



Bounds and limit theorems for a layered queueing model in electric vehicle charging

Angelos Aveklouris¹ · Maria Vlasiou¹ · Bert Zwart^{1,2}

Received: 10 October 2018 / Revised: 13 May 2019
© The Author(s) 2019

Abstract

The rise of electric vehicles (EVs) is unstoppable due to factors such as the decreasing cost of batteries and various policy decisions. These vehicles need to be charged and will therefore cause congestion in local distribution grids in the future. Motivated by this, we consider a charging station with finitely many parking spaces, in which electric vehicles arrive in order to get charged. An EV has a random parking time and a random charging time. Both the charging rate per vehicle and the charging rate possible for the station are assumed to be limited. Thus, the charging rate of uncharged EVs depends on the number of cars charging simultaneously. This model leads to a layered queueing network in which parking spaces with EV chargers have a dual role, of a server (to cars) and a customer (to the grid). We are interested in the performance of the aforementioned model, focusing on the fraction of vehicles that get fully charged. To do so, we develop several bounds and asymptotic (fluid and diffusion) approximations for the vector process which describes the total number of EVs and the number of not fully charged EVs in the charging station, and we compare these bounds and approximations with numerical outcomes.

Keywords Electric vehicle charging · Layered queueing networks · Fluid approximation · Diffusion approximation

Mathematics Subject Classification 60K25 · 90B15 · 68M20

✉ Angelos Aveklouris
a.aveklouris@tue.nl

Maria Vlasiou
m.vlasiou@tue.nl

Bert Zwart
Bert.Zwart@cwi.nl

¹ Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

² Centrum Wiskunde en Informatica, 1090 GB Amsterdam, The Netherlands

1 Introduction

The rise of electric vehicles (EVs) is unstoppable due to factors such as the decreasing cost of batteries and various policy decisions [25]. Currently, the bottlenecks are the ability to charge a battery at a fast rate and the number of charging stations, but this bottleneck is expected to move toward the current grid infrastructure. This is illustrated in [24]; the authors evaluate the impact of the energy transition on a real distribution grid in a field study, based on a scenario for the year 2025. The authors confront a local low-voltage grid with electrical vehicles and ovens and show that charging a small number of EVs is enough to burn a fuse. Additional evidence of congestion is reported in [12]. This paper proposes to model and analyze such congestion by the use of the so-called *layered queueing networks*. Layered networks are specific queueing networks where some entities in the system have a dual role; for example, servers (in our context: parking spaces with EV chargers) become customers to a higher layer (here: the power grid). The use of layered queueing networks allows us to analyze the interaction of two sources of congestion: first, the number of available spaces with charging stations (as not all cars find a space), and second, the amount of available power that the power grid is able to feed to the charging station [24].

We consider a charging station (or parking lot) with finitely many parking spaces. Each space has an EV charger connecting with the power grid. EVs arrive at the charging station randomly in order to get charged. If an EV finds an available space, it enters the parking lot and charging starts immediately. An EV has a random parking time and a random charging time. It leaves the parking lot only when its parking time expires; i.e., it remains at its space without consuming power until its parking time expires if finishing its charge within the parking time. Both the charging rate per vehicle and the charging rate possible for the complete charging station are assumed to be limited. Thus, the charging rate of uncharged EVs depends on the number of cars charging simultaneously. Finally, we assume that all available power is shared at the same rate to all cars that need charging. The available power that can be delivered by the grid is assumed to be constant.

Using queueing terminology, our model can be described as a two-layered queueing network. An EV enters the charging station and connects its battery to an EV charger. In our context, EVs play the role of customers, while EV chargers are the servers. Thus, the system of EVs and EV chargers can be viewed as the first layer. Moreover, EV chargers are connected to the power grid. Thus, at the second layer, active EV chargers act as jobs that are served simultaneously by the power grid, which plays the role of a single server.

This paper focuses on the performance analysis of this system under Markovian assumptions. Specifically, we are interested in finding the fraction of fully charged EVs in the charging station, which is equivalent to the probability that an EV leaves the charging station with a fully charged battery. A mostly heuristic description of some partial results in this paper has appeared in [2]. We first start with the steady-state analysis of the original system, for which we can find explicit bounds for the fraction of fully charged EVs. To do so, we study three special cases of the original system: (i) There is enough power for all EVs, (ii) there are enough parking spaces for all EVs, and (iii) the parking lot is full. In these cases, we are able to find the explicit

joint distribution in steady state of the total number of EVs and the number of not fully charged EVs in the charging station, which we call *the vector process*.

In order to improve the bounds for the fraction of fully charged EVs, we next develop a fluid approximation for the number of uncharged EVs in the parking lot. The mathematical results here are closely related to results on processor-sharing queues with impatience [22]. However, the model here is more complicated as there is a limited number of spaces in the system and fully charged cars may not leave immediately as they are still parked.

We then move to diffusion approximations, working in three asymptotic regimes. First, we consider the Halfin–Whitt regime, in which we prove a limit theorem for the vector process, showing that it converges to a two-dimensional reflected Ornstein–Uhlenbeck (OU) process with piecewise linear drift. Then, we consider an overloaded regime for the process describing the number of total EVs in the system. In this case, the limit reduces to a one-dimensional OU process with piecewise linear drift. Finally, we approximate the vector process by a two-dimensional OU process when the parking times are sufficiently large. The mathematical results here are based on martingale arguments [32].

EVs can be charged in several ways. Our setup can be seen as an example of slow charging, in which drivers typically park their EV and are not physically present during charging (but are busy shopping, working, sleeping, etc). For queueing models focusing on fast charging, we refer to [5,48]. Both papers consider a gradient scheduler to control delays. Next, [47] presents a queueing model for battery swapping while [40] is an early paper on a queueing analysis of EV charging, focusing on designing safety control rules (in terms of voltage drops) with minimal communication overhead.

Despite being a relatively new topic, the engineering literature on EV charging is huge. We can only provide a small sample of the already vast but still emerging literature on EV charging. The focus of [39] is on a specific parking lot and presents an algorithm for optimally managing a large number of plug-in EVs. Algorithms to minimize the impact of plug-in EV charging on the distribution grid are proposed in [38]. In [31], the overall charging demand of plug-in EVs is considered. Mathematical models where vehicles communicate beforehand with the grid to convey information about their charging status are studied in [37]. In [30], EVs are the central object and a dynamic program is formulated that prescribes how EVs should charge their battery using price signals.

In addition, layered queueing networks have been successfully applied in analyzing interactive networks in communication networks and manufacturing systems. These are queueing networks where some entities in the system have a dual role. In such systems, the dynamics in layers are correlated and the service speeds vary over time. Layered queueing networks can be characterized by separate layers (see [36,45]) or simultaneous layers (such as our model) [3]. In the first case, customers receive service with some delay. An application where layered networks with separate layers appear is manufacturing systems, for example, [17,18]. On the other hand, in layered networks with simultaneous layers, customers receive service from the different layers simultaneously. Layered networks with simultaneous layers have applications in communication networks, for example, Web-based multitiered system architectures. In such environments, different applications compete for access to shared infrastruc-

ture resources, both at the software level and at the hardware level. For background, see [42,43].

The paper is organized as follows: In Sect. 2, we provide a detailed model description—in particular, we introduce our stochastic model and we define the system dynamics. Next, in Sect. 3, we present some explicit bounds in steady state for the fraction of fully charged EVs. Section 4 contains several asymptotic approximations. First, a fluid approximation is presented; we then derive diffusion limits and approximations in three asymptotic regimes. Numerical validations are presented in Sect. 5. Finally, all proofs are gathered in Sect. 6.

2 Model

In this section, we provide a detailed formulation of our model and explain various notational conventions that are used in the remainder of this work.

2.1 Preliminaries

We use the following notational conventions: All vectors and matrices are denoted by bold letters. Further, \mathbb{R} is the set of real numbers, \mathbb{R}_+ is the set of nonnegative real numbers and \mathbb{N} is the set of strictly positive integers. For real numbers x and y , we define $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$. Furthermore, \mathbf{I} represents the identity matrix and \mathbf{e} and \mathbf{e}_0 are vectors consisting of 1's and 0's, respectively, the dimensions of which are clear from the context. Also, \mathbf{e}_i is the vector whose i^{th} element is 1 and the rest are all 0.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For $T > 0$, let $\mathcal{D}[0, T]^2 := \mathcal{D}[0, T] \times \mathcal{D}[0, T]$ be the two-dimensional Skorokhod space, i.e., the space of two-dimensional real-valued functions on $[0, T]$ that are right continuous with left limits endowed with the J_1 topology; cf. [11]. Observe that as all candidate limit objects we consider are continuous, we only need to work with the uniform topology. It is well known that the space $(\mathcal{D}[0, T]^2, J_1)$ is a complete and separate metric space (i.e., a Polish metric space) [7]. We denote by $\mathcal{B}(\mathcal{D}[0, T]^2)$ the Borel σ -algebra of $\mathcal{D}[0, T]^2$. We assume that all the processes are defined from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{B}(\mathcal{D}[0, T]^2), \mathcal{D}[0, T]^2)$. Further, we write $X(\cdot) := \{X(t), t \geq 0\}$ to represent a stochastic process and $X(\infty)$ to represent a stochastic process in steady state. Moreover, $\stackrel{d}{=}$ and $\stackrel{d}{\rightarrow}$ denote equality and convergence in distribution (weak convergence). For two random variables X, Y , we write $X \leq_{st} Y$ (stochastic ordering) if $\mathbb{P}(X > a) \leq \mathbb{P}(Y > a)$ for any $a \in \mathbb{R}$. Further, $\Phi(\cdot)$ and $\phi(\cdot)$ represent the cumulative probability function and the probability density function (pdf) of the standard normal distribution, respectively. Finally, let $C_b^2(G)$ denote the space of twice continuously differentiable functions on G such that their first- and second-order derivatives are bounded.

2.2 Model description

We consider a charging station with $K > 0$ parking spaces. Each space has an EV charger which is connected to the power grid. EVs arrive independently at the charging station according to a Poisson process with rate λ . They have a random charging requirement and a random parking time denoted by B and D , respectively. The random variables B and D are assumed to be mutually independent and exponentially distributed with rates μ and ν , respectively. If an EV finishes its charging, it remains at its space without consuming power until its parking time expires. We call these EVs *fully charged EVs*. Thus, EVs leave the system only after their parking time expires, which implies that an EV may leave the system without its battery being fully charged. Furthermore, if all spaces are occupied, a newly arriving EV does not enter the system but leaves immediately. As such, the total number of vehicles in the system can be modeled by an Erlang loss system, though we need a more detailed description of the state space.

We denote by $Q(t) \in \{0, 1, \dots, K\}$ the total number of EVs (charged and uncharged) in the system at time $t \geq 0$, where $Q(0)$ is the initial number of EVs. Further, we denote by $Z(t) \in \{0, 1, \dots, Q(t)\}$ the number of EVs without a fully charged battery at time t and by $Z(0)$ the number of such vehicles initially in the system. Thus, $C(t) = Q(t) - Z(t)$ represents the number of EVs with a fully charged battery at time t .

The power consumed by the parking lot is limited and depends on the number of uncharged EVs at time t . We let it be given by the *power allocation function* $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$,

$$L(Z(t)) := Z(t) \wedge M.$$

We assume that the parameter M is given and that $0 < M \leq K$. For example, the parameter M can depend on the contract between the power grid and the charging station. Alternatively, M can be thought of as the maximum number of EVs the charging station can charge at a maximum rate, where without loss of generality we can assume that the maximum rate is one. The model is illustrated in Fig. 1.

Finally, note that the processes $Q(\cdot)$, $Z(\cdot)$, and $C(\cdot)$ depend on K and M . We write $Q_M^K(\cdot)$, $Z_M^K(\cdot)$, and $C_M^K(\cdot)$, when we wish to emphasize this. It is clear from our context that the two-dimensional process $\{(Q(t), Z(t)) : t \geq 0\}$ is Markov. The transition rates in the interior and on the boundary are shown in Fig. 2.

2.2.1 Alternative model description in the case of infinitely many parking spaces

Here, we give an alternative description of our model in the case where there are infinitely many parking spaces; i.e., $K = \infty$. In this case, the model can be described as a tandem queue with impatient customers; see Fig. 3. EVs arrive at the charging station, which has M servers, and charging starts immediately. There are two possible scenarios. First, an EV gets fully charged during D and moves to the second queue, which has an infinite number of servers. This happens with rate $\mu(Z(t) \wedge M)$. In the second queue, EVs get served with rate $\nu C(t)$. In the second scenario, an EV

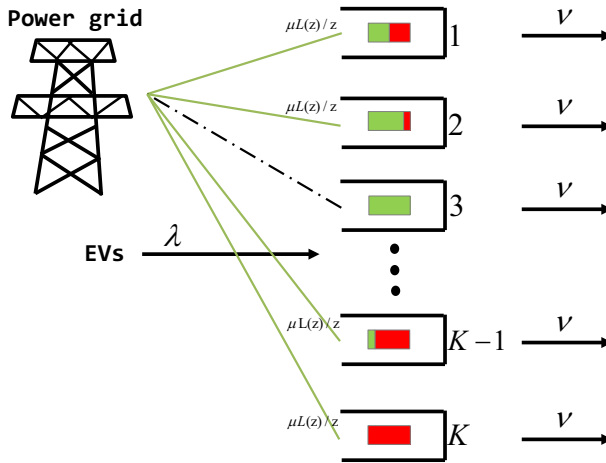


Fig. 1 A charging station with K EV chargers

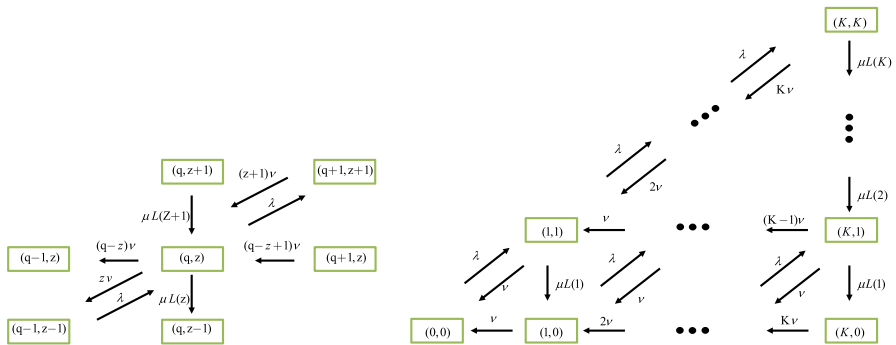
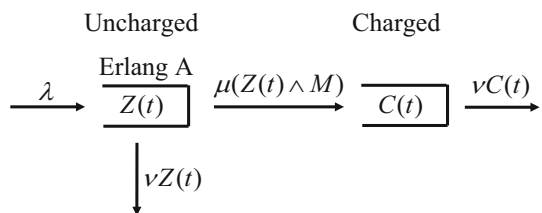


Fig. 2 Transition rates in the interior (left) and on the boundary (right) of the process $\{(Q(t), Z(t)) : t \geq 0\}$

Fig. 3 Model description in the case of infinitely many parking spaces



abandons its charging because its parking time expired (and thus leaves the first queue impatiently); this happens with rate $\nu Z(t)$. Note that the total “rate in” in the system is λ , and the total “rate out” is $\nu(Z(t) + C(t)) = \nu Q(t)$. In other words, $Q(t)$ describes the number of customers in an $M/M/\infty$ queue; i.e., its steady-state distribution is a Poisson distribution with rate λ/ν . As we will see in Proposition 3.3, the process describing the number of uncharged EVs in the system (i.e., $Z(\cdot)$) behaves as a modified Erlang-A queue. The transition rates are shown in Fig. 4.

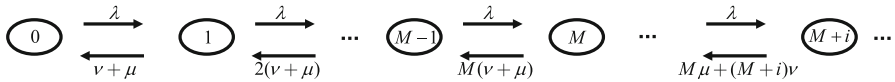


Fig. 4 Transition rates of the process $Z(\cdot)$ (Erlang-A)

2.3 System dynamics

In this section, we introduce the dynamics that describe the evolution of the system. We avoid a rigorous sample-path construction of the stochastic processes, and we refer to [11, 32] for background.

For a constant r , let $N_r(\cdot)$ be a Poisson process with rate r . The total number of EVs in the system at time $t \geq 0$, $Q(t)$, is given by

$$Q(t) = Q(0) + N_\lambda \left(\int_0^t \mathbb{1}_{\{Q(s) < K\}} ds \right) - N_{v,1} \left(\int_0^t Z(s) ds \right) - N_{v,2} \left(\int_0^t C(s) ds \right), \quad (2.1)$$

where $N_\lambda(\cdot)$, $N_{v,1}(\cdot)$, and $N_{v,2}(\cdot)$ are independent Poisson processes. Here, the number of EVs that arrive at the charging station during the time interval $[0, t]$ is given by the process $N_\lambda \left(\int_0^t \mathbb{1}_{\{Q(s) < K\}} ds \right)$, $N_{v,1} \left(\int_0^t Z(s) ds \right)$ is the number of uncharged EVs that depart up to time t , and $N_{v,2} \left(\int_0^t C(s) ds \right)$ counts the departures of fully charged EVs up to time t . Hence, $N_{v,1} \left(\int_0^t Z(s) ds \right) + N_{v,2} \left(\int_0^t C(s) ds \right)$ is the total number of departures until time t (irrespective of whether the EVs are fully charged or not). In other words, and by the properties of the Poisson process, we have

$$N_{v,1} \left(\int_0^t Z(s) ds \right) + N_{v,2} \left(\int_0^t C(s) ds \right) \stackrel{d}{=} N_v \left(\int_0^t Q(s) ds \right), \quad (2.2)$$

and hence (2.1) describes the population in a well-known Erlang loss queue [29].

Another important process is the number of uncharged EVs in the system, $Z(t)$, which can be written in the following form:

$$Z(t) = Z(0) + N_\lambda \left(\int_0^t \mathbb{1}_{\{Q(s) < K\}} ds \right) - N_\mu \left(\int_0^t L(Z(s)) ds \right) - N_{v,1} \left(\int_0^t Z(s) ds \right), \quad (2.3)$$

where $N_\mu \left(\int_0^t L(Z(s)) ds \right)$ is the number of EVs that get fully charged during $[0, t]$ and is independent of the aforementioned Poisson processes.

Finally, the process which describes the number of fully charged EVs is given by

$$C(t) = Q(t) - Z(t) = C(0) + N_\mu \left(\int_0^t L(Z(s)) ds \right) - N_{v,2} \left(\int_0^t C(s) ds \right).$$

Observe that in the case $K = \infty$, (2.3) is reduced to the Erlang-A system [21,49]. All the previous equations hold almost surely and are defined on the same probability space.

It is clear that the vector process $(Q(\cdot), Z(\cdot))$ constitutes a two-dimensional Markov process. In the sequel, we are interested in finding the joint stationary distribution of $(Q(\cdot), Z(\cdot))$ and in deriving the fraction of fully charged EVs. Although the computation of the exact joint distribution does not seem promising, we are able to obtain exact bounds for the fraction of fully charged EVs in the next section.

3 Explicit bounds

The goal of this section is to give explicit results on some performance measures. In an EV charging setting, one may be interested in finding the fraction of EVs that get fully charged. This is an important performance measure from the point of view of both drivers and of the manager of the charging station. Note that the fraction of EVs that get fully charged (in steady-state) equals $P_s = \frac{v\mathbb{E}[C_M^K(\infty)]}{v\mathbb{E}[Q_M^K(\infty)]} = \frac{\mathbb{E}[C_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]} = 1 - \frac{\mathbb{E}[Z_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]}$, since in equilibrium $v\mathbb{E}[C_M^K(\infty)]$ is the departure rate of fully charged EVs and $v\mathbb{E}[Q_M^K(\infty)]$ is the departure rate of all EVs. Further, P_s gives the probability that a vehicle leaves the charging station with a fully charged battery.

For the general model (i.e., for $K < \infty$ and $M < \infty$) given in Sect. 2, define the steady-state probabilities $p(q, z) := \lim_{t \rightarrow \infty} \mathbb{P}(Q_M^K(t) = q, Z_M^K(t) = z)$. For simplicity, we use $p(q, z)$ instead of $p_M^K(q, z)$. These steady-state probabilities are characterized by the following balance equations: For $(q, z) \in \{\mathbb{R}_+^2 : z \leq q\}$, we have that

$$\begin{aligned} (qv\mathbb{1}_{\{q>0\}} + \lambda\mathbb{1}_{\{q<K\}} + \mu L(z)\mathbb{1}_{\{z>0\}})p(q, z) &= \lambda\mathbb{1}_{\{z>0\}}p(q-1, z-1) \\ &+ (z+1)v\mathbb{1}_{\{q<K\}}p(q+1, z+1) + \mu L(z+1)\mathbb{1}_{\{q \neq z\}}p(q, z+1) \\ &+ (q-z+1)v\mathbb{1}_{\{q<K\}}p(q+1, z). \end{aligned} \quad (3.1)$$

A closed-form solution of the balance equations for any K and any M does not seem possible. However, we are able to obtain explicit solutions in some special cases. Below, we derive bounds for P_s based on three different cases: (i) There is enough power for everyone ($M = \infty$), (ii) there are enough parking spaces for everyone ($K = \infty$), (iii) a full parking lot ($Q(t) \equiv K$). In the next proposition, we give upper and lower bounds for the fraction of EVs that get fully charged.

Proposition 3.1 Let $C_M^K(\infty)$ and $Q_M^K(\infty)$ be the number of fully charged EVs and the total number of EVs in steady state for any K and any M . We have that

$$\frac{\mathbb{E}[C_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]} \leq \frac{\mathbb{E}[C_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]} \leq \frac{\mathbb{E}[C_K^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]}. \quad (3.2)$$

Moreover, an additional lower bound is given by

$$\frac{\mathbb{E}[Q_M^K(\infty)] - \mathbb{E}[Z_f(\infty)]}{\mathbb{E}[Q_M^K(\infty)]} \leq \frac{\mathbb{E}[C_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]}, \quad (3.3)$$

where $Z_f(\cdot)$ is defined in Sect. 3.3.

The proof of this proposition makes use of coupling arguments and stochastic ordering of random variables, and it is given in Sect. 6. We now briefly present the solution of the balance equations for the three special cases described above.

3.1 Enough power for everyone

Assume that K is finite and that there is enough power for all EVs to be charged at a maximum rate, i.e., $M = K$. In this case, the allocation function takes the form $L(Z(t)) = Z(t)$, and the balance equations can be solved explicitly and are given below.

Proposition 3.2 Let $K < \infty$ and $M = K$; then, the solution $p_e(\cdot, \cdot)$ to the balance Eq. (3.1) is given by the following (Binomial) distribution:

$$p_e(q, z) := p_Q(q) \frac{q!}{z!(q-z)!} \left(\frac{\mu}{v+\mu} \right)^{q-z} \left(\frac{v}{v+\mu} \right)^z, \quad (3.4)$$

where

$$p_Q(q) := \sum_{z=0}^q p_e(q, z) = \frac{1}{q!} \left(\frac{\lambda}{v} \right)^q p_Q(0). \quad (3.5)$$

Moreover, the probability of an empty system is given by

$$p_e(0, 0) = p_Q(0) = \left(\sum_{i=0}^K \frac{1}{i!} \left(\frac{\lambda}{v} \right)^i \right)^{-1}.$$

3.2 Enough parking spaces for everyone

In the second case, we assume that there are infinitely many parking spaces; i.e., $K = \infty$ and $M < \infty$. In this case, all EVs can find a free position and the process

$Z(\cdot)$ can be modeled as a Markov process itself where its transition rates are given in Fig. 4. We see in the next proposition that the process $Z(\cdot)$ behaves as a modified Erlang-A model with M servers [49]. The main difference here is that EVs can leave the system even if they are in service (i.e., are getting charged).

Proposition 3.3 *For $z = 0, 1, \dots$, let $p_Z(z) := \lim_{t \rightarrow \infty} \mathbb{P}(Z_M^\infty(t) = z)$ be the stationary distribution of the Markov process $\{Z_M^\infty(t), t \geq 0\}$. It is given by*

$$p_Z(z) = \begin{cases} \frac{1}{z!} \left(\frac{\lambda}{\nu + \mu} \right)^z p_Z(0), & \text{if } z \leq M, \\ \frac{1}{M!} \left(\frac{\lambda}{\nu + \mu} \right)^M \prod_{k=M+1}^z \frac{\lambda}{M\mu + k\nu} p_Z(0), & \text{if } z > M, \end{cases}$$

where

$$p_Z(0) = \left(\sum_{j=0}^M \frac{1}{j!} \left(\frac{\lambda}{\nu + \mu} \right)^j + \sum_{j=M+1}^{\infty} \frac{1}{M!} \left(\frac{\lambda}{\nu + \mu} \right)^M \prod_{k=M+1}^j \frac{\lambda}{M\mu + k\nu} \right)^{-1}.$$

3.3 A full parking lot

Finally, we consider the case where the parking lot is always full; i.e., the total number of EVs (uncharged and charged) is equal to the number of parking spaces. Roughly speaking, we assume that the arrival rate is infinite and that we replace (immediately) each departing EV by a newly arriving EV, which we assume to be uncharged. Hence, the total number of EVs always remains constant and it is equal to K . In other words, the original two-dimensional stochastic model reduces to a one-dimensional model. For this model, we find its steady-state distribution below. This result yields an upper bound for the number of uncharged EVs in the original system and hence a lower bound for the fraction of EVs that get fully charged. As we shall see later, the result in this section plays a crucial role in the study of the diffusion limit in the overloaded regime. Also, in the numerics, we see that a modification of the full parking lot case gives a very good approximation for the fraction of fully charged EVs.

Under these assumptions, all newly arriving EVs are uncharged and so it turns out that the process describing the number of uncharged EVs in the system, $\{Z_f(t), t \geq 0\}$, is a birth–death process. In particular, the birth rate is $\nu(K - Z_f(t))$ and the death rate is equal to $\mu(Z_f(t) \wedge M)$. The steady-state distribution of the aforementioned birth–death process is given in the following proposition.

Proposition 3.4 *The steady-state distribution of the Markov process $\{Z_f(t), t \geq 0\}$ is given by*

$$\pi_f(z) = \begin{cases} \left(\frac{\mu}{\nu} \right)^{M-z} \frac{\prod_{j=0}^{M-z-1} (M-j)}{\prod_{j=1}^{M-z} (K-M-j)} \pi_f(M), & \text{if } 0 \leq z < M, \\ \frac{1}{M^{z-M}} \left(\frac{\nu}{\mu} \right)^{z-M} \prod_{j=0}^{z-M-1} (K-M-j) \pi_f(M), & \text{if } M \leq z \leq K, \end{cases}$$

where

$$\pi_f(M) = \left(\sum_{l=0}^{M-1} \left(\frac{\mu}{\nu} \right)^{M-l} \frac{\prod_{j=0}^{M-l-1} (M-j)}{\prod_{j=1}^{M-l} (K-M-j)} + \sum_{l=M}^K \frac{1}{M^{l-M}} \left(\frac{\nu}{\mu} \right)^{l-M} \prod_{j=0}^{l-M-1} (K-M-j) \right)^{-1}.$$

In Sect. 5, we validate these bounds in the three regimes: moderately, critically, and overloaded. As we will see, the bounds are not very close in general. For this reason, we move to asymptotic approximations.

4 Asymptotic approximations

In this section, we present asymptotic approximations. First we focus on the fluid approximation, and then we move to three diffusion approximations. Consider a family of systems indexed by $n \in \mathbb{N}$, where n tends to infinity, with the same basic structure as that of the system described in Sect. 2. To indicate the position in the sequence of systems, a superscript n will be appended to the system parameters and processes. In the remainder of this section, we assume that $\mathbb{E}[Q^n(0)]$ and $\mathbb{E}[Z^n(0)]$ are finite. Finally, the proofs of the limit theorems are based on martingale arguments and are given in Sects. 6.2–6.5. We give a rigorous proof for Theorem 4.5 in Sect. 6.3, and we omit the full details for the other proofs.

4.1 Fluid approximation

Here we study a fluid model, which is a deterministic model that can be thought of as a formal law of large numbers approximation under appropriate scaling. We develop a fluid approximation for finite K , following a similar approach as in [22]. The main differences here are the finitely many servers in the system and that the state space consists of two regions: $\{Z(t) > M\}$ and $\{Z(t) \leq M\}$.

To obtain a nontrivial fluid limit, we assume that the capacity of power in the n^{th} system is given by nM , the arrival rate by $n\lambda$, the number of parking spaces by nK , and we do not scale the time. The fluid scaling of the process describing the number of uncharged EVs in the charging station is given by $\frac{Z^n(\cdot)}{n}$. This scaling gives rise to the following definition of a fluid model.

Definition 4.1 (*Fluid model*) A continuous function $z(t) : \mathbb{R}_+ \rightarrow [0, K]$ is a fluid model solution if it satisfies the ODE

$$z'(t) = \lambda \wedge \nu K - \nu z(t) - \mu(z(t) \wedge M), \quad (4.1)$$

for $t \in [0, t^*)$, where $t^* = \inf\{s \geq 0 : z(s) = 0\}$ and $z(t) \equiv 0$ for $t \geq t^*$.

Note that (4.1) can be written as $z'(t) = R(z(t))$, where $R(\cdot) = \lambda \wedge \nu K - \nu \cdot -\mu(\cdot \wedge M)$. Further, the operator $R(\cdot)$ is Lipschitz continuous in \mathbb{R}_+ , which guarantees that (4.1) has a unique solution. In the proof of Proposition 4.2, we shall see that if the initial state of the fluid model solution is $z(0) \in [0, K]$, then $z(t) \leq K$ for any $t \geq 0$. This last statement ensures that our fluid model is well defined.

Next, we see that the fluid model solution can arise as a limit of the fluid-scaled process $\frac{Z^n(\cdot)}{n}$. The proof of the following proposition is based on martingale arguments and is given in Sect. 6.2.

Proposition 4.1 *If $\frac{Z^n(0)}{n} \xrightarrow{d} z(0)$ and $\frac{Q^n(0)}{n} \xrightarrow{d} K$, then we have that $\frac{Z^n(\cdot)}{n} \xrightarrow{d} z(\cdot)$, as $n \rightarrow \infty$. Moreover, the deterministic function $z(\cdot)$ satisfies (4.1).*

Remark 4.1 We point out that we can extend the previous proposition to the case $\frac{Q^n(0)}{n} \xrightarrow{d} q(0) \leq K$. This requires a modification of Definition 4.1. In particular, we need to replace the quantity $(\lambda \wedge \nu K)t$ by $\lambda t - y(t)$, where $y(t) = \int_0^t \mathbb{1}_{\{q(s)=K\}} dy(s)$ and $q(s)$ is the fluid approximation of an Erlang loss queue, given by $q(s) = \min\{q^* + (q(0) - q^*)e^{-\nu t}, K\}$, with $q^* = \min\{\lambda/\nu, K\}$; see [35, Proposition 6.17 and 6.21]. Then, Proposition 4.1 holds for any $q(0) \leq K$. Note that this assumption does not affect the analysis later, as the invariant point of this extended fluid model is independent of the initial state $q(0)$.

Moreover, the next proposition states that the fluid model solution converges to the unique invariant point as time goes to infinity.

Proposition 4.2 *Let B and D be exponential random variables with rates μ and ν . We have that, for any $z(0) \in [0, K]$, $z(t) \rightarrow z^*$ exponentially fast as $t \rightarrow \infty$. In addition, z^* is given by the unique positive solution to the following fixed-point equation:*

$$z^* = (\lambda \wedge \nu K) \mathbb{E} \left[\min \left\{ D, B \max \left\{ 1, \frac{z^*}{M} \right\} \right\} \right]. \quad (4.2)$$

In the proof of Proposition 4.2, we shall see that if $z(0) = z^*$ then $z(t) = z^*$ for any $t \geq 0$, i.e., z^* is the unique invariant point of (4.1). The point z^* can be viewed as an approximation of the expected number of uncharged EVs in the system for the original (stochastic) model. Observing that the quantity $\mathbb{E} \left[\min \left\{ D, B \max \left\{ 1, \frac{z^*}{M} \right\} \right\} \right]$ is the actual sojourn time of an uncharged EV in the system and that the quantity $(\lambda \wedge \nu K)$ plays the role of the arrival rate, (4.2) can be seen as a version of Little's law. Further, if we allow a processor-sharing discipline and infinity many servers (i.e., $L(\cdot) \equiv 1$ and $K = \infty$), then (4.2) reduces to [22, Eq. (4.1)].

Remark 4.2 We shall see in the proof of Proposition 4.2 that the invariant point z^* has a simpler form than (4.2) but the latter holds much more generally. If the random variables B and D are generally distributed and possibly dependent with $\mathbb{E}[B \wedge D] < \infty$, then (4.2) still holds. The mathematical analysis then requires the use of measure-valued processes, which is beyond of the scope of the current work; for a heuristic approach, see [4]. Thus, we present the proofs only under Markovian assumptions.

To ensure that z^* is indeed a fluid approximation, we show that we can interchange the fluid and the steady-state limits. First, note that $Z(\cdot)$ has a limiting distribution. To see this, observe that $Z(\cdot)$ is bounded almost surely from above by the queue length of an Erlang-A queue with M servers and infinite buffer. Alternatively, we can bound it by the queue length of an $M/G/\infty$ queue. Now, using the same arguments as in [34], we conclude that $Z(\cdot)$ is a regenerative process and that there exists a stationary limit, $Z(\infty)$. The next proposition says that the stationary scaled sequence of random variables converges to the unique invariant point z^* .

Proposition 4.3 *The stationary fluid-scaled sequence of random variables $\frac{Z^n(\infty)}{n}$ is tight and $\frac{Z^n(\infty)}{n} \xrightarrow{d} z^*$, as $n \rightarrow \infty$.*

Note that the arrival rate in an Erlang loss queue is known and it is equal to $\lambda(1 - B(\lambda/\nu, K))$, where $B(\lambda/\nu, K)$ is the blocking probability in a loss system with K servers and traffic intensity λ/ν . Furthermore, $\lambda(1 - B(\lambda/\nu, K))$ is asymptotically exact for our fluid approximation in the sense that $\lambda(1 - B(n\lambda/\nu, nK)) \rightarrow (\lambda \wedge \nu K)$, as $n \rightarrow \infty$. To improve the fluid approximation, we replace $(\lambda \wedge \nu K)$ by $\lambda(1 - B(\lambda/\nu, K))$, leading to

$$z^* = \lambda(1 - B(\lambda/\nu, K)) \mathbb{E} \left[\min \left\{ D, B \max \left\{ 1, \frac{z^*}{M} \right\} \right\} \right]. \quad (4.3)$$

Heuristically, we assume that an EV sees the system in stationarity throughout its sojourn and we use Little's law and a version of the snapshot principle [33].

Having found the fluid approximation for the number of uncharged EVs in the charging station, we derive the fluid approximation for the fraction of EVs that get successfully charged. Let \bar{P}_s denote the probability that an EV leaves the parking lot with a fully charged battery in the fluid model. This is given by $\bar{P}_s = \mathbb{P}(D > B \max\{1, z^*/M\})$, where z^* is the unique solution of (4.3). Under our assumptions, an explicit expression for this probability can be found. That is,

$$\bar{P}_s = \begin{cases} \frac{\mu}{\nu + \mu}, & z^* \leq M, \\ \frac{\mu M}{\lambda(1 - B(\lambda/\nu, K))}, & z^* > M. \end{cases}$$

We now focus on the fluid approximation for the number of uncharged EVs when the parking lot is full (see Sect. 3.3). Analogously to Definition 4.1, we can define a fluid model, which we call $z_f(\cdot)$.

Proposition 4.4 *Assume that the scaled parking spaces and the scaled power capacity are given by $K^n = Kn$ and $M^n = Mn$, respectively. If $\frac{Z_f^n(0)}{n} \xrightarrow{d} z_f(0)$, we have that $\frac{Z_f^n(\cdot)}{n} \xrightarrow{d} z_f(\cdot)$, and $z_f(t) \rightarrow z_f^*$ as n and t go to infinity. Further, the limits can be interchanged and z_f^* is given by the following formula:*

$$z_f^* = \begin{cases} \frac{vK}{v+\mu}, & \text{if } z_f^* \leq M, \\ \frac{vK-\mu M}{v}, & \text{if } z_f^* > M. \end{cases} \quad (4.4)$$

We give a heuristic approach to deriving (4.4), skipping the proof, which can be done by using a similar procedure as in the general case. The intuition behind (4.4) is as follows: Let $S^n(B)$ be the sojourn time (in steady state) of an uncharged EV in the n^{th} system. By Little's law, we have that

$$\mathbb{E}[Z_f^n(\infty)] = v\mathbb{E}[K^n - Z_f^n(\infty)]\mathbb{E}[S^n(B)]. \quad (4.5)$$

By the discussion after Theorem 2.4 in [23], because we observe the system in steady state at time 0, the number of uncharged EVs hardly changes for large enough n . By the snapshot principle, we have that $S^n(B) = B \frac{Z_f^n(\infty)}{Z_f^n(\infty) \wedge M^n} + o(n) = B \left(\frac{Z_f^n(\infty)}{M^n} \vee 1 \right) + o(n)$. Applying the last equation in (4.5) and dividing by n yields

$$\mathbb{E}\left[\frac{Z_f^n(\infty)}{n}\right] = v\mathbb{E}\left[K - \frac{Z_f^n(\infty)}{n}\right]\mathbb{E}\left[B \left(\frac{Z_f^n(\infty)}{Mn} \vee 1 \right) + \frac{o(n)}{n}\right].$$

Now, taking the limit as n goes to infinity leads to

$$z_f^* = \frac{v(K - z_f^*) \left(\frac{z_f^*}{M} \vee 1 \right)}{\mu} = \frac{v(K - z_f^*)z_f^*}{\mu(z_f^* \wedge M)}.$$

Finally, z_f^* is given by the following fixed-point equation:

$$\mu(z_f^* \wedge M) = v(K - z_f^*),$$

and solving this last equation leads to (4.4).

We shall see in the numerical examples in Sect. 5 that the fluid approximation is a good approximation of the fraction of fully charged EVs in most cases. However, especially in the underloaded regime and for small number of EV chargers, the error becomes larger. In the next section, we move to diffusion approximations.

4.2 Diffusion approximations

In this section, we show diffusion limit theorems and diffusion approximations for the process describing the number of uncharged and the total number of EVs in the parking lot (*the vector process*). To do this, we follow the strategy set up in [32] using the martingale representation.

First, we work on the Halfin–Whitt regime (see Sect. 4.2.1). Using the “square-root staffing rule” to scale the system parameters, we extend [32, Theorem 7.1] and we

obtain a limit which is a reflected two-dimensional OU process with piecewise linear drift. Then, we derive an equation which characterizes its steady-state distribution, the so-called basic adjoint relation (BAR). However, it turns out that the computation of the steady-state distribution is a hard problem which is beyond the scope of this paper, and it remains an open problem.

The second asymptotic regime we consider here is an overloaded regime (Sect. 4.2.2). Assuming that the process describing the total number of EVs is in an overloaded regime and using the “square-root staffing rule” to scale the total power capacity in the system, we can show that the scaled vector process converges weakly to a one-dimensional limit. Thus, we can compute its steady-state distribution.

Finally, in Sect. 4.2.3, we focus on the case where the parking times of the EVs are sufficiently large. We give a heavy traffic limit and a two-dimensional approximation for the vector process.

4.2.1 Diffusion approximation in the Halfin–Whitt regime

The main goal in this section is to prove a two-dimensional diffusion limit for the vector process. For $-\infty < \beta, \kappa < \infty$, consider the following scaling:

1. $\lambda^n = n(v + \mu)$,
2. $M^n = \frac{\lambda^n}{v + \mu} + \beta\sqrt{n}$,
3. $K^n = \frac{\lambda^n}{v} + \kappa\sqrt{n}$.

Define a sequence of diffusion-scaled processes $\hat{Q}^n(\cdot) := \frac{Q^n(\cdot) - \frac{\lambda^n}{v}}{\sqrt{n}}$ and $\hat{Z}^n(\cdot) := \frac{Z^n(\cdot) - \frac{\lambda^n}{v + \mu}}{\sqrt{n}}$. The allocation function in the n^{th} system is given by $L^n(Z^n(\cdot)) := Z^n(\cdot) \wedge M^n$. We can then prove the following theorem.

Theorem 4.5 (Diffusion limit in the Halfin–Whitt regime) *If $(\hat{Z}^n(0), \hat{Q}^n(0)) \xrightarrow{d} (\hat{Z}(0), \hat{Q}(0))$ as $n \rightarrow \infty$, then $(\hat{Z}^n(\cdot), \hat{Q}^n(\cdot)) \xrightarrow{d} (\hat{Z}(\cdot), \hat{Q}(\cdot))$. The limit satisfies the following two-dimensional stochastic differential equation:*

$$\begin{aligned} \begin{pmatrix} d\hat{Z}(t) \\ d\hat{Q}(t) \end{pmatrix} &= \begin{pmatrix} b_1(\hat{Z}(t)) \\ b_2(\hat{Q}(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{2(v + \mu)} & 0 \\ 0 & \sqrt{2(v + \mu)} \end{pmatrix} \begin{pmatrix} dW_{\hat{Z}}(t) \\ dW_{\hat{Q}}(t) \end{pmatrix} \\ &\quad - \begin{pmatrix} d\hat{Y}(t) \\ d\hat{Y}(t) \end{pmatrix}, \end{aligned} \quad (4.6)$$

where $b_1(x) = -\mu(x \wedge \beta) - vx$ and $b_2(x) = -vx$. Further, $W_{\hat{Z}}(\cdot)$ and $W_{\hat{Q}}(\cdot)$ are driftless, univariate Brownian motions such that $2(v + \mu)\mathbb{E}[W_{\hat{Z}}(t)W_{\hat{Q}}(t)] = (2v + \mu)t$. In addition, $\hat{Y}(\cdot)$ is the unique nondecreasing nonnegative process such that (4.6) holds and $\int_0^\infty \mathbb{1}_{\{\hat{Q}(t) < \kappa\}} d\hat{Y}(t) = 0$.

Adapting [32, Sect. 7.3], we can show that the last theorem also holds if we allow the arrival process to be a general stochastic process under the assumption that it satisfies the functional central limit theorem.

The proof of Theorem 4.5 is given in Sect. 6.3.1 and is organized as follows:

1. We first establish a continuity result and show the existence and uniqueness of the candidate limit (Proposition 6.1).
2. We then rewrite the system dynamics using appropriate martingales and filtrations; see Eqs. (6.15), (6.16), and Proposition 6.2.
3. Next, we show in Proposition 6.3 that the corresponding fluid-scaled processes converge weakly to deterministic functions.
4. Finally, the proof of Theorem 4.5 is completed by applying the martingale central limit theorem in [20] and Proposition 6.1.

Next, we focus on characterizing the joint steady-state distribution of the limit given by (4.6). Our approach is to find a functional equation which describes the joint steady-state distribution, the so-called *basic adjoint relation*. The next step is to use the BAR in order to obtain a key relation for the moment-generating function of the vector process. The piecewise linear drift and the existence of the reflection in (4.6) make the key relation complicated, and its analysis is beyond of the scope of this paper.

For any $t \geq 0$, we know that $\hat{Z}(t) \in \mathbb{R}$ and $\hat{Q}(t) \leq \kappa$. It is more convenient to transform the previous processes such that $\hat{Z}(t) \in \mathbb{R}$ and $\kappa - \hat{Q}(t) \geq 0$. To do so, we recall that $b_2(x) = -\nu x$. Thus, the diffusion limit can be written in the following integral form—see (4.6):

$$\begin{aligned}\hat{Z}(t) &= \hat{Z}(0) + \int_0^t b_1(\hat{Z}(s))ds + \sqrt{2(\nu + \mu)}W_{\hat{Z}}(t) - \hat{Y}(t), \\ \hat{Q}(t) &= \hat{Q}(0) - \nu \int_0^t \hat{Q}(s)ds + \sqrt{2(\nu + \mu)}W_{\hat{Q}}(t) - \hat{Y}(t),\end{aligned}$$

where $\hat{Y}(\cdot)$ is defined in Theorem 4.5. Multiplying by (-1) , adding and subtracting the terms κ and $\nu\kappa t$ in the last equation, we obtain

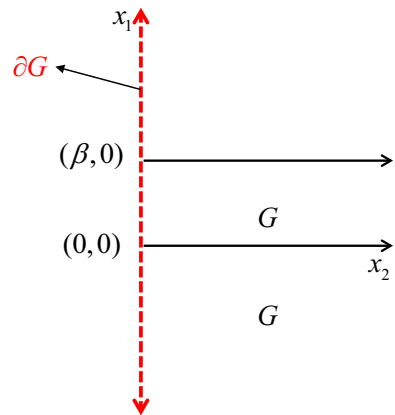
$$\kappa - \hat{Q}(t) = \kappa - \hat{Q}(0) + \nu \int_0^t (\kappa + \hat{Q}(s) - \kappa)ds - \sqrt{2(\nu + \mu)}W_{\hat{Q}}(t) + \hat{Y}(t).$$

Defining $\hat{Q}_\kappa(t) := \kappa - \hat{Q}(t)$ for $t \geq 0$, we have that

$$\hat{Q}_\kappa(t) = \hat{Q}_\kappa(0) + \int_0^t b_\kappa(\hat{Q}_\kappa(s))ds - \sqrt{2(\nu + \mu)}W_{\hat{Q}}(t) + \hat{Y}(t),$$

where $b_\kappa(x) = \nu(\kappa - x)$. The process $\hat{Q}_\kappa(t)$ represents the number of available spots in the parking lot at time $t \geq 0$ (after scaling and after taking the limit as n goes to infinity). Furthermore, $\hat{Y}(t)$ increases if and only if $\hat{Q}_\kappa(t) = 0$. Define $X(\cdot) := (\hat{Z}(\cdot), \hat{Q}_\kappa(\cdot))$, and note that each component of $X(\cdot)$ is a semimartingale. Let $G = \{x := (x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}$. The boundary and the closure of G are given by $\partial G = \{x \in \mathbb{R}^2 : x_2 = 0\}$ and $\bar{G} = G \cup \partial G$, respectively. Now, observe that $X(\cdot) \in \bar{G}$ for any $t \geq 0$. A geometrical representation of the space G and its boundary is shown in the next figure (Fig. 5).

Fig. 5 The space G and its boundary for $\beta > 0$



Before we continue the analysis of deriving the BAR, we note some properties for the process $\hat{Y}(\cdot)$, which is known as the *regulator*. It is known that $(\hat{Q}(\cdot), \hat{Y}(\cdot))$ satisfies a one-dimensional reflection mapping (or one-dimensional Skorokhod problem). The regulator $\hat{Y}(\cdot)$ is continuous, nondecreasing and has the property

$$\int_0^\infty \mathbb{1}_{\{\hat{Q}_\kappa(t) > 0\}} d\hat{Y}(t) = 0,$$

or equivalently, for all $t \geq 0$,

$$\hat{Y}(t) = \int_0^t \mathbb{1}_{\{\hat{Q}_\kappa(s) = 0\}} d\hat{Y}(s).$$

By [11, Theorem 6.1], almost all the paths of the regulator are Lipschitz continuous on the space $\{x(\cdot) \in D(0, \infty), x(0) \geq 0\}$ under the uniform topology, and are hence absolutely continuous. From the latter, it follows that $\hat{Y}(\cdot)$ is of bounded variation. Moreover, by the proof of [46, Theorem 2.2] (and also in [6, Corollary 6]), there exists a (positive) constant w such that

$$\hat{Y}(t) = w \int_0^t \mathbb{1}_{\{\hat{Q}_\kappa(s) = 0\}} ds. \quad (4.7)$$

For more details, we refer to [8, Lemma 3.1] and [28].

In the sequel, we focus on deriving a functional equation which characterizes the steady-state distribution $\pi(\cdot, \cdot)$ of the process $\{X(t), t \geq 0\}$, provided that it exists. To handle the boundary of the space G , we define a measure σ on $(\bar{G}, \mathfrak{B}(\bar{G}))$ given by

$$\sigma(B) = \mathbb{E}^\pi \left[\int_0^1 \mathbb{1}_{\{X(s) \in B\}} d\hat{Y}(s) \right], \quad B \in \mathfrak{B}(\bar{G}). \quad (4.8)$$

Further, it follows by (4.7) that $\sigma(B) \leq w \mathbb{E}^\pi[\hat{Y}(1)] < \infty$, which yields that σ is a finite measure. Moreover, we define it (for simplicity) in \bar{G} , but as $\hat{Y}(\cdot)$ increases only

on the boundary ∂G , the measure concentrates on the boundary. In other words, σ consists of a finite boundary measure.

Using [27, Theorem 1 and Remark 5.2] and Itô calculus, the BAR takes the following form:

$$\int_{\bar{G}} \mathcal{L}f(\mathbf{x})\pi(d\mathbf{x}) - \int_{\bar{G}} \frac{\partial f}{\partial x_1}(\mathbf{x})\sigma(d\mathbf{x}) + \int_{\bar{G}} \frac{\partial f}{\partial x_2}(\mathbf{x})\sigma(d\mathbf{x}) = 0, \quad (4.9)$$

where the boundary measure σ is defined in (4.8) and \mathcal{L} is the second-order operator, i.e.,

$$\begin{aligned} \mathcal{L}f(\mathbf{x}) = & b_1(x_1) \frac{\partial f}{\partial x_1}(\mathbf{x}) + b_\kappa(x_2) \frac{\partial f}{\partial x_2}(\mathbf{x}) + (\nu + \mu) \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}) \\ & + (\nu + \mu) \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x}) - (2\nu + \mu) \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}). \end{aligned}$$

The next step is to derive a key relation between the moment-generating functions of π and σ . Let us define the two-dimensional moment-generating function (MGF) of π ,

$$G^\pi(\boldsymbol{\theta}) := \mathbb{E}^\pi[e^{\boldsymbol{\theta} \cdot \mathbf{X}(\infty)}] = \int_{\bar{G}} e^{\boldsymbol{\theta} \cdot \mathbf{x}} \pi(d\mathbf{x}),$$

for $\boldsymbol{\theta} := (\theta_1, \theta_2) \in \mathbb{R}^2$, and $\boldsymbol{\theta} \cdot \mathbf{x} := \theta_1 x_1 + \theta_2 x_2$. In the same way, we define the one-dimensional MGF of σ ,

$$G^\sigma(\theta_1) := \int_{\bar{G}} e^{\theta_1 x_1} \sigma(d\mathbf{x}).$$

Further, we assume that there exists a set Θ such that $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^2 : G^\pi(\boldsymbol{\theta}) < \infty, G^\sigma(\theta_1) < \infty\}$. Assuming that $\boldsymbol{\theta} \in \Theta$ and adapting [16, Lemma 4.1], we derive the following key relation:

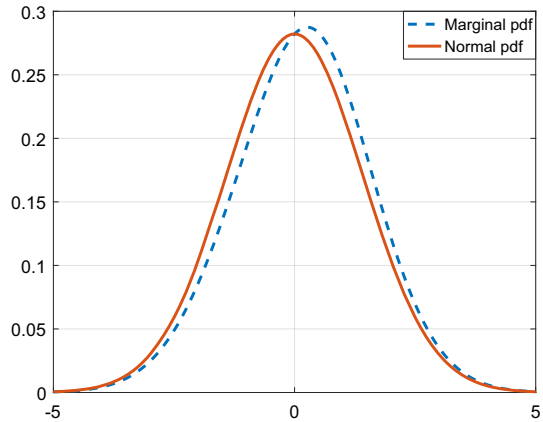
$$\begin{aligned} & -\mu\beta\theta_1 \mathbb{E}^\pi[\mathbb{1}_{\{Z(\infty) > \beta\}} e^{\mathbf{X}(\infty) \cdot \boldsymbol{\theta}}] - \mu\theta_1 \mathbb{E}^\pi[Z(\infty) \mathbb{1}_{\{Z(\infty) \leq \beta\}} e^{\mathbf{X}(\infty) \cdot \boldsymbol{\theta}}] - \nu\theta_1 G_{\theta_1}^\pi(\boldsymbol{\theta}) \\ & - \nu\theta_2 G_{\theta_2}^\pi(\boldsymbol{\theta}) + \gamma(\boldsymbol{\theta}) G^\pi(\boldsymbol{\theta}) + (\theta_2 - \theta_1) G^\sigma(\theta_1) = 0, \end{aligned} \quad (4.10)$$

where $\gamma(\boldsymbol{\theta}) = \nu\kappa\theta_2 + (\nu + \mu)(\theta_1^2 + \theta_2^2) - (2\nu + \mu)\theta_1\theta_2$ and $G_{\theta_i}^\pi(\cdot)$ denotes the derivative with respect to θ_i , $i = 1, 2$.

Equation (4.10) is rather complicated due to the piecewise linear term and the existence of the boundary measure. Although the analysis of (4.10) is beyond the scope of the current paper, we conjecture that the Wiener–Hopf method [13] and boundary value techniques [14] may be applied.

It turns out that (4.10) remains quite complicated even if we assume $K = \infty$, i.e., no boundary measure. Contrary to the one-dimensional case [9], the steady-state distribution of $(Z(\cdot), Q(\cdot))$ cannot be written as a linear combination of two distributions. To see this, define $\pi_-(\mathbf{x})$ to be a bivariate normal distribution with mean

Fig. 6 Marginal pfd of $\hat{Q}(\infty)$ and $\text{Normal}(0, \frac{v+\mu}{v})$ pdf for $\beta = 0$ and $K = \infty$



vector $\mu_- = (0, 0)$ and covariance matrix $\Sigma_- = \begin{pmatrix} 1 & 1 \\ 1 & \frac{v+\mu}{v} \end{pmatrix}$. In addition, let $\pi_+(\mathbf{x})$ be a bivariate normal distribution with mean vector $\mu_+ = (-\mu\beta/v, 0)$ and covariance matrix $\Sigma_+ = \begin{pmatrix} \frac{v+\mu}{2v} & \frac{2v+\mu}{2v} \\ \frac{2v+\mu}{2v} & \frac{2v+\mu}{v} \end{pmatrix}$. The distributions π_- and π_+ correspond to the solution of the Kolmogorov forward equations (or Fokker–Planck equations) of (4.6) with drift function $-(v + \mu)x$ and $-\mu\beta - vx$, respectively. Adapting [9], we define $\pi_\infty(\mathbf{x}) := c_1\pi_-(\mathbf{x})\mathbb{1}_{\{x_1 \leq \beta\}} + c_2\pi_+(\mathbf{x})\mathbb{1}_{\{x_1 > \beta\}}$, where the constants c_1, c_2 are given by [21, Eqs. (3.9), (3.10)]. Namely, we have that

$$c_1 = \left(\Phi(\beta) + \sqrt{\frac{v+\mu}{v}} \exp \left\{ \frac{\mu\beta^2}{2v} \right\} \left(1 - \Phi \left(\sqrt{\frac{v+\mu}{v}} \beta \right) \right) \right)^{-1},$$

$$c_2 = c_1 \sqrt{\frac{v+\mu}{v}} \exp \left\{ \frac{\mu\beta^2}{2v} \right\},$$

where $\Phi(\cdot)$ represents the cumulative probability function of the standard normal distribution. It can be easily verified that π_∞ is indeed a probability distribution, but it does not satisfy the correct marginal distribution of $\hat{Q}(\infty)$ as is shown in Fig. 6. It is well known that $\hat{Q}(\infty)$ follows a normal distribution with zero mean and variance $\frac{v+\mu}{v}$. For a discussion on this topic, see [15]. In the sequel, we move in different asymptotic regimes.

4.2.2 Diffusion approximation in an overloaded parking lot

In this section, we study an overloaded parking lot. First, we show a diffusion limit in the case that the parking lot is always full (see Sect. 3.3). We then show that the diffusion-scaled vector process for the original system collapses to the one-dimensional limit in this case. Our motivation in this section comes from results in [1]. Specifically, the authors there show that under an appropriate scaling (including parameter and timescaling), the number of empty spaces in an overloaded parking lot

behaves like an $M/M/1$ queue. However, here we need a modification of this result by dropping the timescaling.

First, we define the dynamical equation that describes the evolution of the process of the number of uncharged EVs when the parking lot is always full. Let $N_v^f(\cdot)$ and $N_\mu^f(\cdot)$ denote two independent Poisson processes with rates ν and μ , respectively. For any $t \geq 0$, we have that

$$Z_f(t) = Z_f(0) + N_v^f\left(\int_0^t (K - Z_f(s))ds\right) - N_\mu^f\left(\int_0^t Z_f(s) \wedge M ds\right), \quad (4.11)$$

where $Z_f(0) \leq K$ almost surely.

Next, we introduce our asymptotic regime. Take K^n and M^n such that $K^n = nK$ and $M^n = \frac{\nu}{\nu+\mu}K^n + \sqrt{n}\beta$, where $K, \beta \geq 0$. The following proposition gives a diffusion limit for the scaled process describing the number of uncharged EVs, i.e., $Z_f(\cdot)$.

Proposition 4.6 Define the scaled process $\hat{Z}_f^n(\cdot) := \frac{Z_f^n(\cdot) - \frac{\nu}{\nu+\mu}K^n}{\sqrt{n}}$. If $\hat{Z}_f^n(0) \xrightarrow{d} \hat{Z}_f(0)$, then $\hat{Z}_f^n(\cdot) \xrightarrow{d} \hat{Z}_f(\cdot)$, where the limit satisfies the following stochastic differential equation:

$$d\hat{Z}_f(t) = v(\hat{Z}_f(t))dt + \sqrt{\frac{2\nu\mu K}{\nu + \mu}}dW(t).$$

Moreover, the drift function is given by $v(x) = \begin{cases} -(\nu + \mu)x, & \text{if } x \leq \beta, \\ -\nu x - \mu\beta, & \text{if } x > \beta, \end{cases}$ and $W(\cdot)$ is a standard Brownian motion.

Next, we give the steady-state distribution of the process $\hat{Z}_f(\cdot)$. This can be done by following [9, Eq. (3)]. Define the following truncated normal probability density functions:

$$\pi_f^-(x) = \frac{\phi\left(\frac{x}{\sigma_1}\right)}{\sigma_1 \Phi\left(\frac{\beta}{\sigma_1}\right)}, \text{ for } x \leq \beta, \text{ and } \pi_f^+(x) = \frac{\phi\left(\frac{x}{\sigma_2}\right)}{\sigma_1 \left(1 - \Phi\left(\frac{\beta - \frac{\mu\beta}{\nu}}{\sigma_2}\right)\right)}, \text{ for } x > \beta,$$

where $\sigma_1^2 = \frac{1}{\nu+\mu} \frac{\nu\mu K}{\nu+\mu}$ and $\sigma_2^2 = \frac{\mu K}{\nu+\mu}$. Now, the pdf of $\hat{Z}_f(\cdot)$ is given by

$$\pi_f(x) = d_1 \pi_f^-(x) \mathbb{1}_{\{x \leq \beta\}} + d_2 \pi_f^+(x) \mathbb{1}_{\{x > \beta\}}, \quad (4.12)$$

for $x \in \mathbb{R}$. Moreover, the constants are $d_1 = \frac{1}{1+r}$ and $d_2 = 1 - d_1$ with $r = \frac{\sigma_1^2 \pi_f^-(\beta)}{\sigma_2^2 \pi_f^+(\beta)}$.

Having studied the system when it is always full, we now move to the original stochastic model. The first step is to find a relation between the process that gives the number of uncharged EVs and the process that gives the empty parking spaces. Recall that the total number of EVs in the system is given by the following equation:

$$Q(t) = Q(0) + N_\lambda \left(\int_0^t \mathbb{1}_{\{Q(s) < K\}} ds \right) - N_{v,1} \left(\int_0^t Z(s) ds \right) - N_{v,2} \left(\int_0^t (Q(s) - Z(s)) ds \right),$$

and the number of uncharged EVs is

$$Z(t) = Z(0) + N_\lambda \left(\int_0^t \mathbb{1}_{\{Q(s) < K\}} ds \right) - N_\mu \left(\int_0^t Z(s) \wedge M ds \right) - N_{v,1} \left(\int_0^t Z(s) ds \right).$$

Define the stochastic process that describes the number of empty parking spaces in the system, $E(t) := K - Q(t)$, $t \geq 0$. Using the definition of the process $E(\cdot)$, we have that the system dynamics can be rewritten as follows:

$$E(t) = E(0) - N_\lambda \left(\int_0^t \mathbb{1}_{\{E(s) > 0\}} ds \right) + N_{v,1} \left(\int_0^t Z(s) ds \right) + N_{v,2} \left(\int_0^t (K - E(s) - Z(s)) ds \right), \quad (4.13)$$

$$Z(t) = Z(0) + N_\lambda \left(\int_0^t \mathbb{1}_{\{E(s) > 0\}} ds \right) - N_\mu \left(\int_0^t Z(s) \wedge M ds \right) - N_{v,1} \left(\int_0^t Z(s) ds \right). \quad (4.14)$$

By (4.13), it follows that

$$N_\lambda \left(\int_0^t \mathbb{1}_{\{E(s) > 0\}} ds \right) = E(0) - E(t) + N_{v,1} \left(\int_0^t Z(s) ds \right) + N_{v,2} \left(\int_0^t (K - E(s) - Z(s)) ds \right).$$

Applying the last equation in (4.14) yields

$$Z(t) = Z(0) + E(0) - E(t) + N_{v,2} \left(\int_0^t (K - E(s) - Z(s)) ds \right) - N_\mu \left(\int_0^t Z(s) \wedge M ds \right). \quad (4.15)$$

The last relation and an asymptotic bound for the process $E^n(\cdot)$ (see Proposition 6.4) are the core elements we use to prove the main result in this section.

Theorem 4.7 Assume that $\lambda^n = \lambda n$, $K^n = Kn$, $M^n = \frac{v}{v+\mu}K^n + \beta\sqrt{n}$, and $\mathbb{P}(E^n(0) = 0) = 1$. Further, we assume $vK < \lambda$. If $\frac{Z^n(0) - \frac{v}{v+\mu}K^n}{\sqrt{n}} \xrightarrow{d} \hat{Z}_f(0)$, then

$$\frac{Z^n(\cdot) - \frac{v}{v+\mu}K^n}{\sqrt{n}} \xrightarrow{d} \hat{Z}_f(\cdot), \text{ as } n \rightarrow \infty,$$

where the process $\hat{Z}_f(\cdot)$ is given in Proposition 4.6.

A diffusion approximation for the expected number in the original system in an overloaded regime is now given by

$$\mathbb{E}[Z_f(\infty)] \approx \sqrt{K} \mathbb{E}[\hat{Z}_f(\infty)] + \frac{v}{v+\mu}K, \quad (4.16)$$

and, by using (4.12), we have that

$$\mathbb{E}[\hat{Z}_f(\infty)] = d_1 \int_{-\infty}^{\beta} \pi_f^-(x) dx + d_2 \int_{\beta}^{\infty} \pi_f^+(x) dx. \quad (4.17)$$

The asymptotic regime for an overloaded system leads to a one-dimensional approximation. In the next section, motivated by [41], we consider an asymptotic regime where we scale the parking times, which leads to a two-dimensional diffusion approximation.

4.2.3 Diffusion approximation for small parking rates

In this section, we study a diffusion approximation in the case where the parking rate v is “small.” First, we focus on the system with infinitely many parking spaces and we show a heavy traffic limit theorem; see Sect. 2.2.1 for an alternative model description when $K = \infty$. In this case, the limit is a two-dimensional OU process with reflection. Then, making an overloaded assumption (for the uncharged EVs), we derive a two-dimensional OU limit process and we obtain the same limit if we assume a sufficiently large number of parking spaces.

Assume that $K = \infty$. Define the traffic intensity for this model as $\rho := \frac{\lambda}{\mu M}$. Let μ, M be fixed. Further, define $v^n = \frac{1}{n}$ and $\lambda^n = \mu M \left(1 - \frac{c}{\sqrt{n}}\right)$ for some constant c . Note that $\sqrt{n}(1 - \rho^n) = c$, which is our heavy traffic assumption. Moreover, define the diffusion-scaled process as follows:

$$\tilde{Z}^n(t) := \frac{Z^n(nt)}{\sqrt{n}} \quad \text{and} \quad \tilde{Q}^n(t) := \frac{Q^n(nt) - \mu M n}{\sqrt{n}}.$$

The next proposition states a heavy traffic result for the two-dimensional scaled process.

Proposition 4.8 (Heavy traffic) Assume that $(\tilde{Z}^n(0), \tilde{Q}^n(0)) \xrightarrow{d} (\tilde{Z}(0), \tilde{Q}(0))$ as $n \rightarrow \infty$. We have that $(\tilde{Z}^n(\cdot), \tilde{Q}^n(\cdot)) \xrightarrow{d} (\tilde{Z}(\cdot), \tilde{Q}(\cdot))$ and that the limit satisfies the following two-dimensional stochastic differential equation:

$$\begin{pmatrix} d\tilde{Z}(t) \\ d\tilde{Q}(t) \end{pmatrix} = - \begin{pmatrix} c\mu M + \tilde{Z}(t) \\ c\mu M + \tilde{Q}(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{2\mu M} & 0 \\ 0 & \sqrt{2\mu M} \end{pmatrix} \begin{pmatrix} dW_{\tilde{Z}}(t) \\ dW_{\tilde{Q}}(t) \end{pmatrix} + \begin{pmatrix} d\tilde{Y}(t) \\ 0 \end{pmatrix},$$

where $\tilde{Y}(t)$ satisfies the relation $\int_0^\infty \mathbb{1}_{\{\tilde{Z}(t)>0\}} d\tilde{Y}(t) = 0$. Further, $W_{\tilde{Z}}(\cdot)$ and $W_{\tilde{Q}}(\cdot)$ are driftless, univariate Brownian motions such that $\mathbb{E} \left[W_{\tilde{Z}}(t) W_{\tilde{Q}}(t) \right] = t/2$.

Observe that the limit process in the last proposition depends on the reflection at zero, which makes the calculation of the joint distribution hard. We next consider an overloaded regime for the number of uncharged EVs. In this regime, the process $\tilde{Z}^n(\cdot)$ will not reach zero for large enough n [41]. To this end, let λ, μ, M be fixed with $\lambda > \mu M$ and $v^n = 1/n$. Modifying slightly the scaled processes, i.e.,

$$\tilde{Z}_o^n(t) := \frac{Z^n(nt) - (\lambda - \mu M)n}{\sqrt{n}} \quad \text{and} \quad \tilde{Q}_o^n(t) := \frac{Q^n(nt) - \lambda n}{\sqrt{n}},$$

we are able to show the following proposition.

Proposition 4.9 Let $\lambda > \mu M$. If $(\tilde{Z}_o^n(0), \tilde{Q}_o^n(0)) \xrightarrow{d} (\tilde{Z}_o(0), \tilde{Q}_o(0))$ as $n \rightarrow \infty$, then $(\tilde{Z}_o^n(\cdot), \tilde{Q}_o^n(\cdot)) \xrightarrow{d} (\tilde{Z}_o(\cdot), \tilde{Q}_o(\cdot))$. The diffusion limit satisfies the following two-dimensional stochastic differential equation:

$$\begin{pmatrix} d\tilde{Z}_o(t) \\ d\tilde{Q}_o(t) \end{pmatrix} = - \begin{pmatrix} \tilde{Z}_o(t) \\ \tilde{Q}_o(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{\lambda} & -\sqrt{\mu M} & -\sqrt{\lambda - \mu M} & 0 \\ \sqrt{\lambda} & 0 & -\sqrt{\lambda - \mu M} & -\sqrt{\mu M} \end{pmatrix} d\mathbf{W}(t),$$

where $\mathbf{W}(\cdot) = (W_1(\cdot), W_2(\cdot), W_3(\cdot), W_4(\cdot))^T$, with $W_i(\cdot)$ independent standard Brownian motions.

Note that we derive the same limit if we assume that $K < \infty$ and we scale the number of parking spaces in the n^{th} system, K^n , such that $\frac{K^n - \lambda n}{\sqrt{n}} \rightarrow \infty$, as $n \rightarrow \infty$. In this case, the fraction of time that the scaled process $\tilde{Q}_o^n(\cdot)$ spends on the boundary is negligible. This is made rigorous in the following lemma.

Lemma 4.10 For $T > 0$, we have that for any $\epsilon > 0$ there exists n_ϵ such that

$$\mathbb{P} \left(\sup_{t \leq T} \tilde{Q}_o^n(t) < \frac{K^n - \lambda n}{\sqrt{n}} \right) > 1 - \epsilon,$$

for all $n > n_\epsilon$.

Remark 4.3 The sequence $\{\tilde{Q}_o^n(t), t \geq 0\}$ is stochastically bounded as it converges in distribution in $(\mathcal{D}[0, \infty), J_1)$ which is a complete and separate metric space [44, Corollary 3.1]. Then, Lemma 4.10 follows and holds true for any deviating sequence R^n instead of $\frac{K^n - \lambda n}{\sqrt{n}}$. Finally, note that we only need the weak convergence of the process $\tilde{Q}_o^n(\cdot)$ and the fact that the quantity $\frac{K^n - \lambda n}{\sqrt{n}}$ goes to infinity.

The joint steady-state distribution of $(\tilde{Z}_o(\cdot), \tilde{Q}_o(\cdot))$, say $\pi_o(\cdot, \cdot)$, is given by a bivariate normal distribution with mean $\mu = (0, 0)$ and covariance matrix $\Sigma = \begin{pmatrix} \frac{\lambda}{2\lambda - \mu M} & \frac{2\lambda - \mu M}{2\nu} \\ \frac{2\lambda - \mu M}{2\nu} & \frac{\lambda}{\nu} \end{pmatrix}$. Note that Σ is indeed a covariance matrix as it is positive definite. To see this, observe that

$$\det(\Sigma) = \frac{1}{4\nu^2}(4\lambda^2 - 4\lambda^2 - \mu^2 M^2 + 4\lambda\mu M) > \frac{1}{4\nu^2}(-\mu^2 M^2 + 4\mu^2 M^2) > 0,$$

where the first inequality holds by the assumption $\lambda > \mu M$.

Now, for parameters of the original system such that $\lambda > \mu M$, sufficiently “small” ν , and $K > \lambda/\nu$, we suggest the following diffusion approximation:

$$\begin{aligned} \mathbb{E}[Z(\infty)] &\approx \frac{\mathbb{E}[\tilde{Z}_o^n(\infty)]}{\sqrt{\nu}} + \frac{(\lambda - \mu M)}{\nu}, \\ \mathbb{E}[Q(\infty)] &\approx \frac{\mathbb{E}[\tilde{Q}_o^n(\infty)]}{\sqrt{\nu}} + \frac{\lambda}{\nu}. \end{aligned}$$

5 Numerical evaluation

In this section, we validate numerically the previous bounds and approximations for three cases: the moderately ($\lambda < \nu K$), critically ($\lambda = \nu K$), and overloaded ($\lambda > \nu K$) systems. We focus on the expected number of uncharged EVs in the system and the probability that an EV leaves the charging station with a fully charged battery (the success probability). In all the numerical examples, we solve the flow balance equation (3.1) using standard numerical methods and we let $\nu = 1$. Finally, the relative error is calculated by the following formula: $\text{RE} = \frac{|\mathbb{E}[Z(\infty)] - \mathbb{E}[Z^{ap}(\infty)]|}{\mathbb{E}[Z(\infty)]} 100\%$, where $\mathbb{E}[Z(\infty)]$ denotes the expected number of EVs in the original system by solving the two-dimensional Markov process and $\mathbb{E}[Z^{ap}(\infty)]$ denotes the expected number of uncharged EVs for the aforementioned approximations.

First, we evaluate the fluid approximation. Table 1 gives the relative error between the expected number of uncharged EVs for the original system and the fluid approximation given in (4.2) for different values of the number of parking spaces K and for $\mu = 1$. For a given K , we give only the *maximum* relative error for $0 < M \leq K$. As expected, the relative error decreases as λ and K increase. In Tables 2, 3, and 4, we present the relative error between the expected number of uncharged EVs for the original system and the modified fluid approximation given in Eq. (4.3) for different

Table 1 Evaluation of the original fluid approximation for $\mu = 1$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = K$	39.6569	28.5579	23.8308	21.2686	19.3942
$\lambda = 1.2K$	27.9191	18.0935	14.3587	12.0822	10.4540

Table 2 Evaluation of the modified fluid approximation for $\mu = 1/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	5.1261	4.6479	3.3548	2.6049	2.1330
$\lambda = K$	4.4924	2.4775	2.3401	2.2914	2.2800
$\lambda = 1.2K$	3.5314	3.2012	2.3025	1.5766	1.1326

Table 3 Evaluation of the modified fluid approximation for $\mu = 1$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	8.8421	7.2831	6.5797	6.1286	5.7892
$\lambda = K$	11.0904	6.0576	3.5069	2.2082	1.7918
$\lambda = 1.2K$	8.7045	3.7961	3.1936	2.8380	2.5959

Table 4 Evaluation of the modified fluid approximation for $\mu = 3/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	16.6180	14.4757	10.5565	8.2579	6.7973
$\lambda = K$	14.4059	6.0264	4.3725	3.7537	3.3362
$\lambda = 1.2K$	9.6682	6.7387	5.6886	5.0620	4.6325

values of the charging rate: $\mu = 1/2, 1, 3/2$. Not surprisingly, the relative error is much smaller in this case. For high values of λ and K the relative error is approximately 2–5% rather than 10–20%. In addition, the modified fluid approximation seems to be reasonable also in the moderate regime.

Next, we evaluate the approximation in the case of a full parking lot (see Sect. 3.3) and the diffusion approximation in an overloaded regime given by (4.16). To improve the approximations, we directly modify them by replacing the parameter K by the expected total number of EVs in the original system, i.e., $\lambda(1 - B(\lambda/\nu, K))/\nu$. In particular, the modified $\mathbb{E}[Z_f(\infty)]$ is calculated based on the stationary distribution in Proposition 3.4 after replacing K by $\lambda(1 - B(\lambda/\nu, K))/\nu$. Moreover, the modified diffusion approximation in an overloaded regime is given by

$$\mathbb{E}[Z_f(\infty)] \approx \sqrt{\frac{\lambda(1 - B(\lambda/\nu, K))}{\nu}} \mathbb{E}[\hat{Z}_f(\infty)] + \frac{\lambda(1 - B(\lambda/\nu, K))}{\nu + \mu},$$

where $\mathbb{E}[\hat{Z}_f(\infty)]$ is given in (4.17). Tables 5, 6, and 7 give the relative error for the modified $\mathbb{E}[Z_f(\infty)]$ and $\mu = 1/2, 1, 3/2$, respectively. As we expect, it decreases as λ and K increase for all values of the charging rate μ . Furthermore, this approximation results in small relative errors in all regimes ($< 3\%$). The (prelimit) approximation $\mathbb{E}[Z_f(\infty)]$ is better than the modified diffusion approximation in (4.16), as we see in Tables 8, 9, and 10.

In the sequel, we depict the bounds in (3.2), the modified bound in (3.3)—where we replace K by $\lambda(1 - B(\lambda/\nu, K))$ —(dotted line), the modified fluid approximation (4.3) (dashed line), and the modified diffusion approximation in (4.16) (dash-dot line). Note that the modified bound in (4.16) becomes

$$\frac{\lambda(1 - B(\lambda/\nu, K))/\nu - \mathbb{E}[Z_f(\infty)]}{\lambda(1 - B(\lambda/\nu, K))/\nu},$$

Table 5 Evaluation of the modified $\mathbb{E}[Z_f(\infty)]$ for $\mu = 1/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	2.2673	1.6892	1.4294	1.2612	1.1351
$\lambda = K$	1.6397	0.9427	0.7629	0.6589	0.5866
$\lambda = 1.2K$	2.2684	0.6129	0.4266	0.3188	0.2469

Table 6 Evaluation of the modified $\mathbb{E}[Z_f(\infty)]$ for $\mu = 1$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	3.0083	2.2710	1.9940	1.8354	1.7248
$\lambda = K$	1.9747	1.2064	0.8632	0.6492	0.5425
$\lambda = 1.2K$	1.2803	0.6708	0.4649	0.3557	0.2873

Table 7 Evaluation of the modified $\mathbb{E}[Z_f(\infty)]$ for $\mu = 3/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	3.4294	2.5751	2.2303	2.0191	1.8637
$\lambda = K$	1.9995	1.1885	0.9050	0.7513	0.6526
$\lambda = 1.2K$	1.3806	0.7231	0.4886	0.3657	0.2897

Table 8 Evaluation of the modified diffusion approximation in an overloaded regime for $\mu = 1/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	6.0965	4.6910	3.8560	3.2478	2.7785
$\lambda = K$	5.4127	3.2611	2.8785	2.6108	2.3893
$\lambda = 1.2K$	3.5314	3.2012	2.3025	1.5766	1.1326

Table 9 Evaluation of the modified diffusion approximation in an overloaded regime for $\mu = 1$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	12.1493	9.1953	7.8522	7.0228	6.4357
$\lambda = K$	11.7346	8.1938	6.2103	4.7916	3.7020
$\lambda = 1.2K$	10.7606	6.7321	5.1773	4.5167	4.0661

Table 10 Evaluation of the modified diffusion approximation in an overloaded regime for $\mu = 3/2$

	$K = 10$ (%)	$K = 20$ (%)	$K = 30$ (%)	$K = 40$ (%)	$K = 50$ (%)
$\lambda = 0.8K$	19.1007	14.5761	12.2086	10.5932	9.3703
$\lambda = K$	17.3220	10.3962	6.5466	5.6277	5.0069
$\lambda = 1.2K$	14.8738	9.5541	7.9676	7.0158	6.3613

where the modified $\mathbb{E}[Z_f(\infty)]$ is as we discussed earlier. In Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18, the vertical axes give the probability that an EV leaves the parking lot with a fully charged battery (the success probability) and the horizontal axes give the ratio M/K . For each regime, we plot the success probability for $K = 10, 20, 30, 50$ and $\mu = 1$. In the moderate regime, the lower bound ($K = \infty$) is very close for high values of parking spaces. This is not surprising because the time that the process spends on the boundary is negligible in this case. The fluid approximation seems to be quite good in most of the cases. Finally, note that the modified bound (3.3) does not give a lower bound for the original system. However, it seems to be the best approximation for all the cases, even in the moderate regime.

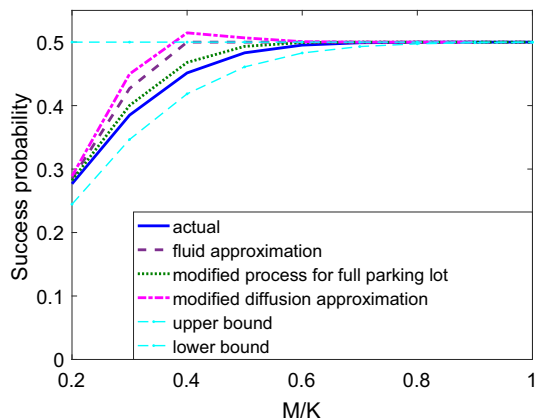
Fig. 7 $K = 10$ and $\lambda = 0.8K$ 

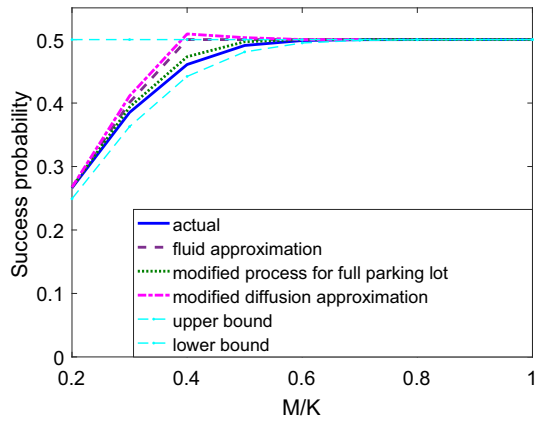
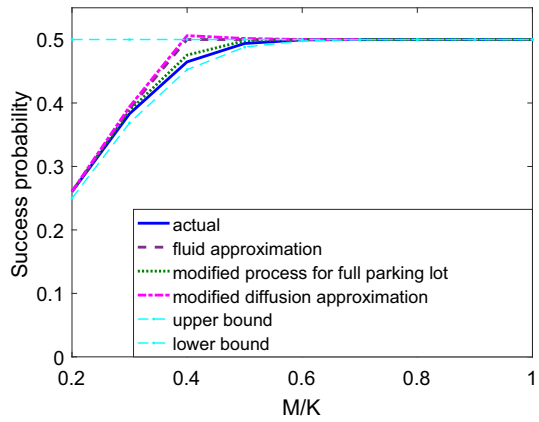
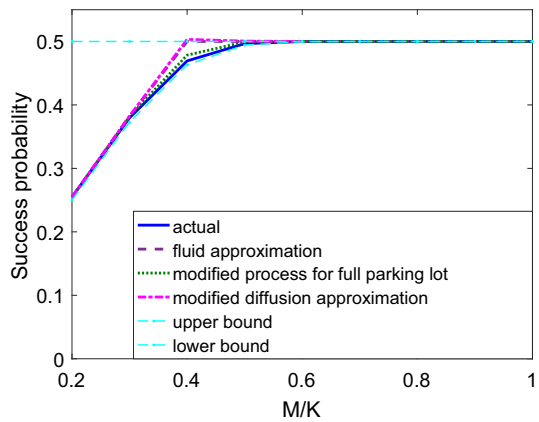
Fig. 8 $K = 20$ and $\lambda = 0.8K$ **Fig. 9** $K = 30$ and $\lambda = 0.8K$ **Fig. 10** $K = 50$ and $\lambda = 0.8K$ 

Fig. 11 $K = 10$ and $\lambda = K$

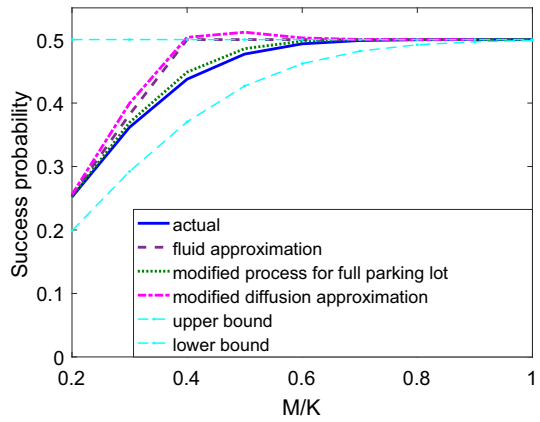


Fig. 12 $K = 20$ and $\lambda = K$

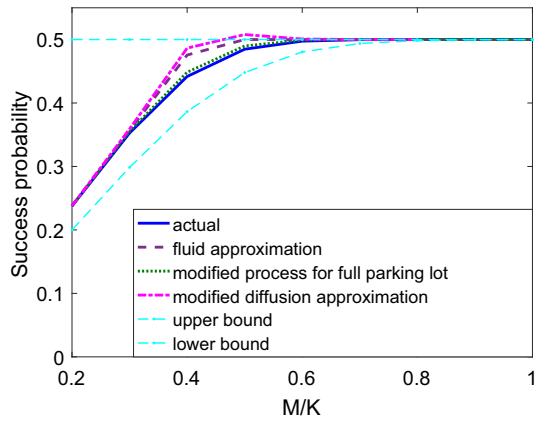


Fig. 13 $K = 30$ and $\lambda = K$

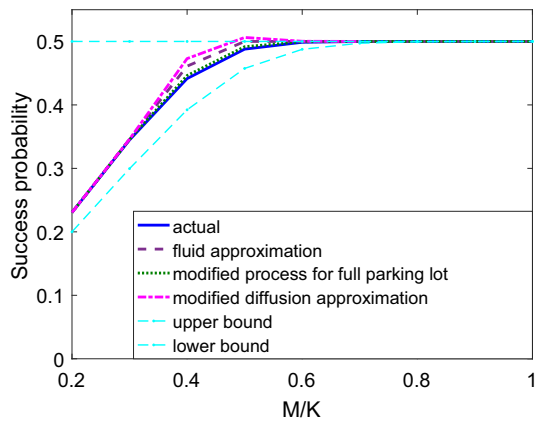


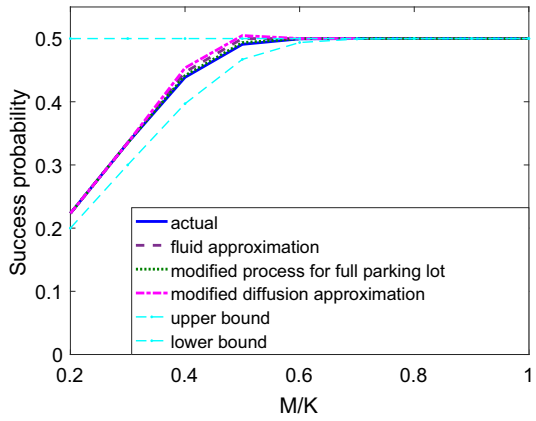
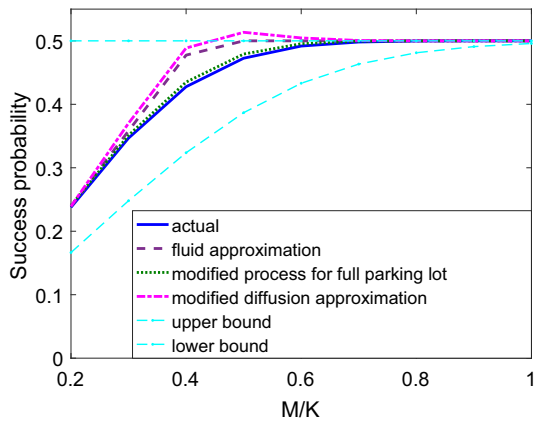
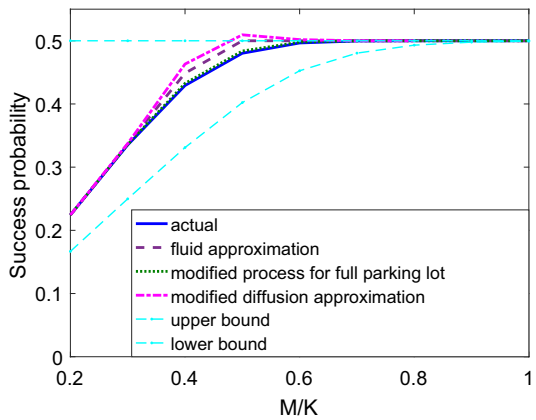
Fig. 14 $K = 50$ and $\lambda = K$ Fig. 15 $K = 10$ and $\lambda = 1.2K$ Fig. 16 $K = 20$ and $\lambda = 1.2K$ 

Fig. 17 $K = 30$ and $\lambda = 1.2K$

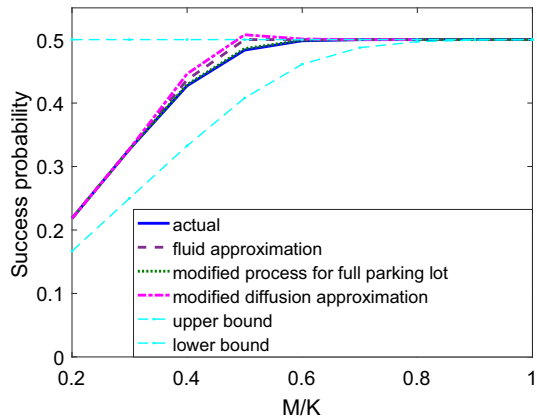
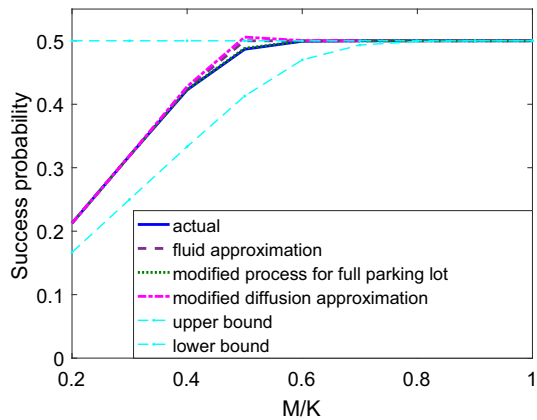


Fig. 18 $K = 50$ and $\lambda = 1.2K$



6 Proofs

6.1 Proofs for Sect. 3

Proof of Proposition 3.1 First, we show the upper bound. Recall that the probability (in stationarity) that an EV leaves with a full battery is given by $P_s = \frac{\mathbb{E}[C_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]}$.

As the parking times do not depend on M , we have that $\mathbb{E}[Q_M^K(\infty)] = \mathbb{E}[Q_K^K(\infty)]$ and hence the upper bound would follow from $\mathbb{E}[C_M^K(\infty)] \leq \mathbb{E}[C_K^K(\infty)]$, which is equivalent to $\mathbb{E}[Z_M^K(\infty)] \geq \mathbb{E}[Z_K^K(\infty)]$ (as the total number of EVs remains the same for both systems). We now proceed to show the last inequality using coupling arguments. Fix a sample path $\omega \in \Omega$. Assume for simplicity that $Z_M^K(0) = Z_K^K(0) \leq M$, and take identical arrival, charging, and parking times for both systems. Note that the invariant distributions do not depend on the initial conditions, and so what we assume on the initial conditions is without loss of generality. Further, define $T := \inf\{t \geq 0 : Z_M^K(t) = M\}$. Note that the arrival process is the same for both systems as

the total number of EVs remains the same for both systems. That is, $Z_M^K(t) = Z_K^K(t)$ for $t \leq T$ since the charging rate is the same for both systems. Moreover, using the fact that $Z_M^K(t) \leq K$ and $M \leq K$, the following inequality holds for any $t > T$:

$$\frac{Z_M^K(t) \wedge M}{Z_M^K(t)} \leq \frac{Z_M^K(t) \wedge K}{Z_M^K(t)} = 1 = \frac{Z_K^K(t) \wedge K}{Z_K^K(t)}.$$

That is, the charging rate in the system (K, M) is always bounded above by the charging rate in the system (K, K) (the latter is always constant and it is equal to one). That is, a departed customer in the system (K, M) has already departed in the system (K, K) . In other words, $Z_M^K(t) \geq Z_K^K(t)$ for any $t \geq 0$, and removing the conditioning on the sample path ω , we derive $Z_M^K(t) \geq_{st} Z_K^K(t)$. By the existence of the stationary distribution, we obtain $Z_M^K(\infty) \geq_{st} Z_K^K(\infty)$ and hence $\mathbb{E}[Z_M^K(\infty)] \geq \mathbb{E}[Z_K^K(\infty)]$ which proves the upper bound in (3.2).

We move now to the proof of the lower bound in (3.2). First, we show that $Z_M^\infty(\infty) \geq_{st} Z_M^K(\infty)$ by using coupling arguments. Fix a sample path $\omega \in \Omega$. Assume that $Z_M^\infty(0) = Z_M^K(0)$, and take identical arrival, charging, and parking times for both systems. Define $T^* = \inf\{t > 0 : Q_M^K(t) = K\}$. It follows that $Z_M^\infty(t) = Z_M^K(t)$ for $t \leq T^*$. After time T^* , the blocked arrivals in the loss queue will enter in the queue with infinitely many parking spaces. That is, $Z_M^\infty(t) \geq Z_M^K(t)$ for all $t \geq 0$. Removing now the conditioning on the sample path ω , we derive $Z_M^\infty(t) \geq_{st} Z_M^K(t)$ for all $t \geq 0$, and by the existence of the stationary distribution, we have that $Z_M^\infty(\infty) \geq_{st} Z_M^K(\infty)$. Hence, we have the following inequality for the sojourn time of an uncharged EV in the system (in stationarity): $S_M^\infty(B) \geq_{st} S_M^K(B)$. That is, $\mathbb{E}[S_M^K(B)] \leq \mathbb{E}[S_M^\infty(B)]$.

By Little's law, we obtain that $\mathbb{E}[S_M^K(B)] = \frac{\mathbb{E}[Z_M^K(\infty)]}{\lambda(1-B(\lambda/v))} = \frac{1}{v} \frac{\mathbb{E}[Z_M^K(\infty)]}{\mathbb{E}[Q_M^K(\infty)]}$ and $\mathbb{E}[S_M^\infty(B)] = \frac{\mathbb{E}[Z_M^\infty(\infty)]}{\lambda} = \frac{1}{v} \frac{\mathbb{E}[Z_M^\infty(\infty)]}{\mathbb{E}[Q_M^\infty(\infty)]}$. Hence, the lower bound in (3.2) follows.

It remains to show (3.3). Let $(Q_\lambda(\cdot), Z_\lambda(\cdot))$ denote the total number of EVs and the number of uncharged EVs if the arrival rate is λ . First, using coupling arguments, we prove that if $\lambda_1 \leq \lambda_2$, then $Q_{\lambda_1}(t) \leq_{st} Q_{\lambda_2}(t)$ and $Z_{\lambda_1}(t) \leq_{st} Z_{\lambda_2}(t)$ for any $t \geq 0$. Assume the following coupling: If an arrival occurs to the system with arrival rate λ_1 , it also occurs to system with arrival rate λ_2 . Hence, as $\lambda_1 \leq \lambda_2$, there are more arrivals in the second system. Further, we assume that all other parameters, i.e., μ, v, M, K , are equal in both systems. Assume that both systems start empty and define $T^{**} = \inf\{t > 0 : Q_{\lambda_2}(t) = K\}$. As in the second system there are more arrivals, we have that $Q_{\lambda_1}(t) \leq Q_{\lambda_2}(t)$ and $Z_{\lambda_1}(t) \leq Z_{\lambda_2}(t)$, for $t \leq T^{**}$. By the Markovian assumptions, we have that the residual charging and parking times are exponential with rates μ and v . That is, at any new event after time T^{**} , we can resample the charging and parking times and hence the probability of a departure in the system with arrival rate λ_1 is higher or equal to the probability of a departure in the system with arrival rate λ_2 . In other words, for $t \geq 0$ and for $x > 0$, $\mathbb{P}(Q_{\lambda_2}(t) \leq x) \leq \mathbb{P}(Q_{\lambda_1}(t) \leq x)$ and $\mathbb{P}(Z_{\lambda_2}(t) \leq x) \leq \mathbb{P}(Z_{\lambda_1}(t) \leq x)$. The last relation is equivalent to $Q_{\lambda_1}(t) \leq_{st} Q_{\lambda_2}(t)$ and $Z_{\lambda_1}(t) \leq_{st} Z_{\lambda_2}(t)$, for $t \geq 0$. In the sequel, we see that $Z_f(\cdot)$ can arise as the limit of $Z_\lambda(\cdot)$ as $\lambda \rightarrow \infty$, assuming that $Q_\lambda(0) \xrightarrow{d} K$ and $Z_\lambda(0) \xrightarrow{d} Z_f(0)$. To see

this, first observe that $Q_\lambda(\cdot) \xrightarrow{d} K$ as $\lambda \rightarrow \infty$. Now, combining (2.1) and (2.3), we have that

$$\begin{aligned} Z_\lambda(t) &= Z_\lambda(0) - Q_\lambda(0) + Q_\lambda(t) + N_{v,2} \left(\int_0^t (Q_\lambda(s) - Z_\lambda(s)) ds \right) \\ &\quad - N_\mu \left(\int_0^t Z_\lambda(s) \wedge M ds \right). \end{aligned}$$

Taking $\lambda \rightarrow \infty$ and using the continuous mapping theorem, we have that $Z_\lambda(\cdot) \xrightarrow{d} Z_\infty(\cdot)$, where

$$\begin{aligned} Z_\infty(t) &= Z_f(0) + N_{v,2} \left(\int_0^t (K - Z_\infty(s)) ds \right) - N_\mu \left(\int_0^t Z_\infty(s) \wedge M ds \right) \\ &\stackrel{d}{=} Z_f(t), \end{aligned}$$

and where the last equality follows by (4.11). Furthermore, $Z_\lambda(\cdot)$ is nondecreasing. That is, $Z_\lambda(t) \leq_{st} Z_f(t)$ for any $t \geq 0$ and, by the existence of the stationary distributions, we obtain $Z_\lambda(\infty) \leq_{st} Z_f(\infty)$. By the last inequality, it follows that $\mathbb{E}[Z_\lambda(\infty)] \leq \mathbb{E}[Z_f(\infty)]$, and hence (3.3). \square

Proof of Proposition 3.2 Note that the distribution $p_Q(q)$ corresponds to the stationary distribution of a one-dimensional Erlang loss system. Furthermore, by [29, Sect. 1.3], we know that

$$p_Q(q) = \frac{1}{q!} \left(\frac{\lambda}{v} \right)^q p_Q(0), \text{ where } p_Q(0) = \left(\sum_{i=0}^K \frac{1}{i!} \left(\frac{\lambda}{v} \right)^i \right)^{-1}.$$

Thus, the probability of an empty system is $p_e(0, 0) = p_Q(0)$.

As it is well known that a solution of the balance equations of a Markov process is unique, we shall show that $p_e(q, z)$ for $z \leq q$ satisfies the flow balance equation (3.1). Then, the proof of the proposition is completed. First, we note the relations between $p_e(q + a, z + b)$ and $p_e(q, z)$ for $a, b \in \{-1, 0, 1\}$. By (3.5), we obtain that

$$p_e(q) = \frac{1}{q(q-1)!} \frac{\lambda}{v} \left(\frac{\lambda}{v} \right)^{q-1} p_Q(0) = \frac{1}{q} \frac{\lambda}{v} p_e(q-1). \quad (6.1)$$

Now, applying the previous relation in (3.4), we have that

$$\begin{aligned} p_e(q-1, z-1) &= q \frac{v}{\lambda} p_Q(q) \frac{(q-1)!}{(z-1)!(q-z)!} \left(\frac{\mu}{v+\mu} \right)^{q-z} \left(\frac{v}{v+\mu} \right)^{z-1} \\ &= z \frac{v}{\lambda} \frac{v+\mu}{v} p_Q(q) \frac{q!}{z!(q-z)!} \left(\frac{\mu}{v+\mu} \right)^{q-z} \left(\frac{v}{v+\mu} \right)^z \\ &= z \frac{v+\mu}{\lambda} p_e(q, z). \end{aligned}$$

Working analogously, we derive the following relations:

$$p_e(q-1, z-1) = z \frac{v+\mu}{\lambda} p_e(q, z), \quad (6.2)$$

$$p_e(q+1, z+1) = \frac{1}{z+1} \frac{\lambda}{v+\mu} p_e(q, z), \quad (6.3)$$

$$p_e(q, z+1) = \frac{q-z}{z+1} \frac{v}{\mu} p_e(q, z), \quad (6.4)$$

$$p_e(q+1, z) = \frac{1}{q-z+1} \frac{\lambda}{v} \frac{\mu}{v+\mu} p_e(q, z). \quad (6.5)$$

Using the above equations and recalling that $L(z) = z$ when $M = K$, the right-hand side of (3.1) for $0 < z, q < K$ and $z \neq q$ can be written as follows:

$$\begin{aligned} & \left(z(v+\mu) + \frac{v(z+1)}{z+1} \frac{\lambda}{v+\mu} + v(z+1) \frac{q-z}{z+1} + \frac{(q-z+1)}{q-z+1} \frac{\lambda\mu}{v+\mu} \right) p_e(q, z) \\ &= \left(z(v+\mu) + \frac{\lambda v}{v+\mu} + (q-z)v + \frac{\lambda\mu}{v+\mu} \right) p_e(q, z) = (qv + \lambda + z\mu) p_e(q, z). \end{aligned}$$

That is, $p_e(q, z)$ satisfies (3.1) for $0 < z, q < K$ and $z \neq q$. To show that $p_e(q, z)$ satisfies (3.1) for $0 < q < K$ and $z = 0$, we apply (3.4) and (6.1) in the right-hand side of (3.1). This leads to

$$\begin{aligned} & (q+1)v p_e(q+1) \left(\frac{\mu}{v+\mu} \right)^{q+1} + v p_e(q+1)(q+1) \left(\frac{\mu}{v+\mu} \right)^q \frac{v}{v+\mu} \\ &+ \mu p_e(q)q \left(\frac{\mu}{v+\mu} \right)^{q-1} \frac{v}{v+\mu} \\ &= \left(\lambda \left(\frac{\mu}{v+\mu} \right)^{q+1} + \lambda \left(\frac{\mu}{v+\mu} \right)^q \frac{v}{v+\mu} + \mu q \left(\frac{\mu}{v+\mu} \right)^{q-1} \frac{v}{v+\mu} \right) p_Q(q) \\ &= \left(\lambda + \mu q \frac{v+\mu}{\mu} \frac{v}{v+\mu} \right) \left(\frac{\mu}{v+\mu} \right)^q p_Q(q) = (\lambda + qv) p_Q(q, 0). \end{aligned}$$

In the same way, we show that the right-hand side of (3.1) for $0 < z < K$ and $q = K$ becomes

$$\begin{aligned} & \left(\lambda K \frac{v}{\lambda} \frac{(K-1)!}{(z-1)!(K-z)!} \left(\frac{\mu}{v+\mu} \right)^{K-z} \left(\frac{v}{v+\mu} \right)^{z-1} \right. \\ & \left. + \mu(z+1) \frac{K!}{(z+1)!(K-z-1)!} \left(\frac{\mu}{v+\mu} \right)^{K-z-1} \left(\frac{v}{v+\mu} \right)^{z+1} \right) p_Q(K). \end{aligned}$$

The last quantity is equal to

$$\left(z v \frac{v + \mu}{v} + (K - z) \mu \frac{v}{v + \mu} \frac{v + \mu}{\mu} \right) \frac{K!}{z!(K - z)!} \left(\frac{\mu}{v + \mu} \right)^{K - z} \left(\frac{v}{v + \mu} \right)^z p_Q(K) \\ = (Kv + z\mu) p_e(K, z).$$

Using again relations (6.2), (6.3), and (6.5), the right-hand side of (3.1) is written, for $q < K$ and $q = z$, as follows:

$$\left(\lambda q \frac{v + \mu}{\lambda} + \frac{v\lambda}{v + \mu} + \frac{\lambda\mu}{v + \mu} \right) p_e(q, q) = \left(q(v + \mu) + \frac{\lambda v}{v + \mu} + \frac{\lambda\mu}{v + \mu} \right) p_e(q, q) \\ = (q(v + \mu) + \lambda) p_e(q, q).$$

Using again relations (6.2)–(6.5), it follows immediately that $p_e(q, z)$ satisfies Eq. (3.1) also for the remaining cases, i.e., $(q, z) = (0, 0)$, $(q, z) = (K, K)$, and finally $(q, z) = (K, 0)$. \square

Proof of Proposition 3.3 We show that $Z_M^\infty(\cdot)$ behaves as a modified Erlang-A queue. Although we adapt the proof in [49, Sect. 6.6.1], we briefly describe it here for completeness. First, we write the flow balance equations for the Markov process $Z_M^\infty(\cdot)$ and then we solve them. The balance equations for the Markov process $Z_M^\infty(\cdot)$ are given by

$$\begin{cases} (\lambda + z(v + \mu))p_Z(z) = \lambda p_Z(z - 1) + (z + 1)(v + \mu)p_Z(z + 1), \\ (\lambda + M\mu + zv)p_Z(z) = \lambda p_Z(z - 1) + (M\mu + (z + 1)v)p_Z(z + 1), \end{cases} \quad (6.6)$$

for $0 < z < M$ and $z \geq M$, respectively. For $z = 0$, we have that

$$\lambda p_Z(0) = (v + \mu)p_Z(1).$$

Using the last equation and (6.6), we derive inductively the following relations:

$$\lambda p_Z(z - 1) = z(v + \mu)p_Z(z), \quad \text{if } z < M,$$

and

$$(M\mu + zv)p_Z(z) = \lambda p_Z(z - 1), \quad \text{if } z \geq M.$$

The balance equations now can be simplified as follows:

$$\begin{cases} \lambda p_Z(z) = (z + 1)(v + \mu)p_Z(z + 1), & \text{if } z < M, \\ \lambda p_Z(z) = (M\mu + (z + 1)v)p_Z(z + 1), & \text{if } z \geq M. \end{cases} \quad (6.7)$$

Observe that we can directly solve the system (6.7). For $z < M$, it is easy to see that

$$p_Z(z) = \frac{1}{z!} \left(\frac{\lambda}{v + \mu} \right)^z p_Z(0). \quad (6.8)$$

We show that, for $z = M$, the solution of (6.7) is also given by the last formula. By the first equation of (6.7) for $z = M - 1$ and (6.8), we obtain the following equation:

$$p_Z(M) = \frac{1}{M} \left(\frac{\lambda}{v + \mu} \right)^M p_Z(M - 1) = \frac{1}{M!} \left(\frac{\lambda}{v + \mu} \right)^M p_Z(0).$$

It remains to find the solution in the case $z > M$. We do so by induction. Note that, by the second equation of (6.7) for $z = M$, we have that

$$p_Z(M + 1) = \frac{\lambda p_Z(M)}{M\mu + (M + 1)v} = \frac{\lambda}{M\mu + (M + 1)v} \frac{1}{M!} \left(\frac{\lambda}{v + \mu} \right)^M p_Z(0).$$

Finally, it is easy to verify that the solution of (6.7) for $z > M$ is given by

$$p_Z(z) = \frac{1}{M!} \left(\frac{\lambda}{v + \mu} \right)^M \prod_{k=M+1}^z \frac{\lambda}{M\mu + kv} p_Z(0).$$

The probability of an empty system (there are not uncharged vehicles in the parking lot) can be found by the normalization condition and it is given by (3.3). Finally, we show that the infinite summation in (3.3) converges. To this end, note that $L(z)\mu + zv \geq z \min\{v, \mu\}$. Applying the last observation in (3.3), we have that

$$\begin{aligned} & \sum_{j=0}^M \frac{1}{j!} \left(\frac{\lambda}{v + \mu} \right)^j + \sum_{j=M+1}^{\infty} \frac{1}{M!} \left(\frac{\lambda}{v + \mu} \right)^M \prod_{k=M+1}^j \frac{\lambda}{M\mu + kv} \\ & \leq \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\lambda}{\min\{v, \mu\}} \right)^j = \exp \left\{ \frac{\lambda}{\min\{v, \mu\}} \right\}. \end{aligned}$$

□

Proof of Proposition 3.4 First, we write the balance equations for the one-dimensional birth–death process $\{Z_f(t), t \geq 0\}$. These are given by

$$\begin{cases} (v(K - z) + \mu z) \pi_f(z) = (v(K - z + 1) \pi_f(z - 1)) + \mu(z + 1) \pi_f(z + 1), \\ (v(K - z) + \mu M) \pi_f(z) = (v(K - z + 1) \pi_f(z - 1)) + \mu M \pi_f(z + 1), \end{cases}$$

for $0 < z < M$ and $M \leq z < K$, respectively. On the boundary, the following equations hold:

$$vK \pi_f(0) = \mu \pi_f(1) \text{ and } \mu M \pi_f(K) = v \pi_f(K - 1).$$

Note that the balance equations can be simplified to

$$\begin{cases} v(K - z)\pi_f(z) = \mu(z + 1)\pi_f(z + 1), & \text{if } 0 \leq z < M, \\ \mu M\pi_f(z) = v(K - z + 1)\pi_f(z - 1), & \text{if } M \leq z \leq K. \end{cases} \quad (6.9)$$

Applying $z = M - 1$ in the first equation of (6.9), we obtain

$$\pi_f(M - 1) = \frac{\mu M}{v(K - M + 1)} p_Z(M),$$

and recursively we have that

$$\pi_f(M - i) = \left(\frac{\mu}{v}\right)^i \frac{\prod_{j=0}^{i-1} (M - j)}{\prod_{j=1}^i (K - M + j)} \pi_f(M), \text{ if } 0 < i \leq M.$$

Changing the variable $z = M - i$ in the last equation yields

$$\pi_f(z) = \left(\frac{\mu}{v}\right)^{M-z} \frac{\prod_{j=0}^{M-z-1} (M - j)}{\prod_{j=1}^{M-z} (K - M - j)} \pi_f(M), \text{ if } 0 \leq z < M.$$

Working analogously, by the second equation of (6.9) we derive

$$\pi_f(M + i) = \frac{1}{M^i} \left(\frac{\mu}{v}\right)^i \prod_{j=0}^{i-1} (K - M - j) \pi_f(M), \text{ if } 0 \leq i \leq K - M,$$

which leads to

$$\pi_f(z) = \frac{1}{M^{z-M}} \left(\frac{v}{\mu}\right)^{z-M} \prod_{j=0}^{z-M-1} (K - M - j) \pi_f(M), \text{ if } M \leq z \leq K.$$

Finally, $\pi_f(M)$ is determined by the normalization equation $\sum_{z=0}^K \pi_f(z) = 1$. \square

6.2 Proofs for Sect. 4.1

Proof of Proposition 4.1 In this proof, we use martingale arguments. Define the filtration $\mathcal{F}_t^n := \sigma(Z^n(0), Q^n(0), Z^n(s), Q^n(s) : 0 \leq s \leq t)$ augmented by including all the null sets for $t \geq 0$ and $n \geq 1$. Applying the fluid scaling to the dynamical equation (2.3), we have that

$$\begin{aligned} Z^n(t) &= Z^n(0) + N_\lambda \left(n \int_0^t \mathbb{1}_{\{Q^n(s) < K\}} ds \right) - N_\mu \left(n \int_0^t \frac{Z^n(s)}{n} \wedge M ds \right) \\ &\quad - N_{v,1} \left(n \int_0^t \frac{Z^n(s)}{n} ds \right). \end{aligned}$$

Defining the operator $\bar{\mathcal{M}}_r^n = \frac{1}{n} (N_r(\cdot) - r \cdot)$ and following [32], we can write

$$\begin{aligned} \frac{Z^n(t)}{n} &= \frac{Z^n(0)}{n} + \bar{\mathcal{M}}_{\lambda}^n(nt) - \bar{\mathcal{M}}_{\mu}^n \left(n \int_0^t \frac{Z^n(s)}{n} \wedge M ds \right) \\ &\quad - \bar{\mathcal{M}}_{v,1}^n \left(n \int_0^t \frac{Z^n(s)}{n} ds \right) - \frac{1}{n} \int_0^t \mathbb{1}_{\{\frac{Q^n(s)}{n}=K\}} dN(\lambda ns) + \lambda t \\ &\quad - \mu \int_0^t \frac{Z^n(s)}{n} \wedge M ds - v \int_0^t \frac{Z^n(s)}{n} ds. \end{aligned}$$

The term $\frac{1}{n} \int_0^t \mathbb{1}_{\{\frac{Q^n(s)}{n}=K\}} dN(\lambda ns)$ denotes the number of EVs that are lost due to finding the system full under the fluid scaling. By [26, Proposition 4.11], this is a weakly convergent sequence. Further, by [35, Propositions 6.17 and 6.21] and by the assumption $\frac{Q^n(0)}{n} \xrightarrow{d} K$, we have that $\frac{Q^n(\cdot)}{n} \xrightarrow{d} q(\cdot) \equiv K$ if $\lambda/\nu \geq K$. On the other hand, if $\lambda/\nu < K$, we have that $q(t) < K$ for $t > 0$. Now, by the proof of [26, Theorem 3.6] (see [26, Eqs. (3.34) and (3.40)]), it turns out that

$$\frac{1}{n} \int_0^t \mathbb{1}_{\{\frac{Q^n(s)}{n}=K\}} dN(\lambda ns) \xrightarrow{d} \max\{\lambda - \nu K, 0\}t.$$

Further, observing that $\bar{\mathcal{M}}_r^n(\cdot)$ are zero mean martingales with respect to the filtration \mathcal{F}_t^n for any $n \in \mathbb{N}$ (cf. [32]) and taking $n \rightarrow \infty$, we derive that $\frac{Z^n(\cdot)}{n} \xrightarrow{d} z(\cdot)$. Moreover, the limit function is characterized by the following functional equation:

$$z(t) = z(0) + \lambda t - \max\{\lambda - \nu K, 0\}t - \mu \int_0^t z(s) \wedge M ds - \nu \int_0^t z(s) ds,$$

which is equivalent to (4.1). \square

Proof of Proposition 4.2 It is not hard to solve the ODE (4.1) explicitly in the two regions, namely

$$z(t) = \begin{cases} \frac{\lambda \wedge \nu K}{\nu + \mu} + \left(z(0) - \frac{\lambda \wedge \nu K}{\nu + \mu} \right) e^{-(\nu + \mu)t}, & \text{if } z(t) \leq M, \\ \frac{\lambda \wedge \nu K - \mu M}{\nu} + \left(z(0) - \frac{\lambda \wedge \nu K - \mu M}{\nu} \right) e^{-\nu t}, & \text{otherwise.} \end{cases} \quad (6.10)$$

Define $z_1 := \frac{\lambda \wedge \nu K}{\nu + \mu}$ and $z_2 := \frac{\lambda \wedge \nu K - \mu M}{\nu}$. First, note that for a given initial state $z(0) \in [0, K]$, we have that $z(t) \leq K$ for any $t \geq 0$. To see this, observe that if $z(t) \leq M$, then by the model assumptions $z(t) \leq M \leq K$. On the other hand, if $z(t) > M$, we have that

$$z(t) = z_2 + (z(0) - z_2) e^{-\nu t} \leq z_2 \leq K,$$

if $z(0) - z_2 \leq 0$ and

$$z(t) = z_2 + (z(0) - z_2) e^{-\nu t} \leq z_2 + z(0) - z_2 = z(0) \leq K,$$

if $z(0) - z_2 \geq 0$. So, for any $t \geq 0$, it follows that $z(t) \leq K$ and hence Definition 4.1 is well defined.

In the sequel, we show that $z(t)$ converges as t goes to infinity to the following point:

$$z = \begin{cases} z_1, & \text{if } z_1 \leq M, \\ z_2, & \text{otherwise.} \end{cases} \quad (6.11)$$

Then, we show that z is unique and, using the Markovian assumptions, we see that it is equivalent to (4.2), and hence $z = z^*$. Also, observe that z^* is an invariant point. Indeed, if we set $z(0) = z^*$, then by (6.10) we have that $z(t) = z^*$ for $t \geq 0$.

First, assume that $z_1 \leq M$. If $z(0) \leq M$, then $z(t) \leq M$ for any $t \geq 0$. To see this, note that if $z(0) - z_1 \leq 0$ then we have that

$$z_1 + (z(0) - z_1) e^{-(\nu+\mu)t} \leq z_1 \leq M.$$

On the other hand, if the quantity $z(0) - z_1$ is positive, we show that there does not exist $t^* > 0$ such that $z(t^*) > M$. Supposing that there exists t^* such that $z(t^*) > M$, we have that

$$z_1 + (z(0) - z_1) e^{-(\nu+\mu)t^*} > M;$$

by assumption $z(0) - z_1 > 0$, so the last inequality leads to

$$t^* < -\frac{1}{\nu + \mu} \ln \left(\frac{M - z_1}{z(0) - z_1} \right).$$

Now, by the assumption $z(0) \leq M$, we obtain $\frac{M-z_1}{z(0)-z_1} \geq 1$ and hence $t^* \leq 0$, which yields a contradiction. Next, we assume that $z(0) > M$. In this case, we show that there exists t_2^* such that $z(t_2^*) \leq M$. First, observe that if $z_1 \leq M$ then $z_2 \leq M$. Hence, $z(0) - z_2 > M - z_2 \geq 0$. Now, we note that

$$z_2 + (z(0) - z_2) e^{-\nu t} \leq M$$

leads to

$$t > -\frac{1}{\nu} \ln \left(\frac{M - z_2}{z(0) - z_2} \right) > 0,$$

due to $z(0) - z_2 > M - z_2$. Let t_2^* be the first time that $z(t) \leq M$ starting from a point $z(0) > M$. Setting now as initial point $z(t_2^*)$, we conclude that if $z_1 \leq M$ then $z(t) \leq M$ for $t \geq t_2^*$ and hence

$$z(t) = z_1 + (z(t_2^*) - z_1) e^{-(\nu+\mu)t}, \text{ for } t \geq t_2^*,$$

which yields

$$|z(t) - z_1| \leq (z(t_2^*) - z_1) e^{-(\nu+\mu)t} \leq (z(0) - z_1) e^{-(\nu+\mu)t}, \text{ for } t \geq t_2^*.$$

This concludes the proof for the case $z_1 \leq M$. The case $z_2 > M$ follows the same logic.

Now, we prove that (4.2) is equal to z . To this end, observe that

$$B \max \left\{ 1, \frac{z^*}{M} \right\} \stackrel{d}{=} B',$$

where B' is an exponential random variable with $\mathbb{E}[B'] = \frac{1}{\mu \max \left\{ 1, \frac{z^*}{M} \right\}}$. We note that (4.2) can be written as

$$\begin{aligned} z^* &= (\lambda \wedge \nu K) \mathbb{E} \left[\min \left\{ D, B \max \left\{ 1, \frac{z^*}{M} \right\} \right\} \right] = (\lambda \wedge \nu K) \mathbb{E} [\min \{ D, B' \}] \\ &= (\lambda \wedge \nu K) \frac{1}{\nu + \mu \max \left\{ 1, \frac{z^*}{M} \right\}}. \end{aligned}$$

Solving the last equation yields $z^* = z$.

To conclude the proof of Proposition 4.2, it remains to show the uniqueness of the invariant point z^* . In other words, we show that (6.11) [and hence (4.2)] has a unique solution. It is not hard to see that if $z_1 < M$, then $z_2 < M$. So, z_2 cannot be the solution of (6.11). That is, z_1 is the unique solution of (6.11). On the other hand, if $z_1 > M$ [i.e., it is not solution of (6.11)], then $z_2 > M$. That is, z_2 is the unique solution of (6.11). Finally, if $z_1 = M$, then we have that $z_2 = z_1 = M$. In any case, (6.11) has a unique solution. \square

Proof of Proposition 4.3 Let $\frac{Z^n(\infty)}{n}$ be the stationary fluid-scaled number of uncharged EVs. We know that $0 \leq Z^n(\infty) \leq K^n$, which yields $\frac{Z^n(\infty)}{n} \leq K$ almost surely. In other words, the sequence of random variables $\frac{Z^n(\infty)}{n}$ is stochastically bounded in \mathbb{R} and hence it is tight. Now, we consider the process $\{Z^n(t), t \geq 0\}$ starting at point $Z^n(\infty)$. That is, $Z^n(t) \stackrel{d}{=} Z^n(\infty)$ for any $t \geq 0$. Since $\frac{Z^n(\infty)}{n}$ is tight for any convergent subsequence, there exists a further subsequence, say $\frac{Z^{\bar{n}}(\infty)}{\bar{n}}$, such that $\frac{Z^{\bar{n}}(\infty)}{\bar{n}} \xrightarrow{d} \bar{z}^*$, as $\bar{n} \rightarrow \infty$. We now have that, for any $t \geq 0$,

$$\frac{Z^{\bar{n}}(t)}{\bar{n}} \stackrel{d}{=} \frac{Z^{\bar{n}}(\infty)}{\bar{n}} \xrightarrow{d} \bar{z}^*, \text{ as } \bar{n} \rightarrow \infty,$$

and so \bar{z}^* is an invariant point. By the uniqueness of the invariant point, we derive $\bar{z}^* = z^*$. This concludes the proof. \square

6.3 Proofs for Sect. 4.2

6.3.1 Proof of Theorem 4.5

We start the analysis by establishing a continuity result, which can be proved by using results in [32].

Proposition 6.1 *Let $t \geq 0$ and $-\infty < \kappa < \infty$. Consider the following system:*

$$\begin{aligned} x_1(t) &= b_1 + g_1(t) + \int_0^t h_1(x_1(s))ds - y(t), \\ x_2(t) &= b_2 + g_2(t) + \int_0^t h_2(x_2(s))ds - y(t), \end{aligned} \quad (6.12)$$

where b_i are positive constants, $h_i : \mathbb{R} \rightarrow \mathbb{R}$ satisfy $h_i(0) = 0$ and are Lipschitz continuous functions for $i = 1, 2$, and $x_2(t) \leq \kappa$. In addition, $y(\cdot)$ is a nondecreasing nonnegative function in $D[0, \infty)$ such that (6.12) holds and $\int_0^\infty \mathbb{1}_{\{x_2(t) < \kappa\}} dy(t) = 0$. Given $b_i \in \mathbb{R}$ and $g_i(\cdot) \in D[0, \infty)$, we have that the system (6.12) has a unique solution $(x_1(\cdot), x_2(\cdot), y(\cdot))$. Moreover, the functions $(x_1(\cdot), x_2(\cdot), y(\cdot))$ are continuous in $D[0, \infty)^3$ if $D[0, \infty)$ is endowed with the uniform topology over bounded intervals or the J_1 topology.

Proof of Proposition 6.1 First, observe that the function $y(\cdot)$ is independent of the function $x_1(\cdot)$. We know by [32, Theorem 7.3] that the second equation of (6.12) has a unique solution $(x_2(\cdot), y(\cdot))$ and that $x_2(\cdot)$ and $y(\cdot)$ are continuous in $D[0, \infty)$ (endowed with the uniform topology over bounded intervals or the J_1 topology). Furthermore, we have that $y(\cdot), g_2(\cdot) \in D[0, \infty)$ which implies $y(\cdot) + g_2(\cdot) \in D[0, \infty)$. The last observation together with [32, Theorem 4.1] implies that the first equation of (6.12) has a unique continuous solution. That is, system (6.12) has a unique solution $(x_1(\cdot), x_2(\cdot), y(\cdot))$ and each function is continuous. \square

In order to continue our analysis, we need to define appropriate filtrations. Take the following filtrations, for $n \geq 1$:

$$\begin{aligned} \mathcal{F}_{t,1}^n &= \sigma \left(Z^n(0), N_{\lambda^n}(s), N_\mu \left(\int_0^s L^n(Z^n(z))dz \right), N_{v,1} \left(\int_0^s Z^n(z)dz \right) \right. \\ &\quad \left. : 0 \leq s \leq t \right) \end{aligned}$$

and

$$\mathcal{F}_{t,2}^n = \sigma \left(Q^n(0), N_{\lambda^n}(s), N_v \left(\int_0^s Q^n(z)dz \right) : 0 \leq s \leq t \right).$$

In the sequel, we work with the filtrations

$$\mathcal{F}_t^n = \sigma \left(\mathcal{F}_{t,1}^n, \mathcal{F}_{t,2}^n \right),$$

augmented by including all the null sets for $t \geq 0$ and $n \geq 1$.

Now, notice that the system dynamics (2.1) and (2.3) can be rewritten in the following form:

$$Q^n(t) = Q^n(0) + N_{\lambda^n}(t) - N_v \left(\int_0^t Q^n(s) ds \right) - Y^n(t),$$

and

$$\begin{aligned} Z^n(t) &= Z^n(0) + N_{\lambda^n}(t) - N_\mu \left(\int_0^t L^n(Z^n(s)) ds \right) - N_{v,1} \left(\int_0^t Z^n(s) ds \right) \\ &\quad - Y^n(t), \end{aligned}$$

where $Y^n(t) = \int_0^t \mathbb{1}_{\{Q^n(s)=K^n\}} dN_{\lambda^n}(s)$. The process $Y^n(t)$ counts all the customers that are lost when all the servers (chargers) are busy up to time t in the n^{th} system. Defining the operator $\mathcal{M}_r(\cdot) := N_r(\cdot) - (r \cdot)$, where “ r ” indicates the rate of the Poisson process $N_r(\cdot)$, the system dynamics take the following form:

$$\begin{aligned} Q^n(t) &= Q^n(0) + \mathcal{M}_{\lambda^n}(t) - \mathcal{M}_v \left(\int_0^t Q^n(s) ds \right) + \lambda^n t - v \int_0^t Q^n(s) ds \\ &\quad - Y^n(t), \end{aligned} \quad (6.13)$$

and

$$\begin{aligned} Z^n(t) &= Z^n(0) + \mathcal{M}_{\lambda^n}(t) - \mathcal{M}_\mu \left(\int_0^t L^n(Z^n(s)) ds \right) - \mathcal{M}_{v,1} \left(\int_0^t Z^n(s) ds \right) \\ &\quad + \lambda^n t - \mu \int_0^t L^n(Z^n(s)) ds - v \int_0^t Z^n(s) ds - Y^n(t). \end{aligned} \quad (6.14)$$

In order to derive appropriate equations (in the prelimit) for the diffusion-scaled processes, subtract and add the terms $n \frac{v+\mu}{v}$, $n \frac{v+\mu}{v} t$ in (6.13) and the terms n , $n\mu t$, nvt in (6.14), and then divide both by \sqrt{n} . Recalling that $L^n(Z^n(t)) = Z^n(t) \wedge M^n$, $\lambda^n = n(v + \mu)$, $M^n = \frac{\lambda^n}{v+\mu} + \beta\sqrt{n}$ and observing that $M^n - n = \beta\sqrt{n}$ and $\lambda^n t - n(v + \mu)t = 0$, we obtain the following equations for the diffusion-scaled processes $\hat{Q}^n(\cdot)$ and $\hat{Z}^n(\cdot)$:

$$\begin{aligned} \hat{Q}^n(t) &= \hat{Q}^n(0) + \hat{\mathcal{M}}_{\lambda^n}^n(t) - \hat{\mathcal{M}}_v^n \left(\int_0^t Q^n(s) ds \right) - v \int_0^t \hat{Q}^n(s) ds - \hat{Y}^n(t), \end{aligned} \quad (6.15)$$

and

$$\begin{aligned}\hat{Z}^n(t) = & \hat{Z}^n(0) + \hat{\mathcal{M}}_{\lambda^n}^n(t) - \hat{\mathcal{M}}_{\mu}^n\left(\int_0^t L^n(Z^n(s))ds\right) - \hat{\mathcal{M}}_{v,1}^n\left(\int_0^t Z^n(s)ds\right) \\ & - \mu \int_0^t (\hat{Z}^n(s) \wedge \beta)ds - \nu \int_0^t \hat{Z}^n(s)ds - \hat{Y}^n(t),\end{aligned}\quad (6.16)$$

where $\hat{\mathcal{M}}_r^n(\cdot) := \frac{\mathcal{M}_r(\cdot)}{\sqrt{n}}$ and the scaling for the process $\hat{Y}^n(\cdot)$ is analogous. The following proposition shows that the processes $\hat{\mathcal{M}}_r(\cdot)$ are martingales.

Proposition 6.2 *Under the assumptions $\mathbb{E}[Z^n(0)] < \infty$ and $\mathbb{E}[Q^n(0)] < \infty$, we have that the processes $\hat{\mathcal{M}}_{\lambda^n}^n(\cdot)$, $\hat{\mathcal{M}}_{\mu}^n(\cdot)$, $\hat{\mathcal{M}}_{v,1}^n(\cdot)$, and $\hat{\mathcal{M}}_v^n(\cdot)$ are square-integrable martingales with respect to the filtration $\mathcal{F}^n := \{\mathcal{F}_t^n, t \geq 0\}$. Their associated predictable quadratic variations, denoted by $\langle \cdot \rangle$, are*

$$\langle \hat{\mathcal{M}}_{\lambda^n}^n(t) \rangle = \frac{\lambda^n}{n}t = (\nu + \mu)t, \quad (6.17)$$

$$\left\langle \hat{\mathcal{M}}_{\mu}^n\left(\int_0^t L^n(Z^n(s))ds\right) \right\rangle = \mu \frac{\int_0^t L^n(Z^n(s))ds}{n}, \quad (6.18)$$

$$\left\langle \hat{\mathcal{M}}_{v,1}^n\left(\int_0^t Z^n(s)ds\right) \right\rangle = \nu \frac{\int_0^t Z^n(s)ds}{n}, \quad (6.19)$$

$$\left\langle \hat{\mathcal{M}}_v^n\left(\int_0^t Q^n(s)ds\right) \right\rangle = \nu \frac{\int_0^t Q^n(s)ds}{n}, \quad (6.20)$$

and $\mathbb{E}[\langle \hat{\mathcal{M}}_r^n(t) \rangle] < \infty$, for $t \geq 0$ and $n \geq 1$.

Proof Fix $n \geq 1$. The result for the process $\hat{\mathcal{M}}_{\lambda^n}^n(\cdot)$ follows immediately by applying [32, Lemma 3.1].

Now, note that by the system dynamics (2.1), (2.3), the fact that $Q^n(t) = Z^n(t) + C^n(t)$ and (2.2), we have that

$$\hat{\mathcal{M}}_v^n\left(\int_0^t Q^n(s)ds\right) = \hat{\mathcal{M}}_{v,1}^n\left(\int_0^t Z^n(s)ds\right) + \hat{\mathcal{M}}_{v,2}^n\left(\int_0^t C^n(s)ds\right), \quad (6.21)$$

where

$$\hat{\mathcal{M}}_{v,2}^n\left(\int_0^t C^n(s)ds\right) = \frac{N_{v,2}^n\left(\int_0^t C^n(s)ds\right) - \nu \int_0^t C^n(s)ds}{\sqrt{n}}.$$

Since $N_{v,1}(\cdot)$ and $N_{v,2}(\cdot)$ are independent Poisson processes, [32, Lemma 3.1] implies that $\hat{\mathcal{M}}_{v,i}^n(\cdot)$ are \mathcal{F}^n -martingales for $i = 1, 2$. Observe that $\hat{\mathcal{M}}_{v,2}^n(\cdot)$ is adapted to the filtration \mathcal{F}^n as the latter contains all the information about the processes $Q^n(\cdot)$ and $Z^n(\cdot)$ for fixed n . This is enough to determine the process $C^n(\cdot)$ at any $t \geq 0$. Using (6.21), the assumptions $\mathbb{E}[Z^n(0)], \mathbb{E}[Q^n(0)] < \infty$, the inequalities

$$\begin{aligned} Z^n(t) &\leq Z^n(0) + N_{\lambda^n}(t), \\ Q^n(t) &\leq Q^n(0) + N_{\lambda^n}(t), \end{aligned} \quad (6.22)$$

and adapting the proof in [32, Lemma 3.4], we obtain that, for fixed $n \geq 1$, the following moment conditions are satisfied:

$$\begin{aligned} \mathbb{E} \left[\int_0^t Q^n(s) ds \right] &< \infty, \quad \mathbb{E} \left[\int_0^t Z^n(s) ds \right] < \infty, \\ \mathbb{E} \left[\int_0^t C^n(s) ds \right] &\leq \mathbb{E} \left[\int_0^t Q^n(s) ds \right] < \infty, \\ \mathbb{E} \left[\int_0^t L^n(Z^n(s)) ds \right] &= \mathbb{E} \left[\int_0^t Z^n(s) \wedge M^n ds \right] < \infty. \end{aligned}$$

Also, again by [32, Lemma 3.4], we derive that the following moments related to the Poisson processes are finite:

$$\begin{aligned} \mathbb{E} \left[N_v^n \left(\int_0^t Q^n(s) ds \right) \right] &< \infty, \quad \mathbb{E} \left[N_{v,1}^n \left(\int_0^t Z^n(s) ds \right) \right] < \infty, \\ \mathbb{E} \left[N_{v,2}^n \left(\int_0^t C^n(s) ds \right) \right] &< \infty, \quad \mathbb{E} \left[N_\mu^n \left(\int_0^t L^n(Z^n(s)) ds \right) \right] < \infty. \end{aligned}$$

Now, the result follows by the conclusion of the proof of [32, Theorem 7.1]. To this end, observe that $\left(\int_0^t Z^n(s) ds, \int_0^t C^n(s) ds, \int_0^t Z^n(s) \wedge M^n ds \right)$ is a \mathcal{F}^n -stopping time. Thus, applying [20, Chapter 2, Theorem 8.7], we derive that

$$\left(\hat{\mathcal{M}}_{\lambda^n}^n(t), \hat{\mathcal{M}}_\mu^n \left(\int_0^t L^n(Z^n(s)) ds \right), \hat{\mathcal{M}}_{v,1}^n \left(\int_0^t Z^n(s) ds \right), \hat{\mathcal{M}}_{v,2}^n \left(\int_0^t C^n(s) ds \right) \right)$$

is an \mathcal{F}^n -martingale. Finally, note that by (6.21), as $\hat{\mathcal{M}}_\mu^n(\cdot)$ is adapted to the filtration \mathcal{F}^n , we obtain that $\hat{\mathcal{M}}_v^n \left(\int_0^t Q^n(s) ds \right)$ is also an \mathcal{F}^n -martingale. \square

In order to apply the martingale central limit theorem, we first need to show that the corresponding fluid-scaled processes converge to a deterministic function (step 3), i.e.,

$$\bar{Z}^n(\cdot) := \frac{Z^n(\cdot)}{n} \xrightarrow{d} e(\cdot), \quad (6.23)$$

and

$$\bar{Q}^n(\cdot) := \frac{Q^n(\cdot)}{n} \xrightarrow{d} \frac{v + \mu}{\mu} e(\cdot), \quad (6.24)$$

where the function $e : [0, \infty) \rightarrow [0, \infty)$ is defined by $e(t) \equiv 1$. The following proposition presents the fluid limit.

Proposition 6.3 If $\bar{Z}^n(0) := \frac{Z^n(0)}{n} \xrightarrow{d} e(0)$ and $\bar{Q}^n(0) := \frac{Q^n(0)}{n} \xrightarrow{d} \frac{\nu+\mu}{\mu}e(0)$, then (6.23) and (6.24) hold as $n \rightarrow \infty$.

Proof We prove the fluid limits using the martingale representations (6.15) and (6.16). If the sequences $\{\hat{Z}^n(\cdot)\}$ and $\{\hat{Q}^n(\cdot)\}$ are stochastically bounded in $D[0, \infty)$, then by [32, Lemma 5.9] we have that

$$\frac{\hat{Z}^n(\cdot)}{\sqrt{n}} \xrightarrow{d} \eta(\cdot) \text{ and } \frac{\hat{Q}^n(\cdot)}{\sqrt{n}} \xrightarrow{d} \eta(\cdot),$$

where the function $\eta : [0, \infty) \rightarrow [0, \infty)$ is defined by $\eta(t) \equiv 0$. Note that the last limits are equivalent to (6.23) and (6.24).

The diffusion-scaled processes $\hat{Z}^n(\cdot)$ and $\hat{Q}^n(\cdot)$ have the martingale representation (6.15) and (6.16). In order to prove that they are stochastically bounded in $D[0, \infty)$, it is enough to show that the corresponding martingales are stochastically bounded in $D[0, \infty)$; see [32, Lemma 5.5]. By [32, Lemma 5.8] the martingales are stochastically bounded if the sequences of their predictable quadratic variations (6.17)–(6.20) are stochastically bounded in \mathbb{R} for each $t \geq 0$. To prove that the sequences of the predictable quadratic variations of the martingales in expressions (6.15) and (6.16) are stochastically bounded in \mathbb{R} , we use (6.22), [32, Lemma 6.2] and the fact that for any $t \geq 0$, $\frac{Q^n(t)}{n} \leq \frac{K^n}{n} < \infty$.

The predictable quadratic variation for the arrival process (6.17) is obviously bounded as it is a deterministic function. For (6.17), we have that

$$\mu \frac{\int_0^t L^n(Z^n(s))ds}{n} \leq \mu t \frac{M^n}{\sqrt{n}} = \mu t \left(1 + \frac{\beta}{\sqrt{n}}\right) \leq \mu t(1 + \beta).$$

The result for (6.19) follows by applying (6.22) and [32, Lemma 6.2]. Finally, applying the inequality $\frac{Q^n(t)}{n} \leq \frac{K^n}{n}$ in (6.20), we obtain

$$\nu \frac{\int_0^t Q^n(s)ds}{n} \leq \nu t \frac{K^n}{n} = \nu t \left(\frac{\nu + \mu}{\nu} + \frac{\kappa}{\sqrt{n}}\right) \leq \nu t \left(\frac{\nu + \mu}{\nu} + \kappa\right).$$

□

Before we move to the final step of the proof of Theorem 4.5, we make a remark about the fluid limit of the number of fully charged EVs in the system, which we need later.

Remark 6.1 Note that the diffusion-scaled process $\{\hat{C}^n(\cdot)\}$ is also stochastically bounded in $D[0, \infty)$ and the fluid limit is given by the difference $\frac{\nu+\mu}{\nu} - 1 = \frac{\mu}{\nu}$.

Now, we are ready to put all the pieces together, leading to the last step of the proof of Theorem 4.5.

Proof of Theorem 4.5 By Proposition 6.3 and the continuous mapping theorem [19, Theorem 1.2], we have that, as $n \rightarrow \infty$,

$$\begin{aligned} \langle \hat{\mathcal{M}}_{\lambda^n}^n(t) \rangle &\rightarrow (v + \mu)t, \quad \left\langle \hat{\mathcal{M}}_{\mu}^n \left(\int_0^t L^n(Z^n(s)) ds \right) \right\rangle \rightarrow \mu t, \\ \left\langle \hat{\mathcal{M}}_{v,1}^n \left(\int_0^t Z^n(s) ds \right) \right\rangle &\rightarrow vt, \quad \left\langle \hat{\mathcal{M}}_v^n \left(\int_0^t Q^n(s) ds \right) \right\rangle \rightarrow (v + \mu)t. \end{aligned}$$

Applying the martingale central limit theorem in [20], we have that

$$\begin{aligned} &\left(\hat{\mathcal{M}}_{\lambda^n}^n(\cdot), \hat{\mathcal{M}}_{\mu}^n(\cdot), \hat{\mathcal{M}}_{v,1}^n(\cdot), \hat{\mathcal{M}}_v^n(\cdot) \right) \xrightarrow{d} \\ &\left(\sqrt{v + \mu} W_{\lambda}(\cdot), \sqrt{\mu} W_{\mu}(\cdot), \sqrt{v} W_{v,1}(\cdot), \sqrt{v + \mu} W_v(\cdot) \right), \end{aligned} \quad (6.25)$$

where $W_{\lambda}(\cdot)$, $W_{\mu}(\cdot)$, $W_{v,1}(\cdot)$, and $W_v(\cdot)$ are (nonindependent) standard Brownian motions. It is essential to observe that, by (6.21) and Remark 6.1, we have that

$$\sqrt{v + \mu} W_v \stackrel{d}{=} \sqrt{v} W_{v,1} + \sqrt{\mu} W_{v,2},$$

where now $W_{\lambda}(\cdot)$, $W_{\mu}(\cdot)$, $W_{v,i}(\cdot)$, $i = 1, 2$, are independent standard Brownian motions. Furthermore, by the properties of Brownian motion, we obtain

$$\sqrt{v + \mu} W_{\lambda} - \sqrt{\mu} W_{\mu} - \sqrt{v} W_{v,1} \stackrel{d}{=} \sqrt{2(v + \mu)} W_{\hat{Z}}, \quad (6.26)$$

$$\sqrt{v + \mu} W_{\lambda} - \sqrt{v} W_{v,1} - \sqrt{\mu} W_{v,2} \stackrel{d}{=} \sqrt{2(v + \mu)} W_{\hat{Q}}, \quad (6.27)$$

where $W_{\hat{Z}}(\cdot)$ and $W_{\hat{Q}}(\cdot)$ are nonindependent standard Brownian motions. Further, we have that

$$\mathbb{E} \left[W_{\hat{Z}}(t) W_{\hat{Q}}(t) \right] = \frac{\mathbb{E} \left[(v + \mu) W_{\lambda}(t)^2 \right] + \mathbb{E} \left[v W_{v,1}(t)^2 \right]}{2(v + \mu)} = \frac{(2v + \mu)}{2(v + \mu)} t.$$

In addition, by [32, Theorem 7.3], we know that $(Q^n(\cdot), Y^n(\cdot))$ satisfies a one-dimensional reflection mapping; see [11, Sect. 6.2] for background on the reflection mapping. That is, $Y^n(\cdot)$ is the unique nondecreasing nonnegative process such that $Q^n(t) \leq K^n$, (6.15) holds and

$$\int_0^{\infty} \mathbb{1}_{\{Q^n(t) < K^n\}} dY^n(t) = 0,$$

which is equivalent to

$$\int_0^{\infty} \mathbb{1}_{\{\hat{Q}^n(t) < \kappa\}} dY^n(t) = 0.$$

Now, combining Theorem 6.1, Proposition 6.2, (6.25), (6.26), and (6.27), we derive that

$$(\hat{Z}^n(\cdot), \hat{Q}^n(\cdot), \hat{Y}^n(\cdot)) \xrightarrow{d} (\hat{Z}(\cdot), \hat{Q}(\cdot), \hat{Y}(\cdot)) \quad \text{in } D[0, \infty)^3,$$

where the vector $(\hat{Z}(\cdot), \hat{Q}(\cdot), \hat{Y}(\cdot))$ is characterized by (4.6). This concludes the proof of Theorem 4.5. \square

6.4 Proofs for Sect. 4.2.2

Proof of Proposition 4.6 Adding and subtracting the terms $\frac{v}{v+\mu}K^n$, $\frac{\mu v}{v+\mu}K^n t$, $\frac{v^2}{v+\mu}K^n t$, and the means of the Poisson processes in (4.11), we have that

$$\begin{aligned} Z_f^n(t) - \frac{v}{v+\mu}K^n &= Z_f^n(0) - \frac{v}{v+\mu}K^n + \left(N_v^f \left(n \int_0^t \left(\frac{K^n - Z_f^n(s)}{n} \right) ds \right) \right. \\ &\quad \left. - v n \int_0^t \left(\frac{K^n - Z_f^n(s)}{n} \right) ds \right) - \left(N_\mu^f \left(\mu n \int_0^t \frac{Z_f^n(s)}{n} \wedge \frac{M^n}{n} ds \right) \right. \\ &\quad \left. - \mu n \int_0^t \frac{Z_f^n(s)}{n} \wedge \frac{M^n}{n} ds \right) + v \int_0^t \left(K^n - Z_f^n(s) + \frac{v}{v+\mu}K^n \right) ds \\ &\quad - \mu \int_0^t \left(Z_f^n(s) - \frac{v}{v+\mu}K^n \right) \wedge \sqrt{n}\beta ds - (v+\mu) \frac{v}{v+\mu}K^n t. \end{aligned}$$

Recalling that $\hat{\mathcal{M}}_r^n(\cdot) := \frac{N_r^f(\cdot) - (r \cdot)}{\sqrt{n}}$ and dividing the last equation by \sqrt{n} , we have that

$$\begin{aligned} \hat{Z}_f^n(t) &= \hat{Z}_f^n(0) + \hat{\mathcal{M}}_v^n \left(n \int_0^t \frac{K^n - Z_f^n(s)}{n} ds \right) - \hat{\mathcal{M}}_\mu^n \left(n \int_0^t \frac{Z_f^n(s)}{n} \wedge \frac{M^n}{n} ds \right) \\ &\quad - v \int_0^t \hat{Z}_f^n(s) ds - \mu \int_0^t \hat{Z}_f^n(s) \wedge \beta ds + \frac{1}{\sqrt{n}} (vK^n - vK^n) t. \end{aligned}$$

Observing that the quantity $\int_0^t \frac{K^n - Z_f^n(s)}{n} ds$ is stochastically bounded and allowing $n \rightarrow \infty$ in the last equation, we derive

$$\begin{aligned} \hat{Z}_f(t) &= -v \int_0^t \hat{Z}_f(s) ds - \mu \int_0^t \hat{Z}_f(s) \wedge \beta ds + \sqrt{v \left(K - \frac{vK}{v+\mu} \right)} W_1(t) \\ &\quad - \sqrt{\mu \left(\frac{vK}{v+\mu} \right)} W_2(t), \end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are (independent) standard Brownian motions. Finally, by the properties of Brownian motion, we get

$$\sqrt{v \left(K - \frac{vK}{v+\mu} \right)} W_1(t) - \sqrt{\mu \left(\frac{vK}{v+\mu} \right)} W_2(t) \stackrel{d}{=} \sqrt{\frac{2v\mu K}{v+\mu}} W(t),$$

where $W(t)$ is a standard Brownian motion. \square

6.4.1 Proof of Theorem 4.7

The rest of this section gives a proof of Theorem 4.7. We first show a bound for the process $E^n(\cdot)$.

Proposition 6.4 *Let $T > 0$. We have that, for any $\epsilon > 0$, there exists n_ϵ such that*

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} E^n(t) \leq L(nT)^{1/4} + L \log(nT) \right) > 1 - \epsilon,$$

for any $n \geq n_\epsilon$, where L is a positive constant.

Before we proceed to the proof of Proposition 6.4, we show a preliminary result.

Lemma 6.5 *Let $E_M(t)$ denote the queue length process in an $M/M/1$ queue at time $t \geq 0$, with arrival rate νK and service rate λ such that $\nu K < \lambda$. For any $T > 0$, we have that*

$$\sup_{0 \leq t \leq nT} E_M(t) \leq L(nT)^{1/4} + L \log(nT),$$

almost surely as $n \rightarrow \infty$, where L is a positive constant.

Proof The proof of this lemma is based on results in [11]. By [11, Theorem 6.16], there exists a reflected (at zero) Brownian motion, $\tilde{E}_M(\cdot)$ with drift $\nu K - \lambda < 0$ such that

$$\sup_{0 \leq t \leq nT} |E_M(t) - \tilde{E}_M(t)| = o \left((nT)^{1/4} \right),$$

or equivalently

$$\sup_{0 \leq t \leq nT} |E_M(t) - \tilde{E}_M(t)| \leq L'(nT)^{1/4}, \quad (6.28)$$

for all $L' > 0$, almost surely as $n \rightarrow \infty$. Further, by [11, Theorem 6.3],

$$\sup_{0 \leq t \leq nT} |\tilde{E}_M(t)| = O(\log(nT)),$$

or alternatively there exists $L > 0$ such that

$$\sup_{0 \leq t \leq nT} |\tilde{E}_M(t)| \leq L \log(nT), \quad (6.29)$$

almost surely as $n \rightarrow \infty$.

Now, using the triangle inequality leads to

$$\begin{aligned} \sup_{0 \leq t \leq nT} E_M(t) &= \sup_{0 \leq t \leq nT} |E_M(t) - \tilde{E}_M(t) + \tilde{E}_M(t)| \\ &\leq \sup_{0 \leq t \leq nT} |E_M(t) - \tilde{E}_M(t)| + \sup_{0 \leq t \leq nT} |\tilde{E}_M(t)|. \end{aligned}$$

Applying (6.28) (by choosing $L' = L$) and (6.29) in the last inequality, we have that

$$\sup_{0 \leq t \leq nT} E_M(t) \leq L(nT)^{1/4} + L \log(nT),$$

almost surely as $n \rightarrow \infty$. \square

Now, we are ready to show Proposition 6.4.

Proof of Proposition 6.4 Let $E_M^n(t)$ denote the queue length process in an $M/M/1$ queue at time $t \geq 0$, with arrival rate $n\nu K$ and service rate $n\lambda$. Using standard coupling arguments, it can be shown that $\sup_{0 \leq t \leq T} E^n(t) \leq \sup_{0 \leq t \leq T} E_M^n(t)$ almost surely.

Further, we have that $\sup_{0 \leq t \leq T} E_M^n(t) \stackrel{d}{=} \sup_{0 \leq t \leq nT} E_M(t)$. Hence, for any $q^n > 0$,

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} E^n(t) \leq q^n \right) \geq \mathbb{P} \left(\sup_{0 \leq t \leq nT} E_M(t) \leq q^n \right).$$

Choosing $q^n = L(nT)^{1/4} + L \log(nT)$ and applying Lemma 6.5, we have that for any $\epsilon > 0$ there exists n_ϵ such that

$$\mathbb{P} \left(\sup_{0 \leq t \leq nT} E_M(t) \leq L(nT)^{1/4} + L \log(nT) \right) > 1 - \epsilon,$$

for $n > n_\epsilon$. This concludes the proof as ϵ is arbitrary. \square

Remark 6.2 Define the sample path set $\mathcal{G}^n \subseteq \Omega$ such that

$$\mathcal{G}^n := \left\{ \omega \in \Omega : \sup_{0 \leq t \leq T} E^n(t) \leq L(nT)^{1/4} + L \log(nT) \right\}.$$

By Proposition 6.4, it follows that $\mathbb{P}(\mathcal{G}^n) \rightarrow 1$, as $n \rightarrow \infty$. In the sequel, we assume that $\omega \in \mathcal{G}^n$.

Proof of Theorem 4.7 Rewriting (4.15) for the n^{th} system and using the assumption $\mathbb{P}(E^n(0) = 0) = 1$ yields

$$Z^n(t) = Z^n(0) - E^n(t) + N_{v,2} \left(\int_0^t (K^n - E^n(s) - Z^n(s)) ds \right) \\ - N_\mu \left(\int_0^t Z^n(s) \wedge M^n ds \right).$$

Let $T > 0$ and $\omega \in \mathcal{G}^n$. We have that

$$0 \leq \frac{E^n(t)}{\sqrt{n}} \leq \sup_{0 \leq t \leq T} \frac{E^n(t)}{\sqrt{n}} \leq \frac{L(nT)^{1/4} + L \log(nT)}{\sqrt{n}} \rightarrow 0,$$

as $n \rightarrow \infty$. The result now follows by adapting the proof of Proposition 4.6. \square

6.5 Proofs for Sect. 4.2.3

Proof of Proposition 4.8 First, we note that it is enough to show the result for $M = 1$. Then, we replace μ by μM ; see [41, Remark 5].

Scale the time by n , and add and subtract the means of the Poisson processes in (2.1) and (2.3) to obtain

$$Q^n(nt) = Q^n(0) + (N_{\lambda^n}(nt) - \lambda^n nt) - \left(N_{v^n} \left(n \int_0^t Q^n(ns) ds \right) \right. \\ \left. - nv^n \int_0^t Q^n(ns) ds \right) + \lambda^n nt - nv^n \int_0^t Q^n(ns) ds,$$

and

$$Z^n(nt) = Z^n(0) + (N_{\lambda^n}(nt) - \lambda^n nt) - \left(N_\mu \left(n \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds \right) \right. \\ \left. - n\mu \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds \right) - \left(N_{v^n,1} \left(n \int_0^t Z^n(ns) ds \right) - nv^n \int_0^t Z^n(ns) ds \right) \\ + \lambda^n nt - n\mu \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds - nv^n \int_0^t Z^n(ns) ds.$$

Define $\bar{Q}^n(t) = Q^n(nt)/n$, $\bar{Z}^n(t) = Z^n(nt)/n$, and recall that $\mathcal{M}_r(\cdot) := N_r(\cdot) - (r \cdot)$. Adding and subtracting the terms μn and $\lambda^n nt$ in the first equation yields

$$Q^n(nt) - \mu n = Q^n(0) - \mu n + \mathcal{M}_{\lambda^n}(nt) - \mathcal{M}_1 \left(n \int_0^t \bar{Q}^n(s) ds \right) - c\mu\sqrt{n}t \\ - \int_0^t (Q^n(ns) - \mu n) ds,$$

and

$$\begin{aligned} Z^n(nt) &= Z^n(0) + \mathcal{M}_{\lambda^n}(nt) - \mathcal{M}_{\mu} \left(n \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) > 0\}} ds \right) \\ &\quad - \mathcal{M}_{1,1} \left(n \int_0^t \tilde{Z}^n(s) ds \right) + \mu \left(1 - \frac{c}{\sqrt{n}} \right) nt - \mu n \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) > 0\}} ds \\ &\quad - \int_0^t (Z^n(ns)) ds. \end{aligned}$$

Dividing the last equations by \sqrt{n} and observing that

$$\mu \left(1 - \frac{c}{\sqrt{n}} \right) nt - \mu n \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) > 0\}} ds = -c\mu\sqrt{n}t + \mu n \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) = 0\}} ds,$$

we obtain

$$\tilde{Q}^n(t) = \tilde{Q}^n(0) + \hat{\mathcal{M}}_{\lambda^n}^n(nt) - \hat{\mathcal{M}}_1^n \left(n \int_0^t \tilde{Q}^n(s) ds \right) - c\mu t - \int_0^t \tilde{Q}^n(s) ds,$$

and

$$\begin{aligned} \tilde{Z}^n(t) &= \tilde{Z}^n(0) + \hat{\mathcal{M}}_{\lambda^n}^n(nt) - \hat{\mathcal{M}}_{\mu}^n \left(n \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) > 0\}} ds \right) \\ &\quad - \hat{\mathcal{M}}_{1,1}^n \left(n \int_0^t \tilde{Z}^n(s) ds \right) - c\mu t + \mu\sqrt{n} \int_0^t \mathbb{1}_{\{\tilde{Z}^n(s) = 0\}} ds - \int_0^t \tilde{Z}^n(s) ds. \end{aligned}$$

Now, taking the limit $n \rightarrow \infty$ and using the reflection mapping [11], we derive

$$\begin{aligned} d\tilde{Q}(t) &= -(c\mu + \tilde{Q}(t))dt + \sqrt{2\mu}dW_{\tilde{Q}}(t), \\ d\tilde{Z}(t) &= -(c\mu + \tilde{Z}(t))dt + \sqrt{2\mu}dW_{\tilde{Z}}(t) + d\tilde{Y}(t), \end{aligned}$$

where $\int_0^\infty \mathbb{1}_{\{\tilde{Z}(t) > 0\}} d\tilde{Y}(t) = 0$. Further, $\mathbb{E} \left[W_{\tilde{Z}}(t) W_{\tilde{Q}}(t) \right] = \frac{\mu M t}{2\mu M} = t/2$. \square

Proof of Proposition 4.9 First, as in Proposition 4.8, we note that it is enough to show the result for $M = 1$ and then replace μ by μM . Scale the time by n , and add and subtract the means of the Poisson processes in (2.1) and (2.3) to obtain

$$\begin{aligned} Q^n(nt) &= Q^n(0) + (N_\lambda(nt) - \lambda nt) - \left(N_{v^n} \left(n \int_0^t Q^n(ns) ds \right) \right. \\ &\quad \left. - nv^n \int_0^t Q^n(ns) ds \right) + \lambda nt - nv^n \int_0^t Q^n(ns) ds, \end{aligned}$$

and

$$\begin{aligned} Z^n(nt) = & Z^n(0) + (N_\lambda(nt) - \lambda nt) - \left(N_\mu \left(n \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds \right) \right. \\ & \left. - n\mu \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds \right) - \left(N_{\nu^n,1} \left(n \int_0^t Z^n(ns) ds \right) - n\nu^n \int_0^t Z^n(ns) ds \right) \\ & + \lambda nt - n\mu \int_0^t \mathbb{1}_{\{Z^n(ns) > 0\}} ds - n\nu^n \int_0^t Z^n(ns) ds. \end{aligned}$$

Adding and subtracting the terms λn and λnt in the first equation and the terms $(\lambda - \mu)n$ and $(\lambda - \mu)nt$ in the second equation yields

$$\begin{aligned} Q^n(nt) - \lambda n = & Q^n(0) - \lambda n + \mathcal{M}_\lambda(nt) - \mathcal{M}_1 \left(n \int_0^t \bar{Q}^n(s) ds \right) \\ & - \int_0^t (Q^n(ns) - \lambda n) ds, \end{aligned}$$

and

$$\begin{aligned} Z^n(nt) - (\lambda - \mu)n = & Z^n(0) - (\lambda - \mu)n + \mathcal{M}_\lambda(nt) \\ & - \mathcal{M}_\mu \left(n \int_0^t \mathbb{1}_{\{\bar{Z}^n(s) > 0\}} ds \right) - \mathcal{M}_{1,1} \left(n \int_0^t \bar{Z}^n(s) ds \right) \\ & + \lambda nt - (\lambda - \mu)nt - \mu n \int_0^t \mathbb{1}_{\{\bar{Z}^n(s) > 0\}} ds \\ & - \int_0^t (Z^n(ns) - (\lambda - \mu)n) ds. \end{aligned}$$

Dividing the last equations by \sqrt{n} and observing that

$$\lambda nt - (\lambda - \mu)nt - \mu n \int_0^t \mathbb{1}_{\{\bar{Z}^n(s) > 0\}} ds = \mu n \int_0^t \mathbb{1}_{\{\bar{Z}^n(s) = 0\}} ds,$$

we obtain

$$\tilde{Q}_o^n(t) = \tilde{Q}_o^n(0) + \hat{\mathcal{M}}_\lambda^n(nt) - \hat{\mathcal{M}}_1^n \left(n \int_0^t \bar{Q}^n(s) ds \right) - \int_0^t \tilde{Q}_o^n(s) ds,$$

and

$$\begin{aligned} \tilde{Z}_o^n(t) = & \tilde{Z}_o^n(0) + \hat{\mathcal{M}}_\lambda^n(nt) - \hat{\mathcal{M}}_\mu^n \left(n \int_0^t \mathbb{1}_{\{\bar{Z}^n(s) > 0\}} ds \right) \\ & - \hat{\mathcal{M}}_{1,1}^n \left(n \int_0^t \bar{Z}^n(s) ds \right) - \mu \sqrt{n} \int_0^t \mathbb{1}_{\{\bar{Z}_o^n(s) = 0\}} ds - \int_0^t \tilde{Z}_o^n(s) ds. \end{aligned}$$

By the overloaded assumption, it follows that $\sqrt{n} \int_0^t \mathbb{1}_{\{\bar{Z}^n(s)=0\}} ds \rightarrow 0$, as $n \rightarrow \infty$ [41]. Now, taking the limit $n \rightarrow \infty$, we derive

$$\begin{aligned} d\tilde{Q}_o(t) &= -\tilde{Q}_o(t)dt + \sqrt{\lambda}dW_1(t) - \sqrt{\mu}dW(t), \\ d\tilde{Z}_o(t) &= -\tilde{Z}_o(t)dt + \sqrt{\lambda}dW_1(t) - \sqrt{\mu}dW_2(t) - \sqrt{\lambda - \mu}dW_3(t), \end{aligned}$$

and we can write $\sqrt{\mu}W(t) \stackrel{d}{=} \sqrt{\lambda - \mu}W_3(t) + \sqrt{\mu}W_4(t)$, where W_i are independent standard Brownian motions. \square

7 Concluding remarks

This paper proposes modelling an electric vehicle charging system by using layered queueing networks. We develop several bounds and approximations for the number of uncharged EVs in the system and the probability that an EV leaves the charging station with a fully charged battery. In the numerical examples, it seems that a modification of the approximation for a full parking lot leads to a good approximation. Further, the fluid approximation seems to be good in most cases and we believe that the diffusion approximation in the Halfin–Whitt regime will improve the fluid approximation. Unfortunately, the exact (or even numerical) solution of (4.9) seems very hard.

From an application standpoint, it is important to remove various model assumptions. If parking and charging times are given by the (possibly dependent) generally distributed random variables B and D , we can develop a measure-valued fluid model by extending [22]. In addition, we can include in the model time-varying arrival rates, multiple EV types, and multiple parking lots, thus extending results in [34]. Moreover, the distribution grid (low-voltage network) plays a crucial role and it should be included in the model; see [4] for a heuristic approach. For another application in EV charging including the distribution grid, see [10], where simulation results are presented for a Markovian model.

Acknowledgements We would like to thank the anonymous reviewer for his/her helpful comments. The research of Angelos Aveklouris is funded by a TOP Grant of the Netherlands Organization for Scientific Research (NWO) through Project 613.001.301. The research of Maria Vlasiou is supported by the NWO MEERVOUD Grant 632.003.002. The research of Bert Zwart is partly supported by the NWO VICI Grant 639.033.413.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Aldous, D.: Some interesting processes arising as heavy traffic limits in an $M/M/\infty$ storage process. *Stoch. Process. Appl.* **22**(2), 291–313 (1986)
2. Aveklouris, A., Nakahira, Y., Vlasiou, M., Zwart, B.: Electric vehicle charging: a queueing approach. *SIGMETRICS Perform. Eval. Rev.* **45**(2), 33–35 (2017)

3. Aveklouris, A., Vlasiou, M., Zhang, J., Zwart, B.: Heavy-traffic approximations for a layered network with limited resources. *Probab. Math. Stat.* **37**(2), 497–532 (2017)
4. Aveklouris, A., Vlasiou, M., Zwart, B.: A stochastic resource-sharing network for electric vehicle charging. *IEEE Trans. Control Net. Syst.* (**accepted**) (2019). [arXiv:1711.05561](https://arxiv.org/abs/1711.05561)
5. Bayram, S., Michailidis, G., Devetsikiotis, M., Granelli, F.: Electric power allocation in a network of fast charging stations. *IEEE J. Sel. Areas Commun.* **31**(7), 1235–1246 (2013)
6. Bertoin, J.: *Lévy Processes*, vol. 121. Cambridge University Press, Cambridge (1996)
7. Billingsley, P.: *Convergence of Probability Measures*, 2nd edn. Wiley, New York (1999)
8. Bo, L., Wang, Y., Yang, X.: First passage times of (reflected) Ornstein–Uhlenbeck processes over random jump boundaries. *J. Appl. Probab.* **48**(3), 723–732 (2011)
9. Browne, S., Whitt, W.: Piecewise-linear diffusion processes. In: Dshalalow, J. (ed.) *Advances in Queueing: Theory, Methods, and Open Problems*, pp. 463–480. CRC Press, Boca Raton (1995)
10. Carvalho, R., Buzna, L., Gibbens, R., Kelly, F.: Critical behaviour in charging of electric vehicles. *New J. Phys.* **17**(9), 095001 (2015)
11. Chen, H., Yao, D.D.: *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*, vol. 46. Springer, New York (2001)
12. Clement-Nyns, K., Haesen, E., Driesen, J.: The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE Trans. Power Syst.* **25**(1), 371–380 (2010)
13. Cohen, J.: The Wiener–Hopf technique in applied probability. *J. Appl. Probab.* **12**(S1), 145–156 (1975)
14. Cohen, J., Boxma, O.J.: *Boundary Value Problems in Queueing System Analysis*, vol. 79. Elsevier, New York (2000)
15. Dai, J., Dieker, A.: Nonnegativity of solutions to the basic adjoint relationship for some diffusion processes. *Queueing Syst.* **68**(3–4), 295 (2011)
16. Dai, J., Miyazawa, M.: Reflecting Brownian motion in two dimensions: exact asymptotics for the stationary distribution. *Stoch. Syst.* **1**(1), 146–208 (2011)
17. Dorsman, J.L., Boxma, O.J., Vlasiou, M.: Marginal queue length approximations for a two-layered network with correlated queues. *Queueing Syst.* **75**(1), 29–63 (2013)
18. Dorsman, J.L., Vlasiou, M., Zwart, B.: Heavy-traffic asymptotics for networks of parallel queues with Markov-modulated service speeds. *Queueing Syst.* **79**(3–4), 293–319 (2015)
19. Durrett, R.: *Stochastic Calculus: A Practical Introduction*, vol. 6. CRC Press, Boca Raton (1996)
20. Ethier, S.N., Kurtz, T.G.: *Markov Processes: Characterization and Convergence*. Wiley, New York (1986)
21. Fleming, P., Stolyar, A., Simon, B.: Heavy traffic limit for a mobile phone system loss model. In: *Proc. 2nd Int'l. Conf. on Telecomm. Syst. Mod. and Analysis*, Nashville (1994)
22. Gromoll, C., Robert, P., Zwart, B., Bakker, R.: The impact of reneging in processor sharing queues. *SIGMETRICS Perform. Eval. Rev.* **34**(1), 87–96 (2006)
23. Gromoll, H.C., Robert, P., Zwart, B.: Fluid limits for processor-sharing queues with impatience. *Math. Oper. Res.* **33**(2), 375–402 (2008)
24. Hoogsteen, G., Molderink, A., Hurink, J.L., Smit, G.J., Kootstra, B., Schuring, F.: Charging electric vehicles, baking pizzas, and melting a fuse in Lochem. *CIRE Open Access Proc. J.* **2017**(1), 1629–1633 (2017)
25. International Energy Agency: *Global EV outlook* (2017). <https://www.iea.org/publications/freepublications/publication/GlobalEVO Outlook2017.pdf/>
26. Kang, W.: Fluid limits of many-server retrial queues with nonpersistent customers. *Queueing Syst.* **79**(2), 183–219 (2015)
27. Kang, W., Ramanan, K.: Characterization of stationary distributions of reflected diffusions. *Ann. Appl. Probab.* **24**(4), 1329–1374 (2014)
28. Karatzas, I., Shreve, S.: *Brownian Motion and Stochastic Calculus*, vol. 113. Springer, New York (1991)
29. Kelly, F., Yudovina, E.: *Stochastic Networks*, vol. 2. Cambridge University Press, Cambridge (2014)
30. Kempker, P., Dijk, N.V., Scheinhardt, W., Berg, H.V.D., Hurink, J.: Optimization of charging strategies for electric vehicles in PowerMatcher-driven smart energy grids. In: *Proceedings of the 9th EAI International Conference on Performance Evaluation Methodologies and Tools* (2015)
31. Li, G., Zhang, X.P.: Modeling of plug-in hybrid electric vehicle charging demand in probabilistic power flow calculations. *IEEE Trans. Smart Grid* **3**(1), 492–499 (2012)
32. Pang, G., Talreja, R., Whitt, W.: Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.* **4**, 193–267 (2007)

33. Reiman, M.: The heavy traffic diffusion approximation for sojourn times in Jackson networks. In: Disney, R., Ott, T. (eds.) *Applied Probability-Computer Science: The Interface Volume II*, pp. 409–422. Birkhauser, Boston, MA (1982)
34. Remerova, M., Reed, J., Zwart, B.: Fluid limits for bandwidth-sharing networks with rate constraints. *Math. Oper. Res.* **39**(3), 746–774 (2014)
35. Robert, P.: *Stochastic Networks and Queues*, vol. 52. Springer, Berlin (2013)
36. Rolia, J.A., Sevcik, K.C.: The method of layers. *IEEE Trans. Softw. Eng.* **21**(8), 689–700 (1995)
37. Said, D., Cherkaoui, S., Khoukhi, L.: Queuing model for EVs charging at public supply stations. *Wirel. Commun. Mobile Comput. Conf.* **1**(5), 65–70 (2013)
38. Sortomme, E., Hindi, M., MacPherson, J., Venkata, S.: Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses. *IEEE Trans. Smart Grid* **2**(1), 198–205 (2011)
39. Su, W., Chow, M.Y.: Performance evaluation of an EDA-based large-scale plug-in hybrid electric vehicle charging algorithm. *IEEE Trans. Smart Grid* **3**(1), 308–315 (2012)
40. Turitsyn, K., Sinitsyn, N., Backhaus, S., Chertkov, M.: Robust broadcast-communication control of electric vehicle charging. In: *2010 1st IEEE Int. Conf. Smart Grid Commun.*, pp. 203–207 (2010)
41. Ward, A.R., Glynn, P.W.: A diffusion approximation for a Markovian queue with reneging. *Queueing Syst.* **43**(1), 103–128 (2003)
42. Van der Mei, R.D., Hariharan, R., Reeser, P.: Web server performance modeling. *Telecommun. Syst.* **16**(3–4), 361–378 (2001)
43. Van der Weij, W., Bhulai, S., Van der Mei, R.: Dynamic thread assignment in web server performance optimization. *Perform. Eval.* **66**(6), 301–310 (2009)
44. Whitt, W.: Proofs of the martingale FCLT. *Probab. Surv.* **4**, 268–302 (2007)
45. Woodside, M., Neilson, J.E., Petriu, D.C., Majumdar, S.: The stochastic rendezvous network model for performance of synchronous client–server-like distributed software. *IEEE Trans. Comput.* **44**(1), 20–34 (1995)
46. Xing, X., Zhang, W., Wang, Y.: The stationary distributions of two classes of reflected Ornstein–Uhlenbeck processes. *J. Appl. Probab.* **46**(3), 709–720 (2009)
47. You, P., Sun, Y., Pang, J., Low, S., Chen, M.: Battery swapping assignment for electric vehicles: a bipartite matching approach. *SIGMETRICS Perform. Eval. Rev.* **45**(2), 85–87 (2017)
48. Yudovina, E., Michailidis, G.: Socially optimal charging strategies for electric vehicles. *IEEE Trans. Autom. Control* **60**(3), 837–842 (2015)
49. Zeltyn, S.: *Call Centers with Impatient Customers: Exact Analysis and Many-Server Asymptotics of the $M/M/n + G$ Queue*. Technion-Israel Institute of Technology, Faculty of Industrial and Management Engineering, Haifa (2005)