

Resilience of Public Transit Systems Under Short Random Service Suspensions: A Bulk-Service Queue Model

Baichuan Mo^a, Li Jin^{b,c,*}, Max Z. Shen^e, Haris N. Koutsopoulos^d, Jinhua Zhao^f

^a*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

^b*UMich Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China*

^c*Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA*

^d*Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA*

^e*Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA 94720*

^f*Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract

This paper proposes a stochastic framework to model the resilience of public transit systems under short random service suspensions. Two aspects regarding resilience are evaluated: 1) system's stability and 2) system performance changes (queue length and waiting time) under random service suspensions. We adopt a bulk-service queue model to formulate the queuing behavior at a station. A Markov chain model is used to model passenger flow dynamics across stations. We model the random service suspension as a two-state (failure and normal) Markov process, where vehicles are assumed to stop in the failure state. With specific assumptions, the distribution of headway and the number of arrival passengers within a headway is derived, which are then used to calculate the mean and variance of queue length and waiting time at each station with analytical formulations. These analytical formulations allow efficient calculation of system performance and quantify the impact of random service suspensions. We also derive the stability condition of the system. The closed-form formula implies that the system is more likely to be unstable with high incident rates, high incident duration, high demand, low service frequency, and low vehicle capacity. The proposed model is implemented on a bus network with sensitivity analysis of different parameters (such as incident rate, incident duration, vehicle capacity, etc.). Results show that the congested stations (i.e., stations with high demand rates) are more vulnerable to random service suspensions. We also validate the theoretical results with a simulation model, showing consistency between two outcomes.

1. Introduction

Public transit (PT) systems play a crucial role in cities worldwide, transporting people to jobs, homes, outings, and a vast variety of other activities. However, PT systems are usually susceptible to unplanned delays and service disruptions, which may come from equipment, weather, passengers, or other internal and external factors.

Short-term service suspension happens frequently in PT systems. According to [Mo et al. \(2021\)](#), there are on average 75 incidents happening in the Chicago urban rail system per day and more than 75% of

*Corresponding author

Email address: li.jin@sjtu.edu.cn (Li Jin)

them are within 5 minutes. Causes for these short-term suspensions can be signal system failure, passenger behavior, and infrastructure problems. For this reason, it is important to recognize how a PT system is affected by these short-term service suspensions.

An important concept related to the system's reaction to incidents is "resilience". Resilience, in the context of managed infrastructure systems, is defined as the endogenous capacity of a system to cope with exogenous perturbations (Pachauri et al., 2014). The quantification of resilience varies from different studies. In this study, we consider two aspects regarding resilience: a) system's ability to stay stable under random service suspensions, and b) system performance changes under random service suspensions. The first aspect requires **stability analysis** for the system. For the second aspect, we calculate the mean and variance of passengers' **queue length** and **waiting time** under random service suspensions.

Queuing behavior at a PT station is usually modeled by the bulk-service queue model (Powell, 1981; Islam et al., 2014; Wang et al., 2014). Bulk service means that customers are served in groups rather than individually. At a PT station, with the arrival of vehicles (e.g., buses or trains), a group of passengers will board (i.e., being served in groups). If the vehicle capacity is less than the number of customers waiting, some customers are left behind (Kahraman and Gosavi, 2011). Most of the previous studies on a bulk service model for PT systems focus at the station level (Selvi and Rosenshine, 1983; Powell, 1985; Wang et al., 2014). Islam et al. (2014) used a Markov model to extend the station-level analysis to the route level. However, these studies all considered PT systems under normal operating condition, the research on the queuing analysis for PT systems under service suspensions are limited. Regarding the service disruption on the bulk-service queue model, Madan (1989) first considered a single channel bulk service queue subject to interruptions. They assumed there are two states (work and repair) in the system and derived the probability generating function (PGF) of queue length using steady-state equations. Following Madan (1989)'s framework, there are many research extensions such as considering more channels (Singh and Ram, 1991), more heterogeneous states (Madan, 1992; Ayyappan and Karpagam, 2020), different service interruption assumptions (Jayaraman et al., 1994), and different repair policies (Tadj and Choudhury, 2009; Tadj et al., 2012). However, all these studies assumed that the service is rendered with a fixed batch size, which is not valid for PT systems where the available vehicle capacity for boarding is a random variable depending on the current vehicle load. Besides, all these studies used steady-state equations to derive multiple PGFs of queue length under different system states (e.g., work and repair). Results are usually mathematically tedious and the queue length and waiting time can only be analyzed with a very small service batch size (e.g., Madan (1989) only analyzed the situation with service batch size equal to 1, which is equivalent to an M/M/1 queue). Finally, previous studies usually consider the breakdown of servers. But there is no trivial way to map the "breakdown of servers" to a PT system with valid real-world assumptions because PT systems do not operate in a way that each station is an independent server.

To fill the research gaps, we propose a bulk queue service-based framework to describe the passenger and vehicle dynamics for a PT system and analyze the system resilience under short random service suspensions. The objective of this study is to derive the stability condition of a PT system and the mean and variance of passengers' queue length and waiting time for each station under random suspensions. This analysis provides important insights into PT systems' resilience and performance changes under service disruptions, which is

helpful for future control and planning strategies.

This work can be seen as an extension of [Powell \(1985\)](#) and [Islam et al. \(2014\)](#)'s works from normal conditions to incident conditions. [Powell \(1985\)](#) proposed a bulk service queue model for transportation terminals (i.e., station-level) with analytical queue length and waiting time formulations under normal conditions and [Islam et al. \(2014\)](#) extended the analysis from station-level to route-level. In this study, we explicitly model the random service suspension in a single-route PT system (in reality, it represents a bus route or one-directional rail line, which is a basic element of more complex PT networks). Different from typical service interruption studies where servers may break down, we assume a **vehicle** in the PT system may suffer from random suspensions. A detailed discussion for this assumption is provided in [Section 3.2](#), where we show how this assumption corresponds to many real-world situations and how it can be seen as the first step toward a general incident representation in PT systems. Under this assumption, we extend [Powell \(1985\)](#) and [Islam et al. \(2014\)](#)'s works to obtain the mean and variance of passengers' queue length and waiting time at each station in the single route PT system by analyzing the headway distribution under random service suspensions. The major contribution of this paper is fourfold:

- This is the first study to explore analytically the bulk-service queuing problem involving short random service suspensions with a special orientation to PT systems. We model the service suspension in PT systems from the analysis of vehicles' speed curves, which is a novel and practical way to consider "server breakdown" in PT systems.
- The headway under random service suspensions is proved to follow a compound Poisson-Exponential distribution in this study, based on which we derive the PGF and corresponding moments of the number of arrival passengers within a headway (these are critical components for the bulk-service queue model). This is a new analytical contribution to the bulk-service queuing theory.
- Based on [Islam et al. \(2014\)](#)'s work, a Markov chain model is implemented to capture the inter-station passenger flow dynamics, which extend the typical bulk-service queuing analysis from the station level to the system level.
- We propose an efficient roots-solving method to find all complex roots within a unit circle for this study's model specification.

The remaining paper is organized as follows. [Section 2](#) reviews the literature on the bulk-service queue model, random service disruptions, and queuing model for public transit systems. [Section 3](#) presents the model settings of single-route system and random services suspensions. [Section 4](#) shows the analysis and derivations of the major results. [Section 5](#) provides numerical examples to illustrate the theoretical results and validate the proposed approach with simulation. [Section 6](#) concludes the paper and outlooks on future research.

2. Literature review

2.1. Bulk queue model

In bulk service queuing models, customers are served in batches, the batches may be of fixed length or variable length. Sometimes, the service rate may depend on the number of customers waiting for service. The motivation of this model rises from manufacturing systems, elevators, transport vehicles, etc.

Bailey (1954) originated the study of bulk queues by considering a system with simple Poisson arrivals to a server that takes, at particular points in time, all waiting customers up to a fixed capacity c . If no customers are waiting, an empty batch departs, implying that the server is never idle. The queue, denoted by $M/G^c/1$ is described using the embedded Markov chain defined at points of service completions. Immediately following Bailey (1954), Downton (1955) obtained the waiting time distribution of bulk service queues by considering random arrivals and random service time distribution. Jaiswal (1960) confirmed the result of Downton (1955). He derived waiting time distribution using the embedded Markov-chain technique.

The general bulk service rule was first introduced by Neuts (1967), where a server, on finishing one batch, may remain idle if there are fewer than m customers waiting for service. Thus all departing batches from the queue have at least m customers, although no more than maximum service capacity.

Along with and after those milestone studies, papers have appeared which can be differentiated on the basis of the type of queue (arrival process, service process, number of servers), what is solved for (queues, waiting times, busy periods, etc.), the time domain of the solution (i.e., steady-state or transient), and the method of solution (transforms or direct numerical methods). One can refer to Chaudhry and Templeton (1983) and Sasikala and Indhira (2016) for a more complete review of the development of bulk service queue models.

2.2. Random service disruption

The subject of queuing systems wherein the service channel is subject to breakdowns is a popular subject that has received a lot of attention for the past fifty years. For a recent survey of the related literature, readers can refer to Krishnamoorthy et al. (2014).

However, most of the research on this topic deals with models where the server serves the customers one at a time. Very few bulk service papers existed. Madan (1989) studied a single-channel queueing system with Poisson arrivals and exponential service in batches of fixed size. Also, both the operative times and the repair times of the service channel are assumed to be exponential. Madan (1992) generalized Madan (1989) to the case where the repairs are performed in two phases. Singh and Ram (1991) also generalized Madan (1989). They considered a system with three identical channels and where the operative time and repair time of all three service channels are distributed exponentially. Jayaraman et al. (1994) considered a single-server queueing system with general bulk service. Arrivals are Poisson processes but alternate between two modes according to whether the server is operational or in the failed state. The duration of the operating periods and repair periods follow exponential and phase-type distributions, respectively. Tadj and Choudhury (2009) analyzed a bulk service queueing system with an unreliable server, Poisson input, and general service and repair times. Tadj et al. (2012) considered a bulk service queueing system where service is rendered to groups of customers of fixed quorum size. Service consists of two consecutive phases and may take a vacation

following the second phase of service. While providing service, the server may break down and a delay period precedes the repair period.

2.3. Queuing model in public transit systems

Queuing theory in PT systems is usually conducted at the station level, aiming at obtaining the mean queue length and waiting time. In the case of regular services where headways are equal, assuming that a) passengers arrive randomly at stops as a Poisson process and b) passengers can be served by the first arriving vehicle, the mean waiting time of passengers ($\mathbb{E}[W]$) is given by:

$$\mathbb{E}[W] = H/2, \quad (1)$$

where H is the service headway and W is the passenger waiting time. This is the most widely used queuing assumption in transit studies (Dial, 1967; Clerq, 1972; Wirasinghe, 1980). However, in the case where service is not completely reliable, the assumption of regular service can be problematic. Numerous models have been proposed to address cases where some degree of irregularity is involved in bus arrivals (Welding, 1957; Osuna and Newell, 1972). A well-known model proposed by Osuna and Newell (1972) with Poisson arrival passengers and stochastic headways shows that

$$\mathbb{E}[W] = \frac{1}{2} \cdot \left[\mathbb{E}[H] + \frac{\text{Var}[H]}{\mathbb{E}[H]} \right], \quad (2)$$

where $\mathbb{E}[H]$ and $\text{Var}[H]$ are the expectation and variance of headway, respectively. In the case of regular services, the variance of the headway is zero and this model reverts to Equation 1.

However, the results in Eq. 1 and 2 do not consider the vehicle capacity (i.e., they assume all passengers can board the first arrival vehicle). In a congested PT system, passengers may be left behind due to limited vehicle capacity, leading to an increase in waiting times (Mo et al., 2020). To model the left behind behavior, the bulk service queue model is applied in PT systems. Powell (1981, 1983, 1985) used a bulk service queue model to evaluate the passenger queue length and waiting time at a public transportation terminal. The closed-form mean and variance for these two variables are derived using a transform method. Rapoport et al. (2010) studied bulk service queues with constant or variable capacity and endogenously determined arrival times. Wang et al. (2014) proposed a bulk service and batch arrival queuing model with reneging behaviors to estimate passengers' waiting for public transport services.

All the aforementioned studies consider the queuing analysis at the station level. The extension of queuing analysis from station to system level is not a trivial problem. First, the boarding and alighting behavior at upstream stations affect the available capacity distribution at downstream stations. Second, there may be headway correlations in the system, leading to different headway distributions for different stations (Marguier, 1985; Hickman, 2001). To address this problem, Islam et al. (2014, 2015) proposed a Markov model to combine the Powell (1981) and Hickman (2001)'s works and used a bulk service model to analyze the PT systems queuing at a route level. However, a limitation in their research is that the calculation of headway correlation does not consider the vehicle capacity (though the capacity constraint is considered in the queuing behavior), resulting in the inconsistency of model assumptions.

Our paper can be seen as an extension of Powell (1985) and Islam et al. (2014)’s work to incorporate the random service suspensions in a PT system with more consistent and reasonable assumptions.

2.4. Service interruption in public transit systems

Studies on service interruption in public transit systems can be categorized into two groups: analysis and control. For the analysis groups, previous research has used a variety of methods to analyze the impact of service disruptions. Of these methods, the three most common are graph theory-based, survey-based, simulation-based, and empirical data-based. Graph theory-based methods usually derive resilience or vulnerability indicators based on the network topology (Yin et al., 2016; Zhang et al., 2018; Xu et al., 2015; Berdica, 2002). These methods are effective for understanding high-level network properties related to incidents. Survey-based methods investigate passenger behavior and opinions during the incident period (Currie and Muir, 2017; Murray-Tuite et al., 2014; Fukasawa et al., 2012; Teng and Liu, 2015; Lin et al., 2018). Passengers’ individual-level behavior is analyzed and understood using econometric models. Simulation-based methods simulate passenger flows on the transit network under incident scenarios (Balakrishna et al., 2008; Suarez et al., 2005; Hong et al., 2018). The empirical data-based method uses the smart card and vehicle location data to analyze real-world incident impacts (Sun et al., 2016; Tian and Zheng, 2018; Mo et al., 2021). These studies can output many metrics of interest such as the vehicle load changes, additional travel delays caused by incidents, distribution of the impact, etc. For the control under service disruption, previous studies have proposed various frameworks including shuttle bus design (Jin et al., 2016; Luo et al., 2019), vehicle holding (O’Dell and Wilson, 1999), integrating local services (Jin et al., 2014), and time table rescheduling (Kroon and Huisman, 2011).

The resilience analysis presented in this paper belongs to the “analysis” category, which aims to obtain the stability condition of a PT system and the mean and variance of passengers’ queue length and waiting time for each station under the short random suspensions with a theoretical orientation. None of the previous studies has used the bulk service model for this type of analysis.

3. Model

3.1. Single-route public transit system and vehicle movement

Consider a single-route PT system with N stations as shown in Figure 1. Vehicles are dispatched from a transportation hub (also referred to as station 0) close to station 1, and travel from station 1 to station N . At a specific station n , we assume that passenger arrival follows a Poisson process with a fixed rate $\lambda^{(n)}$ during our study time of interest. And when a vehicle arrives at station n , each passenger in the vehicle has a probability of $\alpha^{(n)}$ to alight. Thus, the number of alighting passengers at station n follows a binomial distribution. The Poisson arrival and binomial alighting are two widely assumptions in much PT-related literature (Hickman, 2001). In this study, we do not consider renegeing behaviors of passengers (i.e., passengers may leave the system with too much waiting time) because this paper focuses on “short” service suspensions. We assume passengers choose to wait in this case. This assumption can be validated by empirical studies (Sun et al., 2016; Rahimi et al., 2019) showing that passengers start to leave the system when the delay is large enough

(e.g., 30 minutes or more). However, future studies may incorporate balking and renegeing as an extension of this work.

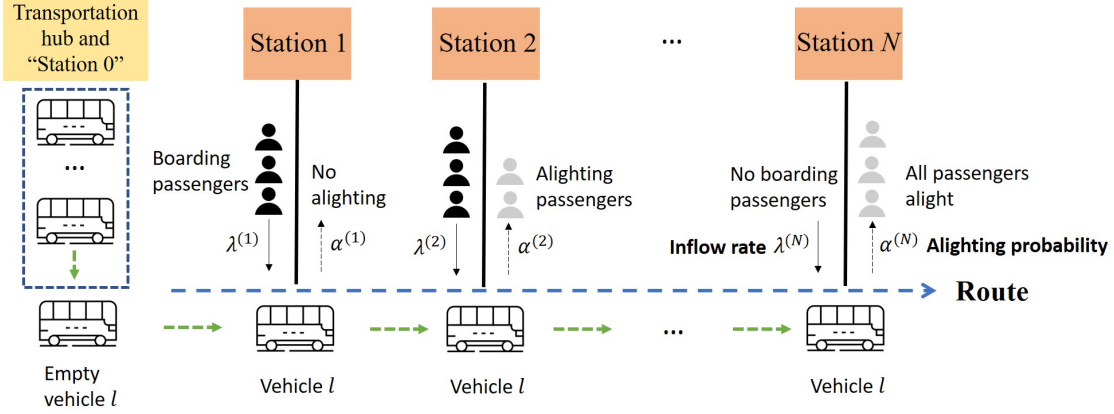


Figure 1: Schematic presentation of a single-route public transit system

Let $l = 1, 2, \dots$ be a superscript denoting the vehicle run number (or vehicle ID). Smaller l means vehicles dispatched at an earlier time. Figure 2 shows the diagram of vehicles and passengers interaction at station n in the time dimension. At station n , let $t_A^{(n,l)}$ be the time that vehicle l arrives at station n , and $t_D^{(n,l)}$ the time that vehicle l departs station n . $H^{(n,l)}$ is the headway between the preceding vehicle $l - 1$ and vehicle l , as they depart stop n (i.e., $H^{(n,l)} = t_D^{(n,l)} - t_D^{(n,l-1)}$). When a vehicle arrives at station n , some of the on-board passengers alight first, then the queuing passengers start to board. Let $Q^{(n,l)}$ be the number of queuing passengers when vehicle l arrives at station n , $R^{(n,l)}$ the number of left behind passengers when vehicle l departs station n , and $Y^{(n,l)}$ the number of arrival passengers between $t_D^{(n,l)}$ and $t_A^{(n,l+1)}$. By definition, we have

$$Q^{(n,l+1)} = R^{(n,l)} + Y^{(n,l)}. \quad (3)$$

In this study, we assume the dwell time (i.e., $t_D^{(n,l)} - t_A^{(n,l)}$) is negligible compared to the vehicle travel time ($t_A^{(n,l+1)} - t_D^{(n,l)}$) such that the number of arrival passengers during the dwell time is zero (same assumption as Powell (1981)). Then, given the headway $H^{(n,l+1)}$, $Y^{(n,l)} | H^{(n,l+1)}$ follows a Poisson distribution of rate $\lambda^{(n)} H^{(n,l+1)}$:

$$Y^{(n,l)} | H^{(n,l+1)} \sim \text{Poi}(\lambda^{(n)} H^{(n,l+1)}). \quad (4)$$

In other words, $Y^{(n,l)}$ can be seen as the number of arrival passengers within a headway (i.e., $H^{(n,l+1)}$).

From the vehicle's perspective, let $S^{(n,l)}$ be the number of available space after passenger alighting from vehicle l at station n , $G^{(n,l)}$ be the number of remaining passengers on vehicle l after alighting at station n . By definition, we have

$$G^{(n,l)} = C - S^{(n,l)}, \quad (5)$$

where C is the maximum capacity of vehicles. Denote $V^{(n,l)}$ as the vehicle load (i.e., number of on-board passengers) when vehicle l departs station n (i.e., the vehicle load when it arrives at station $n + 1$). Then, the number of alighting passengers for vehicle l at station n given $V^{(n-1,l)}$ follows a binomial distribution:

$$V^{(n-1,l)} - G^{(n,l)} \mid V^{(n-1,l)} \sim \text{Bin}(V^{(n-1,l)}, \alpha^n). \quad (6)$$

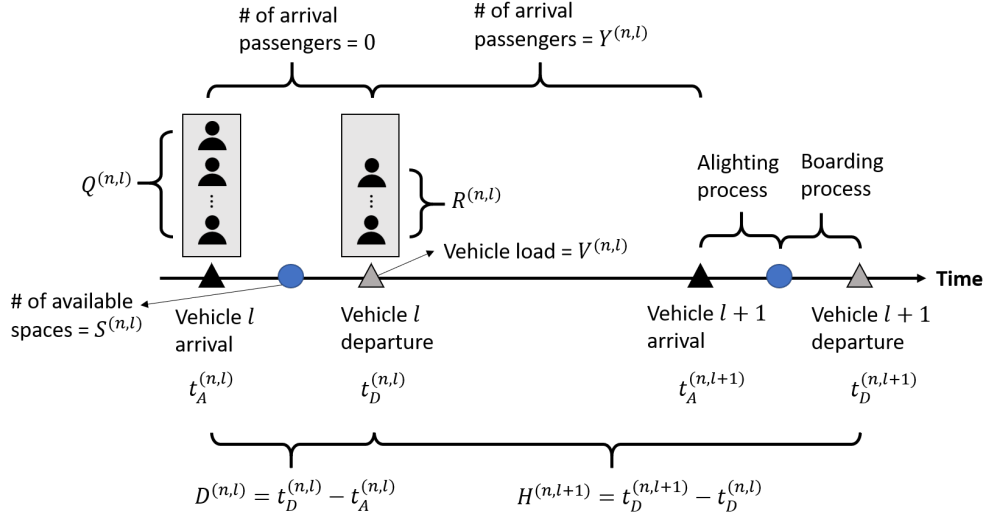


Figure 2: Diagram of vehicles and passengers interaction at station n in the time dimension

3.2. Random service suspensions and vehicle speed curve

Consider that there are random service suspensions when a vehicle travels in the system. Given these disturbances, the speed curve of vehicle l from station n to $n + 1$ can be described as the red line in Figure 3. That is, every random incident causes a speed reduction or stop of the vehicle. In reality, these incidents can be caused by many reasons. For example, in a bus system, they may be caused by traffic congestion or accidents, drivers' or passengers' abnormal behavior, vehicle engine issues, etc. In a rail system, the reasons may be signal failures, infrastructure problems, and drivers' or passengers' abnormal behavior. The speed curve is a general representation of different incidents, interruptions, suspensions, or disruptions for the vehicle's movement.

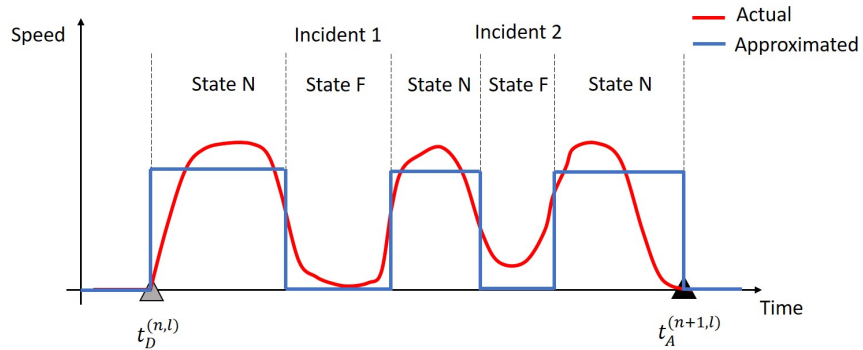


Figure 3: Schematic speed curve of vehicle l traveling from station n to $n + 1$

The actual vehicle speed profile under interruptions can be complicated. To facilitate mathematical modeling, we assume the speed of a vehicle under random interruptions can be approximated by an impulse function (blue line in Figure 3). The impulse function then separates the vehicle into traveling and stopping phases, denoted as state N (normal) and state F (failure), respectively. In state N , a vehicle travels at a constant speed. Once an incident happens, the vehicle stops immediately and enters state F . We assume that the duration of an incident follows an exponential distribution with the rate θ (i.e., mean of $\frac{1}{\theta}$). Besides, for a sufficiently small time interval, Δ , the probability of incident occurrence is $\gamma\Delta$. Then, the state of a vehicle yields a two-state Markov process (Figure 4) with the state space of $\{N, F\}$.

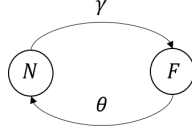


Figure 4: Transition diagram of vehicle states

Approximating the actual speed curve as an impulse function can be seen as the first step toward a general incident representation in PT systems. Actually, any fashions of incident can be represented as a mixture of different types of normal and failure states. The normal and failure states can be defined more diversely with heterogeneous occurrence probabilities and duration for different categories of incidents, which results in a more sophisticated speed curve approximation.

Note that the vehicle speed curve does not introduce any assumptions on spatial correlation or independence. That is, different vehicles' states can be independent or correlated. The model still holds as long as every single vehicle's state follows the aforementioned Markov process. For example, different trains' states may be correlated because the failure comes from the power system, different buses can have independent states as the failure comes from their own engines. Both of these scenarios can still be modeled as above as long as we assume the failure causes every single train/bus to have an impulse-like speed function.

3.3. Headway under random service suspensions

Under the assumption of impulse-function speed, all vehicles have the same fixed travel speed under state N . Therefore, if there were no incident in the system, all stations would have the same deterministic headway (denoted as \bar{H}). This implies that after vehicle $l - 1$ departs station n (i.e., after $t_D^{(n,l-1)}$), vehicle l must travel for \bar{H} time under normal state to reach station n . Considering the stopping time due to incidents, the actual headway of vehicle l at station n is

$$H^{(n,l)} = \bar{H} + I^{(n,l)}, \quad (7)$$

where $I^{(n,l)}$ is the total duration of the incidents happening between $t_D^{(n,l-1)}$ and $t_D^{(n,l)}$ ¹.

¹ $I^{(n,l)}$ by definition should be the total duration of the incidents happening between $t_D^{(n,l-1)}$ and $t_A^{(n,l)}$. However, as we assume the dwell time is negligible compared to vehicle travel time between stations, the number of incidents occurring between the $t_A^{(n,l)}$ and $t_D^{(n,l)}$ is 0

To implement the bulk service queue model, one important assumption is that the headway between different vehicle runs are independent, that is, $\text{Cov}[H^{(n,l)}, H^{(n,l+1)}] = 0, \forall l = 1, 2, \dots$. Notice that $\text{Cov}[H^{(n,l)}, H^{(n,l+1)}] = \text{Cov}[I^{(n,l)}, I^{(n,l+1)}]$ because \bar{H} is a constant. $I^{(n,l)}$ and $I^{(n,l+1)}$ represent the incident duration of vehicle i and $i + 1$ traveling in the same route segment (from station $n - 1$ to n). In reality, if the incidents are caused by road congestion or infrastructure issues, it is possible that the incident durations for two consecutive vehicles passing through the same route segment are correlated. However, in this study, we ignore the correlation because addressing the correlation is not a trivial problem in the bulk service queue model (Powell, 1981) and is beyond the scope of this paper (especially when we also consider the random service suspensions).

4. Analysis

Recall that the objective of this study is to derive the stability condition of a PT system and the mean and variance of passengers' queue length and waiting time for each station under random service suspensions. Figure 5 shows the organization of this section and how the distributions of different random variables (particularly, $S^{(n,l)}, V^{(n,l)}, Q^{(n,l)}$) are calculated. The major calculation consists of three parts:

- Given the distribution of $V^{(n-1,l)}$, calculate the distribution of $S^{(n,l)}$. The details are shown in Section 4.1
- Given the distribution of $S^{(n,l)}$ calculate the distribution of $Q^{(n,l)}$ and the mean and variance of queue length and waiting time at station n . This is a complicated process and is discussed in Section 4.3.
- Given the distribution of $S^{(n,l)}$ and $Q^{(n,l)}$, calculate the distribution of $V^{(n,l)}$, which is discussed in Section 4.2

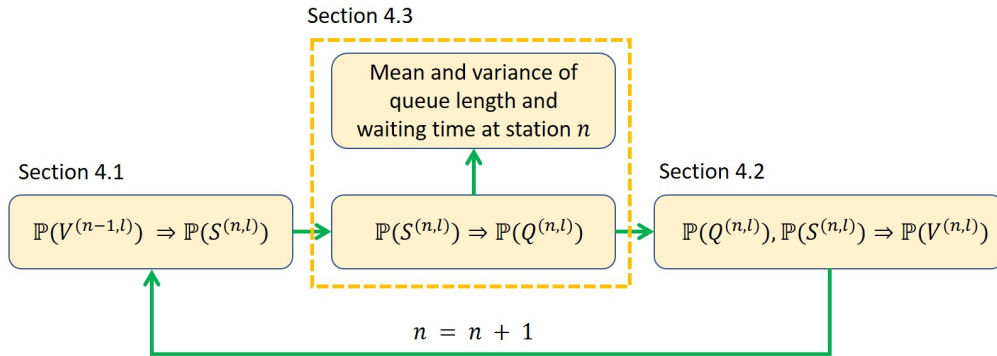


Figure 5: Analysis framework

With the three components, we can derive the distribution of $S^{(n,l)}, Q^{(n,l)}, V^{(n,l)}$ for all $n = 1, \dots, N$ given the distribution of $V^{(0,l)}$ (i.e., vehicle load when vehicle l arrives at the first station, it is always zero by definition). Note that, in this section, we focus on the steady-state distribution of these variables. So we will finally derive the distribution results as $l \rightarrow \infty$.

After obtaining the corresponding distributions, we discuss the stability condition in Section 4.4 and summarize the whole calculation process in Section 4.5.

4.1. Available vehicle space steady-state distribution

In this section, we aim to derive the steady-state distribution of $S^{(n,l)}$ given the steady-state distribution of $V^{(n-1,l)}$. Define $v_k^{(n,l)} := \mathbb{P}(V^{(n,l)} = k)$, $s_k^{(n,l)} := \mathbb{P}(S^{(n,l)} = k)$, and $g_k^{(n,l)} := \mathbb{P}(G^{(n,l)} = k)$ for all $k = 0, 1, \dots, C$. Assuming that the steady state probabilities for all variables exist (the stability condition will be discussed in Section 4.4), we have $v_k^{(n)} := \lim_{l \rightarrow \infty} v_k^{(n,l)} = \mathbb{P}(V^{(n)} = k)$, $s_k^{(n)} := \lim_{l \rightarrow \infty} s_k^{(n,l)} = \mathbb{P}(S^{(n)} = k)$, and $g_k^{(n)} := \lim_{l \rightarrow \infty} g_k^{(n,l)} = \mathbb{P}(G^{(n)} = k)$, where $V^{(n)} = \lim_{l \rightarrow \infty} V^{(n,l)}$, $S^{(n)} = \lim_{l \rightarrow \infty} S^{(n,l)}$, and $G^{(n)} = \lim_{l \rightarrow \infty} G^{(n,l)}$.

Proposition 1. $\forall n = 1, \dots, N$, given the distribution of $V^{(n-1)}$ (i.e., $v^{(n)} := [v_0^{(n-1)}, \dots, v_C^{(n-1)}] \in \mathbb{R}^{C+1}$), the distribution of $S^{(n)}$ (i.e., $s^{(n)} := [s_0^{(n-1)}, \dots, s_C^{(n-1)}] \in \mathbb{R}^{C+1}$) is given as:

$$s_k^{(n)} = g_{C-k}^{(n)} \quad \forall k = 0, 1, \dots, C, \quad (8)$$

where $g^{(n)} := [g_0^{(n)}, \dots, g_C^{(n)}] \in \mathbb{R}^{C+1}$ and

$$g^{(n)} = v^{(n-1)} A^{(n)}, \quad (9)$$

$A^{(n)}$ is a $(C+1)$ by $(C+1)$ matrix with the element in row i and column j equal to a_{ij}^n , and a_{ij}^n is defined as

$$a_{ij}^n = \begin{cases} 1, & \text{if } i = 0 \text{ and } j = 0 \\ \binom{i}{i-j} (\alpha^{(n)})^{i-j} (1 - \alpha^{(n)})^j, & \text{if } i \geq j \text{ and } i, j \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad \forall i, j = 0, 1, \dots, C \quad (10)$$

Proof. When vehicle l arrives at station n , by definition, there are $V^{(n-1,l)}$ number of passengers in the vehicle. Given that there are i passengers on-board after when vehicle l arrives at station n , let the probability that there are j passengers remaining on the vehicle be a_{ij}^n . a_{ij}^n also represents the probability of $i-j$ passengers alighting, which follows a binomial distribution with parameters i and $\alpha^{(n)}$ if $i \geq j$ and $i, j \neq 0$. Hence, a_{ij}^n can be expressed as Eq. 10. Then we have

$$g^{(n,l)} = v^{(n-1,l)} A^{(n)} \quad (11)$$

where $v^{(n-1,l)} = [v_0^{(n-1,l)}, \dots, v_C^{(n-1,l)}] \in \mathbb{R}^{C+1}$, $g^{(n,l)} = [g_0^{(n,l)}, \dots, g_C^{(n,l)}] \in \mathbb{R}^{C+1}$. According to the relationship between $S^{(n,l)}$ and $G^{(n,l)}$ as shown in Eq. 5, the distribution of number of available spaces after alighting is simply

$$s_k^{(n,l)} = g_{C-k}^{(n,l)} \quad \forall k = 0, 1, \dots, C \quad (12)$$

Notice that Eq. 11 and 12 holds for all l . Since we assume the steady state distributions exist, letting $l \rightarrow \infty$ for both sides of Eq. 11 and 12 finishes the proof. \square

4.2. Vehicle load steady-state distribution

In this section, we aim to derive the steady-state distribution of $V^{(n,l)}$ given the steady-state distribution of $G^{(n,l)}$ and $Q^{(n,l)}$. Define $q_k^{(n)} := \lim_{l \rightarrow \infty} q_k^{(n,l)} = \mathbb{P}(Q^{(n)} = k)$, where $Q^{(n)} = \lim_{l \rightarrow \infty} Q^{(n,l)}$ and $q_k^{(n,l)} = \mathbb{P}(Q^{(n,l)} = k)$. Denote the first C elements of the steady-state queue length distribution as $q_{0:C-1}^{(n)}$, where $q_{0:C-1}^{(n)} = [q_0^{(n)}, \dots, q_{C-1}^{(n)}] \in \mathbb{R}^C$.

Proposition 2. $\forall n = 1, \dots, N$, given the distribution of $G^{(n)}$ (i.e., $g^{(n)}$) and $q_{0:C-1}^{(n)}$, the distribution of $V^{(n)}$ can be expressed as:

$$v^{(n)} = g^{(n)} B^{(n)} \quad (13)$$

where $B^{(n)}$ is a matrix with the element in row i and column j equal to $b_{ij}^{(n)}$, and $b_{ij}^{(n)}$ is defined as

$$b_{ij}^{(n)} = \begin{cases} q_{j-i}^{(n)}, & \text{if } 0 \leq i \leq j < C \\ 1 - \sum_{k=0}^{C-i-1} q_k^{(n)}, & \text{if } j = C \text{ and } 0 \leq i < C \\ 1, & \text{if } i = j = C \\ 0, & \text{otherwise} \end{cases} \quad \forall i, j = 0, 1, \dots, C \quad (14)$$

Proof. Let $b_{ij}^{(n,l)}$ be the probability that the load of vehicle l is j after passenger boarding given that there are i passengers on-board after alighting (i.e., $G^{(n,l)} = i$) at station n . Hence, If $0 \leq i \leq j < C$, $b_{ij}^{(n,l)}$ is simply the probability that there are $j - i$ passengers in the queue (such that after boarding there are j passengers on the vehicle):

$$b_{ij}^{(n,l)} = q_{j-i}^{(n,l)}, \quad \text{if } 0 \leq i \leq j < C. \quad (15)$$

If $j = C$ and $0 \leq i < C$, which means the vehicle reaches capacity after boarding. Then $b_{ij}^{(n,l)}$ should be the probability that there are greater than or equal to $C - i$ passengers in the queue (i.e., one minus the probability that there are less than or equal to $C - i - 1$ passengers in the queue). This leads to:

$$b_{ij}^{(n,l)} = 1 - \sum_{k=0}^{C-i-1} q_k^{(n,l)}, \quad \text{if } j = C \text{ and } 0 \leq i < C. \quad (16)$$

When $i = j = C$, we simply have $b_{ij}^{(n,l)} = 1$ because no matter there are passengers waiting in the queue or not, nobody can board as the vehicle is full. With the expression of $b_{ij}^{(n,l)}$, the vehicle load distribution can be calculated as

$$v_j^{(n,l)} = \sum_{i=1}^C g_i \cdot b_{ij}^{(n,l)} \quad \forall j = 0, 1, \dots, C \quad (17)$$

Notice that Eq. 17 holds for all l . As we assume the steady state distributions exist, taking $l \rightarrow \infty$ for

both sides of Eq. 17 and rewriting it to a matrix form finishes the proof. \square

4.3. Queuing analysis at a station

In this section, supposing that we know the distribution of $S^{(n)}$ (i.e., $s^{(n)} = [s_0^{(n)}, \dots, s_C^{(n)}] \in \mathbb{R}^{C+1}$), our goal is to derive $q_{0:C-1}^{(n,l)}$ and the mean and variance of passenger queue length and waiting time.

4.3.1. probability generating function of queue length

We start with deriving the probability generating function (PGF) for $Q^{(n)}$, where $Q^{(n)} = \lim_{l \rightarrow \infty} Q^{(n,l)}$.

Proposition 3. $\forall n = 1, \dots, N$, given the distribution of $S^{(n)}$ (i.e., $s^{(n)}$), the PGF of $Q^{(n)}$ can be expressed as:

$$Q(z) = \frac{\sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} (z^C - z^{C-u+i}) \right]}{\frac{z^C}{\bar{Y}(z)} - \sum_{u=0}^C s_u^{(n)} z^{C-u}}, \quad (18)$$

where $Y(z)$ is the PGF of $Y^{(n)}$ and $Y^{(n)} = \lim_{l \rightarrow \infty} Y^{(n,l)}$ is the number of arrival passengers at station n within a headway at the steady state.

Proof. The proof follows the similar idea in Powell (1981) and is attached in Appendix A. The difference from Powell (1981) is that we consider an arbitrary vehicle capacity distribution $s_0^{(n)}, \dots, s_C^{(n)}$, while in Powell (1981) the capacity is fixed. Note that Powell (1985) provided an equivalent formulation as Eq. 18 with variable vehicle capacities using the transform of $S^{(n)}$. \square

In Eq. 18, there are C unknown variables, $q_0^{(n)}, \dots, q_{C-1}^{(n)}$. Note that $q_C^{(n)}$ does not appear in $Q(z)$ because when $u = C$ and $i = C$, we have $q_C^{(n)} (z^C - z^{C-u+i}) \equiv 0$. To quantify $Q(z)$, Rouché's theorem is used. Let NUM(z) and DEN(z) be the numerator and denominator of $Q(z)$ (i.e., $Q(z) = \frac{\text{NUM}(z)}{\text{DEN}(z)}$). As shown in Powell (1981), one can prove that DEN(z) (i.e., $\frac{z^C}{\bar{Y}(z)} - \sum_{u=0}^C s_u^{(n)} z^{C-u}$) has exactly C complex roots within (or on) the unit circle on a complex plane using Rouché's theorem. Notice that for any $z \in \mathbb{C}$ that satisfies $|z| \leq 1$, where \mathbb{C} is the set of complex numbers, the generating function $Q(z)$ must be analytic. Therefore, if z^* is the root of DEN(z) (i.e., DEN(z^*) = 0), it should also be the root of NUM(z) (i.e., NUM(z^*) = 0) such that $Q(z)$ is analytic (Rudin, 2006). Hence, one can solve $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ from the following two steps:

- **Step 1:** Solve DEN(z) = 0 for C different roots $z_0^*, \dots, z_{C-1}^* \in \mathbb{C}$ that satisfies $|z_i^*| \leq 1, \forall 0 \leq i \leq C-1$. Note that $z = 1$ is always one root of DEN(z). But it does not give information about $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ as NUM(1) = 0 is naturally satisfied. Hence, we adopt the convention that assigning $z_0^* = 1$.
- **Step 2:** Combining NUM(z_i^*) = 0 ($\forall 1 \leq i \leq C-1$) and $Q(1) = 1$, solve for $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ (there are C system equations and C unknown variables). Note that when $z \rightarrow 1$, both NUM(z) and DEN(z) approach 0. Therefore, using L'Hopital's rule we have

$$\lim_{z \rightarrow 1} Q(z) = \lim_{z \rightarrow 1} \frac{\text{NUM}'(z)}{\text{DEN}'(z)} = \frac{\sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} (u-i) \right]}{\bar{S}^{(n)} - \bar{Y}^{(n)}} = 1 \quad (19)$$

where $\bar{S}^{(n)} = \sum_{u=0}^C u s_u^{(n)} = \mathbb{E}[S^{(n)}]$, $\bar{Y}^{(n)} = Y'(1) = \mathbb{E}[Y^{(n)}]$. Eq. 19 is the actual equation used for solving $q_{0:C-1}^{(n)}$ (instead of directly using $Q(1) = 1$).

4.3.2. Queue length distribution

Though $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ can be obtained by solving C system equations as mentioned in Section 4.3.1, we attempt to provide a simpler way to calculate $q_{0:C-1}^{(n)}$ in this section, which is known as matching the polynomial coefficients.

Proposition 4. $\forall n = 1, \dots, N$, given the distribution of $S^{(n)}$ (i.e., $s^{(n)}$), all complex roots of $DEN(z)$ (i.e., z_0^*, \dots, z_{C-1}^*), and $\bar{Y}^{(n)}$, if $s_C^{(n)} > 0$, then $q_{0:C-1}^{(n)}$ can be solved as:

$$q_0^{(n)} = \frac{1}{s_C^{(n)}} (\bar{S}^{(n)} - \bar{Y}^{(n)}) \prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1}, \quad (20)$$

and

$$q_{0:C-1}^{(n)} = \tilde{\eta}^{(n)} (\Lambda^{(n)})^{-1}, \quad (21)$$

where $\tilde{\eta}^{(n)} = [s_C^{(n)} q_0^{(n)} \eta_0^{(n)}, s_C^{(n)} q_0^{(n)} \eta_1^{(n)}, \dots, s_C^{(n)} q_0^{(n)} \eta_{C-1}^{(n)}] \in \mathbb{R}^C$ and

$$\Lambda^{(n)} = \begin{bmatrix} s_C^{(n)} & s_{C-1}^{(n)} & s_{C-2}^{(n)} & \dots & s_1^{(n)} \\ 0 & s_C^{(n)} & s_{C-1}^{(n)} & \dots & s_2^{(n)} \\ \dots & 0 & s_C^{(n)} & \dots & s_3^{(n)} \\ 0 & \dots & 0 & \dots & s_4^{(n)} \\ 0 & 0 & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & s_C^{(n)} \end{bmatrix} \in \mathbb{R}^{C \times C}. \quad (22)$$

$\eta_j^{(n)}$ is the polynomial coefficient of z^j in $\prod_{i=0}^{C-1} \left(1 - \frac{z}{z_i^*}\right)$ (i.e., $\sum_{j=0}^C \eta_j^{(n)} z^j := \prod_{i=0}^{C-1} \left(1 - \frac{z}{z_i^*}\right)$). As z_i^* is specified to station n , a superscript n is added to the coefficients.

Proof. The derivation is shown in Appendix B. □

Note that assuming $s_C^{(n)} > 0$ in Proposition 4 does not lost the generality because otherwise we can reduce C such that $s_C^{(n)} > 0$ always holds.

4.3.3. Analytical formulation of mean and variance of queue length and waiting time

After solving for $q_0^{(n)}, \dots, q_{C-1}^{(n)}$, $Q(z)$ is determined. And the expectation and variance of queue length at station n can be written in Eq. 23 and 24 by definition:

$$\mathbb{E}[Q^{(n)}] = \sum_{k=0}^{\infty} k q_k^{(n)} = \left. \frac{dQ(z)}{dz} \right|_{z=1} \quad (23)$$

$$\text{Var}[Q^{(n)}] = \mathbb{E}[(Q^{(n)})^2] - \mathbb{E}[Q^{(n)}]^2 = \left. \frac{d^2 Q(z)}{dz^2} \right|_{z=1} + \mathbb{E}[Q^{(n)}] - \mathbb{E}[Q^{(n)}]^2. \quad (24)$$

Proposition 5. $\forall n = 1, \dots, N$, given the distribution of $S^{(n)}$ and the expression of $Y(z)$, $\mathbb{E}[Q^{(n)}]$ and $\text{Var}[Q^{(n)}]$ can be calculated as:

$$\mathbb{E}[Q^{(n)}] = \frac{\bar{\bar{S}}^{(n)} + \bar{Y}^{(n)} + (\bar{S}^{(n)} - \bar{Y}^{(n)})[1 + 2(\bar{S}^{(n)} - C)] - (\bar{S}^{(n)} - \bar{Y}^{(n)})^2}{2(\bar{S}^{(n)} - \bar{Y}^{(n)})} + \sum_{i=1}^{C-1} \frac{1}{1 - z_i^*} \quad (25)$$

$$\begin{aligned} \text{Var}[Q^{(n)}] = & \frac{1}{12(\bar{S}^{(n)} - \bar{Y}^{(n)})^2} \left[-4(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)})(\bar{S}^{(n)} - \bar{Y}^{(n)}) + 3(\bar{\bar{S}}^{(n)} + \bar{\bar{Y}}^{(n)})^2 \right. \\ & \left. - [6(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)}) - 1](\bar{S}^{(n)} - \bar{Y}^{(n)})^2 - (\bar{S}^{(n)} - \bar{Y}^{(n)})^4 \right] - \sum_{i=1}^{C-1} \frac{z_i^*}{(1 - z_i^*)^2} \end{aligned} \quad (26)$$

where $\bar{\bar{S}}^{(n)}$ and $\bar{\bar{S}}^{(n)}$ (resp. $\bar{\bar{Y}}^{(n)}$ and $\bar{\bar{Y}}^{(n)}$) are the second and third central moments of $S^{(n)}$ (resp. $Y^{(n)}$).

Proof. The derivation follows the same idea as [Powell \(1981\)](#). Details are mathematical tedious and are thus attached in [Appendix C](#). These results are equivalent to [Powell \(1985\)](#) who considered the general bulk-service queue model. \square

Proposition 6. $\forall n = 1, \dots, N$, given the distribution of $S^{(n)}$ and the expression of $Y(z)$, the mean and variance of waiting time at station n (denoted as $W^{(n)}$) is given as:

$$\mathbb{E}[W^{(n)}] = \frac{\bar{Q}_t^{(n)}}{\lambda^{(n)}} \quad (27)$$

$$\text{Var}[W^{(n)}] = \frac{\bar{\bar{Q}}_t^{(n)} - \bar{Q}_t^{(n)}}{(\lambda^{(n)})^2} \quad (28)$$

where $Q_t^{(n)}$ is the queue length at an arbitrary time point (as opposed to $Q^{(n)}$ which is the queue length at the time of vehicle arrival). $\bar{Q}_t^{(n)}$ and $\bar{\bar{Q}}_t^{(n)}$ are defined as

$$\bar{Q}_t^{(n)} = \mathbb{E}[Q_t^{(n)}] - \bar{Y}^{(n)} + \frac{1}{2} \left(\bar{\bar{Y}}^{(n)} / \bar{Y}^{(n)} + \bar{Y}^{(n)} - 1 \right) \quad (29)$$

$$\bar{\bar{Q}}_t^{(n)} = \text{Var}[Q_t^{(n)}] - \bar{Y}^{(n)} + \frac{1}{12(\bar{Y}^{(n)})^2} \left[4\bar{Y}^{(n)}\bar{\bar{Y}}^{(n)} + 6(\bar{Y}^{(n)})^2\bar{\bar{Y}}^{(n)} - (\bar{Y}^{(n)})^2 + (\bar{Y}^{(n)})^4 - 3(\bar{\bar{Y}}^{(n)})^2 \right] \quad (30)$$

Eq. 27 is the application of Little's law. Proposition 6 is directly obtained from [Powell \(1985\)](#).

Remark 1. The formulation of $\mathbb{E}[Q^{(n)}]$, $\text{Var}[Q^{(n)}]$, $\mathbb{E}[W^{(n)}]$, and $\text{Var}[W^{(n)}]$ in this study are same as [Powell \(1985\)](#) because in his paper the $M/G^{[S]}/1$ bulk queue model was considered, where $G^{[S]}$ represents a general (i.e., arbitrary) bulk-service distribution, which includes the service distribution incorporating random service suspension considered in this study. However, this does not lower the contribution of this paper because to implement these equations, the formulation of $Y(z)$ needs to be specified. And in [Section 4.3.4](#) we will show how random service suspension introduces a new distribution for $Y^{(n)}$, which has not been considered in the literature.

4.3.4. Headway distribution

According to Proposition 4 to 6, to calculate $q_{0:C-1}^{(n)}$ and the mean and variance of queue length and waiting time, it is essential to specify $Y(z)$ (i.e., the PGF of the number of arrival passengers within a headway). According to Eq. 4, taking $l \rightarrow \infty$ gives that $Y^{(n)}|H^{(n)}$ is a Poisson random variable with parameter $\lambda^{(n)}H^{(n)}$. Therefore, we first consider the distribution of $H^{(n)}$ under the random service suspension.

According to the discussion in Section 3.3, the actual headway for vehicle l at station n is $H^{(n,l)} = t_D^{(n,l)} - t_D^{(n,l-1)} = \bar{H} + I^{(n,l)}$, where $I^{(n,l)}$ is the total duration of incident between $t_D^{(n,l-1)}$ and $t_D^{(n,l)}$. Since \bar{H} is a constant, obtaining headway distribution is equivalent to quantifying the distribution of $I^{(n,l)}$.

Proposition 7. *The total incident duration between $t_D^{(n,l-1)}$ and $t_D^{(n,l)}$ follows a compound Poisson-Exponential distribution with Poisson rate $\gamma\bar{H}$ and exponential rate θ . Mathematically,*

$$I^{(n,l)} = \sum_{i=1}^K X_i \quad \text{where } X_i \sim \mathbf{Exp}(\theta) \quad \forall i = 1, \dots, K, \text{ and } K \sim \mathbf{Poi}(\gamma\bar{H}) \quad (31)$$

Proof. After vehicle $l-1$ departs from station n , vehicle l must travel for additional (un-disrupted) \bar{H} time to reach station n . Since the system can only switch to the incident state from the normal state, the number of incident occurrence times, K , follows a Poisson distribution with rate $\gamma\bar{H}$. For the i -th incident, the vehicle stopping time, X_i , follows an exponential distribution with rate θ (i.e., mean $\frac{1}{\theta}$). Therefore, the sum of all incident duration is $I^{(n,l)} = \sum_{i=1}^N X_i$, where $X_i \sim \mathbf{Exp}(\theta) \quad \forall i = 1, \dots, N$, and $N \sim \mathbf{Poi}(\gamma\bar{H})$ \square

Now let us consider the distribution of $Y^{(n)}$. To derive the MGF of $Y^{(n)}$, the following lemma is introduced.

Lemma 1. For two arbitrary random variable X and Y , suppose that

- there is a $\delta > 0$ such that for t in $(-\delta, \delta)$, the MGF of $X|Y$ is $M_{X|Y}(t) = C_1(t)e^{C_2(t)Y}$, where $C_1(t)$ and $C_2(t)$ are finite functions of t that do not depend on Y ,
- and the MGF of Y , $M_Y(\cdot)$, exists and $M_Y[C_2(t)]$ is finite for t in $(-\delta, \delta)$.

Then the MGF of X is given by

$$M_X(t) = C_1(t)M_Y[C_2(t)], \quad -\delta < t < \delta. \quad (32)$$

Proof. The proof of Lemma 1 can be found in Villa and Escobar (2006) Result 1. \square

Proposition 8. *Under the setting of this study, $\forall n = 1, \dots, N$, the PGF of $Y^{(n)}$, $Y(z)$, can be expressed as*

$$Y(z) = M_{Y^{(n)}}(\log z) = e^{\bar{H}\lambda^{(n)}(z-1)} \exp \left[\gamma\bar{H} \left(\frac{\theta}{\theta - \lambda^{(n)}(z-1)} - 1 \right) \right] \quad (33)$$

Proof. As $I^{(n,l)}$ follows a compound Poisson-Exponential distribution (Proposition 7), the moment generation function (MGF) of a compound Poisson-Exponential variable can be written as

$$M_{I^{(n,l)}}(t) = e^{\gamma \bar{H}(\frac{\theta}{\theta-t}-1)} \quad \forall t < \theta \quad (34)$$

Therefore, the MGF for $H^{(n,l)} = \bar{H} + I^{(n,l)}$ is $M_{H^{(n,l)}}(t) = e^{\bar{H}t} e^{\gamma \bar{H}(\frac{\theta}{\theta-t}-1)}$. As this equation holds for all vehicles l , the MGF of $H^{(n)}$ (i.e., $l \rightarrow \infty$) is $M_{H^{(n)}}(t) = e^{\bar{H}t} e^{\gamma \bar{H}(\frac{\theta}{\theta-t}-1)}$.

Recall that $Y^{(n)}|H^{(n)}$ is a Poisson random variable with parameter $\lambda^{(n)}H^{(n)}$. So, the MGF of $Y^{(n)}|H^{(n)}$ is $M_{Y^{(n)}|H^{(n)}}(t) = \exp[\lambda^{(n)}H^{(n)}(e^t - 1)]$. Based on Lemma 1, setting $C_1(t) = 1$ and $C_2(t) = \lambda^{(n)}(e^t - 1)$, we conclude that the MGF of $Y^{(n)}$ is

$$M_{Y^{(n)}}(t) = e^{\bar{H}\lambda^{(n)}(e^t-1)} \exp \left[\gamma \bar{H} \left(\frac{\theta}{\theta - \lambda^{(n)}(e^t - 1)} - 1 \right) \right] \quad (35)$$

Substituting $t = \log z$ in Eq. 35 finishes the proof. □

From Eq. 35, taking corresponding derivatives and with slight manipulation we obtain:

$$\bar{Y}^{(n)} = \frac{\bar{H}\lambda^{(n)}(\gamma + \theta)}{\theta} \quad (36)$$

$$\bar{\bar{Y}}^{(n)} = \frac{\bar{H}\lambda_n(2\gamma\lambda_n + \gamma\theta + \theta^2)}{\theta^2} \quad (37)$$

$$\bar{\bar{\bar{Y}}}^{(n)} = \frac{\bar{H}\lambda_n(6\gamma\lambda_n^2 + 6\gamma\lambda_n\theta + \gamma\theta^2 + \theta^3)}{\theta^3} \quad (38)$$

With the expression of $Y(z)$, the z_0^*, \dots, z_{C-1}^* can be obtained by solving the nonlinear equation $\text{DEN}(z) = 0$. Then $q_{0:C-1}^{(n)}$ and other resilience indicators at station n can be obtained accordingly.

4.3.5. Solving for the roots

Solving for the roots of $\text{DEN}(z)$ is practically difficult as known in the queuing theory literature because typical optimization algorithms end up in finding only one root, while we need to find all C roots within the unit circle. This is especially challenging for $Y(z)$ with complex expressions because the objective function can be highly nonlinear (such as $Y(z)$ in this study). In this study, we propose an interpolation searching algorithm to efficiently find all roots of $\text{DEN}(z)$ within the unit circle.

Notice that $\text{DEN}(z) = 0$ is equivalent to find z_0^*, \dots, z_{C-1}^* , such that

$$\frac{1}{Y(z_k^*)} - S(1/z_k^*) = 0 \Leftrightarrow F(z_k^*) = 1 \quad \forall k = 0, \dots, C-1 \quad (39)$$

where $F(z) := Y(z)S(1/z)$. Taking the logarithm of both sides of Eq. 39 and matching the real and imaginary parts gives:

$$\begin{cases} \text{Re}[\log(H(z))] = 0 \\ \text{Im}[\log(H(z))] = 0 \end{cases} \quad (40)$$

where $\text{Re}[\cdot]$ and $\text{Im}[\cdot]$ represent the real and imaginary part of a complex number. Eq. 40 can be solved efficiently with many optimization algorithms (such as trust-region and Levenberg-Marquardt algorithms). However, as there are C optimal solutions for this problem with $|z^*| \leq 1$, the challenge is how to select different initial values so as to find all satisfied solutions.

Based on some numerical testing, we empirically observe that the distribution of the C solutions has an oval-like shape. Figure 6 shows some examples of the solution distribution with different values of $\rho^{(n)}$ (where $\rho^{(n)} = \bar{Y}^{(n)}/\bar{S}^{(n)}$ is the utilization ratio) and $s^{(n)}$. It is found that the closer $\rho^{(n)}$ is to 1 (resp. 0), the closer the shape of the root distribution is to an ellipse (resp. circle). The value of $s^{(n)}$ (i.e., available capacity distribution) can also slightly affect the root distribution.

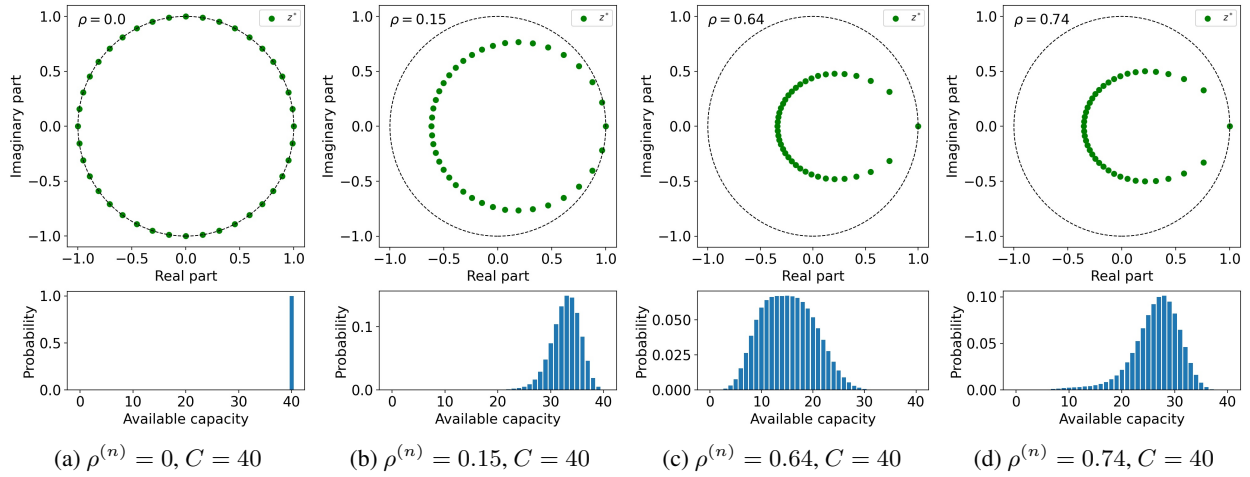


Figure 6: Examples of root distribution

We first express the complex number in polar coordinate system with $z = r \exp[\varphi i]$, where $i = \sqrt{-1}$, r is the length from z to the origin and φ is the angle. Eq. 40 now has C optimal solutions (r_k^*, φ_k^*) for $k = 0, 1, \dots, C - 1$, where $0 \leq r_k^* \leq 1$ and $0 \leq \varphi_k^* < 2\pi$. Note that $z_0^* = 1$ corresponds to $r_0^* = 1$ and $\varphi_0^* = 0$. Another property is that the roots must appear as conjugate pairs. Hence, if (r^*, φ^*) is a root and $0 < \varphi^* < \pi$, then $(r^*, 2\pi - \varphi^*)$ is also a root.

The proposed searching algorithm has two steps. The first step is referred to as “clockwise searching”, which is adapted from the numerical method in Powell (1985). The empirical observation (Figure 6) shows a rough relationship that $r_{k+1}^* - r_k^* \approx r_k^* - r_{k-1}^*$, especially for small $\rho^{(n)}$. This is equivalent to

$$r_{k+1}^* \approx 2r_k^* - r_{k-1}^* \quad (41)$$

Eq. 41 provides a way to determine the initial value for solving the $k + 1$ -th root when the k -th and $k - 1$ -th roots are available. As we already know $r_0^* = 1$ and $\varphi_0^* = 0$, we first empirically set the initial value for solving the second root as $r_1^{\text{Ini}} = 1 - 0.5\rho^{(n)}$ and $\varphi_1^{\text{Ini}} = 3\pi/C$. This is motivated by shape of the root distribution with respect to $\rho^{(n)}$. Then r_1^{Ini} and φ_1^{Ini} is used as the initial value to solve for r_1^* and φ_1^* based on Eq 40. For $k \geq 2$, the initial value for solving the k -th root is set as $r_k^{\text{Ini}} = r_{k-1}^* + (r_{k-1}^* - r_{k-2}^*)$,

$\varphi_k^{\text{Ini}} = \varphi_{k-1}^* + (\varphi_{k-1}^* - \varphi_{k-2}^*)$ according to Eq. 41.

However, only performing step 1 (i.e., [Powell \(1985\)](#)'s method) may not find all C distinct roots. Figure 7 shows some examples of the comparison between roots found in step 1 and all roots. We observe that when $\rho^{(n)}$ is relatively large (i.e., the system is relatively congested), the clockwise searching does not perform well because the approximate relationship in Eq. 41 does not hold. Even when $\rho^{(n)}$ is relatively small, it is also possible that some roots do not perfectly fit the oval-like shape (such as Figure 7a), resulting in the failure of step 1 to find all roots.

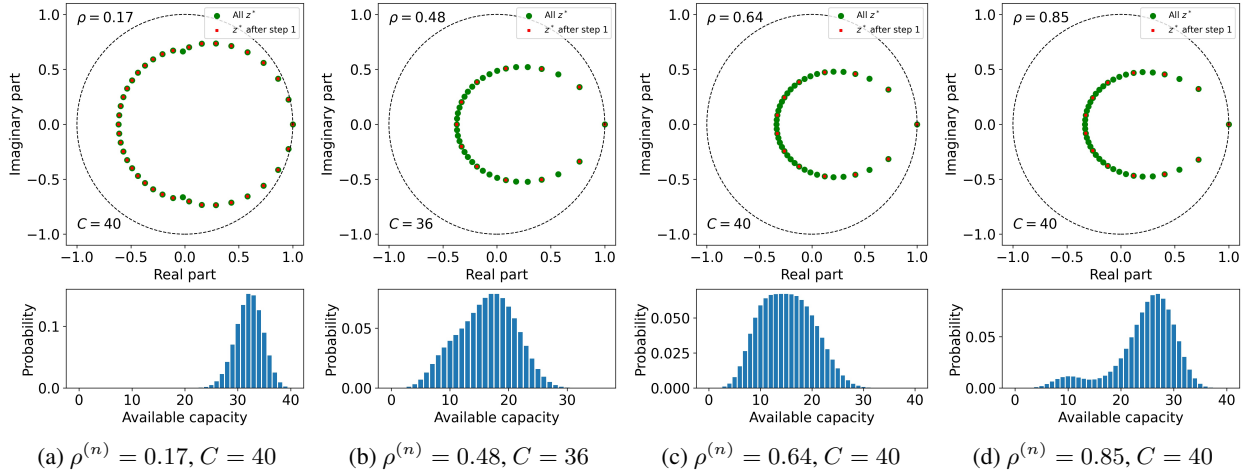


Figure 7: Comparison between roots found in clockwise searching and all roots

Therefore, we propose a second step called “interpolation searching”. Let the set of found roots from step 1 be $\mathcal{Z}^{(0)} = \{(r_0^{(0)}, \varphi_0^{(0)}), (r_1^{(0)}, \varphi_1^{(0)}), \dots, (r_{M_0}^{(0)}, \varphi_{M_0}^{(0)})\}$, where $M_0 = |\mathcal{Z}^{(0)}|$ is the number of found roots from step 1. Without loss of generality, suppose the elements in $\mathcal{Z}^{(0)}$ are clockwise ranked (i.e., $\varphi_0^{(0)} < \varphi_1^{(0)} < \dots < \varphi_{M_0}^{(0)}$). The interpolation searching is described in Algorithm 1. The main idea is to perform interpolation between any two adjacent roots that are already found. The interpolated points are set as initial values and fed into Eq. 40 to solve for new distinct roots. Then we update the found roots set with the new distinct roots or perform more subdivisional (i.e., larger L) interpolation if no distinct roots are found. This process is repeated until there are C distinct roots found. In Algorithm 1, L is a parameter controlling how many points to interpolate between two known roots, and ϵ is a predetermined probability threshold to add randomness into the searching process.

From the results of numerical testing, our algorithm allows us to find desired roots for all testing scenarios in Section 5.1. While the methods of [Powell \(1985\)](#) (i.e., only step 1) and [Wilson \(2014\)](#) (which is used in [Islam et al. \(2015\)](#)) fail to.

4.4. Stability condition

For all the derivations above, we assume that the steady-state distributions of all variables exist. This triggers the discussion about the stability condition, which is also an important indicator of the system’s resilience. At the station level, the stability condition is described in Proposition 9.

Algorithm 1 Interpolation searching

```

1: Initialize  $\mathcal{Z}^{(0)}$ ,  $M_0$ ,  $\epsilon$ . Initialize  $L = 2$ ,  $k = 0$ .
2: while  $M_k < C$  do
3:   Initialize  $\mathcal{Z}^{\text{Ini}}$  as an empty set.
4:   for  $i = 1 : M_k$  do
5:     for  $d = 1 : L - 1$  do
6:        $r^{\text{Ini}} = r_i^{(k)} + d \cdot \frac{r_{i+1}^{(k)} - r_i^{(k)}}{L}$ ;  $\varphi^{\text{Ini}} = \varphi_i^{(k)} + d \cdot \frac{\varphi_{i+1}^{(k)} - \varphi_i^{(k)}}{L}$ 
7:       Draw a random value  $w$  uniformly from  $[0, 1)$ 
8:       if  $w < \epsilon$  then
9:         Draw a random value  $\delta_1$  uniformly from  $[-\frac{|r_{i+1}^{(k)} - r_i^{(k)}|}{2L}, \frac{|r_{i+1}^{(k)} - r_i^{(k)}|}{2L}]$ 
10:         $r^{\text{Ini}} = r^{\text{Ini}} + \delta_1$ 
11:        Draw a random value  $\delta_2$  uniformly from  $[-\frac{|\varphi_{i+1}^{(k)} - \varphi_i^{(k)}|}{2L}, \frac{|\varphi_{i+1}^{(k)} - \varphi_i^{(k)}|}{2L}]$ 
12:         $\varphi^{\text{Ini}} = \varphi^{\text{Ini}} + \delta_2$ 
13:        Add  $(r^{\text{Ini}}, \varphi^{\text{Ini}})$  into  $\mathcal{Z}^{\text{Ini}}$ .
14:   Initialize  $\mathcal{Z}^{\text{temp}}$  as an empty set.
15:   for all  $z^{\text{Ini}}$  in  $\mathcal{Z}^{\text{Ini}}$  do
16:     Solve Eq. 40 using  $z^{\text{Ini}}$  as the initial value and obtaining  $z^*_{\text{temp}}$ . Let its conjugate be  $\bar{z}^*_{\text{temp}}$ .
17:     If  $z^*_{\text{temp}}$  ( $\bar{z}^*_{\text{temp}}$ ) not in  $\mathcal{Z}^{(k)}$ , add it into  $\mathcal{Z}^{\text{temp}}$ , otherwise do nothing.
18:    $\mathcal{Z}^{(k+1)} = \mathcal{Z}^{(k)} \cup \mathcal{Z}^{\text{temp}}$  and rank all elements in  $\mathcal{Z}^{(k+1)}$  clockwise
19:   Denote  $\mathcal{Z}^{(k+1)}$  as  $\{(r_0^{(k+1)}, \varphi_0^{(k+1)}), \dots, (r_{M_{k+1}}^{(k+1)}, \varphi_{M_{k+1}}^{(k+1)})\}$ 
20:    $k = k + 1$ 
21:   if  $M_{k+1} = M_k$  then
22:      $L = L + 1$ 

```

Proposition 9. *Under the setting of this study, the bulk-service queuing system at station n is stable if and only if*

$$\rho^{(n)} = \frac{\bar{Y}^{(n)}}{\bar{S}^{(n)}} = \frac{\lambda^{(n)} \bar{H}}{\sum_{u=0}^C s_u^{(n)} u} + \frac{\gamma}{\theta} \frac{\lambda^{(n)} \bar{H}}{\sum_{u=0}^C s_u^{(n)} u} < 1 \quad (42)$$

where $\rho^{(n)}$ is the utilization ratio for station n .

Proof. The stability condition is equivalent to $\mathbb{P}(Q^{(n)} = 0) = q_0^{(n)} > 0$. In Eq. 20, we notice that $\prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1}$ is always greater than 0 (see Appendix B for details), and $s_C^{(n)} > 0$ is a known condition. Therefore, $q_0^{(n)} > 0$ if and only if $\bar{Y}^{(n)} < \bar{S}^{(n)}$ (i.e., $\rho^{(n)} < 1$), which finishes the proof. \square

Proposition 9 is intuitive as it indicates that station n is stable if the average number of arrival passengers within a headway is smaller than the average available capacity for each arrival vehicle (after alighting). And it is shown that the higher rate of incidents (i.e., larger γ) and higher duration of incidents (i.e., higher $\frac{1}{\theta}$) make the system more likely to be unstable. There are some remarks for Proposition 9.

Remark 2. As $\rho^{(n)}$ depends on $s^{(n)}$ and $s^{(n)}$ depends on the roots (i.e., z_0^*, \dots, z_{C-1}^*) at station n , there is no direct way to judge the stability of station n without iterating the previous $n - 1$ stations. But for the

first station ($n = 1$), we have $s_C^{(1)} = 1$ and $s_u^{(1)} = 0$ for all $u = 0, \dots, C - 1$. Then Eq. 42 reduces to $\rho^{(1)} = \frac{\lambda^{(n)}\bar{H}}{C} + \frac{\gamma}{\theta} \frac{\lambda^{(n)}\bar{H}}{C}$, which enables us to judge the stability directly.

Remark 3. Proposition 9 only discusses the stability at the station level. For the whole system, we can naturally define the system stability condition as that “all stations in the system are stable”. Mathematically, the single-route PT system is stable if and only if $\rho^{(n)} < 1, \forall n = 1, 2, \dots, N$.

Remark 4. It is worth discussing the relationship of stability for station n and $n - 1$. If station $n - 1$ is stable, then $s^{(n)}$ can be calculated as described in Section 4.1, and the stability of station n can be evaluated accordingly. However, if station $n - 1$ is not stable, station n may also be stable because there may be passengers alighting at station n . For this situation, we have $v_C^{(n-1)} = 1$ and $v_k^{(n-1)} = 0$ for all $k = 0, 1, \dots, C - 1$. Then $s^{(n)}$ is purely determined by the alighting rate at station n . It is easy to verify that at this situation $\bar{S}^{(n)} = \alpha^{(n)}C$. And the stability condition is $\rho^{(n)} = \frac{\lambda^{(n)}\bar{H}}{\alpha^{(n)}C} + \frac{\gamma}{\theta} \frac{\lambda^{(n)}\bar{H}}{\alpha^{(n)}C} < 1$.

4.5. Summary of calculation procedure

So far, we have derived the calculation process for all variables of interest. Algorithm 2 summarizes the calculation procedure. We only need to iterate for N stations, which is more efficient and provides more analytical insights than a simulation model.

Algorithm 2 Resilience indicators calculation procedure

- 1: Initialize $v_0^{(0)} = 1$ and $v_k^{(0)} = 0 \quad \forall k = 1, \dots, C$.
 - 2: **for** $n = 1 : N$ **do**
 - 3: $g^{(n)} = v^{(n-1)}A^{(n)} \quad \triangleright$ Eq. 9
 - 4: $s_k^{(n)} = 1 - g_{C-k}^{(n)} \quad \forall k = 0, 1, \dots, C \quad \triangleright$ Eq. 8
 - 5: Calculate $\bar{S}^{(n)}, \bar{\bar{S}}^{(n)}$, and $\bar{\bar{\bar{S}}}^{(n)}$ based on $s^{(n)}$.
 - 6: Calculate $\bar{Y}^{(n)}, \bar{\bar{Y}}^{(n)}$, and $\bar{\bar{\bar{Y}}}^{(n)}$ \triangleright Section 4.3.4
 - 7: **if** $\bar{Y}^{(n)} < \bar{S}^{(n)}$ **then** \triangleright Station n is stable
 - 8: Solve the roots z_0^*, \dots, z_{C-1}^* for the denominator of $Q(z)$ in Eq. 18 \triangleright Section 4.3.1
 - 9: Calculate $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ based on z_0^*, \dots, z_{C-1}^* \triangleright Section 4.3.2
 - 10: Calculate $\mathbb{E}[Q^{(n)}], \text{Var}[Q^{(n)}], \mathbb{E}[W^{(n)}],$ and $\text{Var}[W^{(n)}]$ \triangleright Eq. 25 - 28
 - 11: $v^{(n)} = g^{(n)}B^{(n)} \quad \triangleright$ Eq. 13. $B^{(n)}$ is a function of $q_{0:C-1}^{(n)}$
 - 12: **else** \triangleright Station n is not stable
 - 13: $q_k^{(n)} = 0 \quad \forall k = 0, 1, \dots, C - 1$
 - 14: Set $\mathbb{E}[Q^{(n)}], \text{Var}[Q^{(n)}], \mathbb{E}[W^{(n)}],$ and $\text{Var}[W^{(n)}]$ to infinity
 - 15: $v_C^{(n)} = 1$ and $v_k^{(n)} = 0 \quad \forall k = 0, 1, \dots, C - 1$
-

5. Numerical example

5.1. Experiment design

To test the proposed framework, we implement the method to an example bus route adapted from Islam et al. (2015) and Hickman (2001). There are 10 stations in the bus route and the attributes for each station are

shown in Table 1. Note that passengers left behind (i.e., unable to board the first arrival bus due to capacity limitation) are expected to happen at stations 4 and 5 as the passenger arrival rates are relatively high.

Table 1: Example bus system parameters

Station ID	$\lambda^{(n)}$ (passenger/min)	$\alpha^{(n)}$	Station ID	$\lambda^{(n)}$ (passenger/min)	$\alpha^{(n)}$
1	0.75	0	6	1	0.8
2	1.5	0	7	0.75	0.5
3	0.75	0.1	8	0.5	0.1
4	3	0.25	9	0.2	0.75
5	1.5	0.25	10	0	1

To test the sensitivity of resilience indicators to different parameters, we design the value spaces for C , θ , γ , \bar{H} , and the demand factor (Table 2). Demand factor is a value multiplied to the $\lambda^{(n)}$ in Table 1, representing the change of passenger arrival rates. When the sensitivity testing is conducted for one parameter (e.g., C), other parameters (e.g., θ , γ , \bar{H} , and the demand factor) are set as reference values for comparison.

Table 2: Scenario design

Parameters	Value space	Reference value
C	{36, 40, 44, 48}	40
γ (/min)	{0, 1/240, 1/180, 1/120, 1/60, 1/30, 1/15}	1/60
θ (/min)	{1/2, 1/5, 1/10, 1/20}	1/5
\bar{H} (min)	{2, 4, 6, 8}	6
Demand factor	{0.2, 0.4, 0.6, 0.8, 1}	0.8

5.2. Resilience indicators

The mean and standard deviation of queue length for each station under different testing scenarios are shown in Figure 8. Generally, for all scenarios, the queue length patterns are corresponding to the congestion patterns we expect given the passenger arrival and alighting rates. That is, the expected queue length is relatively higher in stations 4 and 5. The expected queue length at the last station is always zero as its passenger arrival rate is equal to 0.

Figure 8a shows the change in queue length patterns with respect to bus capacity. And we observe that the system is not very sensitive to bus capacity. Increasing bus size from 36 to 48 only decreases the expected queue length by around 0.85 at station 4 (the most congested station). The impact on other stations is even smaller.

Figure 8b shows the impact of incident occurrence rate γ on queue length. When there is no random suspension in the system ($\gamma = 0$), the expected queue length at station 4 is 14.42. As the frequency of incidents increases, the system becomes more congested with longer expected queue length and higher variance. When the incident happens 15 minutes per time on average ($\gamma = 1/15$), the expected queue length at station 4 is increased to 26.66.

Similar results can be observed for the duration of incidents (Figure 8c). When the average incident duration is 2 minutes ($\theta = 1/2$), $\mathbb{E}(Q^{(4)}) = 14.94$. While when the average incident duration is increased to 20 minutes ($\theta = 1/20$), $\mathbb{E}(Q^{(4)})$ is increased to 50.28. The impacts of θ and γ on queue length are both more significant at crowding stations.

The impact of \bar{H} is shown in Figure 8d. As expected, higher headway means a lower service rate and thus a higher expected queue length. As \bar{H} increases from 2 minutes to 8 minutes, the queue length at station 4 increases from 5.24 to 29.92. The impact of the demand factor (Figure 8e) shows the similar patterns. As the demand factor increases from 0.2 to 1.0, the queue length at station 4 increases from 3.90 to 23.19. The impacts of \bar{H} and demand factor are relatively similar for crowding and uncrowded stations.

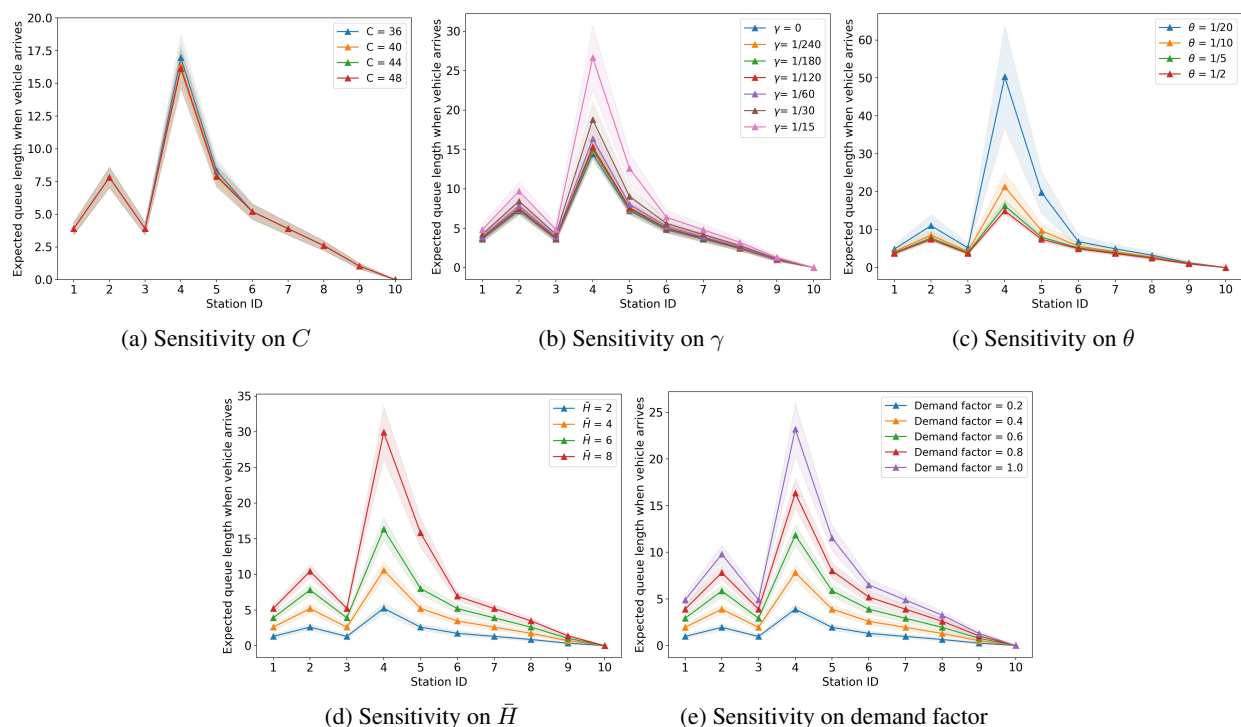


Figure 8: Mean and standard deviation of queue length (the shade part is $0.2 \times$ standard deviation)

Figure 9 shows the mean and standard deviation of passenger waiting time under different testing scenarios. When no left behind happens, the average waiting time for Poisson arrival passengers is shown in Eq. 2. Hence, we observe that in many scenarios, the expected waiting time at the first and last several stations are the same because the available bus capacity at these stations is high and headway distributions are the same for all stations. However, for congested stations, such as stations 4 and 5, high waiting times are observed due to being left behind.

Figure 9a shows the impact of bus size on the waiting time. Similar to the results on queue length, the impact is not very significant. When bus size increases from 36 to 48, the average waiting time at station 4 decreases from 4.20 to 3.85 minutes. As the change of C does not affect the headway distribution, no impact is observed for uncontested stations such as stations 1 and 9.

The impacts of γ and θ on waiting times are shown in Figure 9b and 9c, respectively. As the increase in γ and $1/\theta$ results in an increase in expected headway, the mean waiting times at all stations are increased. But the impacts on crowding stations are more significant. When $\gamma = 0$, there is no incident in the system. All stations have no left behind and their waiting times are all equal to 3 minutes ($\frac{1}{2}\bar{H}$ as no incident means all stations have the same fixed headway). As γ is increased to $1/15$, station 4 has left behind and its waiting time is increased to 8.36 minutes. When θ increases from $1/20$ to $1/2$, the expected waiting time at station 4 increases from 3.19 to 21.95 minutes.

Changes in \bar{H} have the most direct impact on the expected waiting time. When \bar{H} is smaller than 6 minutes, all stations have similar waiting times because there is nearly no left behind. However, when \bar{H} equals to 8 minutes, left behind happens at stations 4 and 5 and their waiting times increase to 8.52 and 9.11 minutes, respectively.

Finally, the impact of demand factors is similar to that of C . Since it does not change the headway distribution, no impact is observed for uncontested stations. With the increase of the demand factor from 0.2 to 1.0, the expected waiting time at the station increases from 3.63 to 4.86 minutes.

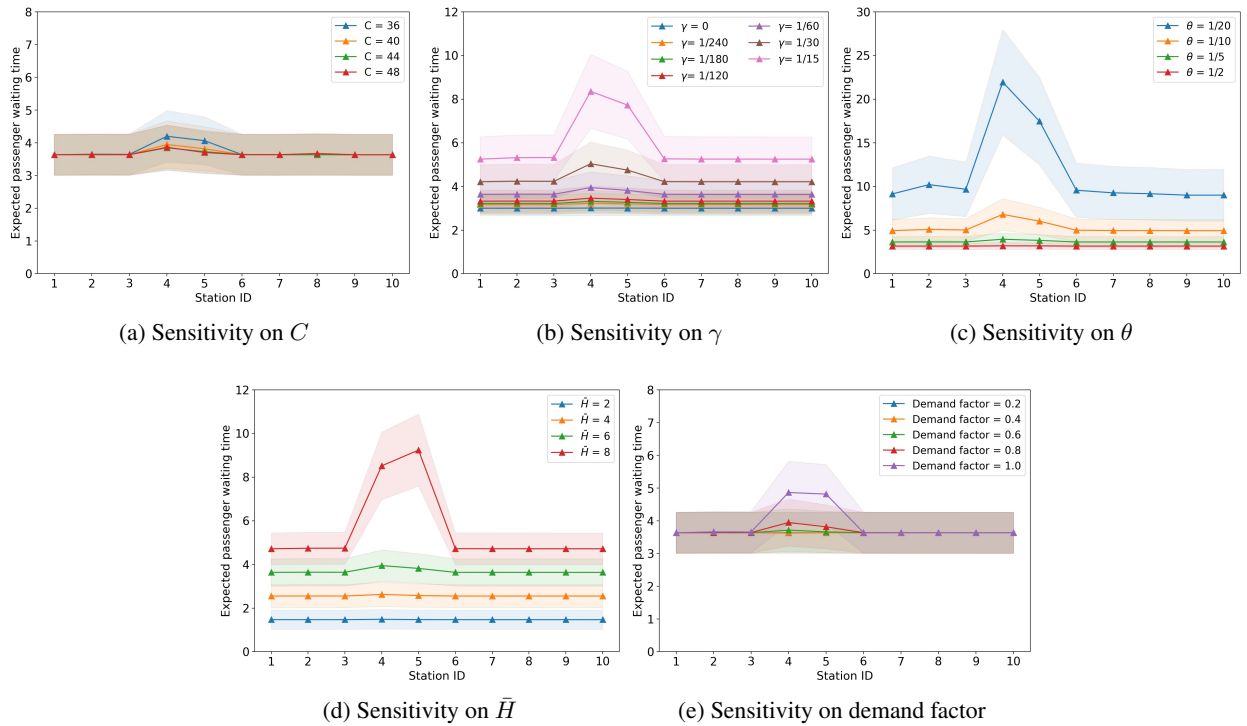


Figure 9: Mean and standard deviation of waiting time (the shade part is $0.2 \times$ standard deviation)

5.3. Comparison between simulation and theoretical results

5.3.1. Simulation model

To validate the theoretical results, we develop a simulation model to calculate the expectation and variance of queue length and waiting time. The simulation procedure is shown in Algorithm 3. For each station n , we generate a Poisson process of arrival passengers. And for each vehicle l , we generate a two-state

Markov process of “normal” and “failure” states. Then, at each time step, the simulation goes through each vehicle, based on the vehicle’s state, deciding the vehicle will move or stay idle. When a vehicle arrives at a station, passengers are boarded based on the first-in-first-serve (FIFS) principle up to the vehicle’s capacity C . Queue lengths at vehicle arrival and passenger waiting times are recorded during the simulation. To ensure the system reaches the steady-state distribution, the first 20% records are dropped.

Algorithm 3 Simulation procedure

- 1: Initialize model parameters: $C, \gamma, \theta, \bar{H}$, Demand factor. Set the total simulation time as T and time step as Δ .
 - 2: Generate Poisson arrival passengers at each station
 - 3: Generate two-state Markov process of “normal” and “failure” states for each vehicle
 - 4: Set time counter $t = 0$
 - 5: **while** $t < T$ **do**
 - 6: **if** $t \% \bar{H} = 0$ **then**
 - 7: Dispatch a new empty vehicle
 - 8: **for** each vehicle in the current vehicle list **do**
 - 9: Update vehicle state
 - 10: **if** vehicle state = “failure” **then**
 - 11: Continue
 - 12: **else**
 - 13: Move vehicle by $\text{speed} \times \Delta$
 - 14: **if** vehicle arrives at a station **then**
 - 15: Record queue length
 - 16: Alight passengers based on the binomial distribution
 - 17: Board passengers based on FIFS principle up to the vehicle capacity
 - 18: Record passenger waiting time
 - 19: $t = t + \Delta$
 - 20: Drop the first 20% records for stability. Calculate $\mathbb{E}[Q^{(n)}], \text{Var}[Q^{(n)}], \mathbb{E}[W^{(n)}],$ and $\text{Var}[W^{(n)}]$ for all stations.
-

5.3.2. Results

We compare the simulation and theoretical results for the reference parameter setting (Table 2). $\Delta = 0.01$ minute and $T = 20,000$ minutes are used. Compare mean and std. for queue length and waiting time. The mean and standard deviation comparisons are shown in Figure 10. We observe that the simulation and theoretical results are well matched. The slight mismatching may be due to the fact of “bus bunching”. That is, in the simulation, bus l is allowed to overtake bus $l - 1$ and arrives at station n earlier. In this situation, the headway will be $t_D^{(n,l-1)} - t_D^{(n,l)}$, instead of $t_D^{(n,l)} - t_D^{(n,l-1)}$ as discussed before. Hence, the headway distribution may be slightly different from Proposition 8. Specifically, as mentioned in Manguier (1985) and Hickman (2001), there exists a headway variance propagation from the upstream to downstream stations, implying that the headway variation in downstream stations is higher. This explains the higher waiting time (Eq. 2) in downstream stations for the simulation model in Figure 10c. Addressing the impact of bus bunching on headway distribution is a challenging job and we reserve it as future work.

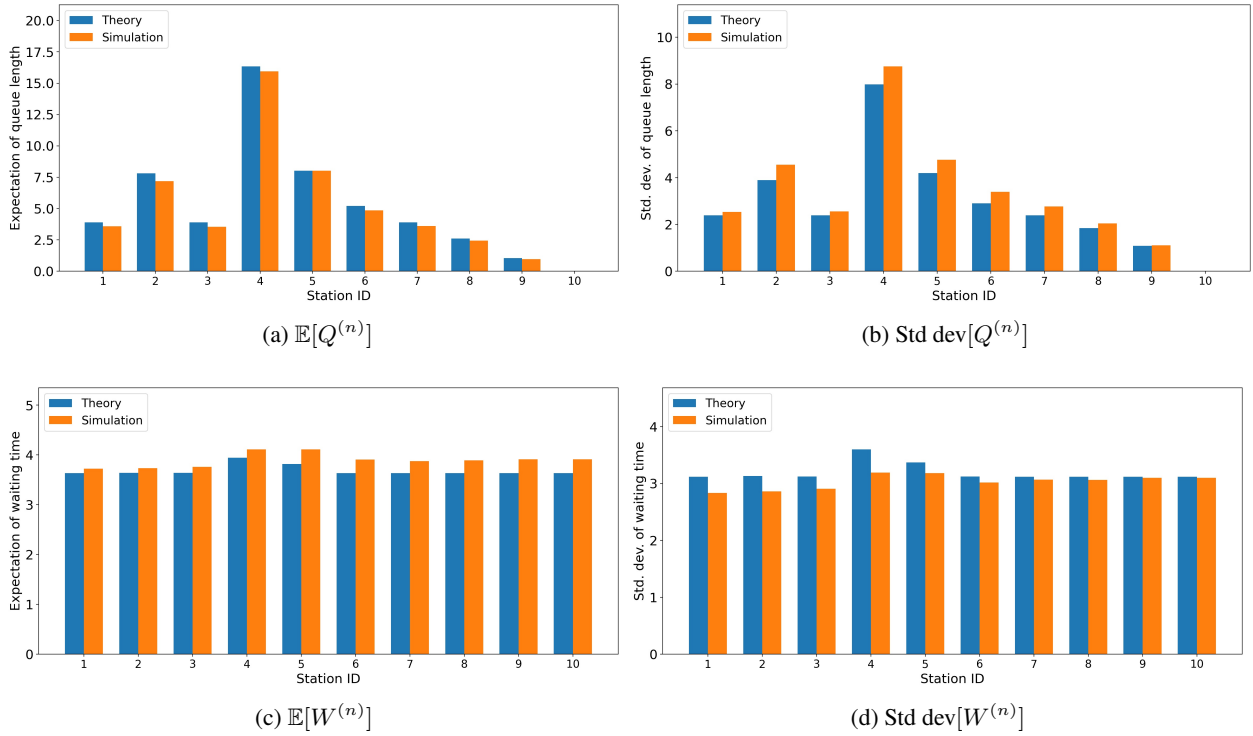


Figure 10: Comparison between simulation and theoretical results (reference scenario)

6. Conclusion and discussion

This paper proposes a stochastic framework to model the resilience of public transit systems under short random service suspensions. Specifically, we analyze the system stability conditions and derive closed-form formulations for the mean and variance of queue length and waiting time at each station. The derived stability conditions are intuitive and imply that the system is more likely to be unstable with high incident rates, high incident duration, high demand, low service frequency, and low vehicle capacity. The proposed model is implemented on an example bus network with sensitivity analysis of different parameters (such as incident rate, incident duration, vehicle capacity, etc.). Results show that the congested stations (i.e., stations with high demand rates) are more vulnerable to random service suspensions. The results are validated with a simulation model, showing consistency between theoretical and simulation outcomes.

The proposed model has several potential implementations. 1) It can facilitate the design and planning of public transit systems with the consideration of random system interruptions, such as the design of headway and vehicle capacity. Moreover, the estimated queue length can be used to evaluate the layout and capacity of congested stations. 2) The model can be used to monitor the system performance and identify critical stations by inputting the historical demand and incident information. 3) The model can support efficient cost-benefit analysis to relieve congestion. Since the model provides close form formulations to waiting time and queue length, it can be used to answer that, to control the waiting time within a threshold, what is the most effective way (e.g., increase vehicle size, decrease headway, or reduce the random suspension rate). In summary, the efficient calculation of the system's resilience indicators can be used in public transit planning,

operation, and management fields.

Future studies can be conducted from the following aspects. First, the model can be extended from line-level to network-level. The main difference between line-level and network-level is the consideration of transfer passengers. A straightforward way is to incorporate the transfer demand as part of the arrival demand. But additional transfer-related parameters (which should be connected with the alighting rate) need to be specified in the model. Second, like many previous random service disruption papers (Section 2.2), the model can be extended to consider partial interruptions (as opposed to fully stopped as assumed in this paper). With partial interruptions, vehicles can still have positive speed at the failure state. And headway distribution needs to be revised. Third, as mentioned before, this paper assumes no balking and renegeing behaviors of passengers. Future studies may consider a more complicated passenger-side behavior to extend this model.

References

- Ayyappan, G., Karpagam, S., 2020. Analysis of a bulk service queue with unreliable server, multiple vacation, overloading and stand-by server. *International Journal of Mathematics in Operational Research* 16, 291–315.
- Bailey, N.T., 1954. On queueing processes with bulk service. *Journal of the Royal Statistical Society: Series B (Methodological)* 16, 80–87.
- Balakrishna, R., Wen, Y., Ben-Akiva, M., Antoniou, C., 2008. Simulation-based framework for transportation network management in emergencies. *Transportation Research Record* 2041, 80–88.
- Berdica, K., 2002. An introduction to road vulnerability: what has been done, is done and should be done. *Transport policy* 9, 117–127.
- Chaudhry, M., Templeton, J.G., 1983. *First course in bulk queues*. .
- Chaudhry, M.L., Madill, B., Briere, G., 1987. Computational analysis of steady-state probabilities of m/g , $b/1$ and related nonbulk queues. *Queueing systems* 2, 93–114.
- Clerq, F., 1972. A public transport assignment method. *Verkeerstechniek, Netherlands* 23.
- Currie, G., Muir, C., 2017. Understanding passenger perceptions and behaviors during unplanned rail disruptions. *Transportation Research Procedia* 25, 4396–4406.
- Dial, R.B., 1967. Transit path finder algorithm. *Highway Research Record* .
- Downton, F., 1955. Waiting time in bulk service queues. *Journal of the Royal Statistical Society: Series B (Methodological)* 17, 256–261.
- Fukasawa, N., Yamauchi, K., Murakoshi, A., Fujinami, K., Tatsui, D., 2012. Provision of forecast train information and consequential impact on decision making for train-choice. *Quarterly Report of RTRI* 53, 141–147.
- Hickman, M.D., 2001. An analytic stochastic model for the transit vehicle holding problem. *Transportation Science* 35, 215–237.
- Hong, L., Gao, J., Zhu, W., 2018. Self-evacuation modelling and simulation of passengers in metro stations. *Safety science* 110, 127–133.
- Islam, M.K., Vandebona, U., Dixit, V.V., Sharma, A., 2014. A bulk queue model for the evaluation of impact

- of headway variations and passenger waiting behavior on public transit performance. *IEEE Transactions on Intelligent Transportation Systems* 15, 2432–2442.
- Islam, M.K., Vandebona, U., Dixit, V.V., Sharma, A., 2015. A model to evaluate the impact of headway variation and vehicle size on the reliability of public transit. *IEEE Transactions on Intelligent Transportation Systems* 16, 1840–1850.
- Jaiswal, N., 1960. Time-dependent solution of the bulk-service queuing problem. *Operations Research* 8, 773–781.
- Jayaraman, D., Nadarajan, R., Sitrarasu, M., 1994. A general bulk service queue with arrival rate dependent on server breakdowns. *Applied mathematical modelling* 18, 156–160.
- Jin, J.G., Tang, L.C., Sun, L., Lee, D.H., 2014. Enhancing metro network resilience via localized integration with bus services. *Transportation Research Part E: Logistics and Transportation Review* 63, 17–30.
- Jin, J.G., Teo, K.M., Odoni, A.R., 2016. Optimizing bus bridging services in response to disruptions of urban transit rail networks. *Transportation Science* 50, 790–804.
- Kahraman, A., Gosavi, A., 2011. On the distribution of the number stranded in bulk-arrival, bulk-service queues of the $m/g/1$ form. *European journal of operational research* 212, 352–360.
- Krishnamoorthy, A., Pramod, P.K., Chakravarthy, S.R., 2014. Queues with interruptions: a survey. *Top* 22, 290–320.
- Kroon, L., Huisman, D., 2011. Algorithmic support for railway disruption management, in: *Transitions Towards Sustainable Mobility*. Springer, pp. 193–210.
- Lin, T., Srikukenthiran, S., Miller, E., Shalaby, A., 2018. Subway user behaviour when affected by incidents in toronto (subwait) survey—a joint revealed preference and stated preference survey with a trip planner tool. *Canadian Journal of Civil Engineering* 45, 623–633.
- Luo, C., Li, X., Zhou, Y., Caunhye, A.M., Alibrandi, U., Aydin, N.Y., Ratti, C., Eckhoff, D., Bojic, I., 2019. Data-driven disruption response planning for a mass rapid transit system, in: *Smart Transportation Systems 2019*. Springer, pp. 205–213.
- Madan, K., 1989. A single channel queue with bulk service subject to interruptions. *Microelectronics Reliability* 29, 813–818.
- Madan, K., 1992. A bulk queueing system with random failures and two phase repairs. *Microelectronics Reliability* 32, 669–677.
- Marguier, P.H.J., 1985. Bus route performance evaluation under stochastic considerations. Ph.D. thesis. Massachusetts Institute of Technology.
- Mo, B., Franque, M.Y.v., Koutsopoulos, H., Attanucci, J., Zhao, J., 2021. Impact of unplanned service disruption on urban rail systems, *Transportation Research Board 100th Annual Meeting*.
- Mo, B., Ma, Z., Koutsopoulos, H.N., Zhao, J., 2020. Capacity-constrained network performance model for urban rail systems. *Transportation Research Record* 2674, 59–69.
- Murray-Tuite, P., Wernstedt, K., Yin, W., 2014. Behavioral shifts after a fatal rapid transit accident: A multinomial logit model. *Transportation research part F: traffic psychology and behaviour* 24, 218–230.
- Neuts, M.F., 1967. A general class of bulk queues with poisson input. *The Annals of Mathematical Statistics* 38, 759–770.

- Osuna, E., Newell, G.F., 1972. Control strategies for an idealized public transportation system. *Transportation Science* 6, 52–72.
- O’Dell, S.W., Wilson, N.H., 1999. Optimal real-time control strategies for rail transit operations during disruptions, in: *Computer-aided transit scheduling*. Springer, pp. 299–323.
- Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., et al., 2014. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.
- Powell, W.B., 1981. Stochastic delays in transportation terminals: New results in the theory and application of bulk queues. Ph.D. thesis. Massachusetts Institute of Technology.
- Powell, W.B., 1983. Bulk service queues with deviations from departure schedules: The problem of correlated headways. *Transportation Research Part B: Methodological* 17, 221–232.
- Powell, W.B., 1985. Analysis of vehicle holding and cancellation strategies in bulk arrival, bulk service queues. *Transportation Science* 19, 352–377.
- Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., Auld, J., 2019. Analysis of transit users’ waiting tolerance in response to unplanned service disruptions. *Transportation Research Part D: Transport and Environment* 77, 639–653.
- Rapoport, A., Stein, W.E., Mak, V., Zwick, R., Seale, D.A., 2010. Endogenous arrivals in batch queues with constant or variable capacity. *Transportation Research Part B: Methodological* 44, 1166–1185.
- Rudin, W., 2006. *Real and complex analysis*. Tata McGraw-hill education.
- Sasikala, S., Indhira, K., 2016. Bulk service queueing models—a survey. *International Journal of Pure and Applied Mathematics* 106, 43–56.
- Selvi, A., Rosenshine, M., 1983. A queueing system for airport buses. *Transportation Research Part B: Methodological* 17, 427–434.
- Singh, I., Ram, C., 1991. Three-server bulk service queue with service interruptions and exponential repairs. *Microelectronics Reliability* 31, 257–259.
- Suarez, P., Anderson, W., Mahal, V., Lakshmanan, T., 2005. Impacts of flooding and climate change on urban transportation: A systemwide performance assessment of the boston metro area. *Transportation Research Part D: transport and environment* 10, 231–244.
- Sun, H., Wu, J., Wu, L., Yan, X., Gao, Z., 2016. Estimating the influence of common disruptions on urban rail transit networks. *Transportation Research Part A: Policy and Practice* 94, 62–75.
- Tadj, L., Choudhury, G., 2009. A quorum queueing system with an unreliable server. *Applied mathematics letters* 22, 1710–1714.
- Tadj, L., Choudhury, G., Rekab, K., 2012. A two-phase quorum queueing system with bernoulli vacation schedule, setup, and n-policy for an unreliable server with delaying repair. *International Journal of Services and Operations Management* 12, 139–164.
- Teng, J., Liu, W.R., 2015. Development of a behavior-based passenger flow assignment model for urban rail transit in section interruption circumstance. *Urban Rail Transit* 1, 35–46.
- Tian, X., Zheng, B., 2018. Using smart card data to model commuters’ responses upon unexpected train delays, in: *2018 IEEE International Conference on Big Data (Big Data)*, IEEE. pp. 831–840.

- Villa, E.R., Escobar, L.A., 2006. Using moment generating functions to derive mixture distributions. *The American Statistician* 60, 75–80.
- Wang, Y., Guo, J., Ceder, A.A., Currie, G., Dong, W., Yuan, H., 2014. Waiting for public transport services: Queueing analysis with balking and renegeing behaviors of impatient passengers. *Transportation Research Part B: Methodological* 63, 53–76.
- Welding, P., 1957. The instability of a close-interval service. *Journal of the operational research society* 8, 133–142.
- Wilson, H., 2014. Complex zeros of functions. <https://www.mathworks.com/matlabcentral/fileexchange/25178-complex-zeros-of-functions>. MATLAB Central File Exchange. Retrieved December 6, 2020.
- Wirasinghe, S., 1980. Nearly optimal parameters for a rail/feeder-bus system on a rectangular grid. *Transportation Research Part A: General* 14, 33–40.
- Xu, X., Chen, A., Jansuwan, S., Heaslip, K., Yang, C., 2015. Modeling transportation network redundancy. *Transportation research procedia* 9, 283–302.
- Yin, H., Han, B., Li, D., Wang, Y., 2016. Evaluating disruption in rail transit network: a case study of beijing subway. *Procedia Engineering* 137, 49–58.
- Zhang, D.m., Du, F., Huang, H., Zhang, F., Ayyub, B.M., Beer, M., 2018. Resiliency assessment of urban rail transit networks: Shanghai metro as an example. *Safety Science* 106, 230–243.

Appendices

Appendix A. Derivation of probability generating function of $Q^{(n)}$

From the relationship in Eq. 3, we have

$$q_k^{(n,l+1)} = \sum_{i=0}^k r_i^{(n,l)} y_{k-i}^{(n,l)} \quad (\text{A.1})$$

where $r_k^{(n,l)} := \mathbb{P}(R^{(n,l)} = k)$ and $y_k^{(n,l)} := \mathbb{P}(Y^{(n,l)} = k)$ for all non-negative integers k . Let $S^{(n,l)}$ be the number of available spaces for train l when it arrives at station n . Given that $S^{(n,l)} = u$, if u is greater than or equal to $Q^{(n,l)}$, all passengers can board and there is no left-behind. Then we have

$$r_0^{(n,l)} \Big|_{S^{(n,l)}=u} = \mathbb{P}(u \geq Q^{(n,l)}) = \sum_{k=0}^u q_k^{(n,l)} \quad (\text{A.2})$$

where $r_k^{(n,l)} \Big|_{S^{(n,l)}=u} = \mathbb{P}(R^{(n,l)} = k \mid S^{(n,l)} = u)$. If u is less than $Q^{(n,l)}$, only u passengers can board and there are $Q^{(n,l)} - u$ number of left-behind passengers. So

$$r_i^{(n,l)} \Big|_{S^{(n,l)}=u} = \mathbb{P}(Q^{(n,l)} - u = i) = q_{u+i}^{(n,l)} \quad i = 1, 2, \dots \quad (\text{A.3})$$

Based on Eq. A.2 and A.3, Eq A.1 can be reformulated as

$$q_k^{(n,l+1)} = \sum_{u=0}^C s_u^{(n,l)} \left(\sum_{i=0}^u q_i^{(n,l)} y_k^{(n,l)} + \sum_{i=u+1}^{u+k} q_i^{(n,l)} y_{k-i+u}^{(n,l)} \right) \quad (\text{A.4})$$

Assume the steady state probabilities for all variables exist, we have $\lim_{l \rightarrow \infty} q_k^{(n,l)} = q_k^{(n)}$, $\lim_{l \rightarrow \infty} s_k^{(n,l)} = s_k^{(n)}$, $\lim_{l \rightarrow \infty} y_k^{(n,l)} = y_k^{(n)}$. Taking the limit of l for both sides of Eq. A.4 leads to

$$q_k^{(n)} = \sum_{u=0}^C s_u^{(n)} \sum_{i=0}^u q_i^{(n)} y_k^{(n)} + \sum_{u=0}^C s_u^{(n)} \sum_{i=u+1}^{u+k} q_i^{(n)} y_{k-i+u}^{(n)} \quad (\text{A.5})$$

Assume the probability generating function (PGF) for $Q^{(n)}$, $R^{(n)}$ and $Y^{(n)}$ are $Q(z)$, $R(z)$, and $Y(z)$, respectively, where $Q^{(n)}$, $R^{(n)}$ and $Y^{(n)}$ and the steady state random variables of $Q^{(n,l)}$, $R^{(n,l)}$ and $Y^{(n,l)}$. Hence,

$$Q(z) = \sum_{k=0}^{\infty} q_k^{(n)} z^k \quad (\text{A.6})$$

$$R(z) = \sum_{k=0}^{\infty} r_k^{(n)} z^k = \sum_{k=0}^{\infty} z^k \sum_{u=0}^C s_u^{(n)} \cdot r_k^{(n)} \Big|_{S^{(n)}=u} \quad (\text{A.7})$$

$$Y(z) = \sum_{k=0}^{\infty} y_k^{(n)} z^k \quad (\text{A.8})$$

Substituting Eq. A.2 and A.3 into A.7 results in

$$R(z) = \sum_{u=0}^C s_u^{(n)} \sum_{i=0}^u q_i^{(n)} + \sum_{k=1}^{\infty} z^k \sum_{u=0}^C s_u^{(n)} q_{k+u}^{(n)} \quad (\text{A.9})$$

$$= \sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} + \frac{1}{z^u} Q(z) - \frac{1}{z^u} \sum_{i=0}^u q_i^{(n)} z^i \right] \quad (\text{A.10})$$

Notice that $Q^{(n)} = R^{(n)} + Y^{(n)}$ (this is obtained by taking the limit of l for Eq. 3). Since $R^{(n)}$ and $Y^{(n)}$ are independent, we have

$$Q(z) = R(z)Y(z) \quad (\text{A.11})$$

Combining Eq. A.10 and A.11 obtains

$$\begin{aligned}
Q(z) &= \frac{Y(z) \sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} \left(1 - \frac{z^i}{z^u}\right) \right]}{1 - \sum_{u=0}^C s_u^{(n)} \frac{Y(z)}{z^u}} \\
&= \frac{\sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} (z^C - z^{C-u+i}) \right]}{\frac{z^C}{Y(z)} - \sum_{u=0}^C s_u^{(n)} z^{C-u}}
\end{aligned} \tag{A.12}$$

Appendix B. Derivation of $q_{0:C-1}^{(n)}$ by matching polynomial coefficients

Though $q_0^{(n)}, \dots, q_{C-1}^{(n)}$ can be obtained by solving C system equations as mentioned in Section 4.3, we attempt to provide a more direct way to calculate $q_{0:C-1}^{(n)}$ in this section.

Using the fact that the numerator of $Q(z)$ is in the polynomial order of C , $Q(z)$ can be reformulated in terms of z_1^*, \dots, z_{C-1}^* as:

$$Q(z) = \frac{(z-1) \prod_{i=1}^{C-1} (z - z_i^*) \sum_{u=0}^C s_u^{(n)} \sum_{i=0}^u q_i^{(n)}}{\frac{z^C}{Y(z)} - \sum_{u=0}^C s_u^{(n)} z^{C-u}} \tag{B.1}$$

When $z \rightarrow 1$, both $\text{NUM}(z)$ and $\text{DEN}(z)$ approach 0. We also have the fact that $\lim_{z \rightarrow 1} Q(z) = 1$. Therefore, using L'Hopital's rule we have

$$\lim_{z \rightarrow 1} Q(z) = 1 = \lim_{z \rightarrow 1} \frac{\text{NUM}'(z)}{\text{DEN}'(z)} = \frac{\prod_{i=1}^{C-1} (1 - z_i^*) \sum_{u=0}^C s_u^{(n)} \sum_{i=0}^u q_i^{(n)}}{\sum_{u=0}^C s_u^{(n)} u - Y'(1)} \tag{B.2}$$

$$\Rightarrow \sum_{u=0}^C s_u^{(n)} \sum_{i=0}^u q_i^{(n)} = \frac{\sum_{u=0}^C s_u^{(n)} u - Y'(1)}{\prod_{i=1}^{C-1} (1 - z_i^*)} \tag{B.3}$$

Define $\bar{Y}^{(n)} := Y'(1) = \mathbb{E}[Y^{(n)}]$ as the mean number of arrival passengers within a headway at station n , $\bar{S}^{(n)} := \sum_{u=0}^C s_u^{(n)} u$ as the mean number of available spaces in an arriving bus at station n . Substituting Eq. B.3 into B.1, $Q(z)$ can be rewritten as

$$Q(z) = \frac{(\bar{S}^{(n)} - \bar{Y}^{(n)})(z-1) \prod_{i=1}^{C-1} \frac{z - z_i^*}{1 - z_i^*}}{\frac{z^C}{Y(z)} - \sum_{u=0}^C s_u^{(n)} z^{C-u}} \tag{B.4}$$

Comparing Eq. B.4 and 18, let the numerators of two equations be equal, we have

$$(\bar{S}^{(n)} - \bar{Y}^{(n)})(z-1) \prod_{i=1}^{C-1} \frac{z - z_i^*}{1 - z_i^*} = \sum_{u=0}^C s_u^{(n)} \left[\sum_{i=0}^u q_i^{(n)} (z^C - z^{C-u+i}) \right] \tag{B.5}$$

As the LHS and RHS of Eq. B.5 are both polynomials about z , the coefficients of each polynomial in z must

be equal. By matching the coefficients of z^0 , we have

$$(-1)(\bar{S}^{(n)} - \bar{Y}^{(n)}) \prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} = -q_0^{(n)} s_C^{(n)} \quad (\text{B.6})$$

which leads to

$$q_0^{(n)} = \frac{1}{s_C^{(n)}} (\bar{S}^{(n)} - \bar{Y}^{(n)}) \prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} \quad (\text{B.7})$$

Note that $\prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1}$ is always greater than 0 because 1) when C is odd, as the complex roots appear as conjugates, $\prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} > 0$. 2) when C is even, besides $z_0^* = 1$, there exists another real root on the negative real axis (denoted as $z_{\frac{C}{2}}^*$, where $-1 \leq z_{\frac{C}{2}}^* < 0$). So, we have $\frac{z_{\frac{C}{2}}^*}{z_{\frac{C}{2}}^* - 1} > 0$, which leads to $\prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} > 0$.

To validate Eq. B.5, consider the fixed capacity situation where $s_C^{(n)} = 1$ and $\bar{S}^{(n)} = C$. Then Eq. B.5 reduces to

$$q_0^{(n)} \Big|_{s_C^{(n)}=1} = (C - \bar{Y}^{(n)}) \prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} \quad (\text{B.8})$$

This is the same as Chaudhry et al. (1987). Now we will derive $q_{1:C-1}^{(n)}$. Observing that the numerator of Eq. B.4 can be rewritten as

$$\begin{aligned} (\bar{S}^{(n)} - \bar{Y}^{(n)})(z - 1) \prod_{i=1}^{C-1} \frac{z - z_i^*}{1 - z_i^*} &= \frac{1}{s_C^{(n)}} (\bar{S}^{(n)} - \bar{Y}^{(n)}) \prod_{i=1}^{C-1} \frac{z_i^*}{z_i^* - 1} \prod_{i=1}^{C-1} \frac{z_i^* - z}{z_i^*} (z - 1) s_C^{(n)} \\ &= s_C^{(n)} q_0^{(n)} (z - 1) \prod_{i=1}^{C-1} \left(1 - \frac{z}{z_i^*}\right) \\ &= -s_C^{(n)} q_0^{(n)} \prod_{i=0}^{C-1} \left(1 - \frac{z}{z_i^*}\right) \end{aligned} \quad (\text{B.9})$$

Define $\prod_{i=0}^{C-1} \left(1 - \frac{z}{z_i^*}\right) := \sum_{j=0}^C \eta_j z^j$, where η_j is the polynomial coefficient of z^j . For the RHS of Eq. B.5, the polynomial coefficient of z^{C-k} is $-\sum_{u=k}^C s_u^{(n)} q_{u-k}^{(n)}$. And from Eq. B.9, the polynomial coefficient of z^{C-k} is $-s_C^{(n)} q_0^{(n)} \eta_{C-k}$. Matching the coefficient of the same order of z leads to

$$s_C^{(n)} q_0^{(n)} \eta_{C-k} = \sum_{u=k}^C s_u^{(n)} q_{u-k}^{(n)} \quad k = 1, 2, \dots, C - 1 \quad (\text{B.10})$$

To validate Eq. B.10, consider the fixed capacity situation where $s_C^{(n)} = 1$ and $s_k^{(n)} = 0, \forall 0 \leq k < C$. then Eq. B.10 reduces to

$$q_{C-k}^{(n)} = q_0^{(n)} \eta_{C-k} \quad k = 1, 2, \dots, C-1 \quad \text{if } s_C^{(n)} = 1 \quad (\text{B.11})$$

which is the same as [Chaudhry et al. \(1987\)](#).

Eq. [B.10](#) can be expressed in a matrix form by adding $s_C^{(n)} q_0^{(n)} \eta_0^{(n)} = s_C^{(n)} q_0^{(n)}$ (note that $\eta_0^{(n)} = 1$ by definition):

$$\tilde{\eta}^{(n)} = q_{0:C-1}^{(n)} \Lambda^{(n)} \quad (\text{B.12})$$

where $\tilde{\eta}^{(n)} = [s_C^{(n)} q_0^{(n)} \eta_0^{(n)}, s_C^{(n)} q_0^{(n)} \eta_1^{(n)}, \dots, s_C^{(n)} q_0^{(n)} \eta_{C-1}^{(n)}] \in \mathbb{R}^C$ and

$$\Lambda^{(n)} = \begin{bmatrix} s_C^{(n)} & s_{C-1}^{(n)} & s_{C-2}^{(n)} & \dots & s_1^{(n)} \\ 0 & s_C^{(n)} & s_{C-1}^{(n)} & \dots & s_2^{(n)} \\ \dots & 0 & s_C^{(n)} & \dots & s_3^{(n)} \\ 0 & \dots & 0 & \dots & s_4^{(n)} \\ 0 & 0 & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & s_C^{(n)} \end{bmatrix} \in \mathbb{R}^{C \times C} \quad (\text{B.13})$$

As $s_C^{(n)} > 0$ is a known condition, the triangular matrix $\Lambda^{(n)}$ is invertible. Thus, we have

$$q_{0:C-1}^{(n)} = \tilde{\eta}^{(n)} (\Lambda^{(n)})^{-1} \quad (\text{B.14})$$

Appendix C. Derivation of queue length mean and variance

Here we try to provide analytical formulations of $\mathbb{E}[Q^{(n)}]$ and $\text{Var}[Q^{(n)}]$. The key is to find $Q'(1)$ and $Q''(1)$. The derivation follows the similar idea in [Powell \(1981\)](#).

Let $A(z) = \frac{(\bar{S}^{(n)} - \bar{Y}^{(n)})(z-1)}{\bar{Y}^{(n)}(z) - \sum_{u=0}^C s_u^{(n)} z^{C-u}}$ and $B_i(z) = \frac{z - z_i^*}{1 - z_i^*}$, then $Q(z) = A(z) \prod_{i=1}^{C-1} B_i(z)$. Based on the fact that $B_i(1) = 1$ and $Q(z) = 1$, we must have $A(1) = 1$. Hence,

$$\begin{aligned} Q'(1) &= A'(1)B_1(1)\dots B_{C-1}(1) + A(1)B_1'(1)\dots B_{C-1}(1) + \dots + A(1)B_1(1)\dots B_{C-1}'(1) \\ &= A'(1) + \sum_{i=1}^{C-1} B_i'(1) \end{aligned} \quad (\text{C.1})$$

Since $B_i'(1) = \frac{1}{1-z_i^*}$, the problem now becomes finding $A'(1)$. Again, let $A(z) = \frac{A_1(z)}{A_2(z)}$. Then,

$$A'(z) = \frac{A_1'(z)A_2(z) - A_1(z)A_2'(z)}{(A_2(z))^2} \quad (\text{C.2})$$

Notice that when $z \rightarrow 1$, the numerator and denominator of $A'(z)$ approach 0 (because $A_1(1) = 0$ and

$A_2(1) = 0$). Therefore, applying L'Hopital's rule yields:

$$A'(z) = \frac{A_1''(z)A_2(z) - A_1(z)A_2''(z)}{2A_2(z)A_2'(z)} \quad (\text{C.3})$$

Again we have 0/0 when $z \rightarrow 1$ because $A_1''(z) = 0$ and $A_2(1) = 0$. Applying L'Hopital's rule once more gives:

$$A'(z) = \frac{-A_1'(z)A_2''(z) - A_1(z)A_2'''(z)}{2A_2'(z)A_2'(z) + 2A_2(z)A_2''(z)} \quad (\text{C.4})$$

Substituting $z = 1$ leads to

$$A'(1) = \frac{-A_1'(1)A_2''(1)}{2(A_2'(1))^2} \quad (\text{C.5})$$

Based on the fact that $Y(1) = 1$, $Y'(1) = \bar{Y}^{(n)}$, $Y''(1) = \mathbb{E}[(Y^{(n)})^2] - \bar{Y}^{(n)}$, we have

$$A_1'(1) = \bar{S}^{(n)} - \bar{Y}^{(n)} \quad (\text{C.6})$$

$$A_2'(1) = \frac{Cz^{C-1}Y(z) - Y'(z)z^C}{Y(z)^2} - \sum_{u=0}^C (C-u)s_u^{(n)}z^{C-u-1} \Big|_{z=1} = \bar{S}^{(n)} - \bar{Y}^{(n)} \quad (\text{C.7})$$

$$\begin{aligned} A_2''(1) &= \frac{C(C-1)z^{C-2}}{Y(z)} - \frac{2Y'(z)Cz^{C-1}}{Y(z)^2} - \frac{Y''(z)z^C}{Y(z)^2} + \frac{2Y'(z)^2z^C}{Y(z)^3} - \sum_{u=0}^C (C-u)(C-u-1)s_u^{(n)}z^{C-u-2} \Big|_{z=1} \\ &= C(C-1) - 2\bar{Y}^{(n)}C - \mathbb{E}[(Y^{(n)})^2] + 2(\bar{Y}^{(n)})^2 + \bar{Y}^{(n)} - C^2 + C + 2C\bar{S}^{(n)} - \bar{S}^{(n)} - \mathbb{E}[(S^{(n)})^2] \\ &= -2\bar{Y}^{(n)}C - \mathbb{E}[(Y^{(n)})^2] + 2(\bar{Y}^{(n)})^2 + \bar{Y}^{(n)} + 2C\bar{S}^{(n)} - \bar{S}^{(n)} - \mathbb{E}[(S^{(n)})^2] \end{aligned} \quad (\text{C.8})$$

Substituting Eq. C.6, C.7, and C.8 into Eq. C.5 results in

$$A'(1) = \frac{2\bar{Y}^{(n)}C + \mathbb{E}[(Y^{(n)})^2] - 2(\bar{Y}^{(n)})^2 - \bar{Y}^{(n)} - 2C\bar{S}^{(n)} + \bar{S}^{(n)} + \mathbb{E}[(S^{(n)})^2]}{2(\bar{S}^{(n)} - \bar{Y}^{(n)})} \quad (\text{C.9})$$

Therefore, we have

$$\mathbb{E}[Q^{(n)}] = \frac{2\bar{Y}^{(n)}C + \mathbb{E}[(Y^{(n)})^2] - 2(\bar{Y}^{(n)})^2 - \bar{Y}^{(n)} - 2C\bar{S}^{(n)} + \bar{S}^{(n)} + \mathbb{E}[(S^{(n)})^2]}{2(\bar{S}^{(n)} - \bar{Y}^{(n)})} + \sum_{i=1}^{C-1} \frac{1}{1 - z_i^*} \quad (\text{C.10})$$

To validate this formulation, let us consider a fixed capacity situation with $s_C^{(n)} = 1$. Then $\bar{S}^{(n)} = C$, $\mathbb{E}[(S^{(n)})^2] = C^2$. Then Eq. C.10 reduces to

$$\mathbb{E}[Q^{(n)}] \Big|_{s_C^{(n)}=1} = \frac{C - C^2 + 2\bar{Y}^{(n)}C + \mathbb{E}[(Y^{(n)})^2] - 2(\bar{Y}^{(n)})^2 - \bar{Y}^{(n)}}{2(C - \bar{Y}^{(n)})} + \sum_{i=1}^{C-1} \frac{1}{1 - z_i^*} \quad (\text{C.11})$$

which is equivalent to [Powell \(1981\)](#)'s.

According to [Eq. 24](#), the key to obtain $\text{Var}[Q^{(n)}]$ is to calculate $Q''(1)$. Taking the logarithm of $Q(z) = A(z) \prod_{i=1}^{C-1} B_i(z)$ gives

$$\log Q(z) = \log A(z) + \sum_{i=1}^{C-1} \log B_i(z) \quad (\text{C.12})$$

Taking derivatives of both sides leads to

$$\frac{Q'(z)}{Q(z)} = \frac{A'(z)}{A(z)} + \sum_{i=1}^{C-1} \frac{B'_i(z)}{B_i(z)} \quad (\text{C.13})$$

Taking derivatives again:

$$\frac{Q''(z)}{Q(z)} - \frac{Q'(z)^2}{Q(z)^2} = \frac{A''(z)}{A(z)} - \frac{A'(z)^2}{A(z)^2} + \sum_{i=1}^{C-1} \left(\frac{B''_i(z)}{B_i(z)} - \frac{B'_i(z)^2}{B_i(z)^2} \right) \quad (\text{C.14})$$

Solving for $Q''(z)$ and letting $z = 1$ gives:

$$Q''(1) = \mathbb{E}[Q^{(n)}]^2 + A''(1) - A'(1)^2 + \sum_{i=1}^{C-1} (B''_i(1) - B'_i(1)^2) \quad (\text{C.15})$$

Notice that $B''_i(1) = 0$ ($\forall i = 1, \dots, C-1$) and $\mathbb{E}[Q^{(n)}] = Q'(1)$. Substituting [Eq. C.1](#) and [C.15](#) into [Eq. 24](#) gives

$$\text{Var}[Q^{(n)}] = A''(1) - A'(1)^2 + A'(1) + \sum_{i=1}^{C-1} (B'_i(1) - B'_i(1)^2) \quad (\text{C.16})$$

Now we only need to solve for $A''(1)$. The process is similar to finding $A'(1)$. Applying L'Hopital's rule five times to [Eq. C.4](#) and substituting $z = 1$ leads to

$$A''(1) = \frac{-2A'_2(1)A''_2(1) + 3A''_2(1)^2}{6A'_2(1)} \quad (\text{C.17})$$

Notice that the derivation process uses $A''_1(z) = 0$, $A_1(1) = 0$, $A_2(1) = 0$, and $A'_1(1) = A'_2(1)$. Details are

omitted due to the tedious mathematical manipulation. To obtain $A_2'''(1)$, taking derivative of Eq. C.8 gives:

$$\begin{aligned}
A_2'''(1) &= \left[\frac{C(C-1)(C-2)z^{C-3}}{Y(z)} - \frac{3Y'(z)C(C-1)z^{C-2}}{Y(z)^2} - \frac{3Y''(z)Cz^{C-1}}{Y(z)^2} + \frac{4Y'(z)^2Cz^{C-1}Y(z)}{Y(z)^4} \right. \\
&\quad - \frac{Y'''(z)z^C}{Y(z)^2} + \frac{2Y(z)Y'(z)Y''(z)z^C}{Y(z)^4} + \frac{4Y''(z)Y'(z)z^C + 2Cz^{C-1}Y'(z)^2}{Y(z)^3} - \frac{6Y(z)Y'(z)^3z^C}{Y(z)^6} \\
&\quad \left. - \sum_{u=0}^C (C-u)(C-u-1)(C-u-2)s_u^{(n)}z^{C-u-3} \right]_{z=1} \\
&= C(C-1)(C-2) - 3\bar{Y}^{(n)}C(C-1) - 3Y''(1)C + 6(\bar{Y}^{(n)})^2C - Y'''(1) + 6\bar{Y}^{(n)}Y''(1) \\
&\quad - 6(\bar{Y}^{(n)})^3 - (C^3 - 3C^2 + 2C) + (2 + 3C^2 - 6C)\bar{S}^{(n)} + (3 - 3C)\mathbb{E}[(S^{(n)})^2] + \mathbb{E}[(S^{(n)})^3]
\end{aligned} \tag{C.18}$$

Notice that $Y'''(1) = \mathbb{E}[(Y^{(n)})^3] - 3\mathbb{E}[(Y^{(n)})^2] + 2\bar{Y}^{(n)}$. Hence,

$$\begin{aligned}
A_2'''(1) &= 3C^2\bar{S}^{(n)} - 3C^2\bar{Y}^{(n)} - 6C\bar{S}^{(n)} - 3C\mathbb{E}[(S^{(n)})^2] - 3C\mathbb{E}[(Y^{(n)})^2] + 6C(\bar{Y}^{(n)})^2 + 6C\bar{Y}^{(n)} + 2\bar{S}^{(n)} \\
&\quad + 3\mathbb{E}[(S^{(n)})^2] + \mathbb{E}[(S^{(n)})^3] + 6\mathbb{E}[(Y^{(n)})^2]\bar{Y}^{(n)} + 3\mathbb{E}[(Y^{(n)})^2] - \mathbb{E}[(Y^{(n)})^3] - 6(\bar{Y}^{(n)})^3 - 6(\bar{Y}^{(n)})^2 - 2\bar{Y}^{(n)}
\end{aligned} \tag{C.19}$$

Substituting Eq. C.7, C.8, and C.19 into Eq. C.17 results in

$$\begin{aligned}
A''(1) &= \left[6C^2(\bar{S}^{(n)})^2 - 12C^2(\bar{S}^{(n)})(\bar{Y}^{(n)}) + 6C^2(\bar{Y}^{(n)})^2 - 6C(\bar{S}^{(n)})\mathbb{E}[(S^{(n)})^2] - 6C(\bar{S}^{(n)})\mathbb{E}[(Y^{(n)})^2](\bar{Y}^{(n)})^2 \right. \\
&\quad + 12C(\bar{S}^{(n)}) + 6C\mathbb{E}[(S^{(n)})^2](\bar{Y}^{(n)}) + 6C\mathbb{E}[(Y^{(n)})^2](\bar{Y}^{(n)}) - 12C(\bar{Y}^{(n)})^3 - (\bar{S}^{(n)})^2 - 2(\bar{S}^{(n)})\mathbb{E}[(S^{(n)})^3] \\
&\quad - 12(\bar{S}^{(n)})\mathbb{E}[(Y^{(n)})^2](\bar{Y}^{(n)}) + 2(\bar{S}^{(n)})\mathbb{E}[(Y^{(n)})^3] + 12(\bar{S}^{(n)})(\bar{Y}^{(n)})^3 + 2(\bar{S}^{(n)})(\bar{Y}^{(n)}) \\
&\quad + 3\mathbb{E}[(S^{(n)})^2]^2 + 6\mathbb{E}[(S^{(n)})^2]\mathbb{E}[(Y^{(n)})^2] - 12\mathbb{E}[(S^{(n)})^2](\bar{Y}^{(n)})^2 + 2\mathbb{E}[(S^{(n)})^3](\bar{Y}^{(n)}) + 3\mathbb{E}[(Y^{(n)})^2]^2 \\
&\quad \left. - 2\mathbb{E}[(Y^{(n)})^3](\bar{Y}^{(n)}) - (\bar{Y}^{(n)})^2 \right] / 6(\bar{S}^{(n)} - \bar{Y}^{(n)})
\end{aligned} \tag{C.20}$$

Now with Eq. C.20 and C.9 we have

$$\begin{aligned}
A''(1) - A'(1)^2 + A'(1) &= \left[(\bar{S}^{(n)})^2 - 4\bar{S}^{(n)}\mathbb{E}[(S^{(n)})^3] - 24\bar{S}^{(n)}\mathbb{E}[(Y^{(n)})^2]\bar{Y}^{(n)} + 4\bar{S}^{(n)}\mathbb{E}[(Y^{(n)})^3] \right. \\
&\quad + 24\bar{S}^{(n)}(\bar{Y}^{(n)})^3 - 2\bar{S}^{(n)}\bar{Y}^{(n)} + 3\mathbb{E}[(S^{(n)})^2]^2 + 6\mathbb{E}[(S^{(n)})^2]\mathbb{E}[(Y^{(n)})^2] - 12\mathbb{E}[(S^{(n)})^2](\bar{Y}^{(n)})^2 + 4\mathbb{E}[(S^{(n)})^3]\bar{Y}^{(n)} \\
&\quad \left. + 3\mathbb{E}[(Y^{(n)})^2]^2 + 12\mathbb{E}[(Y^{(n)})^2](\bar{Y}^{(n)})^2 - 4\mathbb{E}[(Y^{(n)})^3]\bar{Y}^{(n)} - 12(\bar{Y}^{(n)})^4 + (\bar{Y}^{(n)})^2 \right] / 12(\bar{S}^{(n)} - \bar{Y}^{(n)})^2
\end{aligned} \tag{C.21}$$

Slight manipulation of Eq. C.21 leads to

$$\begin{aligned}
& A''(1) - A'(1)^2 + A'(1) \\
&= \frac{-4(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)})(\bar{S}^{(n)} - \bar{Y}^{(n)}) + 3(\bar{\bar{S}}^{(n)} + \bar{\bar{Y}}^{(n)})^2 - [6(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)}) - 1](\bar{S}^{(n)} - \bar{Y}^{(n)})^2 - (\bar{S}^{(n)} - \bar{Y}^{(n)})^4}{12(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)})^2}
\end{aligned} \tag{C.22}$$

Observe that $B'_i(1) - B'_i(1)^2 = \frac{-z_i^*}{(1-z_i^*)^2}$. Therefore, substituting Eq. C.21 into C.16 gives the final results:

$$\begin{aligned}
\text{Var}[Q^{(n)}] &= \frac{-4(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)})(\bar{S}^{(n)} - \bar{Y}^{(n)}) + 3(\bar{\bar{S}}^{(n)} + \bar{\bar{Y}}^{(n)})^2 - [6(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)}) - 1](\bar{S}^{(n)} - \bar{Y}^{(n)})^2 - (\bar{S}^{(n)} - \bar{Y}^{(n)})^4}{12(\bar{\bar{S}}^{(n)} - \bar{\bar{Y}}^{(n)})^2} \\
&\quad - \sum_{i=1}^{C-1} \frac{z_i^*}{(1-z_i^*)^2}
\end{aligned} \tag{C.23}$$