase further still. However, this earlier model adopts a simplified assumption regarding the vehicle service process: each of the two booths in a tandem set presumably serves equally-sized vehicle batches. This idealization fails to unveil the full potential of tandem layouts.

To see this, we formulate analytical models of capacity for tandem layouts in Section 2, and compare our capacity estimates with those from the earlier tandem-booth model in an appendix. Capacity models for the other two alternative layouts are formulated in Section 3. Comparisons of model outputs show that tandem layouts can provide the highest capacities and require the least longitudinal space; see Section 4.
Other considerations, including those that favor the other two alternatives are discussed as well. Findings are summarized in Section 5.

2 Tandem-Booth Capacities

We define a checkpoint’s capacity to be the maximum rate that vehicles can discharge from it. We therefore assume that a vehicle queue is always present at the checkpoint’s entrance.

Consider a set of $c \geq 2$ service booths placed in series, where Fig. 1a shows two sets of booths, each with $c = 2$. Absent buffers for storing queued vehicles between each consecutive pair of tandem booths, the vehicle-processing capacity of a tandem set will almost always be less than $c$ times the capacity of a single booth. This is because tandem service creates two types of blocking effects. First, a vehicle being served at an upstream booth in the tandem set can temporarily block queued vehicles from reaching any vacant booth(s) downstream. Second, a vehicle being served at a downstream booth can temporarily block a vehicle that is ready to depart from an upstream one. Both effects squander capacity, and are quantified in Section 2.1.

These negative effects can be diminished by inserting a vehicle-holding buffer between each consecutive pair of tandem booths. The added capacity gained from a buffer with space for holding a single vehicle is explored in Section 2.2. The impacts of multi-vehicle storage buffers are examined in Section 2.3.

2.1 Tandem Layout Without a Buffer

We assume that the service times for vehicles at an arbitrary booth, $S_i (i = 1, 2, ..., c$, and recall that $c$ is the number of tandem booths in a travel lane), are independent and identically distributed (i.i.d.) random variables. We normalize the mean service time to be 1 and define the normalized capacity of the set of tandem booths to be $Q_{TDM,c}$.

We define a regenerative point as the instant in time when all booths in a tandem set have been emptied of vehicles. (Of course, the booths will be filled by vehicles from the queue immediately thereafter.) A cycle is defined as the interval between two successive regenerative points. The capacity of a set of tandem booths therefore equals the average number of vehicles served in a cycle, $\bar{N}$, divided by the average duration of the cycle, $\bar{T}$:

$$Q_{TDM,c} = \lim_{t \to \infty} \frac{\text{number of vehicles served in } t}{t} = \frac{E[\text{number of vehicles served in a cycle}]}{E[\text{cycle length}]} = \frac{\bar{N}}{\bar{T}}. \quad (1)$$

When vehicle storage space between the tandem booths, $b$, is zero, meaning that there is no buffer, it has been shown in Hall and Daganzo (1983), and in Gu et al. (2011) that

$\bar{N} = c$, and

$$\bar{T} = \int_{t=0}^{\infty} (1 - (F_S(t))^c) dt,$$
where \( F_S(t) \) is the cumulative distribution function of vehicle service time at a booth. Thus,

\[
Q_{TDM,c}(b = 0) = \frac{c}{\int_{t=0}^{1} (1 - (F_S(t))^c)\,dt}.
\]

(2)

The dashed curves in Fig. 2 present \( Q_{TDM,c}(0) \) as a function of the coefficient of variation in service time, \( C_S \), for \( c = 2-4 \).\(^1\) The curves were constructed for the case of gamma-distributed service time.\(^2\) The solid horizontal line in Fig. 2 presents the normalized capacity for a single booth, \( Q_{SGL} \), and is included for comparison.

![Normalized capacity](image)

Fig. 2 \( Q_{TDM,c}(0) \) for \( c = 2-4 \) and \( Q_{SGL} \) versus \( C_S \)

Inspection of the figure reveals that \( Q_{TDM,c}(0) = c \cdot Q_{SGL} \) only when \( C_S = 0 \). In this idealized case, \( c \) vehicles can always be served in batch mode without blocking each other. Otherwise, \( C_S \) has a negative effect on \( Q_{TDM,c}(0) \). We see that when \( C_S = 1 \), blocking reduces total capacity by 33\%, 45\%, and 52\% for \( c = 2-4 \), respectively, as compared against the case of \( C_S = 0 \).

\(^1\) The coefficient of variation is defined as the ratio of the standard deviation to the mean.

\(^2\) Previous studies (Correa et al., 2004; Kim, 2009; Han et al., 2012; Boronico and Siegel, 1998) used various types of distributions to model vehicle dwell times at toll booths and border crossings. In this paper we choose to use the gamma distribution because it is more general, i.e., the \( C_S \) of a gamma distribution can be set to a range of values. We note that another general distribution, namely the Weibull distribution, has often been used in capacity analysis in traffic engineering. Yet the Weibull distribution has primarily been used to model the perception and reaction times of car drivers (Glassco, 1999; Ci et al., 2009; Liang et al., 2011). There is no clear evidence that Weibull is superior to gamma in modeling vehicle service times at toll booths or border crossings (Correa et al., 2004). Moreover, our analysis shows that other assumed service-time distributions, including the Weibull distribution, yield results that are similar to those that will be presented here.
2.2 Tandem Layout with a Single Vehicle Buffer

The blocking effects and attendant capacity losses can be reduced by inserting a vehicle-storage buffer between a pair of tandem booths. For simplicity, we will from now on assume that \( c = 2 \). This section examines the case where the inserted buffer can hold only a single queued vehicle, i.e. \( b = 1 \). This buffer can temporarily store: i) a vehicle that has finished being served by the upstream booth while the downstream booth is still busy, or ii) a vehicle that is waiting to be served by the downstream booth.

Recall that a vehicle queue is assumed to be always present at the upstream booth. We similarly define a regenerative point as the instant in time when both booths and the buffer have all been emptied of vehicles. Cycles, \( \bar{N} \), and \( T \) are defined accordingly. And Equation (1) still holds.

We use \( Q_{TDM} \) instead of \( Q_{TDM,c} \) because \( c \) is fixed as 2. This capacity depends in part on what we shall designate as \( p \), the number of “prefilled” vehicles that enter the buffer at the very start of each cycle and then initially wait there to be served by the downstream booth. When \( b = 1 \), \( p = 0 \) or 1. These two possibilities for \( p \) are separately examined below.

For the case of \( p = 0 \), the first vehicle to enter a buffer during that cycle is one that has already been served at the upstream booth and is waiting in the buffer to depart the checkpoint. Fig. 3 presents the transitions in the state of the system (the vehicle occupancies at the booths and in the buffer) during a cycle when \( p = 0 \). The upstream and downstream booths in a tandem set are depicted by the two solid circles in each rectangular box, and the buffer is depicted by the intermediate dashed circle. The occupation of any one of these elements by a vehicle is denoted by a number inside the corresponding circle: the numbers ascend with each new vehicle that enters the system. A negative sign following a number indicates that the vehicle has not completed its service; while a positive sign means that the vehicle is ready to depart the system but is presently blocked from doing so. If the Boolean expression in a diamond is true, the system state goes through the arrow marked “Y”, and otherwise through the arrow marked “N”. The \( S_i \) \((i = 1, 2, 3)\) now denotes the service time of vehicle \( i \).

Fig. 3 reveals that when \( S_1 < S_2 \), two vehicles are served in the cycle; otherwise, the number is three. Thus,

---

3 When \( c > 2 \), a buffer can be inserted between each consecutive pair of booths. However in this case, the control rules for the use of the buffers are so complicated that toll violations may become virtually inevitable. Thus configurations of \( c > 2 \) are probably impractical when buffers are to be used.

4 A sought-after valve of \( p \) can be realized by controlling motorists via information signs, light curtains, or treadles (Rubenstein, 1983).
\[ \bar{N} = 2 \cdot Pr\{S_1 < S_2\} + 3 \cdot Pr\{S_1 > S_2\} = 2.5, \]

since the \( S_i \) are i.i.d., such that \( Pr\{S_1 < S_2\} = Pr\{S_1 > S_2\} = 0.5 \).

Also, \( \bar{T} = \bar{N} \cdot \bar{S} - \bar{T}_S = \bar{N} - \bar{T}_S \), where \( T_S \) is the temporal portion of the cycle during which both booths are serving vehicles. (Recall that the average vehicle service time, \( \bar{S} \), is normalized to be 1.) From Fig. 3, we have: \( T_S = S_1 \) if \( S_1 < S_2 \); \( T_S = S_1 \) if \( S_2 < S_1 < S_2 + S_3 \); and \( T_S = S_2 + S_3 \) if \( S_1 > S_2 + S_3 \). Thus,

\[ T_S = \min\{S_1, S_2 + S_3\}. \]

The normalized capacity of the tandem booths is therefore:

\[ Q_{TDM}(b = 1, p = 0) = \frac{\bar{N}}{\bar{T}} = \frac{\bar{N}}{\bar{N} - \bar{T}_S} = \frac{1}{1 - \frac{\bar{T}_S}{\bar{N}}} = \frac{1}{1 - E[\min(S_1, S_2 + S_3)]/2.5}. \quad (3) \]

When \( p = 1 \), the start of the cycle is characterized by a vehicle that is waiting in the buffer to be served by the downstream booth. Three vehicles therefore collectively occupy the booths and buffer at the cycle’s start. The transition diagram for this case is shown in Fig. 4.

---

**Fig. 4** The transition diagram of states in a cycle for \( b = 1, p = 1 \)

From Fig. 4, we have:

\[ \bar{N} = 3 + Pr\{S_1 + S_2 > S_3\}. \]

noting that the fourth vehicle will be served in the cycle if and only if \( S_1 + S_2 > S_3 \); and

\[ \bar{T}_S = E[\min(S_1 + S_2, S_2 + S_3, S_3 + S_4)]. \]

The reader can verify the above equation using an argument similar to what was used for the case of \( p = 0 \).

Thus, the normalized capacity in this case is:
Knowing the distribution of vehicle service time, the expectations and probabilities explicit in (3) and (4) can be computed by numerically integrating the probability density functions, or by other approximation methods such as Monte Carlo simulation.

Fig. 5 presents the normalized capacities (for \( p = 0 \) and \( p = 1 \)) as functions of \( C_S \) and for gamma-distributed service time. The \( Q_{SGL} \) and \( Q_{TDM,2}(0) \) are also displayed for comparison purposes. The figure shows that adding a buffer does not always improve capacity. Note that when \( p = 0 \) and \( C_S < 0.4 \), a buffer diminishes capacity; i.e., \( Q_{TDM}(1,0) < Q_{TDM,2}(0) \).

Fig. 3 reveals why this is the case: when \( S_1 > S_2 \), vehicle 2 moves into the buffer, such that vehicle 3 enters the upstream booth and for a time \( S_3 \) thereafter blocks queued vehicles from reaching the downstream booth. If \( S_1 \) and \( S_2 \) have similar values (i.e., if \( C_S \) is small), this blockage can persist for greater duration and cause a more substantial reduction in capacity.

This undesirable state of affairs can be avoided if \( p = 1 \). Fig. 5 reveals that \( Q_{TDM}(1,1) \geq \max \{Q_{TDM,2}(0), Q_{TDM}(1,0)\} \) for \( C_S \in [0,1] \). This finding speaks in favor of applying limited control to the tandem booths so that the buffer is always prefilled by an un-served vehicle at the start of each cycle.

Fig. 5 also reveals that the capacity improvement that can be achieved by a buffer, \( Q_{TDM}(1,1) - Q_{TDM,2}(0) \), grows as \( C_S \) increases. Previous studies indicate that the \( C_S \) at border crossings typically lies between 0.5 and 0.8 (Lin and Lin, 2001; Booz Allen & Hamilton, 2000). In these cases, Fig. 5 shows that the added capacity brought to a tandem layout by a buffer is about 8~10\% of a single booth’s capacity. This kind of increase can greatly relieve checkpoint congestion.
One could further opt to control the use of the buffer in more general fashion. With this option, control would be applied (e.g. via a changeable message sign or traffic signal) whenever a vehicle in the upstream booth finishes its service when the buffer is empty but the downstream booth is still busy. The upstream vehicle would be allowed to enter the buffer (and thereby other vehicles would be potentially blocked from entering the downstream booth) only if the remaining service time estimated for the downstream booth is sufficiently long. This remaining time could be readily estimated using: i) the elapsed service time downstream when the upstream vehicle has finished its service; and ii) the service time distribution.

However, we find that this general form of control has only a modest effect on capacity, even for buffer size, $b > 1$. The evidence is furnished in Appendix A of this manuscript. Thus, the limited form of control to achieve $p = 1$ seems to be sufficient and is simple to implement.

### 2.3 Tandem Layout with Multi-Vehicle Buffer

Intuitively, a greater buffer size, $b$, can add even further capacity to a tandem layout. The capacity actually obtained will depend in part on $p$, the number of prefilled vehicles in the enlarged buffer at the start of each cycle.

For a tandem layout with a buffer size, $b$, and a prefilled number, $p$, we have the following results:

$$
\bar{N}(b, p) = p + 2 + \sum_{i=p+2}^{b+p+2} \Pr\{\sum_{j=1}^{i} S_{j} > \sum_{j=p+2}^{i} S_{j}\} + \\
\sum_{i=b+3}^{b+p+1} \Pr\{\sum_{j=1}^{i} S_{j} > \max\{\sum_{j=p+1}^{i} S_{j}, \sum_{j=b+3}^{i} S_{j} + S_{1}, \sum_{j=b+4}^{i} S_{j} + \sum_{j=1}^{2} S_{j}, \ldots, S_{i} + \sum_{j=1}^{i-b+2} S_{j}\}\} ; \quad (5)
$$

$$
\bar{T}_{S}(b, p) = E\left[\min\{\sum_{j=1}^{i} S_{j}, \sum_{j=2}^{i} S_{j}, \sum_{j=3}^{i} S_{j}, \ldots, \sum_{j=p+2}^{i} S_{j}\}\right] . \quad (6)
$$

Proofs of these results are furnished in Appendix B. With the outcomes from (5) and (6), normalized capacities can be computed as $Q_{TDM}(b, p) = \frac{1}{1 - \bar{T}_{S}(b, p)/\bar{N}(b, p)}$.

Fig. 6a presents these capacities for the case of $b = 2$ and for gamma-distributed service time. Curves are shown for all possible values of $p$. The figure reveals that the $p$ that makes capacity maximal depends upon $C_{S}$. As examples, we see that capacity is maximized with $p = 2$ when $C_{S} \leq 0.3$, and with $p = 1$ when $0.3 \leq C_{S} < 1$. Note the outer envelope of all curves, shaded light gray in the figure and denoted $Q_{TDM}(2)$, since $b = 2$. Thus, we see that as $C_{S}$ increases it is better to leave some of the buffer space open at the start of each cycle.

We obtain similar findings for other buffer sizes. As another example, Fig. 6b presents capacity curves for the case of $b = 4$. The lightly-shaded outer envelope of capacity is now denoted $Q_{TDM}(4)$.

As an important aside, the earlier model of tandem-booth capacity (Hall and Daganzo, 1983) assumed that each of the two booths in a tandem set will serve a batch of $m$ vehicles each cycle. To avoid excessive blocking, the buffer was assumed to be very underutilized except when $C_{S}$ is small. As a result, the earlier model tends to under-predict tandem-booth capacity, as we show in Appendix C.
3 Capacities of Other Alternative Layouts

Consider the case of a checkpoint with \( M \) single booths placed in parallel fashion. We normalize the width of each booth along with its attendant travel lane as 1, and denote \( w_b \) as the (dimensionless) portion of this total width occupied by the booth. Thus, both the normalized capacity and the lateral space requirement for this checkpoint are \( M \). We next explore how additional service booths can be deployed within this lateral space by staggering the booths (Section 3.1) or by branching them (Section 3.2). Capacity models are formulated for each of these two configurations.
3.1 Staggered Layout

As Fig. 1b shows, sets of staggered booths can be laid-out in parallel fashion. We denote \( k \) \((k \geq 2)\) as the number of staggered booths in each parallel set; in Fig. 1b, \( k = 3 \). The lateral width of each booth set is \( k - (k - 1)w_b \). Thus \( \left\lfloor \frac{M}{k - (k - 1)w_b} \right\rfloor^- \) sets can be arranged within the lateral space required for \( M \) parallel booths, where \( \left\lfloor \cdot \right\rfloor^- \) is the floor operator, denoting the nearest integer value that does not exceed the argument inside. The lateral space left over is:

\[
\begin{align*}
    w_R &= M - \left\lfloor \frac{M}{k - (k - 1)w_b} \right\rfloor^- (k - (k - 1)w_b). \\
\end{align*}
\]

(7)

Within this remainder space, an incomplete set \( \{\frac{w_R - w_b}{1 - w_b}, 0\} \) staggered booths could be installed. Thus, the total number of booths that can be arranged within this lateral space is:

\[
\begin{align*}
    k \left\lfloor \frac{M}{k - (k - 1)w_b} \right\rfloor^- + \max \left\{ \frac{w_R - w_b}{1 - w_b}^- , 0 \right\} ;
\end{align*}
\]

and the normalized capacity per unit lateral space brought by these staggered booths is:

\[
\begin{align*}
    Q_{STGG} &= \frac{1}{M} \left( k \left\lfloor \frac{M}{k - (k - 1)w_b} \right\rfloor^- + \max \left\{ \frac{w_R - w_b}{1 - w_b}^- , 0 \right\} \right). \\
\end{align*}
\]

(8)

where the subscript “STGG” denotes a staggered layout.

Since (8) depends on the checkpoint’s total lateral width, \( M \), \( Q_{STGG} \) cannot be readily compared with the capacity of a single set of tandem booths. Hence, we develop an upper bound on \( Q_{STGG} \), denoted \( \hat{Q}_{STGG} \), by assuming that \( M \) is an integer multiple of the width of a staggered set of \( k \) booths; i.e., \( w_R = 0 \). This upper bound is:

\[
\begin{align*}
    \hat{Q}_{STGG} &= \frac{k}{k - (k - 1)w_b}. \\
\end{align*}
\]

(9)

This bound is independent of \( M \) and is close to \( Q_{STGG} \). For example, when \( w_b = 0.33 \sim 0.41 \) (the range cited in the literature; see Schaufler, 1997) and \( M \geq 5 \), the upper bound typically matches \( Q_{STGG} \) to within 10%.

The longitudinal space required for the staggered layout can be calculated from its geometry. Given are: \( k \), \( w_b \), the normalized length of a booth (relative to the normalized width of the booth and its lane), and the lane shift rate, defined as the ratio between the lateral displacement of a lane and the longitudinal distance along which the lane is shifted (see Fig. 1b). The lane shift rate depends on the design speed of approaching vehicles, driver sight distance, etc. We assume that: \( k = 2; w_b = 0.35 \); the normalized booth length is 1.6; and the lane shift rate is 0.18. The unitless longitudinal space required, \( L_{STGG} \), is then \( 0.35/0.18 + 1.6 \times 2 = 5.14 \). This is equal to 23.2 meters, if a normalized unit length (i.e., a booth width plus its lane’s width) is assumed to be 4.5 meters.

---

\( ^5 \) The lane shift rate comes from the shift rate of a toll plaza’s approach zone suggested by the Illinois State Toll Highway Authority (Schaufler, 1997).
3.2 Branch Layout

Let \( l \geq 2 \) denote the number of branches into which a branched lane is split; see Fig. 1c where the median lane ramifies into \( l = 2 \) lanes downstream, each served by a booth. Note that \( l \) should not be too large, otherwise the branch booths would be starved by the feeding lane upstream. Let \( x \) be the number of regular lanes (i.e., lanes that are not branched out) at the checkpoint, and \( y \) the number of lanes that are branched out; in Fig. 1c, \( x = 4 \) and \( y = 1 \). To maximize the capacity of a branch layout, \( x, y \) should be the optimal solution to the following integer program:

\[
\text{max} \quad x + yl \\
\text{s.t.} \quad x + y(1 - w_b) \leq M \\
\qquad \quad x(1 - w_b) + yl \leq M \\
\qquad \quad x, y \geq 0 \text{ are integers.}
\]

To obtain a closed-form solution, we approximate the above by removing the integer constraint (10d) and solving the resulting linear program. Fig. 7 shows the objective function (solid line) and the feasible region (shaded area) of this linear program. The two constraints, (10b) and (10c), are displayed as dashed and dotted lines, respectively.

![Fig. 7 The optimal solution of the approximated linear program](image)

By recognizing that the slopes of the objective function and the two constraints satisfy \( \frac{1-w_b}{l} < \frac{1}{l} < \frac{1}{1-w_b} \), we find that the optimal solution to the linear program is the solution to the following equations:

\[
\begin{align*}
& x + y(1 - w_b) = M \\
& x(1 - w_b) + yl = M,
\end{align*}
\]

and thus the solution is:
This solution yields an approximation to (and the upper bound on) the normalized capacity of branch booths:

\[
\begin{align*}
\hat{x} &= \frac{M(l+w_b-1)}{l-(1-w_b)^2} \\
\hat{y} &= \frac{Mw_b}{l-(1-w_b)^2}.
\end{align*}
\]  

where the subscript “BRCH” denotes a branch layout.

Numerical analysis shows that when \( w_b = 0.33 \sim 0.41 \) and \( M \geq 5 \), the error of this upper bound is usually below 10%.

The longitudinal space required for the branch layout can also be calculated from geometry. We assume that: \( l = 2; w_b = 0.35 \); the normalized booth length is 1.6; and the lane shift rate is 0.18. The unitless longitudinal space required, \( L_{BRCH} \), is therefore \( \frac{1}{2} \times (2 - (1 - 0.35))/0.18 + 1.6 \times 2 = 6.95 \), which is equal to 31.3 meters, assuming again a normalized unit length of 4.5 meters.

### 4 Comparisons

Fig. 8 shows the normalized capacities of tandem, staggered, and branch layouts versus the normalized longitudinal space available at the checkpoint for the inputs specified in the figure’s caption.

First, the normalized capacities of a tandem layout with no buffer for a high \( C_S = 1 \) (solid, thin curve in Fig. 8) and for a lower \( C_S = 0.53 \) (dashed, thin curve) are obtained from the models furnished in Section 2.1. These capacities increase as the number of booths, \( c \), increases. The \( c \) is constrained by the available longitudinal space. As an example, a tandem layout with \( b = 0 \) and \( c = 2 \) requires a longitudinal space of at least 3.2.

Second, the normalized capacities of a 2-booth tandem layout with a buffer for \( C_S = 1 \) (solid bold curve in Fig. 8) and 0.53 (dashed bold curve) are obtained from the models furnished in Sections 2.2 and 2.3. These capacities increase as buffer size, \( b \), increases; and \( b \) is also constrained by the available longitudinal space. For example, a tandem layout with \( c = 2 \) and \( b = 1 \) requires a longitudinal space of at least 4.8.

Finally, the normalized capacities of a staggered layout (double solid curve) are obtained from (9). These capacities increase with \( k \). Thus, we choose the maximum \( k \) that is feasible given the longitudinal space. As examples, \( k = 2 \) is maximum when the longitudinal space is between 5.2 and 8.7; and \( k = 3 \) is maximum when that space is between 8.7 and 12.2. We plot the normalized capacities of a branch layout (dotted curve) from (13) in similar fashion. These increase with \( l \), which is also constrained by the longitudinal space.
Fig. 8 Comparison of the normalized capacities of tandem, staggered, and branch configurations; we assume that: the length of a booth and of a single-vehicle buffer are each 1.6; \( w_b = 0.35 \); and the lane shift rate is 0.18.

Fig. 8 shows that for given longitudinal (and lateral) space constraints, the capacities furnished by tandem layouts always exceed those of the staggered- and branch-booth designs. This is true even if \( C_S \) is very high, which is detrimental to tandem designs. Recall too that (9) and (13) furnish upper bounds on the capacities of staggered and branch layouts.

Another consideration when choosing a configuration is the average percentage of time that each booth is in use. All else equal, a low utilization level means that more booths (and thus perhaps more human operators) are required to achieve some target for checkpoint capacity. Fig. 9 shows utilization levels per booth (i.e., the normalized capacity of each layout divided by its number of booths) for various layouts and as functions of \( C_S \).

The solid horizontal line in the figure shows that the utilization levels for single-booth, staggered and branch layouts are all 100\%, independent of \( C_S \). The levels for tandem layouts, shown with the dashed curves, are lower than this, save for the idealized case of \( C_S = 0 \). Note how the utilization levels for tandem layouts decrease as \( C_S \) increases, and are lowest in the absence of buffers. We see that these levels are especially low when \( c \geq 3 \), \( b = 0 \) and \( C_S \) is large.

In light of the above, an agency might opt to use tandem layouts only when its booth operations can be automated (e.g. as in the case of electronic toll collection), or when the cost of human labor is low. Of course, the issue of booth utilization should be weighed against the relatively high capacities that tandem layouts can achieve.
5 Conclusions

The paper has presented models for estimating the vehicle-processing capacities at checkpoints with tandem, staggered, and branch configurations. These layouts can be suitable alternatives when adding parallel service booths is infeasible due to lateral space constraints at the checkpoint. The layouts can also be alternatives to strategies such as reversible lanes (Schaufler, 1997), and could be deployed in conjunction with ITS technologies, including automated toll collection schemes and motorist information systems (e.g., Nozick et al., 1998; Khan, 2010).

Among the three layouts, we find that tandem booths without buffers tend to furnish the highest capacities, but have the lowest utilization levels per booth. Buffers can improve the utilization levels, particularly when their use is controlled so that they are filled with suitable numbers of vehicles at the start of each cycle. More elaborate forms of control are found not to yield sizable capacity gains, however; see Appendix A.

The above findings are worth considering when booth service is to be provided by human operators. If labor costs are high, staggered or branch layouts might be used instead.

We further find that tandem-booth capacities diminish as $C_S$ increases. This suggests that grouping vehicles by class in ways that reduce $C_S$ across each tandem set of booths can further improve capacities. Grouping strategies might include the segregation of commercial vehicles from regular passenger vehicles.

Finally, the tandem-booth models formulated in this paper ignore vehicle move-up times, e.g., from the queue to a booth, from an upstream booth to a buffer, etc. These added times may be negligible at border crossings where service times tend to be long. At other checkpoint types, such as toll plazas, our findings nonetheless furnish useful insights regarding the effects of tandem booths and the addition of buffers. Furthermore, our present models can be readily adjusted by including correction terms to account for move-up times.
Acknowledgements

Funding for this work was provided by the Volvo Research and Educational Foundations and the University of California Transportation Center.

Appendix A

Scenario with a General Form of Control when $b = 1$

We examine here the cases of $p = 0$ or $1$ with a general form of control and compare these with the limited form of control that ensures that $p = 1$. For each case, a control decision needs to be made when i) the vehicle dwelling at the upstream booth has just finished its service, ii) the downstream booth is still occupied, and iii) the buffer is empty. We call the time instant when i-iii) are all satisfied a control point. The decision would be either to prohibit the vehicle at the upstream booth from entering the buffer, or to usher it in.

We will first develop the condition under which it is advantageous to prohibit the upstream vehicle from entering the buffer at a control point. To this end, we denote $n_{\text{out}}$ and $\bar{T}_{\text{out}}$ as the number of vehicles served in the cycle and the expected cycle duration, respectively, when a decision is made to prohibit the upstream vehicle from entering the empty buffer; and $n_{\text{in}}$ and $\bar{T}_{\text{in}}$ if the opposite decision is made. At a control point, we will observe the elapsed service times for all the vehicles already served or in service in the cycle. These elapsed service times will be used to make the control decision, as $n_{\text{out}}, \bar{T}_{\text{out}}, n_{\text{in}}, \bar{T}_{\text{in}}$ can be written as functions of these times. We further assume that a cycle with given values of these elapsed service times occurs with probability $\delta$; and that for all other cycles (that occur with probability $1 - \delta$), the average vehicle flow is $q = \bar{n}/\bar{\ell}$, where $\bar{n}$ and $\bar{\ell}$ are the average number of vehicles served per cycle and the average cycle length for all other cycles, respectively.

\[
\frac{1 - \delta)\bar{n} + \delta n_{\text{out}}}{(1 - \delta)\bar{n} + \delta n_{\text{in}}} \geq \frac{1 - \delta)\bar{n} + \delta n_{\text{out}}}{(1 - \delta)\bar{n} + \delta \bar{T}_{\text{out}}}
\]

This is equivalent to:

\[
(1 - \delta)(\bar{n}\bar{T}_{\text{in}} + n_{\text{out}}\bar{\ell}) + \delta n_{\text{out}}\bar{T}_{\text{in}} \geq (1 - \delta)(\bar{n}\bar{T}_{\text{out}} + n_{\text{in}}\bar{\ell}) + \delta n_{\text{in}}\bar{T}_{\text{out}}
\]

For continuously distributed vehicle service times, $\delta$ is generally close to 0. Thus we ignore the second terms on both sides of the above inequality, and obtain:

\[
n_{\text{out}}\bar{\ell} - \bar{n}\bar{T}_{\text{out}} \geq n_{\text{in}}\bar{\ell} - \bar{n}\bar{T}_{\text{in}}
\]

Dividing both sides above by $\bar{\ell}$, we obtain the condition for blocking the upstream vehicle:

\[
\Delta n = (n_{\text{out}} - q\bar{T}_{\text{out}}) - (n_{\text{in}} - q\bar{T}_{\text{in}}) \geq 0
\]

(A1)

Note that $(n_{\text{out}} - q\bar{T}_{\text{out}})$ and $(n_{\text{in}} - q\bar{T}_{\text{in}})$ are the extra numbers of vehicles that would be served in the cycle (compared to the average level) when the two control decisions are made. Thus, $\Delta n$ represents the additional number of vehicles served if the upstream vehicle is prohibited from using the buffer, as compared to the counter decision.
From (A1) we see that $q$ is needed to make the correct control decision. When $\delta \to 0$, $q$ is approximately the vehicle capacity of the tandem booths after applying control, and is therefore unknown. However, the limited form of control described in Section 2.2 provides a lower bound on $q$. For simplicity, we will use this lower bound. This will yield a good approximation to the optimal form of control when the lower bound is close to the real value of $q$ (and we shall see momentarily that it is true).

We now develop $n_{\text{out}}$, $T_{\text{out}}$, $n_{\text{in}}$, and $T_{\text{in}}$ as functions of the (elapsed) vehicle service times at the control point.

When $p = 0$, let $\tau$ denote the service time of vehicle 2 (see again Fig. 3). Given $S_1 > \tau$, $\tau$ is also the elapsed service time of vehicle 1 at the control point. We thus have:

$$n_{\text{in}} = 3, \quad (A2)$$

$$n_{\text{out}} = 2, \quad (A3)$$

$$T_{\text{in}}(\tau) = E[\max \{S_1, S_3 + \tau|S_1 > \tau\} - \tau] = \tau + \int_0^\infty \left(1 - \frac{F_S(t) - F_S(\tau)}{1-F_S(\tau)} F_S(t - \tau)\right)dt, \quad (A4)$$

$$T_{\text{out}}(\tau) = E[S_1|S_1 > \tau] = \frac{\int_0^\infty F_S(t) dt}{1-F_S(\tau)}, \quad (A5)$$

where $f_S(t)$ is the probability density function of the vehicle service time. Equation (A4) is true because:

$$Pr\{T_{\text{in}}(\tau) < t\} = Pr\{\max \{S_1, S_3 + \tau\} < t|S_1 > \tau\}$$

$$= Pr\{S_1 < t|S_1 > \tau\} \cdot Pr\{S_3 + \tau < t\}$$

$$= \begin{cases} 0, & t \in (0, \tau) \\ \frac{F_S(t) - F_S(\tau)}{1-F_S(\tau)} F_S(t - \tau), & t \in (\tau, \infty) \end{cases}$$

and then

$$\bar{T}_{\text{in}}(\tau) = \int_0^\infty (1 - Pr\{T_{\text{in}}(\tau) < t\}) dt$$

$$= \tau + \int_\tau^\infty \left(1 - \frac{F_S(t) - F_S(\tau)}{1-F_S(\tau)} F_S(t - \tau)\right)dt.$$

The proof of (A5) is straightforward and therefore not furnished here.

Substituting (A2)-(A5) into (A1), we can find the range of $\tau$ for which the upstream vehicle should be prohibited from entering the buffer. Given this range of $\tau$, the $\bar{N}$ and $\bar{L}$ can be calculated either via Monte Carlo simulation, or by integrating over the distribution function of $\tau$. The capacity when applying this control, $Q_{TDMS}(1,0)$, can therefore be obtained from Equation (1).

When $p = 1$, the procedure is similar. Now we denote $\tau_1$ and $\tau_3$ as the service times of vehicle 1 and 3, respectively; see again Fig. 4. Given $\tau_1 < \tau_3 < \tau_1 + S_2$, a decision is to be made as to whether or not vehicle 3 is to be prohibited from entering the buffer. We have:

$$n_{\text{in}} = 4, \quad (A6)$$
Consideration unveils the validity of (A9), and (A8) is true because:

\[
\begin{align*}
T_{in}(\tau_1, \tau_3) &= E[\max\{\tau_1 + S_2, \tau_3 + S_4\} | S_2 > \tau_3 - \tau_1] = \tau_3 + \int_{\tau_3}^{\infty} \left(1 - \frac{F_S(t - \tau_1) - F_S(\tau_3 - \tau_1)}{1 - F_S(\tau_3 - \tau_1)} F_S(t - \tau_3)\right) dt,
\end{align*}
\]

(A7)

\[
\begin{align*}
T_{out}(\tau_1, \tau_3) &= E[\tau_1 + S_2 | S_2 > \tau_3 - \tau_1] = \tau_1 + \int_{\tau_3 - \tau_1}^{\infty} \frac{t f_S(t) dt}{1 - F_S(\tau_3 - \tau_1)}.
\end{align*}
\]

(A8)

The capacity when applying the control, \(Q_{TDM,G}(1,1)\), can be calculated in a similar way to the case of \(p = 0\).

Fig. A1 presents the tandem capacities under general control, \(Q_{TDM,G}(1,0)\) and \(Q_{TDM,G}(1,1)\), as thin, solid curves. The normalized capacities for the non-control and limited-control cases are shown with thicker, dashed curves. Note that \(Q_{TDM,G}(1,0)\) is an approximate outer envelope of \(Q_{TDM,2}(0)\) and \(Q_{TDM}(1,0)\). Further note that \(Q_{TDM,G}(1,1)\) is superimposed with \(Q_{TDM}(1,1)\). Hence, any capacity gains brought by the general form of control are negligible. For a tandem layout with a single-vehicle buffer, the limited form of control described in Section 2.2 seems to be sufficient.

![Graph showing normalized capacity comparison](image-url)
Further note the this general form of control only affects the number of vehicles that might be served by the upstream booth when the downstream booth is serving its last vehicle in a cycle (i.e., vehicle 1 when \( p = 0 \), or vehicle 2 when \( p = 1 \)). As a result, the relative capacity increase via general control would be even less when a larger buffer is used (such that more vehicles can be served in a cycle). Therefore, there seems to be little need to explore the general form of control for \( b > 1 \).

**Appendix B**

**Proofs of Equations (5) and (6) in Section 2.3**

For a pair of tandem booths with buffer size \( b \) and \( p \) prefilled vehicles, the first \( p + 1 \) vehicles in the cycle are served by the downstream booth, and the other vehicles by the upstream booth. This is because when the first \( p + 1 \) vehicles finish their service, either the cycle ends (if the upstream booth is idle) or the downstream booth is blocked for the rest of the cycle by another vehicle that occupies the upstream booth. The number of vehicles served by the upstream booth can be any integer value between 1 and \( b + 1 \). The latter number can be achieved when the buffer is completely filled by served vehicles. Thus, the total number of vehicles served in the cycle is no less than \( p + 2 \), and no more than \( b + p + 2 \). Let \( R_i \) (\( i = 1, 2, \ldots, b + p + 2 \)) denote the time when vehicle \( i \) enters a booth (the beginning of the cycle is time 0), and \( P_i \) (\( i = p + 3, p + 4, \ldots, b + p + 2 \)) denote the probability that vehicle \( i \) is served during the cycle. We first calculate \( \mathcal{N} \) as shown below.

Consider two vehicle batches in the cycle: i) the batch from vehicle 1 to \( p + 1 \); and ii) the batch from vehicle \( p + 2 \) to \( b + 2 \) (if vehicle \( b + 2 \) is served in the cycle) or the last vehicle that is served in the cycle. Note that the first batch is served continuously by the downstream booth (i.e., that booth remains busy until it finishes serving the batch); and that the second batch is served continuously by the upstream booth. By virtue of this, we have

\[
P_{i+1} = Pr\{\sum_{j=1}^{p+1} S_j > \sum_{j=p+2}^{i} S_j\} \text{ for } i = p + 2, \ldots, b + 2.
\]

Since vehicle \( b + 2 \) cannot enter the buffer until vehicle 1 has departed, we have

\[
R_{b+3} = \max \{\sum_{j=p+2}^{b+2} S_j, S_1\}.
\]

So,

\[
P_{b+4} = Pr\left\{\sum_{j=1}^{p+1} S_j > R_{b+3} + S_{b+3}\right\} = Pr\left\{\sum_{j=1}^{p+1} S_j > \max \{\sum_{j=p+2}^{b+3} S_j, S_{b+3} + S_1\}\right\}.
\]

Similarly,

\[
R_{b+4} = \max \{R_{b+3} + S_{b+3}, S_1 + S_2\} = \max \{\sum_{j=p+2}^{b+3} S_j, S_{b+3} + S_1 + S_2\};
\]

\[
P_{b+5} = Pr\left\{\sum_{j=1}^{p+1} S_j > R_{b+4} + S_{b+4}\right\} = Pr\left\{\sum_{j=1}^{p+1} S_j > \max \{\sum_{j=p+2}^{b+4} S_j, S_{b+3} + S_{b+4} + S_1 + S_{b+4} + S_1 + S_2\}\right\}.
\]

Following the same logic, we have

\[
P_{i+1} = Pr\{\sum_{j=1}^{p+1} S_j > \max \{\sum_{j=p+2}^{i} S_j, \sum_{j=b+3}^{i} S_j + S_1, \sum_{j=b+4}^{i} S_j + \sum_{j=1}^{i} S_j, \ldots, S_i + \sum_{j=1}^{i-b-2} S_j\}\};
\]

18
\[ R_{i+1} = \max \{ \sum_{j=p+2}^{i} S_j, \sum_{j=b+3}^{i} S_j + S_1, \sum_{j=b+4}^{i} S_j + \sum_{j=1}^{2} S_j, \ldots, \sum_{j=1}^{b-1} S_j \} \text{ for } i = b + 3, \ldots, b + p + 1. \]

So,

\[ N(b, p) = p + 2 + \sum_{i=p+3}^{b+p+2} P_i \]

\[ = p + 2 + \sum_{i=p+2}^{b+p+1} P r\{ \sum_{j=1}^{p+1} S_j > \sum_{j=p+2}^{i} S_j \} + \sum_{i=b+3}^{b+p+1} P r\{ \sum_{j=1}^{p+1} S_j > \max \{ \sum_{j=p+2}^{i} S_j, \sum_{j=b+3}^{i} S_j + S_1, \sum_{j=b+4}^{i} S_j + \sum_{j=1}^{2} S_j, \ldots, \sum_{j=1}^{b-1} S_j \} \}. \]

To calculate \( T_S \), we consider a hypothetical cycle that is identical to the real one, except that in the hypothetical one, vehicles \( p + 2 \) to \( b + p + 2 \) are all served at the upstream booth, even if the downstream booth has long been empty. For this hypothetical cycle, we find the cycle length, \( T' \), to be

\[ T' = \max \{ R_{b+p+2} + S_{b+p+2}, \sum_{j=1}^{p+1} S_j \} \]

\[ = \max \{ \sum_{j=p+2}^{b+p+2} S_j, \sum_{j=b+3}^{b+p+2} S_j + S_1, \sum_{j=b+4}^{b+p+2} S_j + \sum_{j=1}^{2} S_j, \ldots, \sum_{j=1}^{p+1} S_j \}. \]

Note that in the hypothetical cycle, the time duration when both booths are busy is the same as \( T_S \) of the real cycle. We have:

\[ T_S = b + p + 2 - E[T'] \]

\[ = E[\min \{ \sum_{j=1}^{p+1} S_j, \sum_{j=2}^{b+p+2} S_j, \sum_{j=3}^{b+3} S_j, \ldots, \sum_{j=p+2}^{b+p+2} S_j \}] . \]

**Appendix C**

**Comparison of Proposed Tandem Model with the Batch Processing Model of Hall and Daganzo (1983)**

For comparison, we first calculate the capacity of the Hall and Daganzo model with a buffer of size \( b \). In that model, each booth serves a batch of vehicles in succession during a cycle, and the two batches served by both booths have the same size, \( m \). Both booths are emptied before the next batches of vehicles can proceed. It is assumed that vehicles are served without blocking each other. This assumption is approximately true only when \( b \geq m - 1 + 2C_S\sqrt{m-1} \) (see Hall and Daganzo, 1983). Thus, we have:

\[ m \leq 1 + b + 2C_S^2 - 2C_S\sqrt{b + C_S^2} . \]

So, the maximum batch size is

\[ m = \left[ 1 + b + 2C_S^2 - 2C_S\sqrt{b + C_S^2} \right]^- , \quad (\text{C1}) \]

where \([ \quad ]^-\) is the floor operator, denoting the nearest integer value that does not exceed the argument inside.

Equation (C1) shows that \( m \) is a function of \( C_S \) for any given \( b \). For each \( m \), the batch processing model is equivalent to a hypothetical tandem model with no buffer, in which a hypothetical “vehicle” represents a batch of \( m \) vehicles, and the hypothetical “vehicle service time” is the sum of \( m \) vehicles’ service time.
Thus, the capacity of this batch processing model, $Q_{TDM,B}(b)$, can be obtained by (2), where $F_S(t)$ is replaced by the CDF of the sum of $m$ vehicles’ service time.

The resulting capacity curves for $b = 2$ and 4 (dotted curves) are compared with the curves of our proposed model in Figs. C1a and b, respectively. Both figures show that the outcomes of the batch processing model are comparable with those of our model only when $C_S$ is low;\(^6\) but the buffer is largely underutilized in the batch processing strategy when $C_S$ is high. For example, when $C_S = 1$, 21% of a single booth’s capacity is underutilized for both $b = 2$ and 4. This is because too many empty buffers are required in the batch processing model to avoid blockages between vehicles. This limits the batch size that can be served by each booth.

\(^6\) Due to the assumption of the batch processing model, sometimes $Q_{TDM,B}(b)$ looks slightly higher than $Q_{TDM}(b)$, as shown in Fig. C1.
References


