E-Companion to:
Intermodal hub network design with probabilistic service level constraints

Mario José Basallo Triana
Jean-François Cordeau
Navneet Vidyarthi

A Proof of Proposition 2

Let $x = (\mu_{a_1} - \lambda_{a_1})v_a$ and $y = (\mu_{a_2} - \lambda_{a_2})v_a$, where μ_{a_1}, μ_{a_2} and v_a are constants. Variables x and y are an affine transformation of the original variables λ_{a_1} and λ_{a_2}, respectively. It is known that composition with an affine transformation preserves quasiconcavity (Boyd and Vandenberghe 2004, page 102). The total sojourn time distribution $W(x, y) = 1-(xe^{-y} - ye^{-x})/(x-y)$ is quasiconcave for $x, y > 0$ if the function

$$f(x, y) = \frac{xe^{-y} - ye^{-x}}{x - y} \tag{A.1}$$

is quasiconvex. For $\theta \in [0, 1]$, $(x_1, y_1) \geq 0$ and $(x_2, y_2) \geq 0$, function f is quasiconvex if and only if (Boyd and Vandenberghe 2004, page 98)

$$\max\{f(x_1, y_1), f(x_2, y_2)\} \geq f\left[\theta(x_1, y_1) + (1-\theta)(x_2, y_2)\right]. \tag{A.2}$$

Now, the function f is strictly decreasing, as can be verified from its first partial derivatives. Then, if w.l.o.g. $(x_1, y_1) \leq (x_2, y_2)$, Inequality (A.2) holds and f is quasiconvex. However, it remains to prove quasiconvexity when (x_1, y_1) and (x_2, y_2) do not constitute a monotonic pair.

To fully prove the quasiconvexity, we restrict the function f to the line $y = a + bx$, where $a > 0$ and $b < 0$ are parameters chosen to represent the case when the pair (x_1, y_1) and (x_2, y_2) is not monotonic. Then f is quasiconvex if the univariate function $f(x, a+bx)$ given in Equation (A.3) is unimodal with a minimum for all $a > 0$ and $b < 0$. This condition of unimodality with a minimum in the univariate functions directly implies quasiconvexity as a consequence of Equation (A.2).

$$f(x, a+bx) = \frac{xe^{-(a+bx)} - (a+bx)e^{-}x}{x - (a+bx)}, \qquad \text{for } a > 0, b < 0, x \geq 0. \tag{A.3}$$

To prove the unimodality of $f(x, a+bx)$, it is sufficient to show that its first derivative changes sign at most once over the interval $[0, \infty)$, and the change of sign if it occurs, goes from negative to positive as x increases. Taking the first derivative we get

$$f'(x, a+bx) = \frac{e^{-[a+(2+b)x]}}{[a+(b-1)x]^2}\left\{e^{(b-1)x}e^a\left[(1-b)bx^2 + (1-2b)ax + a - a^2\right] + b(b-1)x^2 + abx - a\right\}. \tag{A.4}$$

The term $e^{-[a+(2+b)x]}/\left[a+(b-1)x\right]^2$ is positive and does not alter the sign of the previous expression. An indeterminate situation occurs at the singularity line $y = a + bx = x$; however, by monotonicity, f is quasiconvex along this line. Then, to analyze the sign of $f'(x, a + bx)$ it is sufficient to examine the function

$$g(x) = e^{(b-1)x}e^{a}\left[(1-b)bx^2 + (1-2b)ax + a - a^2\right] + b(b-1)x^2 + abx - a, \tag{A.5}$$

which has the following properties:

Property 1: Function g change of concavity exactly once along the interval $[0,\infty)$ from concave to convex for x increasing.

Property 2: Function g has at most 2 roots on $[0,\infty)$. One of the roots is the point $x = a/(1-b)$, which satisfy

a) A local maximum for $b > -1$,

b) An inflection point for $b = -1$,

c) A local minimum for $b < -1$.

Proof of Property 1:

The first derivative of g is

$$g'(x) = e^{(b-1)x}e^{a}(b-1)\left[(1-b)bx^2 + (1-2b)ax + a - a^2\right] + e^{(b-1)x}e^{a}\left[2(1-b)bx + (1-2b)a\right] + 2b(b-1)x + ab. \tag{A.6}$$

The second derivative of g is

$$g''(x) = e^{(b-1)x}e^{a}\left\{(1-b)^3bx^2 + (b-1)^2\left[a - 2(2+a)b\right]x + (1+a)(b-1)\left[(1-b)a - 2b\right]\right\} + 2b(b-1). \tag{A.7}$$

The sign of g'' determines the type of concavity of g. Given that the terms $e^{(b-1)x}e^{a}$ and $2b(b-1)$ are strictly positive, the sign of g'' is only determined by the factor

$$(1-b)^3bx^2 + (b-1)^2\left[a - 2(2+a)b\right]x + (1+a)(b-1)\left[(1-b)a - 2b\right]. \tag{A.8}$$

This concave parable has the vertex point

$$\left(\frac{a - 2(2+a)b}{2b(b-1)}, \left[a^2 + 4a(b-1)b + 8b^2\right]\frac{b-1}{4b}\right) > 0, \qquad \text{for } a > 0, b < 0,$$

and roots at

$$x_1 = \frac{a(1-2b) - 4b - \sqrt{a^2 - 4ab + 4(2+a)b^2}}{2b(b-1)} > 0,$$
$$x_2 = \frac{a - (4+2a)b + \sqrt{a^2 - 4ab + 4(2+a)b^2}}{2b(b-1)} > 0.$$

The strict positivity of x_1 is not so obvious. By recalling that $a > 0$ and $b < 0$ we have

$$\begin{aligned} a(1-2b) - 4b &> \sqrt{a^2 - 4ab + 4(2+a)b^2}, \\ \left[a(1-2b) - 4b\right]^2 &> a^2 - 4ab + 4(2+a)b^2, \\ 4a^2b^2 - 4a^2b + 12ab^2 - 4ab + 8b^2 &> 0. \end{aligned}$$

Then, Parable in (A.8) is negative in the region $[0, x_1) \cup (x_2 \infty)$ and positive in the interval (x_1, x_2). Now we observe that

$$g''(0) = e^a \{(1+a)(b-1)[(1-b)a - 2b]\} + 2b(b-1) < 0.$$

This strict inequality is valid since

$$\begin{aligned} e^a \{(1+a)(b-1)[(1-b)a - 2b]\} + 2b(b-1) &< 0 \\ e^a \left(-a^2b^2 + 2a^2b - 3ab^2 - 2b^2 - a^2 + 4ab + 2b - a\right) + 2b^2 - 2b &< 0, \\ e^a \left(-a^2b^2 + 2a^2b - 3ab^2 - a^2 + 4ab - a\right) + 2b^2(1-e^a) + 2b(e^a - 1) &< 0, \end{aligned}$$

This analysis implies that g'' is negative at $x = 0$ and turns positive at some point on the real line as x increases. Next, we show that once $g''(x)$ is positive, it remains positive as x increases implying that g'' changes sign only once.

We determine the local optima of g'' by analyzing its derivative

$$g'''(x) = -(b-1)^2 e^a e^{(b-1)x} \left\{(b-1)^2 bx^2 + (b-1)[2(3+a)b - a]x + (2+a)[a(b-1) + 3b]\right\}.$$

Solving for x in the equation $g'''(x) = 0$ we have the following critical points for g''

$$\begin{aligned} x_1' &= \frac{(1-2b)a - 6b - \sqrt{a^2 - 4ab + 4(3+a)b^2}}{2b(b-1)} > 0, \\ x_2' &= \frac{a - 6b - 2ab + \sqrt{a^2 - 4ab + 4(3+a)b^2}}{2b(b-1)} > 0. \end{aligned}$$

Since Parable (A.8) is concave and has a positive vertex then $g''(x_1') > 0$. Also, considering the exponential decay factor $e^{(b-1)x}$ in Equation (A.7), we can conclude that x_1' is a local maximum, and x_2' is a local minimum of g''. It is also true that $g''(x_2') > 0$:

$$g''(x_2') = \exp\left[\frac{a - 6b + \sqrt{a^2 - 4ab + 4(3+a)b^2}}{2b}\right] \left\{(b-1)\left[-2b + \sqrt{a^2 - 4ab + 4(3+a)b^2}\right]\right\} + 2b(b-1), \tag{A.9}$$

$$\geq \frac{1}{\frac{a-6b+\sqrt{a^2-4ab+4(3+a)b^2}}{-2b} + 1} \left\{(b-1)\left[-2b + \sqrt{a^2 - 4ab + 4(3+a)b^2}\right]\right\} + 2b(b-1), \tag{A.10}$$

$$= \frac{-2b}{a - 6b + \sqrt{a^2 - 4ab + 4(3+a)b^2} - 2b} \left\{(b-1)\left[-2b + \sqrt{a^2 - 4ab + 4(3+a)b^2}\right]\right\} + 2b(b-1), \tag{A.11}$$

$$> \frac{-2b}{\sqrt{a^2 - 4ab + 4(3+a)b^2} - 2b} \left\{(b-1)\left[-2b + \sqrt{a^2 - 4ab + 4(3+a)b^2}\right]\right\} + 2b(b-1), \tag{A.12}$$

$$= 0, \tag{A.13}$$

where Inequality (A.10) is valid since $1/(x+1) \geq e^{-x}$, for $x \geq 0$.

The exponential decay term $e^{(b-1)x}$ *attracts* the graph of g'' to the asymptotic limit of $2b(b-1) > 0$ as $x \to \infty$. Furthermore, since that g'' is positive at its local minimum x_2', this implies that once g'' attains a positive value it remains positive as x increases due to the dominance of the exponential term. Therefore g'' changes sign only once on $[0, \infty)$. Figure A.1 illustrates the behavior of function g'', concluding the proof of **Property 1**.

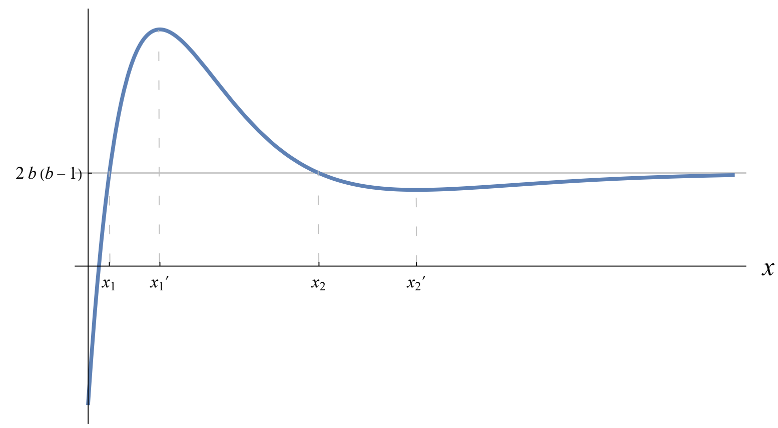

Figure A.1: Graph of $g''(x)$.

Proof of Property 2:

The evaluation of g and its derivatives at point $x = a/(1-b)$ gives

$$g\left(\frac{a}{1-b}\right) = 0, \qquad g'\left(\frac{a}{1-b}\right) = 0, \qquad g''\left(\frac{a}{1-b}\right) = a(b^2-1).$$

Then, this point is a root and a critical point of g. When $b > -1$ it is a local maximum, when $b = -1$ it is an inflection point, and when $b < -1$ it is a local minimum. To see why g has at most two roots consider that $x = a/(1-b)$ is a local minimum (maximum), then at this point g does not cross the x-axis. However, given that g changes from concave to convex exactly once as x increases, this implies the existence of a second root at which g effectively crosses the x axis. There cannot be more than two roots; otherwise, this would imply that g changes its concavity more than once, which is impossible according to our previous analysis. Now consider that $x = a/(1-b)$ is an inflection point since this point determines a change of concavity and the first derivative of g evaluated at this point is 0, this implies that g has a single root at which the function crosses the x axis. See Figure A.2. This concludes the proof of **Property 2**.

From the previous analysis, we conclude that $f'(x, a+bx)$ in Equation (A.4) change sign at most once on $[0, \infty)$, and this change of sign is from negative to positive. This implies that function $f(x, a+bx)$ is unimodal with a minimum. Then function $f(x, y)$ is quasiconvex in general, concluding the proof of Proposition 2.

B Analysis of Conjecture 1

Using the same notation as in the previous appendix, the total sojourn time distribution is concave if function f is convex. The convexity of $f(x, y)$ can be assessed by determining whether the

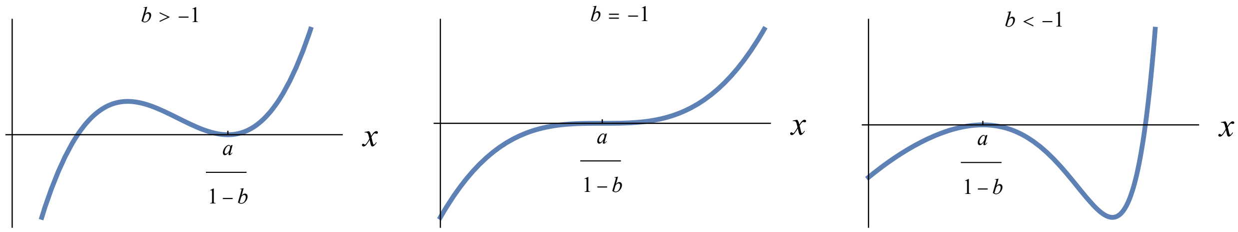

Figure A.2: Graph of $g(x)$.

corresponding Hessian matrix $H_f(x,y)$ is positive (semi)definite in the domain of f. To this end, we use the leading principal minors criteria. Then, H_f is positive (semi)definite if all its leading principal minors are non-negative. The minor of order 1 of H_f is:

$$|H_f(x,y)|_1 = \frac{ye^{-x-y}\left\{2e^x - \left[(1+x-y)^2+1\right]e^y\right\}}{(x-y)^3}.$$

Since f is symmetric, it is enough to show that $|H_f(x,y)|_1$ is non-negative in the region defined by $x \geq y, x > 0$, and $y > 0$. According to this, $|H_f(x,y)|_1$ is non-negative if and only if $2e^x - \left[(1+x-y)^2+1\right]e^y \geq 0$, which is equivalent to $2e^{x-y} - 1 - (1+x-y)^2 \geq 0$. Let $z = x - y \geq 0$, the left-hand side of the previous inequality becomes $2e^z - 1 - (1+z)^2$. This expression has a minimum value of 0, which is obtained when $z = 0$ or $x = y$. This result can be verified using the first and second derivative criteria. Then, the inequality holds, and we conclude that $|H_f(x,y)|_1$ is non-negative everywhere in the domain of f.

On the other hand, the minor of order 2 of H_f is:

$$|H_f(x,y)|_2 = \frac{e^{-2(x+y)}\left\{\begin{array}{c}-e^{2y}\left[(1+x)^2-2xy\right] - e^{2x}\left[(1+y)^2-2xy\right] + \\ e^{x+y}\left[2+2x+2y-2xy-(x-y)^2xy\right]\end{array}\right\}}{(x-y)^4}.$$

In this case, $|H_f(x,y)|_2$ can be negative in the domain of f, as it is shown in Figure B.1. According to this, f is not convex everywhere in its domain. Explicitly defining a region where f is convex is interesting. Note that the second leading principal minor is non-negative if and only if:

$$-e^{2y}\left[(1+x)^2-2xy\right] - e^{2x}\left[(1+y)^2-2xy\right] + e^{x+y}\left[2+2x+2y-2xy-(x-y)^2xy\right] \geq 0. \quad \text{(B.1)}$$

We conjecture that the previous inequality is satisfied in the region $y \geq 1/x$ and $x > 0$. We show that $|H_f(x,y)|_2$ is non-negative along the curve $y = 1/x$. Replacing $y = 1/x$ in Inequality (B.1) and simplifying we get:

$$\frac{e^{-\frac{1}{x}}\left(e^x - e^{1/x}x^2\right)\left[e^{x-\frac{1}{x}}\left(x - \frac{1}{x} - 2\right) + x - \frac{1}{x} + 2\right]}{x^3} \geq 0. \quad \text{(B.2)}$$

By the symmetry of f and given that $y \geq 1/x$, we note that $x \geq 1$ in the region of interest. In this sense, each factor in the numerator of the left-hand side of Inequality (B.2) is non-negative, as can be easily verified using the first and second derivative criteria. Then, the inequality holds, and the strict equality is obtained only when $x = 1$. This allows us to conclude that H_f is positive (semi-)definite on the curve $y = 1/x$.

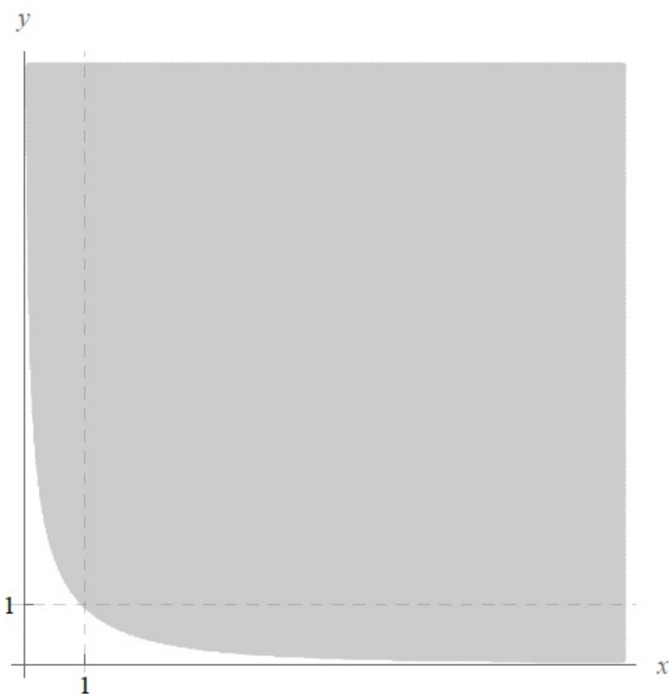

Figure B.1: The shaded area shows the region where $|H_f(x,y)|_2$ is non-negative, or, equivalently, the region where H_f is positive (semi)definite.

Based on the conjecture that f is convex (equivalently: $\tilde{W}$ is concave) for $y \geq 1/x$ and $x > 0$, and given that $\tilde{W}(x,y)$ is a strictly increasing function, it is of interest to find the maximum of $\tilde{W}(x,y)$ constrained to $y = 1/x$. We have the following univariate optimization problem:

$$\max_{x>0} \quad \tilde{W}\left(x, \frac{1}{x}\right) = 1 - \frac{e^{-x} - e^{-\frac{1}{x}}x^2}{1 - x^2},$$

which has solution $\tilde{W}^* = 1 - 2/e$, $x^* = 1$. This suggests that a lower bound for the minimum percentile of the total sojourn time distribution at which such distribution remains concave is $(1-2/e)100 \approx 26.4$. Given that in practical applications, the desired service level α is much higher than $1 - 2/e$, probabilistic service level constraints can be considered to be convex for relevant practical applications.

C A closed form expression for f^a

Consider the case when $\mu_{a_1} - \lambda_{a_1} \neq \mu_{a_2} - \lambda_{a_2}$, from Equation 7 we know that at the α-level set the following relation must hold:

$$1 - \frac{(\mu_{a_1} - \lambda_{a_1})e^{-(\mu_{a_2} - \lambda_{a_2})\tau}}{(\mu_{a_1} - \lambda_{a_1}) - (\mu_{a_2} - \lambda_{a_2})} - \frac{(\mu_{a_2} - \lambda_{a_2})e^{-(\mu_{a_1} - \lambda_{a_1})\tau}}{(\mu_{a_2} - \lambda_{a_2}) - (\mu_{a_1} - \lambda_{a_1})} = \alpha.$$

After some algebraic manipulations, the previous expression can be rewritten as:

$$\left[e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)\right]\lambda_{a_1}+(\mu_{a_2}-\lambda_{a_2})e^{-\mu_{a_1}\tau}e^{\lambda_{a_1}\tau}=\left[e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)\right]\mu_{a_1}+(1-\alpha)\left(\mu_{a_2}-\lambda_{a_2}\right). \tag{C.1}$$

Let

$$\begin{aligned}
d &= e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha),\\
b &= (\mu_{a_2}-\lambda_{a_2})e^{-\mu_{a_1}\tau},\\
c &= \left[e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)\right]\mu_{a_1}+(1-\alpha)\left(\mu_{a_2}-\lambda_{a_2}\right).
\end{aligned}$$

Equation (C.1) is rewritten as $d\lambda_{a_1}+be^{\lambda_{a_1}\tau}=c$, and we want to solve for λ_{a_1} this equation. Using algebraic manipulations we have $a\lambda_{a_1}+be^{\lambda_{a_1}t}=c \Longrightarrow (b/a)te^{\lambda_{a_1}t}=t(c/a)-\lambda_{a_1}t \Longrightarrow (b/a)t=\left((c/a)t-\lambda_{a_1}t\right)e^{-\lambda_{a_1}t} \Longrightarrow (b/a)te^{(c/a)t}=\left(c/a-\lambda_{a_1}\right)te^{\left(c/a-\lambda_{a_1}\right)t}$. Taking the Lambert W function to both sides of the previous equation we obtain $\mathcal{W}\left((b/a)te^{(c/a)t}\right)=\left((c/a)-\lambda_{a_1}\right)t$. After solving for λ_{a_1} we obtain

$$\lambda_{a_1}=f^a(\lambda_{a_2})=\frac{c}{d}-\frac{1}{\tau}\mathcal{W}_r\left(\tau\frac{b}{d}e^{\tau\frac{c}{d}}\right), \tag{C.2}$$

where $\mathcal{W}_r(\cdot)$ is the Lambert W function evaluated at branch $r\in\{-1,0\}$. Selecting the appropriate branch in the Lambert W function is not trivial. The criterion to select r is the analysis of the singularity at the α-level set of the total sojourn time distribution. Let $(\lambda^*_{a_1},\lambda^*_{a_2})$ be such a singularity. The following equations must be satisfied at the α-level set:

$$1-e^{-(\mu_{a_1}-\lambda^*_{a_1})\tau}-(\mu_{a_1}-\lambda^*_{a_1})\tau e^{-(\mu_{a_1}-\lambda^*_{a_1})\tau}=\alpha, \qquad 1-e^{-(\mu_{a_2}-\lambda^*_{a_2})\tau}-(\mu_{a_2}-\lambda^*_{a_2})\tau e^{-(\mu_{a_2}-\lambda^*_{a_2})\tau}=\alpha.$$

The solution for $\lambda^*_{a_1}$ and $\lambda^*_{a_2}$ in previous equations is:

$$\lambda^*_{a_1}=\frac{1}{\tau}\left[1+\mu_{a_1}\tau+\mathcal{W}_{-1}\left(\frac{\alpha-1}{e}\right)\right], \qquad \lambda^*_{a_2}=\frac{1}{\tau}\left[1+\mu_{a_2}\tau+\mathcal{W}_{-1}\left(\frac{\alpha-1}{e}\right)\right]. \tag{C.3}$$

To understand why branch $\mathcal{W}_{-1}$ should be used, note that the first equation can be written as $(\lambda^*_{a_1}-\mu_{a_1})\tau=1+\mathcal{W}_{-1}\left[(\alpha-1)/e\right]$, given that $\lambda^*_{a_1}<\mu_{a_1}$ the left-hand side of such equation is strictly negative, so it is the right-hand side. Then, we must have $\mathcal{W}_r\left[(\alpha-1)/e\right]<-1$. Given that $-1/e<(\alpha-1)/e<0$, the previous inequality is satisfied only when the branch $r=-1$ is considered. According to the previous analysis, we have found that the correct branch selection in the Lambert W function in Equation (C.2) can be done according to the following rule:

$$r=\begin{cases}-1, & \text{if } \lambda_{a_2}\le\lambda^*_{a_2},\\ 0, & \text{if } \lambda_{a_2}>\lambda^*_{a_2}.\end{cases}$$

Function f^a has asymptotes at $\lambda_{a_1} = \ln(1-\alpha)/\tau + \mu_{a1}$ and $\lambda_{a_2} = \ln(1-\alpha)/\tau + \mu_{a2}$. An explicit expression for f^a and its first derivative is

$$f^a(\lambda_{a_2}) = \mu_{a_1} + \frac{(1-\alpha)(\mu_{a_2}-\lambda_{a_2})}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)} - \frac{1}{\tau}\mathcal{W}_r\left\{\frac{(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\mathrm{Exp}\left[\frac{(1-\alpha)(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\right]\right\}, \tag{C.4}$$

$$\begin{aligned} f_1^a(\lambda_{a_2}) = {} & \frac{(1-\alpha)^2 - (1-\alpha)\left[1+(\mu_{a_2}-\lambda_{a_2})\tau\right]e^{-(\mu_{a_2}-\lambda_{a_2})\tau}}{\left[e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)\right]^2} + \\ & \frac{\mathcal{W}_r\left\{\frac{(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\mathrm{Exp}\left[\frac{(1-\alpha)(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\right]\right\}}{1+\mathcal{W}_r\left\{\frac{(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\mathrm{Exp}\left[\frac{(1-\alpha)(\mu_{a_2}-\lambda_{a_2})\tau}{e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)}\right]\right\}} \times \\ & \frac{\left\{(1-\alpha) - \left[1+(\mu_{a_2}-\lambda_{a_2})\tau\right]e^{-(\mu_{a_2}-\lambda_{a_2})\tau}\right\}\left\{e^{-(\mu_{a_2}-\lambda_{a_2})\tau} - (1-\alpha)\left[1-(\mu_{a_2}-\lambda_{a_2})\tau\right]\right\}}{-\tau(\mu_{a_2}-\lambda_{a_2})\left[e^{-(\mu_{a_2}-\lambda_{a_2})\tau}-(1-\alpha)\right]^2}. \end{aligned} \tag{C.5}$$

D Proof of Proposition 4

Consider parameter τ and let $g(x, y \mid \tau)$ be a function of variables x and y defined as follows

$$g(x, y \mid \tau) = \frac{xe^{-\tau y} - ye^{-\tau x}}{x-y}.$$

The total service time distribution $W_{V_a}(\tau \mid \lambda_{a_1}, \lambda_{a_2})$ is related to function g as $W_{V_a}(\tau \mid \lambda_{a_1}, \lambda_{a_2}) = 1 - g(\mu_{a_1} - \lambda_{a_1}, \mu_{a_2} - \lambda_{a_2})$. This corresponds to translations and reflections of function g, where the origin of coordinates $(0, 0)$ is translated to the point (μ_{a_1}, μ_{a_2}). Given that W_{V_a} is an affine transformation of g, the homothetic properties of the level sets of g are preserved after the transformation. We will show that the α-levels sets of $g(x, y \mid \tau)$ and $g(x, y \mid \tau')$, for $0 < \tau, 0 < \tau' \neq \tau$, are homothetic with respect to the origin. This is equivalent to showing that any ray from the origin cuts each α-level set at points with equal slope.

Assume that $\tau = a\tau'$, for $a > 0$. According to this, we have the following relation

$$\begin{aligned} g(x, y \mid \tau') &= \frac{xe^{-\tau' y} - ye^{-\tau' x}}{x-y}, \\ &= \frac{xe^{-(\tau/a)y} - ye^{-(\tau/a)x}}{x-y}, \\ &= \frac{(x/a)e^{-\tau(y/a)} - (y/a)e^{-\tau(x/a)}}{(x/a)-(y/a)}, \\ &= g(x/a, y/a \mid \tau). \end{aligned} \tag{D.1}$$

At the α-level set we have $g(x, y \mid \tau') = g(x/a, y/a \mid \tau) = \alpha$. Consider points (x_0, y_0) and $(x_0/a, y_0/a)$, which are collinear to the origin, or are contained in a ray from the origin. If point (x_0, y_0) is in the α-level set of $g(x, y \mid \tau')$, then point $(x_0/a, y_0/a)$ is also in the α-level set of $g(x, y \mid \tau)$. In this sense, one level set can be interpreted as the projection from the origin of the other level set, and the origin of coordinates is also referred to as the homothetic center of g.

Given that W_{V_a} transforms g in a way that the origin of coordinates is translated to the point (μ_{a_1}, μ_{a_2}), we conclude that the α-level sets of $W_{V_a}(\tau \mid \lambda_{a_1}, \lambda_{a_2})$ and $W_{V_a}(\tau' \mid \lambda_{a_1}, \lambda_{a_2})$ are homothetic concerning the point (μ_{a_1}, μ_{a_2}). We refer to point (μ_{a_1}, μ_{a_2}) as the homothetic center of a family of α-level sets of W_{V_a} defined for distinct values of parameter τ.

E Total sojourn time distribution for $M/M/s$ queues

Let

$$
\begin{aligned}
x &= s_{a_1}\mu_{a_1} - \lambda_{a_1}, \\
y &= s_{a_2}\mu_{a_2} - \lambda_{a_2}, \\
p_{0,a_1} &= \left[\frac{\left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{x}{x+\lambda_{a_1}}} + \sum_{n=0}^{s_{a_1}-1} \left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^n \right]^{-1}, \\
W_{q,a_1}(0) &= 1 - \frac{\left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{x}{x+\lambda_{a_1}}} p_{0,a1}, \\
A_{a_1} &= \frac{s_{a_1}x}{(s_{a_1}-1)x - \lambda_{a_1}} - \frac{x+\lambda_{a_1}}{(s_{a_1}-1)x - \lambda_{a_1}} W_{q,a_1}(0), \\
B_{a_1} &= \frac{1 - W_{q,a_1}(0)}{\frac{s_{a_1}x}{x+\lambda_{a_1}} - 1}.
\end{aligned}
$$

For computing $p_{0,a_2}, W_{q,a_2}, A_{a_2}$, and B_{a_2}, the s_{a_1}, x and λ_{a_1} should be replaced by s_{a_2}, y and λ_{a_2}, respectively. Using the convolution integral, we obtain the following total sojourn time distribution:

$$
\begin{aligned}
W_{V_a}(\tau_a) = A_{a_1}A_{a_2} &\left\{ \frac{s_{a_2}(x+\lambda_{a_1})}{s_{a_2}(x+\lambda_{a_1}) - s_{a_1}(y+\lambda_{a_2})} \left[1 - e^{-\left(\frac{y+\lambda_{a_2}}{s_{a_2}}\right)\tau}\right] - \right. \\
&\left. \frac{s_{a_1}(y+\lambda_{a_1})}{s_{a_2}(x+\lambda_{a_1}) - s_{a_1}(y+\lambda_{a_2})} \left[1 - e^{-\left(\frac{x+\lambda_{a_1}}{s_{a_1}}\right)\tau}\right] \right\} - \\
A_{a_1}B_{a_2} &\left\{ \frac{s_{a_1}y}{s_{a_1}y - x - \lambda_{a_1}} \left[1 - e^{-\left(\frac{x+\lambda_{a_1}}{s_{a_1}}\right)\tau}\right] - \frac{x+\lambda_{a_1}}{s_{a_1}y - x - \lambda_{a_1}} \left(1 - e^{-y\tau}\right) \right\} - \\
A_{a_2}B_{a_1} &\left\{ \frac{s_{a_2}x}{s_{a_2}x - y - \lambda_{a_2}} \left[1 - e^{-\left(\frac{y+\lambda_{a_2}}{s_{a_2}}\right)\tau}\right] - \frac{y+\lambda_{a_2}}{s_{a_2}x - y - \lambda_{a_2}} \left(1 - e^{-x\tau}\right) \right\} + \\
B_{a_1}B_{a_2} &\left[\frac{x}{x-y}\left(1 - e^{-y\tau}\right) - \frac{y}{x-y}\left(1 - e^{-x\tau}\right) \right]. \qquad \text{(E.1)}
\end{aligned}
$$

The previous distribution has a singularity line $s_{a_1}\mu_{a_1} - \lambda_{a_1} = s_{a_2}\mu_{a_2} - \lambda_{a_2}$, and the point $(\lambda^*_{a_1}, \lambda^*_{a_2})$ is denoted as the singularity point at the α-level set $E_\alpha^{W_{V_a}}$. Contrary to the $M/M/1$ case, the previous distribution is symmetric along the singularity line only if each hub has the same number of servers, i.e., $s_{a_1} = s_{a_2}$. In this case, it is not possible to obtain a closed-form functional expression for the α-level set and can be computed using numerical methods. This is described in the next section and in Appendix F. It is difficult to prove the convexity or quasiconvexity of the total sojourn time and this remains a conjecture.

Proof of Proposition 5

First, for a constant $a > 0$, note the validity of the following identities

$$\begin{aligned}
ax &= s_{a_1} a\mu_{a_1} - a\lambda_{a_1},\\
ay &= s_{a_2} a\mu_{a_2} - a\lambda_{a_2},\\
p_{0,a_1} &= \left[\frac{\left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{x}{x+\lambda_{a_1}}} + \sum_{n=0}^{s_{a_1}-1}\left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^n\right]^{-1} = \left[\frac{\left(\frac{a\lambda_{a_1}}{ax+a\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{ax}{ax+a\lambda_{a_1}}} + \sum_{n=0}^{s_{a_2}-1}\left(\frac{a\lambda_{a_1}}{ax+a\lambda_{a_1}}\right)^n\right]^{-1},\\
W_{q,a_1}(0) &= 1 - \frac{\left(\frac{\lambda_{a_1}}{x+\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{x}{x+\lambda_{a_1}}} p_{0,a1} = 1 - \frac{\left(\frac{a\lambda_{a_1}}{ax+a\lambda_{a_1}}\right)^{s_{a_1}}}{s_{a_1}!\frac{ax}{ax+a\lambda_{a_1}}} p_{0,a1},\\
A_{a_1} &= \frac{s_{a_1}x}{(s_{a_1}-1)x - \lambda_{a_1}} - \frac{x+\lambda_{a_1}}{(s_{a_1}-1)x-\lambda_{a_1}} W_{q,a_1}(0) = \frac{s_{a_1}ax}{(s_{a_1}-1)ax - a\lambda_{a_1}} - \frac{ax+a\lambda_{a_1}}{(s_{a_1}-1)ax - a\lambda_{a_1}} W_{q,a_1}(0),\\
B_{a_1} &= \frac{1-W_{q,a_1}(0)}{\frac{s_{a_1}x}{x+\lambda_{a_1}} - 1} = \frac{1-W_{q,a_1}(0)}{\frac{s_{a_1}ax}{ax+a\lambda_{a_1}} - 1}.
\end{aligned}$$

From the first two equations it is clear that multiplying x (y) by the constant a has the same effect on λ_{a_1} (λ_{a_2}). The rest of the proof follows from a similar reasoning to the proof of Proposition 4.

F Numerical evaluation of $\lambda_{a_1} = f^{ka}(\lambda_{a_2})$

The idea is to compute λ_{a_1} for a given value of λ'_{a_2}, with the condition that $(\lambda_{a_1}, \lambda'_{a_2}) \in E_\alpha^{S_{\tau_k}}(a)$. In this sense, λ_{a_1} is the solution of the equation

$$S_{T_{ka}}(\tau_k \mid \lambda_{a_1}, \lambda'_{a_2}) = \alpha. \tag{F.1}$$

Note that the total service time distribution is evaluated at the service time requirement τ_k, which is the value of interest for establishing probabilistic service level constraints. In other words, we want to find the root of the function $S_{T_{ka}}(\tau_k \mid \bullet, \lambda'_{a_2}) - \alpha$. We use Brent's root finding algorithm to solve for λ_{a_1} in Equation (F.1). By the monotonicity property of $S_{T_{ka}}(\tau_k \mid \bullet, \lambda'_{a_2})$, we know that at most one root can be found in the interval $(0, \mu_{a_1})$. No solution may exist for the previous equation. In that case, we say that λ'_{a_2} is not feasible. We use the following notation to write function $f^{ka}(\lambda_{a_2})$:

$$f^{ka}(\lambda_{a_2}) = \mathrm{Root}\left[S_{T_{ka}}(\tau_k \mid \bullet, \lambda_{a_2}) - \alpha\right]. \tag{F.2}$$

It will also be necessary to compute the first derivative of f^{ka}, denoted as f_1^{ka}. We use the finite difference method to obtain a derivative. Just as f^a, f^{ka} has a singularity at point $\left(\lambda^*_{ka_1}, \lambda^*_{ka_1}\right)$. There is no closed-form expression for such a singularity, and it is computed using numerical methods, see Appendix H.

G Proof of Proposition 6

The total service time distribution for commodities k and l is:

$$\int_0^{\tau_k} W_{V_a}\left(\tau_k - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{ka}}(x)dx \quad \text{and} \quad \int_0^{\tau_l} W_{V_a}\left(\tau_l - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{la}}(x)dx,$$

respectively. Let $\tau_l = a\tau_k, a \geq 1$, and consider an arbitrary point $(\lambda_{a_1}, \lambda_{a_2}) \in D$, then

$$\begin{aligned}
\int_0^{\tau_k} W_{V_a}\left(\tau_k - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{ka}}(x)dx &= \int_0^{a\tau_k} W_{V_a}\left(\tau_k - \frac{u}{a} \mid \lambda_{a_1}, \lambda_{a_2}\right) \cdot \frac{1}{a} \cdot g_{U_{ka}}\left(\frac{u}{a}\right) du, \\
&= \int_0^{\tau_l} W_{V_a}\left(\frac{\tau_l - u}{a} \mid \lambda_{a_1}, \lambda_{a_2}\right) \hat{g}_{U_{ka}}(u)du, \\
&\leq \int_0^{\tau_l} W_{V_a}\left(\tau_l - u \mid \lambda_{a_1}, \lambda_{a_2}\right) \hat{g}_{U_{ka}}(u)du,
\end{aligned}$$

where the first equality comes from the substitution $u = ax$, the term $\hat{g}_{U_{ka}}$ in the second equality corresponds to the density function $g_{U_{ka}}$ scaled by a factor a. Now, we show that the following inequality holds:

$$\int_0^{\tau_l} W_{V_a}\left(\tau_l - x \mid \lambda_{a_1}, \lambda_{a_2}\right) \hat{g}_{U_{ka}}(x)dx \leq \int_0^{\tau_l} W_{V_a}\left(\tau_k - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{la}}(x)dx$$

Consider

$$\begin{aligned}
\int_0^{\tau_l} W_{V_a}\left(\tau_k - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{la}}(x)dx - &\int_0^{\tau_l} W_{V_a}\left(\tau_l - x \mid \lambda_{a_1}, \lambda_{a_2}\right) \hat{g}_{U_{ka}}(x)dx, \\
&= \int_0^{\tau_l} W_{V_a}\left(\tau_l - u \mid \lambda_{a_1}, \lambda_{a_2}\right) \left[g_{U_{la}}(u) - \hat{g}_{U_{ka}}\right] du, \\
&= \int_0^{\tau_l} w_{V_a}\left(\tau_l - u \mid \lambda_{a_1}, \lambda_{a_2}\right) \left[G_{U_{la}}(u) - \hat{G}_{U_{ka}}(u)\right] du, \\
&= \int_0^{\tau_l} w_{V_a}\left(\tau_l - u \mid \lambda_{a_1}, \lambda_{a_2}\right) \left[G_{U_{la}}(u) - G_{U_{ka}}\left(\frac{u}{a}\right)\right] du, \\
&\geq 0.
\end{aligned}$$

The second equality is the result of the application of integration by parts. This allows us to conclude that

$$\int_0^{\tau_k} W_{V_a}\left(\tau_k - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{ka}}(x)dx \leq \int_0^{\tau_l} W_{V_a}\left(\tau_l - x \mid \lambda_{a_1}, \lambda_{a_2}\right) g_{U_{la}}(x)dx.$$

Then $S_{T_{ka}}(\tau_k \mid \lambda_{a_1}, \lambda_{a_2}) \leq S_{T_{la}}(\tau_l \mid \lambda_{a_1}, \lambda_{a_2})$, completing the proof.

H Asymptotic analysis for the approximation error in an $M/M/1$ queue

From Section 3.1, and Figure 2, it is known that the α-level set of the total sojourn time distribution E_α^W shows an asymptotic behavior. We consider this behavior to analyze the asymptotic error that results from computing the total service time distribution by assuming a constant transport time. The following analysis is done for the $M/M/1$ queue only.

Let us use the usual substitution $x = \mu_{a_1} - \lambda_{a_1}$ and $y = \mu_{a_2} - \lambda_{a_2}$. For the asymptotic analysis, we let variables x and y be defined on the interval $[0, \infty)$. The total service time distribution is:

$$S_{T_{ka}}(\tau_k \mid x, y) = \int_0^{\tau_k} \left(1 - \frac{xe^{-(\tau_k - v)y} - ye^{(\tau_k - v)x}}{x - y}\right) g_{U_{ka}}(v)\, dv.$$

Note that we use τ_k instead of $\tilde{\tau}_{ka}$ since we are computing the exact value of the total service time distribution. To compute the asymptotic limit for the y variable proceed as follows

$$\begin{aligned}\lim_{x\to\infty}\int_0^{\tau_k}\left(1-\frac{xe^{-(\tau_k-v)y}-ye^{(\tau_k-v)x}}{x-y}\right)g_{U_{ka}}(v)\,dv &= \int_0^{\tau_k}\lim_{x\to\infty}\left(1-\frac{xe^{-(\tau_k-v)y}-ye^{(\tau_k-v)x}}{x-y}\right)g_{U_{ka}}(v)\,dv,\\ &= \int_0^{\tau_k}\left(1-e^{-(\tau_k-v)y}\right)g_{U_{ka}}(v)\,dv.\end{aligned}$$

The first equation is valid whenever $g_{U_{ka}}$ is bounded, as expected in practice. The result of this limit is the convolution of an exponential distribution with rate parameter y and the transportation time PDF, i.e. the total service time distribution when one of the hubs (in this case hub a_1) has an infinite service rate. Now, given that we are interested in obtaining the asymptotic limit of the α-level set, we want the previous limit to satisfy

$$\int_0^{\tau_k}\left(1-e^{-(\tau_k-v)y}\right)g_{U_{ka}}(v)\,dv=\alpha,$$

where the left-hand side is strictly increasing in y. Then the value of y satisfying this equation corresponds to the asymptotic limit of E_α^S. The asymptotic limit for x is the same as for y. Let L_{ka} denote such a limit, whose value depends on the specific form of the PDF of transportation time.

Now to compute the asymptotic error over E_α^S of the approximated total service time distribution $\tilde{S}_{T_{ka}}(\tau_k \mid x,y)$ we first obtain the asymptotic limits for $E_\alpha^{\tilde{S}}$. Taking the limit concerning the x variable we obtain

$$\lim_{x\to\infty}\tilde{S}_{T_{ka}}(\tau_k \mid x,y)=\lim_{x\to\infty}W_{V_a}(\tilde{\tau}_{ka}\mid x,y)=1-e^{-\tilde{\tau}_{ka}y}. \tag{H.1}$$

Then the asymptotic error over E_α^S, which we denote as $\mathscr{E}_{\uparrow E_\alpha^S}$, is given by

$$\begin{aligned}\mathscr{E}_{\uparrow E_\alpha^S} &= \left|1-e^{-\tilde{\tau}_{ka}L_{ka}}-\alpha\right|,\\ &= \left|1-e^{\frac{L_{ka}}{L_{ka}^*}\left[1+\mathcal{W}_{-1}\left(\frac{\alpha-1}{e}\right)\right]}-\alpha\right|.\end{aligned} \tag{H.2}$$

The second equality results from the direct substitution of $\tilde{\tau}_{ka}$ in Equation (13), where (L_{ka}^*, L_{ka}^*) is the singularity at E_α^S, which is computed by solving for y the following equation

$$\int_0^{\tau_k}\left[1-e^{-(\tau_k-v)y}-(\tau_k-v)ye^{-(\tau_k-v)y}\right]g_{U_{ka}}(v)\,dv=\alpha. \tag{H.3}$$

I Proof of Proposition 9

We first show that $\tilde{\Delta}_l^{\mathbf{r}}\geq 0$. Let $[l]=e(ka)$ and $[l-1]=e(k'a)$. By Equation (30), $\tilde{\delta}_{[l]}^{\mathbf{r}}=\tilde{\delta}_{ka}^{\mathbf{r}}=L_{a_1}-\lambda_{ka_1}^{\mathbf{r}}+|s_a^{\mathbf{r}}|\left(L_{a_2}-\lambda_{ka_2}^{\mathbf{r}}\right)\geq 0$. According to Proposition 4, the point $\left(\lambda_{ka_1}^{\mathbf{r}},\lambda_{ka_2}^{\mathbf{r}}\right)$ is the projection from the homothetic center (μ_{a_1},μ_{a_2}) of the point $\left(\lambda_{k'a}^{\mathbf{r}},\lambda_{k'a_2}^{\mathbf{r}}\right)$, which is computed as

$$\left(\lambda_{ka_1}^{\mathbf{r}},\lambda_{ka_2}^{\mathbf{r}}\right)=\left[\mu_{a_1}+\left(\lambda_{k'a_1}^{\mathbf{r}}-\mu_{a_1}\right)\frac{\tilde{\tau}_{k'a}}{\tilde{\tau}_{ka}},\mu_{a_2}+\left(\lambda_{k'a_2}^{\mathbf{r}}-\mu_{a_2}\right)\frac{\tilde{\tau}_{k'a}}{\tilde{\tau}_{ka}}\right].$$

By Proposition 8, $\tilde{\tau}_{ka} \leq \tilde{\tau}_{k'a}$, then $\lambda^{\mathbf{r}}_{ka_1} \leq \lambda^{\mathbf{r}}_{k'a_1}$ and $\lambda^{\mathbf{r}}_{ka_2} \leq \lambda^{\mathbf{r}}_{k'a_2}$. This implies that $\tilde{\delta}^{\mathbf{r}}_{e(ka)} \geq \tilde{\delta}^{\mathbf{r}}_{e(k'a)}$ and $\tilde{\Delta}^{\mathbf{r}}_{e(ka)} = \tilde{\Delta}^{\mathbf{r}}_{[l]} \geq 0$.

Assume that $v_{[l]} = 1$, for $l = 1, 2, \ldots, l'$. From the perspective cut in Equation (27) we have

$$\begin{aligned}
L_{a_1} - \lambda_{a_1} + |s^{\mathbf{r}}_a|(L_{a_2} - \lambda_{a_2}) &\geq \left[L_{a_1} - \lambda^{\mathbf{r}}_{[k']a_1} + |s^{\mathbf{r}}_a| \left(L_{a_2} - \lambda^{\mathbf{r}}_{[k']a_2}\right)\right] v_{[l']} \\
&= \tilde{\delta}^{\mathbf{r}}_{[l']} v_{[l']}, \\
&= \sum_{l=1}^{|\mathscr{L}_a|-1} \tilde{\delta}^{\mathbf{r}}_{[l]} \left(v_{[l]} - v_{[l+1]}\right) + \tilde{\delta}^{\mathbf{r}}_{|\mathscr{L}_a|} v_{|\mathscr{L}_a|}, \\
&= \tilde{\delta}^{\mathbf{r}}_{[1]} v_{[1]} + \sum_{l=2}^{|\mathscr{L}_a|} \left(\tilde{\delta}^{\mathbf{r}}_{[l]} - \tilde{\delta}^{\mathbf{r}}_{[l-1]}\right) v_{[l]} = \sum_{l=1}^{|\mathscr{L}_a|} \tilde{\Delta}^{\mathbf{r}}_{[l]} v_{[l]} = \sum_{l \in \mathscr{L}_a} \tilde{\Delta}^{\mathbf{r}}_{l} v_l,
\end{aligned}$$

where the second equality is valid due to the incremental nature of binary variables $v_{[l]}$. Reorganizing terms we get

$$\lambda_{a_1} + |s^{\mathbf{r}}_a| \lambda_{a_2} + \sum_{l \in \mathscr{L}_a} \tilde{\Delta}^{\mathbf{r}}_l v_l \leq L_{a_1} + |s^{\mathbf{r}}_a| L_{a_2}. \tag{I.1}$$

Multiplying L_{a_1} and L_{a_2} by z_{a_1} and z_{a_2}, respectively, in the right-hand side of the previous is valid and allows to obtain Inequality (31), as desired.

J Cutting plane algorithms for model $M2$

See Algorithm 1.

Algorithm 1: Cutting plane generation at integer solutions of the branching tree of $M2$

Data: $\mathscr{B}, \mathscr{L}_o, \tilde{\varepsilon}, \alpha, (\bar{\mathbf{x}}, \bar{\mathbf{v}}, \bar{\mathbf{z}})$, parameters related to the total service time distribution.

for $o = (m, n) \in \mathscr{B}$ **do**
    if $\bar{z}_m = 1$ *and* $\bar{z}_n = 1$ **then**
        $\lambda_m \leftarrow \sum_{k \in K} \sum_{\substack{a \in \mathscr{A}_k, \\ m \in a}} w_k \bar{x}_{ka}$ and $\lambda_n \leftarrow \sum_{k \in K} \sum_{\substack{a \in \mathscr{A}_k, \\ n \in a}} w_k \bar{x}_{ka}$;
        for $l \in \mathscr{L}_o$ **do**
            if $\bar{v}_l = 1$ *and* $W_{V_o}(\tilde{\tau}_{lo} \mid \lambda_m, \lambda_n) < \alpha - \tilde{\varepsilon}$ **then**
                $(\lambda^{\mathbf{r}}_{lm}, \lambda^{\mathbf{r}}_{ln}) \leftarrow \left[\tilde{f}^{lo}\left(\min\left\{\lambda_n, \tilde{\lambda}^{\max}_{ln}\right\}\right), \min\left\{\lambda_n, \tilde{\lambda}^{\max}_{ln}\right\}\right]$;
                Add cut (31);
                break;
            end
        end
    end
end

K Proof of Proposition 10

Phase type inter-arrival time distribution

Let us first describe the properties of the arrivals and service time distributions. For further details on phase-type distributions see Chapter 3 of Bladt and Nielsen (2017). The PDF and mean value of inter-arrival times at hub m are respectively

$$\text{PDF:} \quad f_m(t) = \boldsymbol{\alpha}_m e^{t\mathbf{T}_m}(-\mathbf{T}_m\mathbf{1}), \tag{K.1}$$

$$\text{Mean:} \quad \frac{1}{\lambda_m} = -\boldsymbol{\alpha}_m\mathbf{T}_m^{-1}\mathbf{1}, \tag{K.2}$$

where $\mathbf{1}$ refer to the vector of the appropriate size with 1s as its elements. For its part, the PDF and mean value of service times at hub m are respectively

$$\text{PDF:} \quad b_m(t) = \boldsymbol{\beta}_m e^{t\mathbf{S}_m}(-\mathbf{S}_m\mathbf{1}), \tag{K.3}$$

$$\text{Mean:} \quad \frac{1}{\mu_m} = -\boldsymbol{\beta}_m\mathbf{S}_m^{-1}\mathbf{1}. \tag{K.4}$$

We use the usual substitution $x = \mu_{a_1} - \lambda_{a_1}$ and $y = \mu_{a_1} - \lambda_{a_1}$ and show homotheticity with the origin as the homothetic center, as we have done in the previous propositions. First, consider some constant a and note that the validity of the following identities:

$$\begin{aligned} ax = a\mu_{a_1} - a\lambda_{a_1} &= -\frac{1}{\boldsymbol{\beta}_{a_1}\left(a\mathbf{S}_{a_1}\right)^{-1}\mathbf{1}} + \frac{1}{\boldsymbol{\alpha}_{a_1}\left(a\mathbf{T}_{a_1}\right)^{-1}\mathbf{1}}, \\ ay = a\mu_{a_2} - a\lambda_{a_2} &= -\frac{1}{\boldsymbol{\beta}_{a_2}\left(a\mathbf{S}_{a_2}\right)^{-1}\mathbf{1}} + \frac{1}{\boldsymbol{\alpha}_{a_2}\left(a\mathbf{T}_{a_2}\right)^{-1}\mathbf{1}}. \end{aligned} \tag{K.5}$$

Clearly, multiplying x (y) by a factor a can be understood as the effect of scaling matrices $\mathbf{S}_{a_1}$ and $\mathbf{T}_{a_1}$ ($\mathbf{S}_{a_2}$ and $\mathbf{T}_{a_2}$) by the same factor. As in previous propositions, this is a core idea for the proof.

We recall that the distribution of the waiting time in the queue at hub m is computed as

$$W_{q,m}(t) = 1 - \sum_{n=0}^{\infty} d_{n,m}\frac{(\theta_m t)^n}{n!}e^{-\theta_m t}. \tag{K.6}$$

From Lucantoni (1985) and Ramaswami (1985), we have that $\theta_m = \max_i\{(-\mathbf{S}_m)_{ii}\} \geq 0$, i.e., θ_m is the maximum among the negative diagonal elements of matrix $\mathbf{S}_m$. The computation of probabilities $d_{n,m}$ is done through a recursive process using the following equations:

$$d_{n,m} = c_m\boldsymbol{\alpha}_m\left(\mathbf{I} - \mathbf{R}_m\right)^{-1}\mathbf{R}_m\mathbf{H}_{n,m}\mathbf{1}, \tag{K.7a}$$

$$c_m^{-1} = \boldsymbol{\alpha}_m\left(\mathbf{I} - \mathbf{R}_m\right)^{-1}\mathbf{1}, \tag{K.7b}$$

$$\mathbf{R}_m = \sum_{n=0}^{\infty}\gamma_{n,m}\mathbf{H}_{n,m}, \tag{K.7c}$$

$$\mathbf{H}_{0,m} = \mathbf{I}, \qquad \mathbf{H}_{n+1,m} = \mathbf{H}_{n,m}\mathbf{P}_m + \mathbf{R}_m\mathbf{H}_{n,m}\mathbf{P}_m^{\circ}\boldsymbol{\alpha}_m, \qquad n \geq 0, \tag{K.7d}$$

$$\mathbf{P}_m = \frac{1}{\theta_m}\mathbf{S}_m + \mathbf{I}, \qquad \mathbf{P}_m^{\circ} = \mathbf{1} - \mathbf{P}_m\mathbf{1}, \tag{K.7e}$$

$$\gamma_{n,m} = \int_0^{\infty}\frac{(\theta_m t)^n}{n!}e^{-\theta_m t}f_m(t)\,dt, \tag{K.7f}$$

where $\mathbf{I}$ refers to the identity matrix of appropriate size. The following result is required to proceed with the proof.

Proposition 1. *The probabilities $d_{n,m}$ are insensitive to scalar multiplications on matrices $\mathbf{S}_m$ and $\mathbf{T}_m$.*

Proof. Proof Note that $d_{n,m}$ depends on the matrices $\mathbf{S}_m$ and $\mathbf{T}_m$ only through Equations (K.7e) and (K.7f), respectively. From Equation (K.7e) $\mathbf{P}_m$ is insensitive to scalar multiplications on $\mathbf{S}_m$ since

$$\mathbf{P}_m = \frac{1}{\theta_m}\mathbf{S}_m + \mathbf{I} = \frac{1}{(a\theta_m)}(a\mathbf{S}_m) + \mathbf{I},$$

where a is a constant. Now we proceed to prove the insensitivity of $\gamma_{n,m}$ in Equation (K.7f), given that the PDF $f_m(t)$ is of phase type (Equation (K.1)), we solve the integral for $\gamma_{m,m}$ as follows:

$$\begin{aligned}
\gamma_{n,m} &= \int_0^\infty \frac{(\theta_m t)^n}{n!} e^{-\theta_m t} \boldsymbol{\alpha}_m e^{t\mathbf{T}_m}(-\mathbf{T}_m\mathbf{1})\,dt, \\
&= \frac{\theta_m^n}{n!}\boldsymbol{\alpha}_m \left(\int_0^\infty e^{-\theta_m t} t^n e^{t\mathbf{T}_m}\,dt\right)(-\mathbf{T}_m\mathbf{1}), \\
&= \frac{\theta_m^n}{n!}\boldsymbol{\alpha}_m \mathcal{L}\left\{t^n e^{t\mathbf{T}_m}\right\}(-\mathbf{T}_m\mathbf{1}), \\
&= \theta_m^n \boldsymbol{\alpha}_m \left(\theta_m\mathbf{I} - \mathbf{T}_m\right)^{-n-1}(-\mathbf{T}_m\mathbf{1}),
\end{aligned}$$

where the expression $\mathcal{L}\{\cdot\}$ on the third equality corresponds to the Laplace transform. This result is well-defined since $\theta_m \geq 0$, see Theorem 3.1.19 of Bladt and Nielsen (2017). Insensitivity is clear since

$$\begin{aligned}
\gamma_{n,m} &= \theta_m^n \boldsymbol{\alpha}_m \left(\theta_m\mathbf{I} - \mathbf{T}_m\right)^{-n-1}(-\mathbf{T}_m\mathbf{1}), \\
&= \theta_m^n \boldsymbol{\alpha}_m \left(\theta_m\mathbf{I} - \mathbf{T}_m\right)^{-n-1}(-\mathbf{T}_m\mathbf{1})\frac{a^{n+1}}{a^{n+1}}, \\
&= (a\theta_m)^n \boldsymbol{\alpha}_m \left[(a\theta_m)\mathbf{I} - (a\mathbf{T}_m)\right]^{-n-1}\left[-(a\mathbf{T}_m)\mathbf{1}\right],
\end{aligned}$$

which completes the proof. □

The total sojourn time distribution in hub arc $a = (a_1, a_2)$ is the convolution of the total waiting and service times at hubs a_1 and a_2, i.e., $W_{V_a} = W_{q,a_1} * w_{q,a_2} * b_{a_1} * b_{a_2}$. Since service times in each hub are of phase type, it is known that its convolution is also of phase type (see Theorem 3.1.26 of Bladt and Nielsen (2017)), the total service time PDF in a 2-hub path is given by

$$b_a = b_{a_1} * b_{a_2} = \boldsymbol{\alpha}_a e^{t\mathbf{S}_a}(-\mathbf{S}_a\mathbf{1}),$$

where

$$\boldsymbol{\alpha}_a = (\alpha_{a_1}, 0) \qquad \text{and} \qquad \mathbf{S}_a = \begin{pmatrix} \mathbf{S}_{a_1} & (-\mathbf{S}_{a_1}\mathbf{1})\boldsymbol{\alpha}_{a_2} \\ \mathbf{0} & \mathbf{S}_{a_2} \end{pmatrix}.$$

Now, to compute the waiting time PDF at hub a_2 we note that there might be a non-zero probability that a customer waits at $t = 0$ since $W_{q,a_2}(0) = 1 - d_{0,a_2}$. Taking this into account, the waiting time PDF at hub a_2 is

$$w_{q,a_2}(t) = (1 - d_{0,m})\delta(t) + d_{0,m}\frac{\partial W_{q,a_2}}{\partial t} = (1 - d_{0,m})\delta(t) + d_{0,m}\sum_{n=0}^{\infty} d_{n,a_2}\frac{(\theta_{a_2}t - n)(\theta_{a_2}t)^n}{tn!}e^{-\theta_{a_2}t}.$$

Assume that $\tau = a\tau'$, for $a > 0$. The total sojourn time distribution for time requirement τ satisfy

$$
\begin{aligned}
W_{V_a}(\tau \mid \mathbf{T}_{a_1}, \mathbf{T}_{a_2}, \mathbf{S}_{a_1}, \mathbf{S}_{a_2}) &= (W_{q,a_1} * w_{q,a_2}) * b_a, && \text{(K.8a)}\\
&= \int_0^\tau \int_0^\tau \left[1 - \sum_{n=0}^{\infty} d_{n,a_1} \frac{(\theta_{a_1} r)^n}{n!} e^{-\theta_{a_1} r}\right] \times \\
&\quad \left\{(1-d_{0,m})\delta(\tau - s) + d_{0,m} \sum_{n=0}^{\infty} d_{n,a_2} \frac{[\theta_{a_2}(\tau-s) - n]\,[\theta_{a_2}(\tau - s)]^n}{(\tau - s) n!} e^{-\theta_{a_2}(\tau - s)}\right\} \times \\
&\quad \left[\boldsymbol{\alpha}_a e^{(\tau - r)\mathbf{S}_a}(-\mathbf{S}_a \mathbf{1})\right] dr\, ds, && \text{(K.8b)}\\
&= \int_0^{a\tau'} \int_0^{a\tau'} \left[1 - \sum_{n=0}^{\infty} d_{n,a_1} \frac{\left(a\theta_{a_1} \frac{r}{a}\right)^n}{n!} e^{-\left(a\theta_{a_1}\frac{r}{a}\right)}\right] \times \\
&\quad \left\{(1-d_{0,m})\delta\left[a\left(\tau' - \frac{s}{a}\right)\right] + d_{0,m} \sum_{n=0}^{\infty} d_{n,a_2} \frac{\left[a\theta_{a_2}\left(\tau' - \frac{s}{a}\right) - n\right]\left[a\theta_{a_2}\left(\tau' - \frac{s}{a}\right)\right]^n}{a\left(\tau' - \frac{s}{a}\right) n!} e^{-a\theta_{a_2}\left(\tau' - \frac{s}{a}\right)}\right\} \times \\
&\quad \left[\boldsymbol{\alpha}_a e^{(\tau' - \frac{r}{a})a\mathbf{S}_a}(-\mathbf{S}_a \mathbf{1})\right] dr\, ds, && \text{(K.8c)}\\
&= \int_0^{\tau'} \int_0^{\tau'} \left[1 - \sum_{n=0}^{\infty} d_{n,a_1} \frac{(a\theta_{a_1} v)^n}{n!} e^{-\left(a\theta_{a_1} v\right)}\right] \times \\
&\quad \left\{(1-d_{0,m})\frac{\delta\left(\tau' - u\right)}{a} + d_{0,m} \sum_{n=0}^{\infty} d_{n,a_2} \frac{\left[a\theta_{a_2}\left(\tau' - u\right) - n\right]\left[a\theta_{a_2}\left(\tau' - u\right)\right]^n}{a\left(\tau' - u\right) n!} e^{-a\theta_{a_2}(\tau' - u)}\right\} \times \\
&\quad \left[\boldsymbol{\alpha}_a e^{(\tau' - v)a\mathbf{S}_a}(-\mathbf{S}_a \mathbf{1})\right] \times a^2 dv\, du, && \text{(K.8d)}\\
&= \int_0^{\tau'} \int_0^{\tau'} \left[1 - \sum_{n=0}^{\infty} d_{n,a_1} \frac{(a\theta_{a_1} v)^n}{n!} e^{-\left(a\theta_{a_1} v\right)}\right] \times \\
&\quad \left\{(1-d_{0,m})\delta\left(\tau' - u\right) + d_{0,m} \sum_{n=0}^{\infty} d_{n,a_2} \frac{\left[a\theta_{a_2}\left(\tau' - u\right) - n\right]\left[a\theta_{a_2}\left(\tau' - u\right)\right]^n}{\left(\tau' - u\right) n!} e^{-a\theta_{a_2}(\tau' - u)}\right\} \times \\
&\quad \left[\boldsymbol{\alpha}_a e^{(\tau' - v)a\mathbf{S}_a}(-a\mathbf{S}_a \mathbf{1})\right] dv\, du, && \text{(K.8e)}\\
&= W_{V_a}(\tau' \mid a\mathbf{T}_{a_1}, a\mathbf{T}_{a_2}, a\mathbf{S}_{a_1}, a\mathbf{S}_{a_2}), && \text{(K.8f)}
\end{aligned}
$$

where Equation (K.8d) results from the substitution $v = r/a$ and $u = s/a$, and by noting that $\delta(at) = \delta(t)/a$, for $a > 0$. From Equation (K.8e) it is noted that

$$
a\mathbf{S}_a = \begin{pmatrix} a\mathbf{S}_{a_1} & (-a\mathbf{S}_{a_1}\mathbf{1})\boldsymbol{\alpha}_{a_2} \\ \mathbf{0} & a\mathbf{S}_{a_2} \end{pmatrix},
$$

which implies Equation (K.8f). From Identities in (K.5) and at the γ_r-level, the previous result also implies $W_{V_a}(\tau \mid x, y) = W_{V_a}(\tau' \mid ax, ay) = \gamma_r$ proving homotheticity in the x-y plane with homothetic center $(0, 0)$. The level sets in the λ_{a_1}-λ_{a_2} plane have homothetic center (μ_{a_1}, μ_{a_2}), which results from the corresponding translations and reflections of the sojourn function.

Gamma inter-arrival time distribution

The PDF and mean value of gamma inter-arrival times with shape parameter l_m and scale parameter $1/k_m$ at hub m are respectively

$$\text{PDF:} \quad f_m(t) = \frac{e^{-k_m t} k_m^{l_m} r^{l_m - 1}}{\Gamma(l_m)}, \tag{K.9}$$

$$\text{Mean:} \quad \frac{1}{\lambda_m} = \frac{l_m}{k_m}, \tag{K.10}$$

We note the validity of the following identities

$$\begin{aligned} ax = a\mu_{a_1} - a\lambda_{a_1} &= -\frac{1}{\boldsymbol{\beta}_{a_1} \left(a\mathbf{S}_{a_1}\right)^{-1} \mathbf{1}} + \frac{ak_{a_1}}{l_{a_1}}, \\ ay = a\mu_{a_2} - a\lambda_{a_2} &= -\frac{1}{\boldsymbol{\beta}_{a_2} \left(a\mathbf{S}_{a_2}\right)^{-1} \mathbf{1}} + \frac{ak_{a_2}}{l_{a_2}}. \end{aligned} \tag{K.11}$$

The inter-arrival time distribution only affects the waiting time distribution in Equation (K.6) through the probabilities $d_{n,m}$. Then, the proof of homotheticity for a general inter-arrival time distribution reduces to the proof of insensitivity of probabilities $d_{n,m}$, similarly to Proposition 1. Now, probabilities $d_{n,m}$ are affected by the inter-arrival time distribution only through Equation (K.7f), which is

$$\begin{aligned} \gamma_{n,m} &= \int_0^\infty \frac{(\theta_m t)^n}{n!} e^{-\theta_m t} f_m(t)\, dt, \\ &= \frac{\theta_m^n}{n!} \mathcal{L}\left\{t^n f_m(t)\right\}(\theta_m), \\ &= \frac{\theta_m^n}{n!} \frac{k_m^{l_m} \Gamma(l_m + n)}{(k_m + \theta_m)^{l_m + n} \Gamma(l_m)}. \end{aligned}$$

Insensitivity of $\gamma_{n,m}$ and consequently $d_{n,m}$ is clear since

$$\begin{aligned} \gamma_{n,m} &= \frac{\theta_m^n}{n!} \frac{k_m^{l_m} \Gamma(l_m + n)}{(k_m + \theta_m)^{l_m + n} \Gamma(l_m)}, \\ &= \frac{\theta_m^n}{n!} \frac{k_m^{l_m} \Gamma(l_m + n)}{(k_m + \theta_m)^{l_m + n} \Gamma(l_m)} \frac{a^{l_m} a^n}{a^{l_m + n}}, \\ &= \frac{(a\theta_m)^n}{n!} \frac{(ak_m)^{l_m} \Gamma(l_m + n)}{(ak_m + a\theta_m)^{l_m + n} \Gamma(l_m)}. \end{aligned}$$

L Parameters for the probabilistic service level constraints

This corresponds to parameters required for establishing probabilistic service level constraints but is not available in standard datasets. These parameters are generated as described below:

- *Direct transportation time*: Standard datasets contain cost or distance information between different nodes, but no information associated with transportation times. We generate transport time data by using a travel speed defined as $v = 0.2\bar{\mu}\bar{C}$, where $\bar{\mu}$ and $\bar{C}$ are the average processing rate and the average direct transportation distance or cost, respectively. The motivation for defining the travel speed in this way is that there are no large differences between the total sojourn time distributions and the transportation time PDFs. The expected direct or truck-only transportation time between nodes i and j is computed as C_{ij}/v, where C_{ij} corresponds to the direct transportation distance or cost.

- *Intermodal transportation time*: The intermodal transportation time includes the collection, transfer, and distribution travel times. We assume that collection and distribution are done at the same travel speed, but the transfer, or transportation between hub nodes, is done at a slower speed. In this way, the expected transportation time for transport path ka is $E[U_{ka}] = E[U_{(i,j)(a_1,a_2)}] = \frac{1}{v}\left(c_{ia_1} + \eta c_{a_1a_2} + c_{a_2j}\right)$, where η is a time correction factor used to account for the differences in travel speeds between trucks and trains. We set $\eta = 1.5$, implying that trains (inter-hub vehicles) are slower than trucks (hub-and-spoke vehicles). This factor corresponds to the average ratio between the travel speeds of the different types of vehicles.

- *PDF of transportation time*: In all cases, the PDF of transportation time follows a gamma distribution with a fixed squared coefficient of variation of $SCV = 0.5$. Consider, for example, transportation path ka with an expected travel time of $E[U_{ka}]$. The shape parameter of the transportation time PDF for path ka is $a_{ka} = 1/SCV$ and the scale parameter is $b_{ka} = E[U_{ka}]/a_{ka}$. The gamma distribution is frequently used to describe travel time variability. The gamma distribution is attractive since it has positive support and can be easily determined from the mean travel time and SCV.

- *Service time requirement*: The maximum service time requirement for commodity k is defined as follows:
$$\tau_k = rG_k^{-1}\left(\ell \mid a_k, b_k\right), \tag{L.1}$$
where G_k^{-1} is the inverse of the cumulative gamma distribution function for the direct transportation time of commodity k, $a_k = 1/SCV$ and scale parameter $b_k = (C_k/v)(1/a_k)$. We set $\ell = 0.7$ and the values of factor r are selected from the set $r \in \{2, 3, 4\}$, where a small value of r leads to tight service time requirements, whereas a large value of r leads to loose service time requirements.

- *Service levels.* Service levels are selected from the set $\alpha \in \{0.80, 0.85, 0.90, 0.95, 0.99\}$.

M Accuracy of the homothetic approximation

We analyze the accuracy in the computation of the service level when the homothetic approximation $\lambda_{a_1} = \tilde{f}^{ka}(\lambda_{a_2})$ is considered instead of the actual function $\lambda_{a_1} = f^{ka}(\lambda_{a_2})$. The point $\left(\tilde{f}^{ka}(\lambda_{a_2}), \lambda_{a_2}\right)$ can be interpreted as an approximation to the point $\left(f^{ka}(\lambda_{a_2}), \lambda_{a_2}\right) \in E_\alpha^{S_{T_{ka}}}$. Hence, $S_{T_{ka}}\left(\tau_k \mid \tilde{f}^{ka}(\lambda_{a_2}), \lambda_{a_2}\right) = \tilde{\alpha}$ is an approximation to the actual service level α. To analyze the error $\alpha - \tilde{\alpha}$, we compute $\tilde{\alpha}$ from a sample of 100 points for $\tilde{f}^{ka}(\lambda_{a_2})$ evenly distributed in the interval $\lambda_{a_2} \in [0, \lambda_{ka_2}^{\max}(a_1)]$. The computations are done for all combinations of commodities and hub arcs on the 10 and 20-node instances of the AP and COL datasets, respectively. Figure M.1

shows that the average error is of the order of 10^{-3}, the maximum error is 0.004, and the minimum error is -1.0×10^{-6}.

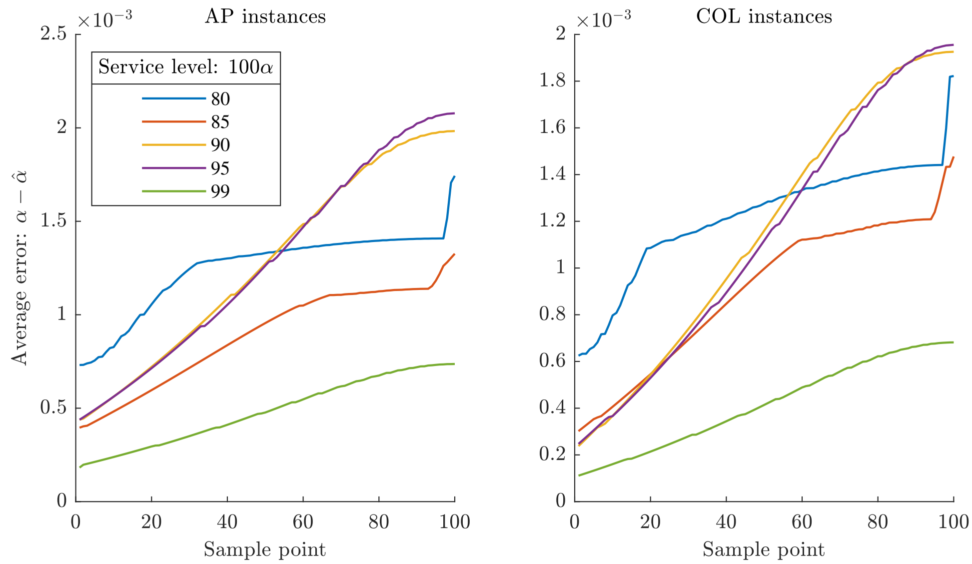

Figure M.1: Approximation error. Average errors are listed in ascending order.

N Additional sensitivity analysis

Effect of varying the service level on modal shift: The *modal shift* represents the fraction of flow captured by the intermodal (rail-road) transportation and is calculated as $\left(\sum_{k\in K}\sum_{a\in\mathcal{A}_k} w_k x_{ka}\right)\times 100\%/\sum_{k\in K} w_k$. Figure N.1 shows the effect of increasing service level requirements on the modal shift. As expected, as the service level requirement increases, the volume of flows through intermodal transportation decreases. For $r = 3$ and $r = 4$, the modal shift decreases at a diminishing rate when the service level α increases. This decrease in modal shift is more pronounced at elevated service levels and strict service time requirements. Conversely, for $r = 2$, the modal shift curve exhibits an inverted S-shape pattern. In this scenario, the modal shift rapidly decreases up to a certain point, after which it decreases more gradually, reaching a modal shift of zero where the intermodal transportation alternative becomes non-viable. The modal shift demonstrates an opposite behavior compared to the total cost.

Effect of considering splittable or non-splittable commodities:

All the results obtained so far pertain to the model where commodities can be split into shipments over several transport paths. Non-splittable orders refer to the case where a commodity cannot be split into several paths (i.e., $x_{ka} \in \{0,1\}$). Table 1 summarize the results for the splittable and non-splittable cases. The gap between the optimal solutions of the splittable and non-splittable case is denoted as “Gap” in the second column of the tables. It is noted that the gap is small, which is common in hub network design models because most flow variables tend to satisfy integrality constraints in splittable formulations. The solution time of the formulation with non-splittable

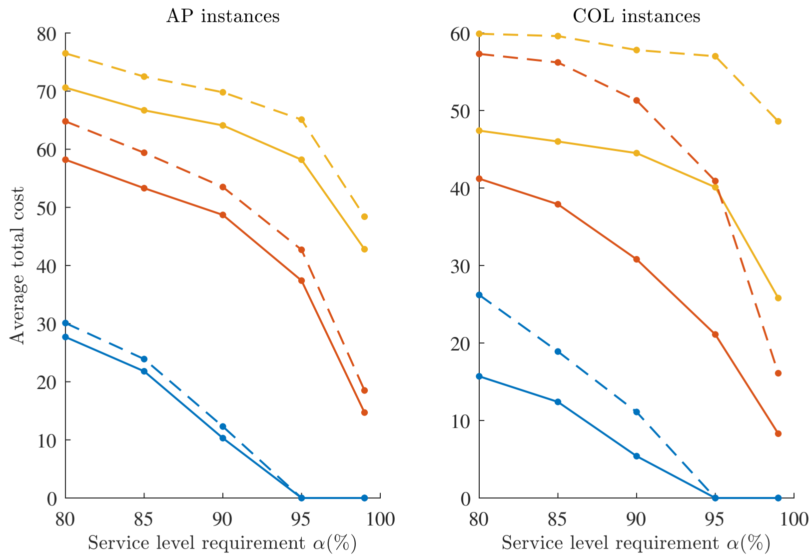

Figure N.1: Average modal shift. Solid lines are the results with probabilistic service level constraints. Dashed lines are the results without probabilistic service level constraints.

Table 1: Summary comparison between the splittable and non-splittable results (average results).

Data set	Network size (# nodes)	Gap (%)	CPU time Non-splittable	CPU time Splittable	Service lvel Non-splittable	Service lvel Splittable
AP	10	0.01	1.1	0.5	96.01	96.02
	20	0.04	250.4	129.6	96.38	96.47
	25	0.02	7,072.9	2,606.0	96.58	96.57
COL	10	0.11	6.2	5.4	94.56	94.50
	20	0.01	959.1	715.2	95.51	95.47
	25	0.01	6,339.3	6,559.6	95.72	95.72

commodities is usually higher than that of the splittable formulation, because of the increased number of binary variables. Lastly, the difference between the hub network structure and the service level offered by the network in each situation is almost negligible in most cases. These characteristics of both formulations can be further exploited in solution algorithms by focusing on solving the splittable formulation and then imposing the integrality condition of the non-splittable case (Contreras et al. 2012).

References

Bladt M, Nielsen B (2017) *Matrix-Exponential Distributions in Applied Probability*. Probability Theory and Stochastic Modelling (Springer US), ISBN 9781493970490.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press), ISBN 978-0521833783.

Contreras I, Cordeau JF, Laporte G (2012) Exact solution of large-scale hub location problems with multiple capacity levels. *Transportation Science* 46(4):439 – 459.

Lucantoni DM (1985) Efficient algorithms for solving the non-linear matrix equations arising in phase-type queues. *Communications in Statistics. Stochastic Models* 1(1):29 – 51.

Ramaswami V (1985) Stationary waiting time distribution in queues with phase type service and in quasi-birth-and-death processes. *Communications in Statistics. Stochastic Models* 1(2):125 – 136.